dc_2020_22

MTA doktori értekezés

Membrán ABC fehérjék működése és gyógyszer-kölcsönhatásai a 3D bioinformatika tükrében

Dr. Hegedűs Tamás

Semmelweis Egyetem, Biofizikai és Sugárbiológiai Intézet
ELKH-SE Biofizikai Virológia Kutatócsoport

Budapest, 2022

Köszönetnyilvánítás

Szüleimnek köszönöm az erkölcsi alapokat, a kitartásra nevelést, s a tudás alapjait (amikor az eltérő számrendszereket nem értettem meg, akkor a fehér asztalra zsákocskákat rajzoltak, amikbe babszemeket tettünk – még nem sejtettük, hogy lesz olyan időszaka életemnek, amikor készségszinten használom a kettes és a tizenhatos számrendszert).

Köszönöm oktatóimnak, hogy felkeltették érdeklődésem a tudományok iránt, különösen Gyimesi István, Pázmándy György, Balog Tamás, Görbe László, és Urbanek Rudolf piarista tanáraimnak.

Köszönöm Sarkadi Balázsnak, Homolya Lászlónak és Váradi Andrásnak, hogy elindítottak a kutatói pályán, s azóta is kritikus szeretettel követik munkámat. Hálás vagyok a Biomembrán közösség valamennyi munkatársának. Tudományos pályám elindításában és irányadásában jelentős szerepe volt az Észak-Karolinai Egyetem kutatóinak (Ric Boucher, Jack Riordan és Nikolay Dokholyan) is.

Különösen nagyon nagy köszönettel tartozom közvetlen munkatársaimnak (kutatók, PhD hallgatók, MSc diákok), akik az elmúlt években rövidebb-hosszabb ideg mellettem dolgoztak és együtt haladtunk előre. Többek között Tordai Hedvignek, Padányi Ritának, Gyimesi Gergelynek és Lór Krisztinek. A Biofiziai Intézet vezetőjének, Kellermayer Miklósnak és munkatársainak támogatása nélkül csoportomat nehéz lett volna működtetni. Csoportom fenntartásában elévülhetetlenül nagy érdeme van a TKI munkatársainak is, különösen Berzéné Pénzes Ilonának.

Köszönöm az együttgondolkodás lehetőségét és a baráti támogatást egyműködő partnereimnek, Lukács Gergelynek, Enyedi Ágnesnek, Markus Geislernek, és Helmut Grubmüllernek. Eredményeinket nem tudtuk volna elérni a Cygnea Bt. (Magyar Csaba), a Boblem Kft. (Babics Balázs), a Wigner GPU Laboratórium (Barnaföldi Gergely), és a KIFÜ (Fekete Attila) munkatársainak szoftver- és hardver-kérdésekben nyújtott segítsége nélkül. Kutatásinkhoz nélkülözhetetlen volt többek között az OTKA/NKFIH és az Észak-Amerikai Cisztás Fibrózis Alapítvány (CFF) támogatása és különböző ösztöndíjak, mint például az MTA Bolyai János Kutatási Ösztöndíj.

Köszönöm a megértést vagy a türelemmel párosult meg-nem-értést közvetettebb munkatársaim mellett barátaimnak (Fábry Gyuri, Andor Csaba, Janitsáry Iván, Tüdős Balázs, Krafft Walter, Somogyi András) és családomnak. Különösen Ilonának, Fanninak, Eminek és Palkónak, mert valószínűleg ők érzik legjobban a kutatói lét hátrányait.

Soli Deo gloria.

Tartalomjegyzék

Rövidítések jegyzéke

A rövidítések feloldása általában az első előfordulásuk mellett található, míg a fontosabb rövidítések feloldását itt is feltüntetettem.

ABC: *ATP Binding Cassette*, ATP-kötő kazetta
ADME-Tox: *absorption, distribution, metabolism, excretion, and toxicity*
ATP: Adenozin-trifoszfát
BCRP: *breast cancer resistance protein*, ABCG2
CFTR: *cystic fibrosis transmembrane conductance regulator*, cisztás fibrózis transzmembrán konduktancia szabályozó
CH: *coupling helix*, kapcsoló spirál, kapcsoló hélix
COG: *center of geometry,* geometriai középpont
FRET: *Fluorescence Resonance Energy Transfer*
ICD: intracelluláris domén
NBD: *nucleotide binding domain*, nukleotidkötő domén
NMR: *nuclear magnetic resonance*, mágneses magrezonancia
TH: transzmembrán hélix
PKA: protein kináz A
TM: transzmembrán vagy transzmembrán hélix
TMD: transzmembrán domén
Pgp: P-glikoprotein, ABCB1
RMSD: *root-mean-square deviation,* átlagos négyzetes eltérés négyzetgyöke
zf: *zebrafish*, zebrahal

1. Bevezetés

1.1. Transzmembrán ABC fehérjék

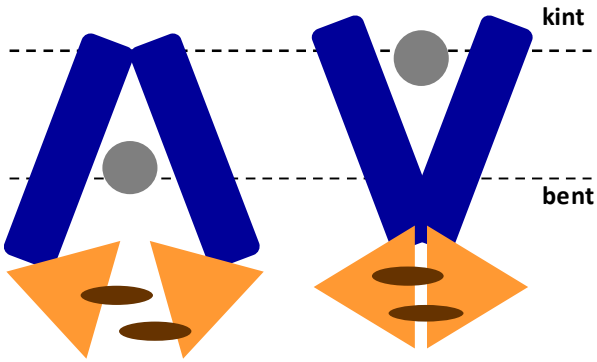
Az ABC (*ATP Binding Cassette*; ATP-kötő kazetta) membránfehérjék minden élőlényben fontos szerepet játszanak különféle vegyületek biológiai membránokon keresztül történő átjuttatásában[1,2]. Baktériumokban vannak olyan ABC importer fehérjék, amelyek tápanyagok külső térből a belső térbe történő szállításáért felelősek[3]. Az eukarióta ABC transzporterek főleg exporterként működnek, a sejtekből a sejten kívüli térbe történő transzportfolyamatokban játszanak szerepet[4,5]. Változatos anyagokat, többek között pigmenteket, hormonokat, és lipideket szállíthatnak. Számos ABC fehérje a sejt szintű kemoimmunitási rendszer részeként multidrog transzporterként működik, fő feladatuk a környezetből érkező mérgező anyagok sejtből történő kipumpálása[6].

Az emberi ABC fehérjék hibás működése vagy termelődése különböző betegségek kialakulásához vezethet, mint például a cisztás fibrózis, a Pseudoxanthoma elasticum, a Dubin-Johnson-szindróma, a II. típusú cukorbetegség és az adrenoleukodisztrófia[1,7–11]. A multidrog ABC transzporterek részt vesznek rákos sejtek kemoterápiás kezeléssel szemben mutatott ellenállóképességének kialakításában is[4,12–15]. Ezek az ABC fehérjék (pl. ABCB1/MDR1/Pgp, ABCC1/MRP1, és ABCG2/BCRP) alacsony szubsztrátspecifitással képesek a sejtből különböző kémiai tulajdonságokkal rendelkező vegyületeket eltávolítani, ezáltal a daganatsejten belüli gyógyszerkoncentrációt a hatásos szint alatt tartani.

1.2. ABC fehérjék szerkezete és működése

Az ABC fehérjék funkcionális egysége két transzmembrán domént (TMD) és két nukleotid-kötő domént (NBD) tartalmaz, amelyek lehetnek egy (ún. teljes transzporterek), két (ún. fél transzporterek), vagy több polipeptid láncon kódolva[4,16]. A két NBD jelenléte azért nélkülözhetetlen, mert a transzporter két ATP-kötő helyét együtt alakítják ki: az egyik NBD-ben található Walker motívum és a másik NBD-ben található *ABC signature* (ujjlenyomat) szekvenciával ill. fordítva[17,18]. A transzport folyamatokat általában e helyek ATP-kötése és/vagy hidrolízise látja el energiával[17,18]. Ez alól kivétel például a humán ABCC7/CFTR klorid csatorna[8,9] és a szulfonilurea receptorok (ABCC8/SUR1 és ABCC9/SUR2)[19]. Ezekben a fehérjékben az ATP/ADP kötődés ioncsatorna nyitást és zárást szabályoz.

Az ATP hiányában megoldott „apo” röntgenszerkezetek "alul-nyitott" konformációt mutatnak, amelyben az NBD-k és a TM hélixek intracelluláris vége távol van egymástól, így a



1. ábra: A szubsztrát (szürke) a lipidből vagy a citoszólból be tud lépni az alul-nyitott szerkezetbe, s olyan konformációs változásokat gerjeszt, amelyek eredményeként az NBD-k (narancssárga) szorosan kapcsolódnak. Ennek hatására a TMD-k (kék) intracelluláris vége záródik, s extracellulárs végük nyitódik, ami lehetővé teszi a szubsztrát távozását.

központi kötőzseb a sejt belsejéből hozzáférhető, ami lehetővé teszi a szubsztrát bekötődését (1. ábra)[18,20]. ATP-kötés hatására a két NBD között szoros kölcsönhatás alakul ki, "dimerizálódnak". Ekkor a TM hélixek intracelluláris vége közel kerül egymáshoz, míg az extracelluláris végek távolodnak egymástól. Ez az átfordulás eredményezi az úgynevezett "alul-zárt és felül-nyitott" konformáció kialakulását, miközben a szubsztrát az extracelluláris térbe tud távozni. Azaz a katalitikus ciklus során a fehérje központi kötőzsebe váltakozó módon érhető el vagy az intracelluláris, vagy az extracelluláris oldal felől. Bár ez az *alternativ access* mechanizmus a legelfogadottabb (1. ábra), számos más működési lehetőséget is vázoltak, mint amilyen a porszívó (*vacuum cleaner model*) és kreditkártya (*credit card swipe model*) mechanizmus[21–23]. Az eltérő mechanizmusok abból fakadhatnak, hogy míg az NBD-k konzerváltsága nagyon magas, addig a TMD-k szekvenciája nem, így szerkezetük és a működési mechanizmusuk is nagyon eltérő lehet. Például, míg a Pgp-szerű szerkezetek alul-nyitott és alul-zárt konformációi vizuálisan is határozottan különböznek, addig az ABCG2-szerű szerkezetek e két konformációja nem tér el annyira egymástól[24].

2. Célkitűzések

A CFTR fehérje mutációi miatt alakul ki a kaukázusi populációban a leggyakoribb és igen súlyos öröklődő betegség, a cisztás fibrózis[8,9]. Célunk, hogy megértsük a mutációknak a CFTR fehérje szerkezetére és dinamikájára kifejtett hatását, hatásmechanizmusát, hogy ezekre alapozva a hatások korrigálására gyógyszereket lehessen fejleszteni. Mivel a fehérjék szervezetünkben 37°C-on nem egyetlen szerkezetet alakítanak ki, hanem funkciójuk ellátása során dinamikusan konformációk sokaságát veszik fel, ezért igen fontos mozgásuk megismerése is. A kísérletesen meghatározott statikus CFTR szerkezetek egyike sem tartamaz klorid vezetésre alkalmas nyitott csatornát (Cikk-3). Ezért a gyógyszer-jelölt vegyületek hatásmechanizmusának, a csatorna nyitásának és zárásának, illetve a klorid transzport folyamatának megértéséhez *in silico* dokkolást és molekuladinamikai szimulációkat végeztünk

(Cikk-1, Cikk-2, Cikk-3). Vizsgálataink együttműködő partnerünk kísérleti eredményeinek atomi szintű értelmezését is segítették (Cikk-1, Cikk-2), tanulmányaink hatékonyabb terápiás fejlesztésekhez járultak és járulhatnak hozzá.

Az ABCG2 (BCRP: *Breast Cancer Resistance Protein*) fehérje igen fontos multidrog transzporter. Gátlásával ugyan nem tudjuk felfüggeszteni a rákos sejtek multidrog-rezisztenciáját, mivel csak egy a kemoimmunitási rendszert alkotó számos fehérje közül[6], ennek ellenére szubsztrátfelismerésének megismerése nagyon fontos. Több mint 100 gyógyszer szervezeten belüli eloszlását (ADME-Tox tulajdonságait) befolyásolja (https://go.drugbank.com/bio_entities/BE0001067), hatásosságukat csökkenti, vagy toxicitásukat növeli[6,25]. Az ABCG2 egyes mutációi befolyásolhatják a fehérje feltekeredését és membrán lokalizációját illetve magát a transzportot, így kóros folyamatokat (pl. köszvényt) okozhatnak[18–20]. A szubsztrátfelismerés, a transzport folyamatok, s a mutációk hatásának megértéséhez molekuladinamikai szimulációkat végeztünk (Cikk-4, Cikk-5). Ehhez kiindulásként nélkülözhetetlen volt a fehérje szerkezeti modelljének felépítése is, mert kísérletes ABCG2 szerkezet csak később vált elérhetővé (Cikk-4).

A CFTR és az ABCG2 fehérjékkel végzett munka során számos, általánosan ABC fehérjékkel vagy transzmembrán fehérjékkel kapcsolatos kérdés is felmerült. Annak érdekében, hogy az ezekre adott megoldásaink ne csak saját projektjeinkben tudjanak hasznosulni, ezért ezeket is közöltük, illetve publikus webapplikációkon keresztül elérhetővé tettük (Cikk-6 – Cikk-10).

3. Módszerek

Az alkalmazott módszerek lényegét itt foglalom össze, míg az adott kutatásban alkalmazott szoftverek, algoritmusok és paraméterek fontosabb részleteit az eredmények részekbe illesztettem, teljes részletességgel közleményeinkben találhatók meg.

Fehérje szerkezetek modellezése. Vizsgálatainkban elsősorban kísérletes szerkezeteket alkalmazunk. A dinamikus, nem feloldott, hiányzó részeket a Modeller[26] hurok-modellező algoritmusával építjük fel. Ha nem elérhető kísérletesen meghatározott szerkezet, akkor a fehérje koordinátáit homológia modellezéssel, a Modeller[26] segítségével hozzuk létre. Általában 100 darab modellt készítünk, s ezekből a DOPA pontszám alapján választjuk a legmegfelelőbbet. Az AlphaFold Fehérje Szerkezeti Adatbázisból[27] is töltünk le szerkezeteket illetve lokálisan is futtatjuk az AlphaFold[28] programot a paraméterek finomhangolása érdekében.

In silico dokkolás. Ha tehetjük, akkor az Autodock programot használjuk kismolekulák fehérjén található kötőhelyének meghatározásához[29]. Ez azonban csak olyan esetben használható, amikor a kötőhely körülbelüli helye már ismert. Ellenkező esetben a pontatlanabb pontozófüggvénnyel rendelkező, ám nagyobb teljesítményű Autodock Vina programot alkalmazzuk[30]. A dokkolásokat lehetőségeink függvényében molekuladinamikai szimulációkkal ellenőrizzük.

Molekula dinamika. Szimulációinkban CHARMM36 vagy CHARMM36m erőteret használunk[31,32]. Ezért a kismolekulák paraméterezését a CgenFF programon alapulva végezzük, a CHARMM-GUI web alkalmazással[33,34]. Általában szimulációs rendszereinket is (fehérje, lipidkettősréteg, stb.) a CHARMM-GUI segítségével építjük fel. Az energiaminimalizálást és az egyensúlyba hozást lokálisan futtatjuk, míg az erőforrásigényes tényleges szimuláció (*production run)* nagyteljesítményű GPU szervereken, HPC környezetben (KIFÜ, Wigner Tudományos Számítási Laboratórium, Max Planck Intézet) fut. Szimlációra a GROMACS szoftvert használjuk[35].

Adatelemzés. Az adatok feldolgozásához elsősorban saját Python szkripteket használunk. Ehhez a leggyakrabban alkalmazott könyvtárak: NumPy, matplotlib, és az MDAnalysis[36–38]. Szerkezetek vizualizációjához és elemzéséhez PyMOL-t (Schrödinger, LLC), míg MD trajektóriák analíziséhez VMD programot használunk[39].

Web alkalmazások. Python alapú keretrendszereket alkalmazunk (TurboGears, Django, és FastAPI), amelyek lehetővé teszik az adat-, a logikai-, és a vizualizációs-réteg szétválasztását. Az esetek többségében az adatok relációs adatbázisban (MySQL vagy PostgreSQL) kerülnek tárolásra, míg az adat és logika közti kapcsolatot objektum-relációs megfeleltetéssel (ORMapper, pl. SQLAlchemy[40]) hozzuk létre.

Hardver. Adatelemző szerverünk 32 fizikai CPU maggal és 256 GB memóriával rendelkezik. Ezen kívül 3 db GPU szerveren (egyenként 2-4 GPU kártya: A6000, P6000, RTX 2080Ti, és GXT 1080Ti) tudunk lokálisan számolásokat futtatni, amelyek közül kettőben 96 GB RAM és 4 TB SSD van, lehetővé téve az AlphaFold futtatásokat. Összesen 32 TB redundáns tárolókapacitás áll rendelkezésünkre.

4. Eredmények

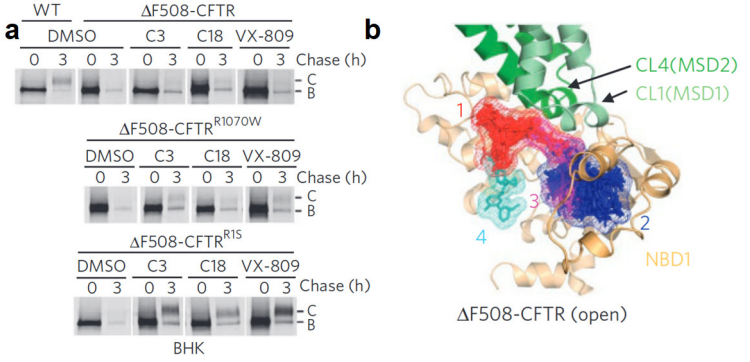
4.1. A CFTR fehérje szerkezetének, dinamikájának és CF-gyógyszer jelöltek vizsgálata

A cisztás fibrózist (CF) a cisztás fibrózis transzmembrán konduktancia szabályozó (CFTR) fehérje mutációi okozzák, amelyek e kloridcsatorna funkcionális expressziójának hiányát eredményezi hámsejtek apikális membránjában[8,9,41]. Az elmúlt évtizedekben jelentős erőforrásokat fordítottak a leggyakoribb mutánsok, mint amilyen a ΔF508 és a G551D, funkcionális expressziójának helyreállítására[42–44]. A betegek jelentős részében (>80%) a ΔF508 mutáció homo- vagy heterozigóta formában megtalálható. A mutáció hatására a CFTR fehérje jelentős része a szintézis után azonnal lebomlik. Emellett, ahogy korábban bemutattam[41], még ha ki is juttatjuk a CFTR ΔF508 fehérjét a sejt membránjába, fiziológiás hőmérsékleten az nem működőképes. Azaz nem csak a fehérje feltekeredését kell kijavítani korrektor molekulákkal, hanem a működését is serkenteni kell potenciátor vegyületekkel[45]. Ezzel szemben a G551D mutáció csak a fehérje működését befolyásolja, a feltekerdését és a sejtmembránba történő kijutását nem[46]. Fontos hangsúlyozni, hogy a CFTR fehérje nem aktív transzporter, az ATP kötődéséből és hasításából eredő konformációs változások a klorid-csatorna nyitását és záródását szabályozzák[9,47,48]. Érdekes módon a szabályozásnak több rétege van, s általánosságban elmondható, hogy fiziológiás körülmények között a rendezetlen domén PKA-általi foszforilációja nélkül a csatorna vagy nem, vagy csak kismértékben nyit[49–51].

4.1.1. Mechanizmus-alapú korrektor kombináció a ΔF508-CFTR kijavítására

(Cikk-1: Mechanism-based corrector combination restores ΔF508-CFTR folding and function)

A ΔF508-CFTR hatékony megmentéséhez az irodalomban leírt másodlagos mentő mutációk hatása alapján stabilizálni kell mind az NBD1-et, mind az NBD és TM domének közötti kapcsolatot[52–54]. Laboratóriumi kísérletekben a CFTR fehérje feltekeredését segítő legígéretesebb korrektor vegyület, a VX-809 a ΔF508-CFTR



2. ábra: (a) Vegyületek hatása a ΔF508-CFTR érési hatékonyságára, metabolikus pulse-chase kísérletekkel mérve. B: éretlen core-glikozilált forma, C: érett komplex-glikozilált forma. **(b)** VX-809 molekulával végzett *in silcio* dokkolás azt mutatja, hogy legjobb kötőhelye az NBD1 és a vele kölcsönható TMD csatoló hélixek (CL1 és CL4) között van.

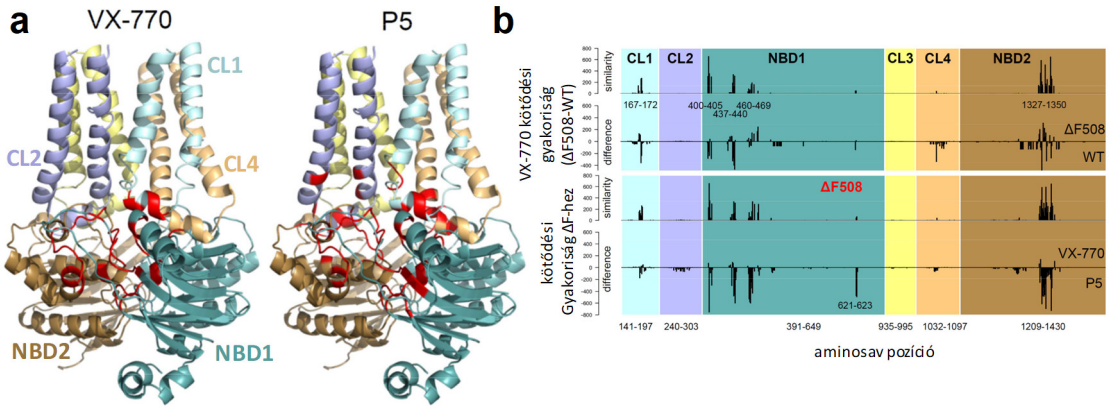
sejtfelszíni mennyiségét a vad típus szintjének kevesebb, mint 15%-ára növeli[55]. Klinikai vizsgálatokban a vegyületet önmagában alkalmazva azonban a betegek állapota nem javult jelentősen és a vegyület hatásmechanizmusa sem volt ismert[55,56]. Ezért megvizsgáltuk, hogy milyen hatással van a fehérjére a VX-809 (2015-ben Lumakaftor néven került forgalomba) és egyéb gyógyszerjelölt molekulák. A számos kísérletből egy metabolikus *pulse-chase* kísérlet eredményeit emeltem ki (2. ábra, a), melynek során a sejteket 20 percig radioaktívan jelölt metionin aminosavat tartalmazó oldatban növesztettük (G. Lukács). Az ez idő alatt termelődött fehérjék radioaktívan jelölődtek, s sorsukat gél elektroforézist követően autoradiográfiával tanulmányoztuk. A ΔF508-CFTR fehérjéből még korrektor vegyületek (C3, C18, és VX-809) hatására sem keletkezett jelentősebb mennyiségű érett forma (*Band C*). Az NBD1 és a TMD2 közötti kapcsolatot helyreálltó szupresszor mutáció (R1070W[54]) jelenlétében sem volt nagy hatása a korrektor molekuláknak a mutáns érésére. Ezzel szemben, ha a korrektorokat az NBD1 stabilitását növelő R1S szupresszor mutációk mellett alkalmaztuk, akkor az érett forma (*Band C*) jelentős növekedését tudtuk megfigyelni. Eredményeink arra utalnak, hogy ezek a korrektor molekulák nem az NBD1-en, hanem az NBD1-TMD1/2 interfészen hatnak. Ezt *in silico* dokkolásaink is alátámasztották (2. ábra, b). Összefoglalva, *in vitro, in vivo* és *in silico* vizsgálatok segítségével a rendelkezésre álló korrektorokat hatásuk helye alapján három csoportba soroltuk. Az I. osztály, beleértve a VX-809 molekulát, az NBD1-TMD1/2 interfészt célozza, a II. osztály az NBD2-t vagy annak az NBD1-vel való kapcsolófelületét, a III. osztály pedig a ΔF508-NBD1-et stabilizálja nem specifikusan. Vizsgálataink idején még csak aspecifikusan ható NBD1 stabilizáló vegyületek (pl. glicerin) voltak ismeretek. Ezért gyakorlati szempontból az egyik legfontosabb üzenetünk az volt, hogy olyan ΔF508-CFTR konstrukcióval kell gyógyszerjelöltek keresését végrehajtani, amelyben az NBD1-TMD1/2 interfész stabilizált (pl. R1070W), ezáltal ki tudjuk válogatni azokat a vegyületeket, amelyek csak az NBD1 stabilitását növelik.

4.1.2. CFTR potenciátorok csökkentik a ΔF508-CFTR funkcionális expresszióját

(Cikk-2: Some gating potentiators, including VX-770, diminish ΔF508-CFTR functional expression)

A CFTR ioncsatorna nyitódását befolyásoló G551D mutációt hordozó CF-betegek kezelése a VX-770 potenciátorral (Ivakaftor) nagymértékben helyreállítja a csatorna aktivitását, terápiás hatása jelentős[44]. A ΔF508-CFTR betegek kezelése a VX-809 (Lumakaftor) korrektorral önmagában nem eredményezte diagnosztikus paraméterek szignifikáns javulását[56]. Mivel ismert, hogy ennek a mutánsnak is sérült a kapuzása, felmerült, hogy a VX-809 korrektor és a

VX-770 potenciátor együttes alkalmazása jobban javíthatja a betegek állapotát[57]. Ezért együttműködő partnereim különböző kísérleti rendszerekben vizsgálták a VX-770 önmagában és korrektorokkal együtt történő alkalmazásának hatását a ΔF508-CFTR fehérje stabilitására. Eredményeink arra utaltak, hogy a VX-770 potenciátor destabilizálja a CFTR fehérjét, s míg ez a destabilizáció eredményezheti a csatorna nyitásának helyreállását, a VX-809 korrektor hatását gátolja. Ezzel szemben a P5 potenciátor hatékonyan növelte a foszforilált CFTR működését viszont nem változtatta meg a fehérje stabilitását. Ez arra utalt, hogy a P5 hatásmechanizmusa különbözik a VX-770 hatásmechanizmusától, ezért hasznos vegyület lehet a ΔF508-CFTR kapuzási hibájának kijavítására. A hatásmechanizmus megértésének érdekében molekuladinamikai szimulációkat és *in silico* dokkolást kombináltunk (3. ábra). Magasabb hőmérsékleten (321 K) végzett MD szimulációkkal részlegesen széttekeredett konformációkat nyertünk. A konformációkat RMSD alapján klasztereztünk és a centroidokhoz külön-külön dokkoltuk a VX-770 és a P5 molekulákat. Számításaink azt mutatták, hogy a P5 kötődik a CL2 régióhoz és egy, az NBD1 β-alegységében, a Walker A hélix közvetlen szomszédságában található hurokhoz, míg a VX-770 nem lép kölcsönhatásba ezekkel a régiókkal. A fő különbség a β-alegységhez történő kötődésben van, s ez a megfigyelés felhasználható új potenciátorok fejlesztéséhez. Tanulmányunk fő üzenete, hogy a VX-770 a funkciót a fehérje destabilizálásán keresztül javítja, ami viszont ellentétes hatású a feltekeredésében és stabilitásában sérült ΔF508-CFTR fehérje korrektorokkal történő megmentésével. Ezért szükséges a meglevő



3. ábra: (a) Az *in silico* dokkolás eredményeinek összefoglalása a CFTR intracelluláris régiójának szerkezetén. A VX-770 és P5 kötőhelyeket piros szín jelöli. Türkiz és lila: első és második intracelluláris hurok (CL1 és CL2), sötét türkiz: NBD1, sárga és narancs: harmadik és negyedik intracelluláris hurok (CL3 és CL4), barna: NBD2. **(b)** Felső sor: A VX-770 ΔF508 és WT fehérjékhez történő kötődésének összehasonlítása kvantitatív módon. Adott aminosavval mutatott kölcsönhatási gyakoriságban megmutatkozó hasonlóságot és a különbséget a következőképpen számoltuk: $((f_{\Delta F508}+f_{WT})\text{-abs}(f_{\Delta F508}-f_{WT}))/2$ és $f_{\Delta F508}-f_{WT}$. A két fehérje kötőhelyei átfednek. Alsó sor: A VX-770 és P5 ΔF508-hoz történő kötődését is így hasonlítottuk össze. Bár a P5 kötőhelyei jelentősen hasonlítanak a VX-770 kötőhelyeihez (magas csúcsok a hasonlósági grafikonon), azonban jelentős különbségek is megfigyelhető, mint pl. 621-623 régió a különbség grafikonon.

potenciátorok optimalizálása vagy új potenciátorok fejlesztése kombinált terápiák hatékony alkalmazásához.

4.1.3. A CFTR klorid-csatornájának azonosítása

(Cikk-3: Discovering the chloride pathway in the CFTR channel)

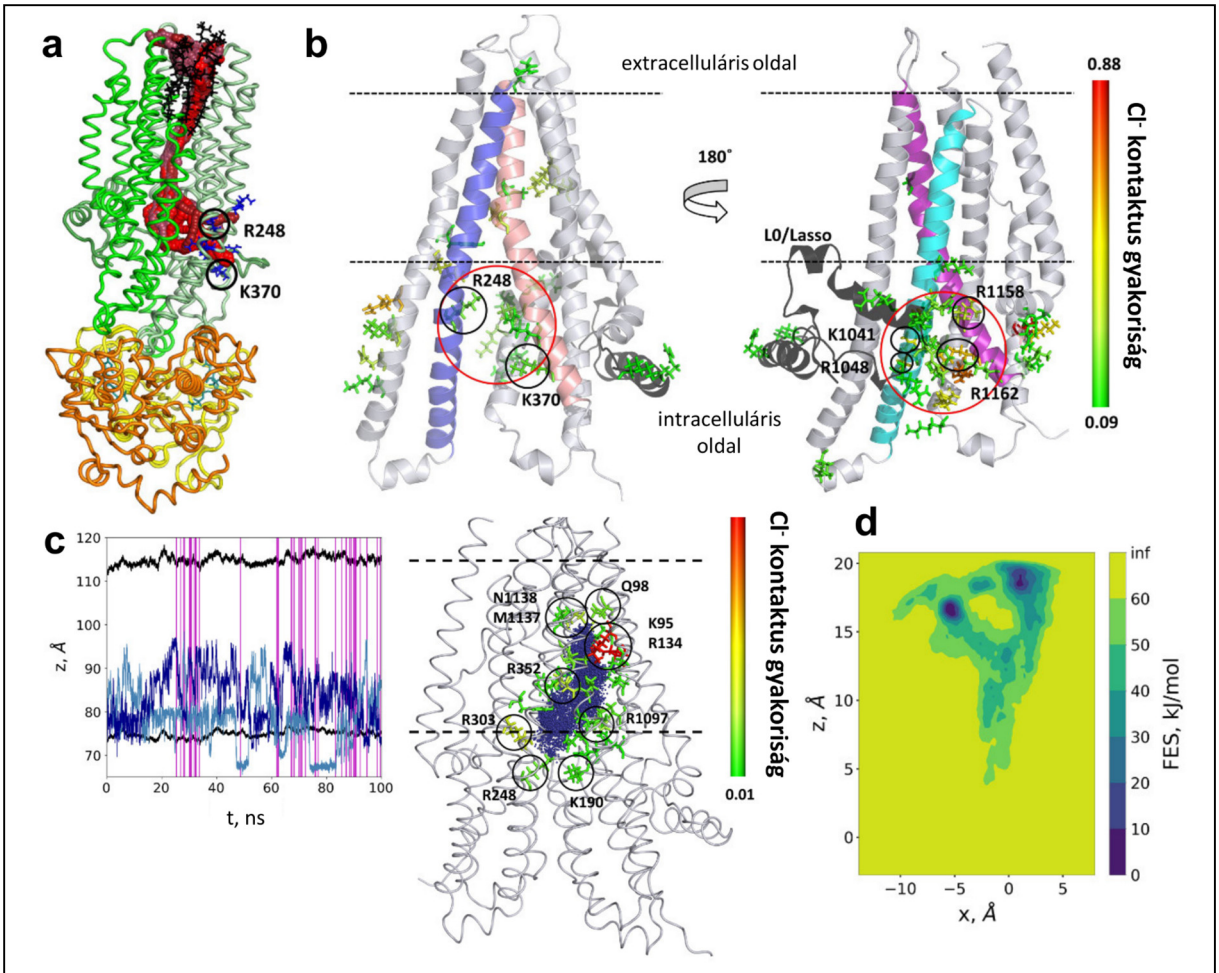
A CFTR szerkezetének meghatározása aktív állapotban, foszforilált R doménnel és ATP jelenlétében elvileg lehetővé tette a kloridvezetés szerkezeti hátterének vizsgálatát[58]. Azonban a foszforilált és ATP-kötött zebrahal zfCFTR szerkezetének krio-EM szerkezete (PDBID: 5w81) nem mutat klorid-vezetésre alkalmas útvonalat[58]. Annak érdekében, hogy a klorid-vezetéshez szükséges geometriával, azaz 3,6 Å átmérőnél nagyobb csatornával rendelkező CFTR-konformációkat tudjunk azonosítani, a zfCFTR szerkezettel egyensúlyi molekuladinamikai (MD) szimulációkat végeztünk. A szimulációkban azonosítottuk azokat a konformációkat, amelyek nyitott transzmembrán régióval rendelkeztek. A 22 egyensúlyi szimuláció közül csak egyetlen mutatott nyitott csatornákat tartalmazó konformációkat is (54/10000 konformáció). Az azonosított útvonalak a TM-régióban nagymértékben hasonlítottak, csak az intracelluláris részekben tértek el egymástól (4. ábra, a). Az intracelluláris oldalon két nyílást azonosítottunk, amelyeket pozitív töltésű aminosavak vesznek körül és a vezetésben betöltött szerepüket kísérletek is alátámasztják[59]. A szimulációinkban azonosított csatornabélelő aminosavak többségéről szintén kísérletes adatok mutatják, hogy részt vesznek a csatorna képzésében[59–61].

A vezetésben fontos szerepet játszó aminosavak azonosítása. Az ionok az MD szimulációkban sokkal mozgékonyabbak, mint a nagyméretű fehérjék. Ezért a szimulációkban a kloridionok aminosavakkal való érintkezései értékes információt szolgáltatnak a kloridionok kölcsönhatási helyeiről. Ezért először azonosítottuk az kloridionok és a fehérje kontaktusait a leghosszabb (100 ns hosszú) egyensúlyi szimulációinkban (n=6). Az egyes aminosavak kloridionokkal való érintkezését ($d < 4$ Å) a szerkezetre vetítettük, és a kontaktus gyakorisága szerint színeztük az oldalláncokat (4. ábra, b). Intenzív kölcsönhatások voltak megfigyelhetők a TM10 és TM12 által meghatározott belépési helyen és a TM4 és TM6 közötti nyílásnál, amelyek megfeleltek a szimulációkban azonosított intracelluláris pórusoknak (4. ábra, a).

Az egyensúlyba hozás során a kloridionok és a vízmolekulák kitöltötték a fehérje belső üregét. A nyitott konformációkat tartalmazó trajektóriában a csatorna belsejében két kloridion volt megfigyelhető, amelyek a membrán normál mentén léptek be a fehérje belső terébe. A várakozásoknak megfelelően a kloridionok taszították egymást, és nem maradtak egymás közelében (4. ábra, c). Továbbá, ezek az ionok nem haladtak át egy szűk keresztmetszeti régión,

amely z = 95 Å körül kezdődik. Feltehetően a nyitott konformációk alacsony valószínűsége miatt egyik kloridion sem volt megfelelő helyzetben a nyitás pillanatában, ezért nem tudtunk ionátmenetet megfigyelni. Ezen kloridionok és a pórusképző aminosavak kölcsönhatási gyakoriságát is kiszámoltuk (4. ábra, c). Megfigyeltük, hogy a kloridionok több időt töltöttek a pozitív töltésű oldalláncok, nevezetesen a K95, R134, K190, R248 (K a zfCFTR-ben), R303, R352 és R1097 közelében, amelyek a nagy belső üreget illetve az intracelluláris nyílást bélelik[59–61].

Amikor összehasonlítottuk a nyitott és zárt konformációk kontaktus térképeit (cikk, S7. ábra), megfigyelhető volt, hogy a csatorna extracelluláris nyílásánál (a TM 1, 6 és 12 között) elhelyezkedő I344 (M a zfCFTR-ben) és az N1138 (L a zfCFTR-ben) aminosavak elzárták a klorid útvonalat. Mutagenezises kísérletek alapján ezt az extracelluláris pórusnyílást körülvevő aminosavak, nevezetesen az F337, S341 és L102 is befolyásolják a csatorna kapuzását[62,63].



4. ábra: (a) Az MD szimulációban megfigyelt nyitott csatornák egy szerkezetre vetítve (piros golyók). Zöld és türkiz: TMD1/2, sárga és narancs: NBD1/2, fekete: lipid. **(b)** Egyensúlyi szimulációkban megfigyelhető kloridion és aminosav kölcsönhatási helyek. Piros kör: belépési hely; fekete kör: fontosabb kölcsönható aminosavak. **(c)** A fehérje belsejébe bejutó két kloridion nem jut át a kb. 100 Å-nál található szűkületen. Kék vonalak: kloridionok pozíciója, lila függőleges vonalak: csatornanyitás, fekete vonalak: membrán kettősréteg szélei. Jobb oldalt: az egyik kloridion eloszlása és a vele kölcsönható aminosavak a szimuláció során. **(d)** Metadinamikai szimuláció segítségével számoltuk ki a kloridion átjutását jellemző szabadenergia-felületet.

Összehasonlításunk olyan régiókat mutatott ki, amelyek nyitott és zárt állapotban eltérő kontaktusokkal rendelkeznek és így alloszterikus kommunikációs helyként a konformációs változások továbbításában lehet szerepük. Ennek vizsgálatára az alloszterikus útvonalakat gráfelméleti módszerek alkalmazásával jellemeztük[64] (Cikk-3, 6. ábra). Érdekes módon az extracelluláris régió aminosavainak az F508-cal kölcsönható CL4 régióval volt legerősebb az allosztérikus kapcsolata.

Metadinamikai szimulációk. Mivel a hagyományos MD szimulációink során a kloridionok áthaladását a csatornán nem tudtuk megfigyelni, metadinamikai számításokat végeztünk a szűk keresztmetszetű területen átvezető útvonal feltárására és az átjutás potenciálfelületének számítására[65]. A metadinamika megkönnyíti a lokális minimumokból való kilépést azáltal, hogy az egyes reakciókoordinátákon (kollektív változók, CV) előzményfüggő Gauss-potenciálokat halmoz fel, ami eltéríti a szimulációt az energiavölgyből[65]. A trajektóriából kiválasztottuk azt a konformációt, amelyben az egyik klorid a legközelebb volt a szűk keresztmetszeti régióhoz. Ezt a konformációt alkalmaztuk egy hosszú, jól-temperált metadinamikai szimulációban, ahol az eltérítés a kloridion és négy szomszédos Cα atom tömegközéppontja közötti távolságán alapult. A 2D szabadenergia felületeket x/z mentén számoltuk ki a távolság-CV x, y és z komponenseinek felhasználásával (4. ábra, d). Ebből az x/z oldalnézeti vetületből az derült ki, hogy az útvonal a szűk keresztmetszeti régió után kettévált, s ezeken az útvonalakon nem volt jelentős energetikai akadály az ion távozásához.

Ionátjutás kiegyenesített TM8 esetén. A TM8 hélixben található törés a legtöbb CFTR szerkezetben jelen van (PDB ID: 5uak, 6o2p, 6msm, 6o1v, 5uar, 5w81)[58,66–68], s a kutatók azt gondolják, hogy ez okozza a CFTR csatorna egyediségét az ABC család aktív transzportereihez képest. A csirke CFTR szerkezetben azonban nincs ilyen hélix-törés (PDB ID: 6d3s és 6d3r)[69] és hasonló konformációt más ABC szerkezetekben sem észleltek. Annak érdekében, hogy megvizsgáljuk ennek a törésnek a jelentőségét, az 5w81 szerkezetben a TM7 és TM8 hélixeket az MRP1 szerkezete (PDB ID: 5uj9[70]) alapján homológiamodellezéssel megváltoztattuk, azaz a TM8 hélixet kiegyenesítettük (Cikk-9, 5. ábra). Ezzel a korrigált szerkezettel és a korábban használt paraméterekkel futtattunk egyensúlyi szimulációkat. A hat szimulációból ötben tudtunk megfigyelni nyitott konformációt, az 5w81 szerkezettel végzett egyensúlyi szimulációban megfigyelt 0,004% gyakorisághoz képest jóval magasabb, 3,74%-os valószínűséggel. Ezek a számításaink arra utalnak, hogy a CFTR fehérje TM8 hélixének törése a fluorinált Fos-Choline-8 vegyület alkalmazásából fakadó szerkezetmeghatározási műtermék lehet[66], amelynek kialakulásához jelentősen hozzájárulhat a TM8 nagy mozgékonysága is[71].

4.1.4. Összefoglalás

Eredményeink hozzájárultak egy olyan módszer fejlesztéséhez, amely hatékonyabb CFTR korrektor molekulák szűrését teszi lehetővé, s olyan atomi szintű adatokat szolgáltattak, amelyek CF elleni gyógyszermolekulák hatásmechanizmusának megismeréséhez járulnak hozzá. Továbbá atomi szinten azonosítottuk és jellemeztük a kloridionok legvalószínűbb átjutási útvonalát.

4.2. Az ABCG2 fehérje szerkezete, dinamikája, és működése

A humán ABCG2 fehérje egyik legfontosabb feladata szervezetünkben a mérgező anyagokkal szembeni védelem. Xenobiotikumokat pumpál ki a sejtekből a májban, a tejmirigyekben és a placentában vissza az anya szervezetébe, a vér-agy gátban vissza a véráramba, valamint őssejtek védelmét is ellátja fejlődésük bizonyos szakaszában[5,72,73]. Mivel a húgysav kiválasztásában is részt vesz, ezért mutációi hajlamosíthatnak köszvényre[74]. Mindezek miatt működésének és szubsztrát-specifitásának megismerése igen fontos, amihez nélkülözhetetlen atomi szintű szerkezetének ismerete is. A Pgp-szerű ABC transzporterek szerkezetének megismerésétől (2006.) tíz évet kellett várni a humán féltranszportek ABCG2-szerű szerkezetének meghatározásáig[75,76].

4.2.1. Az ABCG2 fold üzenete homológia modellezés alapján

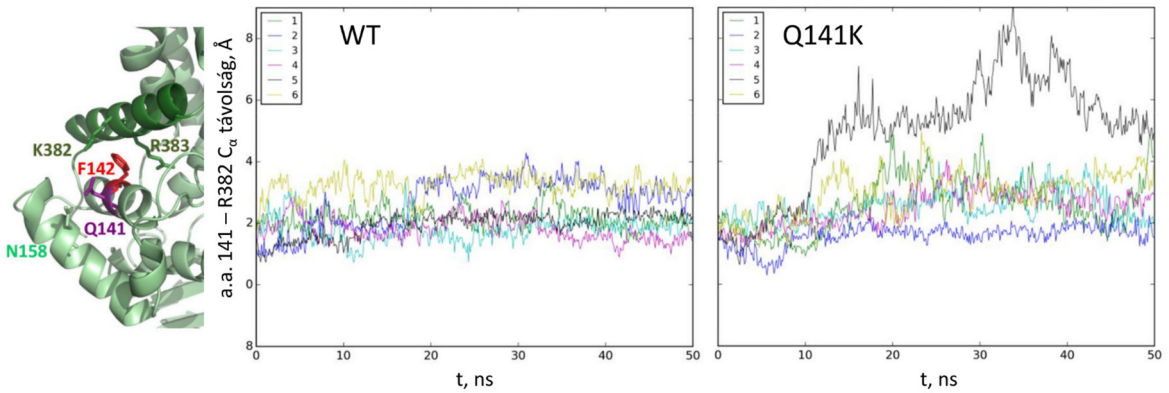
(Cikk-4: Jump into a new fold - A homology based model for the ABCG2/BCRP multidrug transporter)

Az ABCG fehérjék szekvenciájuk alapján külön alcsaládot alkotnak az ABC fehérjék családján belül. Ebben az alcsaládban a transzmembrán régiót alkotó aminosavak száma jelentősen kevesebb a többi alcsalád TM aminosavainak számához képest. Ez alapján 2010-ben már javasoltunk egy olyan ABCG2 szerkezeti modellt (ABC meeting 2010, Innsbruck, Ausztria), amelyben az TM domének intracelluláris régiói rövidek, s ezáltal az NBD-k a membrán kettősréteghez közel helyezkednek el. Ez a szerkezet hasonlít számos bakteriális ABC importer szerkezetére[77,78]. Bár a fehérje szekvenciális tulajdonságai kizártak minden más lehetőséget, pusztán ennek figyelembevételével a szerkezetet nem lehetett megbízhatóan modellezni. Ezért használtak a korai ABCG2 homológia modellek templátjaként egy Pgp-szerű szerkezettel rendelkező bakteriális exportert (Sav1866)[79,80]. 2016 tavaszán vált elérhetővé az első ABCG alcsaládba tartozó transzporter, az ABCG5/ABCG8 heterodimer kristályszerkezete[76], melynek szekvenciája már kellően konzervatív az alcsalád többi tagjának a modellezéséhez.

A gondos, ClustalW-vel[81] végzett és módosított szekvenciaillesztés után száz ABCG2 homológia modellt hoztunk létre a Modeller[26] segítségével. Ezekből a DOPA pontszám alapján kiválasztottuk a legmegfelelőbb szerkezetet, amit lipid kettősrétegbe illesztettünk, energiaminimalizáltunk, és MD szimulációkkal egyensúlyba hoztunk (Cikk-4, Módszerek). Ezt az optimalizált szerkezetet használtuk, amíg az ABCG2 kísérletes szerkezete nem vált elérhetővé. A homológia modellezés során felmerült a kérdés, hogy mennyire szolgálhat egy heterodimer jó templátként egy homodimer fehérjéhez. Egyrészt az ABCG2 szekvencia hasonlósága magas mind az ABCG5 mind az ABCG8 szekvenciájával (teljes egyezés: 27% és 26%, hasonlóság: 48% és 44%, amely értékek ABC transzmembrán fehérjék esetén megfelelőnek számítanak homológia modell építéséhez), és a heterodimerben az ABCG5 és az ABCG8 szerkezete igen hasonló egymáshoz. Másrészt fiziológiás körülmények között egy homodimerben található protomer szerkezetek sem egyeznek meg tökéletesen az idő minden pillanatában. Az MD szimulációval és dokkolással kapott eredményeink magasfokú értelmezhetősége is arra utalt, hogy homológia modellünk jó pontosságú. Ez az első kísérletes szerkezet megjelenése után be is bizonyosodott, ugyanis a szerzők kiszámították a modellünk és az általuk meghatározott szerkezet közötti különbséget, és megállapították, hogy az eltérés minimális (RMSD = 2Å)[82]. A két szerkezet közötti lényegesebb különbség az egyik TM hélix egy aminosavval történő elcsúszása, illetve az alacsony szekvenciális hasonlóság miatt nehezen modellezhető extracelluláris hurkok eltérő konformációja volt.

A homológia és később a kísérletes szerkezeti modell alapján, esetenként MD szimulációkkal, számos mutáció hatását próbáltuk értelmezni[83–85], mint például a V12M, K360del, K86M, E211Q és a Q141K variáció. A Q141K variáns, amely bizonyos népcsoportokban akár 15-30 %-os gyakorisággal is előfordul, gátolja a fehérje feltekeredését és csökkenti plazmamembrán lokalizációját[86,87]. Az ABCG2 szekvenciában szomszédos F142 aminosav analóg helyen található a CFTR F508 aminosavával. Modellünkben az F142 oldallánca a transzmembrán domén felé néz, ezért a CFTR F508 oldallánc fehérjében betöltött szerepe és a pozíció konzerváltsága alapján azt gondoltuk, hogy ez is az NBD és a TMD közötti kapcsolatot stabilizálja. Feltételezésünk szerint a Q141K csere viszont intradomén kölcsönhatásokat tesz tönkre, specifikusan a Q141 és N158 közötti kölcsönhatást (5. ábra).

Szerkezeti modellünk felvetette a lehetőségét annak is, hogy a Q141K is hozzájárulhat az NBD/TMD kapcsolat gyengítéséhez. Ennek az az alapja, hogy ez a szerkezeti interfész eltér a CFTR hasonló régiójának a szerkezetétől. A CFTR esetén az F508 hidrofób kölcsönhatásokat alakít ki, míg az ABCG2 esetén az F142 aminosavat a transzmembrán régió összekapcsoló-hélixében található két pozitív töltésű aminosav (K382, K383) fogja közre. Ennek további



5. ábra: Az ABCG2 NBD és TMD interfész szerkezete. Az F142-t két pozitívan töltött aminosav fogja közre (K382, K383). A Q142 kölcsönhatásban van az N158-val. A Q142K mutáció elsősorban a 142. és 382. pozícióban okoz változást. A két pozícióban található aminosav C_α atomjának a távolsága a Q141K mutánsban sokkal jobban ingadozik, mint a vad típusban (WT), amit hat-hat független MD szimuláció trajektóriából számolt adatok mutatnak.

vizsgálatára molekuladinamikai szimulációkat végeztünk, amelyekben a teljes hosszúságú vad típusú és a Q141K mutáns fehérjék szerkezetének mozgását vizsgáltuk. A szimulációk során egyik modellben sem változott a 141. és 158. pozíció közötti intradomén távolság (5. ábra és Cikk-4, S5. ábra). Ezzel szemben a K141 és K382 aminosavak távolsága erős ingadozást mutatott (5. ábra), ami arra utal, hogy a szerkezetben bekövetkező mozgások miatt a két lizin egymáshoz közel kerülhet annyira, hogy taszító pozitív-pozitív kölcsönhatások érvényesüljenek, amelyek gyengítik az NBD/TMD kapcsolatot.

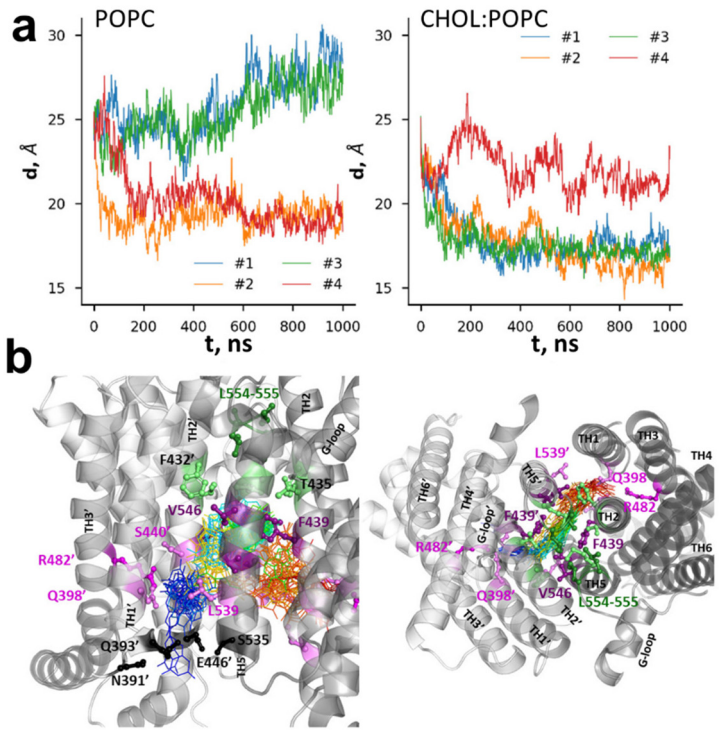
4.2.2. Az ABCG2 fehérje szabályozása és transzport működésének jellemzése

(Cikk-5: The transport pathway in the ABCG2 protein and its regulation revealed by molecular dynamics simulations)

A rendkívül fontos kísérletes ABCG2 szerkezetek adják a szerkezeti vizsgálatok alapját[20,82,88,89], azonban természetüknél fogva nem tartalmaznak információt a fehérje dinamikájáról, azaz a transzport leírására csak limitáltan alkalmazhatók. A transzport folyamatot mind biokémiai, mind számításos módszerekkel már sokat vizsgálták[90–94], azonban az összes kísérleti és a legtöbb *in silico* vizsgálat csak a transzlokációs útvonal egy kisebb részére összpontosított, illetve az ABCG2-vel végzett molekuladinamikai szimulációk nem tartalmaztak szubsztrátmolekulát. Ezért mi olyan MD-szimulációkat végeztünk a lipid kettősrétegbe ágyazott ABCG2 fehérjével, amelynek során egy fiziológiás szubsztrát, a húgysav is jelen volt. Célunk a teljes transzlokációs útvonal jellemzése volt, kiemelve a központi szubsztrátkötő zseb dinamikus változását és az ehhez a kötőhelyhez vezető hozzáférési útvonalak fontos szerepét.

A koleszterinmolekulák hatása a fehérje dinamikájára. Korábban kísérletesen bemutatták, hogy az ABCG2 működését a koleszterin jelentősen befolyásolja[95,96]. Ezt együttműködés keretében mi is vizsgáltuk, s a hatás atomi szintű értelmezéséhez különböző lipidösszetétű membránba ágyazott ABCG2 transzporterrel végeztünk MD szimulációkat. A biológiai membránok *in vivo* koleszterin tartalma nagymértékben függ a sejttípustól és az organellumtól, kb. 20% és 50% között változik[97]. Annak érdekében, hogy az ABCG2 koleszterinkötő helyeit szimulációinkban telítsük, s elkerüljük az *all-atom* szimulációkban a diffúzió által okozott korlátokat, a CHARMM-GUI segítségével olyan szimulációs rendszereket építettünk, amelyek 50%-50% POPC-koleszterin, 50%-50% POPC-szitoszterin vagy 100% POPC molekulát tartalmaztak[33,34,97]. A POPC-szitoszterint tartalmó rendszert kontrollként használtunk, mivel ez a szterin más hatással van az ABCG2 működésére mint a koleszterin[95,96]. A kettősrétegben a szterinmolekulák kiindulási helyzete véletlenszerű volt, s a lipidek egyenletesen oszlottak el. Fontos megjegyezni, hogy mindkét szterol hasonló hatással volt a kettősréteg fizikai tulajdonságaira, például a vastagságra és a lipid-sűrűségre (Cikk-5, S3. ábra). Ezekben a szimulációkban a befelé néző (*inward-facing*) ABCG2 szerkezetet (PDB ID: 6hco) használtuk[98].

Várható volt, hogy a lipidösszetétel a TM hélixek mozgására lesz hatással. Ezért a TM-hélixek határain lévő aminosavak Cα atomjainak helyzetét összegyűjtöttük a szimulációk trajektórijáiból és 2D eloszlást számoltunk az x, y komponensekből, amit hisztogramon ábrázoltunk. Ezt elvégeztük külön-külön az intracelluláris és extracelluláris végek eloszlására. Érdekes módon, koleszterin jelenlétében a hélixek alsó része a központi tengely felé mozdult el és



6. ábra: (a) A TM2 és TM2' intracelluláris végei (a.a. 450) közötti távolság változása 4-4 szimuláció során. **(b)** Egy húgysav (szubsztrát) molekula mozgása rávetítve a kiindulási fehérjeszerkezetre (oldal- és felülnézeti kép). Kék-zöld-sárga-piros színek az idő haladását jelzik, miközben a szubsztrát az 2. kötőzsebből (R482') átkerül a központi kötőzsebbe (sötétlila aminosavak) majd a túloldali 2. zsebbe (R482).

mozgásuk csökkent mértéket mutatott. A szitoszterin is csökkentette a hélix-mozgásokat az egyik szimulációban, azonban a hélixek nem mozdultak a központ felé. A koleszterin legjellemzőbb hatása a TM2 és a TM2' zárása volt, amelyet e hélixek intracelluláris végei közötti távolság csökkenése jellemzett (mindkét protomer V450 Cα atomjai között számolva, 6. ábra, a). Tiszta POPC kettősrétegbe helyezve az ABCG2-t a TM2 és a TM2' vagy egyáltalán nem mutatott záródást, vagy a végek csak kisebb mértékben közeledtek egymáshoz, mint a koleszterin jelenlétében végzett szimulációkban. A szubsztrát-hozzáférést és a kötőzseb térfogatát a hélixek dinamikus fluktuációi befolyásolják, amit az MDpocket szoftverrel vizsgáltunk[99]. A 0,75 vagy annál nagyobb gyakorisággal jelen lévő üregeket elemeztük, és a kiindulási szerkezet zsebeihez hasonlítottuk (Cikk-5: 3. ábra, a-c). Két szimulációban egy nagy üreget figyeltünk meg, amely a 2. zsebet, a 2.' zsebet és a 3. centrális üreget foglalta magába. Ez a komplex üreg a membrán kettősrétege felé (Cikk-5: 3. ábra, b) és a citoszol felé (Cikk-5: 3. ábra, c) egyaránt nyitott volt. A TM2 citoszol felőli záródása azonban megakadályozta az oldalsó nyitást (Cikk-5: 3. ábra, d), és a központi üreg (Site 3) citoszol felőli hozzáférését is korlátozta (Cikk-5: 3. ábra, e).

A transzlokációs útvonal vizsgálata egyensúlyi szimulációkkal. A szubsztrát-transzlokáció részleteinek azonosítására egyensúlyi szimulációkat végeztünk, a befelé nyitott ABCG2 szerkezetet (PDBID: 6hco)[98] POPC/koleszterin kettősrétegbe ágyaztuk húgysav jelenlétében, ami az ABCG2 fiziológiás szubsztrátja[74,100]. Mivel szimulációs időskálánk viszonylag rövid volt (0,5-1 μs), a szubsztrátkötés és transzlokáció valószínűségét úgy növeltük, hogy 30 db húgysavmolekulát helyeztünk a szimulációs dobozba. A 0,5 μs hosszúságú szimulációk egyike során a TM-hélixek citoplazmatikus végei közötti távolság olyan mértékben csökkent, hogy eltűnt a szubsztrát belépési pontja. Egy másik 0,5 μs hosszú szimuláció során a 30-ból egy húgysavmolekula bejutott a 2.' helyre (a B lánc TM1, TM2 és TM3 közötti zsebébe). A belépés a részben szeparált NBD-k közötti térből történt, s a belépési pontot az N391', Q393', E446' és S535 aminosavak határolták (6. ábra, b). A 2. ' zsebben a szubsztrát a Q398', S440', S443', R482' és L539 aminosavakkal lépett kölcsönhatásba a legnagyobb gyakorisággal, majd tovább haladt az extracelluláris tér felé a TM2, TM2', TM5 és TM5' hélix közötti 3. zsebbe. Itt a leggyakoribb szubsztrátinterakciós partnerek közé az F439, T542 és a V546 aminosav tartozott a dimer mindkét feléből, valamint kisebb gyakorisággal az N436 is. A húgysavmolekula az extracelluláris tér felé is tovább haladt, de a Leu-szelepet (Leu 554 és Leu 555 aminosavakat) nem érte el. Végül a húgysav a mozgási irányát az intracelluláris tér felé változtatta, és a belépési ponttal ellentétes oldalon érte el a 2. kötőzsebet. A leggyakoribb kölcsönhatások

(Q398, S440, S443, R482 és L539') ugyanazoknak az aminosavaknak feleltek meg, amelyeket a szemközti 2.' zsebben is megfigyeltünk.

A Leu-szelepen történő átjutás jellemzése metadinamikával. Várható volt, hogy a húgysav nem jut a külső térbe, mivel a befelé-nyitott konformáció a záródást mutató mozgások után még mindig nyitottabb az ismert ATP-kötött, zárt konformációkhoz képest. Ezért az ATP-kötött, befelé-zárt ABCG2 szerkezettel is indítottunk egyensúlyi szimulációkat úgy, hogy a húgysav molekula a Leu-szelep közelében helyezkedett el (Cikk-5: S7. ábra). Átjutást így sem tudtunk megfigyelni, ezért a szubsztrát extracelluláris térbe történő lépését gyorsított szimulációkkal, metadinamikával vizsgáltuk, amellyel az átjutás energetikájának leírására is lehetőség nyílik[101,102].

A metadinamikai MD szimulációk során, ha a szubsztrát hosszabb ideig tartózkodott egy helyen, akkor a potenciálisenergia-felületet úgy változtattuk meg, hogy az extracelluláris irányba történő mozgás esetén csökkenjen a rendszer energiája. Mivel a hosszabb metadinamikai szimulációkban nem sikerült konvergenciát elérnünk, mert a húgysav nem találta meg a visszavezető utat a központi zsebbe, az átjutás energetikáját nem tudtuk pontosan leírni. Ezért több rövid szimulációt végeztünk el addig a pontig, amikor a húgysav áthaladt a Leu-szelepen. A szubsztrát kölcsönhatásba lépett az egyik protomer N601' és P602' és a másik protomer N604 és Y605 aminosavaival (Cikk-5: 5. ábra). A kijárat körüli aminosavak közé tartozott az 556-559, 616-618, valamint az ellentétes TM2 extracelluláris vége (T421 és N425). Annak érdekében, hogy betekintést nyerjünk a szubsztrát áthaladásának energetikájába, minden egyes szimulációban kiszámoltuk a rendszerbe pumpált energiát addig az időpontig, amikor a húgysavmolekula áthaladt a Leu-szelepen (Cikk-5: 5. ábra). A hat szimulációból háromban ez a szabadenergiabevitel 7-13 kcal/mol között volt, ami összevethető az ATP-molekula hidrolíziséből felszabaduló energiával. Fontos megjegyeznünk, hogy MD szimulációkban nem hasíthatunk ATP-t, ezért használtunk metadinamikát energiabevitelre. Továbbá ez nem azt jelenti, hogy egy szubsztrát átjutásához egy ATP molekula hasítására van szükség, hanem azt mutatja, hogy mivel az energia mértéke összemérhető volt az ATP-hasításának energiájával, a zárt konformáció, amit egyébként sokan a transzport lépés utáni konformációnak gondolnak, alkalmas volt a transzport vizsgálatához.

4.2.3. Összefoglalás

A homológia modellel végzett vizsgálataink és kísérleteink arra utaltak, hogy az ABCG2 NBD/TMD interfész szerkezete más mint a CFTR megfelelő régiójának szerkezete. Azaz a Q141K, köszvényt is okozó variáns megmentésére a CFTR megmentésére alkalmazott

módszerektől eltérő megközelítést érdemes alkalmazni. A homológia modellhez történő dokkolásokkal azonosítottuk egy fiziológiás szubsztrát feltételezett átjutási útvonalát. Mivel a dokkolás ilyen esetekben kevésbé megbízható, az útvonal megerősítésére immár kísérletes szerkezetekkel MD szimulációkat végeztünk, amelyek áttörést jelenthetnek a szubsztrát-transzport folyamatának és koleszterin általi szabályozásának megértésében. Bioinformatikai elemzésünk a központi kötőhelyen kívüli drogkötő zsebeket tárt fel (2. zseb), továbbá egy dinamikus transzportútvonalat, amely alkalmas eltérő szerkezetű szubsztrátok szállítására. Valószínűsíthető továbbá, hogy a membrán koleszterintartalma a fehérje citoplazmatikus régióinak zárását segíti elő, így gyorsítva a transzportot.

4.3. Membrán (ABC) fehérjékhez kapcsolódó (web) alkalmazásaink

4.3.1. Adatbázis ABC fehérjék mutációinak összehasonlító elemzéséhez

(Cikk-6: ABCMdb: a database for the comparative analysis of protein mutations in ABC transporters, and a potential framework for a general application; http://abcm2.hegelab.org)

Az ABC fehérjék mutációi számos betegség kialakulásában játszanak szerepet[5,8,85]. Betegségekkel összefüggő mutációk tanulmányozása mellett számos *in vitro* mutagenezises vizsgálat található az irodalomban, amelyben egyes aminosavak szubsztrátfelismerésben, domének közötti kommunikációban, ATP-kötésben és hidrolízisben betöltött szerepét vizsgálták. Egyes aminosavak cseréje megkönnyítheti teljes ABC-transzporterek vagy azok egyes doménjeinek fehérjeexpresszióját és szerkezetének meghatározását is. Mutagenezises vizsgálatok tervezésekor fontos előre tudni, hogy hoztak-e már létre hasonló konstrukciókat, vagy leírtak-e már olyan mutánsokat, amelyek ugyanazokat a vizsgálni kívánt régiókat vagy kölcsönhatásokat befolyásolják. Ha egy ABC-fehérjében egy aminosav szerepét már részletesen jellemezték, akkor a homológia és az aminosav-konzerváltság figyelembevételével, a bioinformatika egyik paradigmája alapján, következtetni lehet más ABC-transzporterekben a homológ helyen levő aminosav szerepére is[103,104].

Az irodalomban leírt mutációk elérésének megkönnyítésére létrehoztuk az ABC-fehérje mutációs adatbázist (ABCMdb, ABCM2), amelyhez a mutációkat publikációk teljes szövegéből (pdf) nyertünk ki automatikus szövegbányászattal. Minden egyes adott emberi ABC fehérje általánosan használt neveivel a PubMed adatbázisban keresést végeztünk, a cikkek teljes szövegét tartalmazó pdf fájlokat a *pdfetch* programmal (https://code.google.com/archive/p/pdfetch) automatikus módon letöltöttük. A mutációk azonosítására létrehoztunk egy *MutationFinder*-alapú[105] munkafolyamatot, amelynek a

bemenete egy pdf file, a kimenete pedig tartalmazza a mutációt és azokat a mondatokat, amelyekben a mutáció említésre került. Az ABCC6 és a CFTR fehérje automatikusan azonosított mutációit összehasonlítottuk ezen fehérjék publikus, betegek szekvenálási adatait tartalmazó mutációs adatbázisaiban található bejegyzésekkel. Az ABCC6 fehérje mutációi okozzák a *pseudoxanthoma elasticum* (PXE) betegséget[7], amelynek során kalcium halmozódik fel elasztikus rostokban, s ez kóros elváltozást idéz elő elsősorban a bőrben, a szemben és a keringési rendszerben. Tanulmányunk írásakor a PXE adatbázis 183 *missense* és *nonsense* ABCC6 mutációt tartalmazott, melyeknek 87%-át megtaláltuk, továbbá 48 olyan egyedi mutációt is kinyertünk, amely a PXE adatbázisában nem szerepelt. A CFTR1 adatbázis (http://www.genet.sickkids.on.ca) 900 *missense* és *nonsense* mutációjából 562 darabot nyertünk vissza cikkekből, a fennmaradó 338 eset vizsgálatából pedig arra következtettünk, hogy ezek többsége nem publikált klinikai vizsgálatból származik. Az általunk használt *MutFinder*-alapú módszer 791 olyan mutációt is talált, amelyek nem voltak jelen ezekben az adatbázisokban, s a mutáns fehérjéket *in vivo* vagy *in vitro* módon vizsgáló publikációkból származtak.

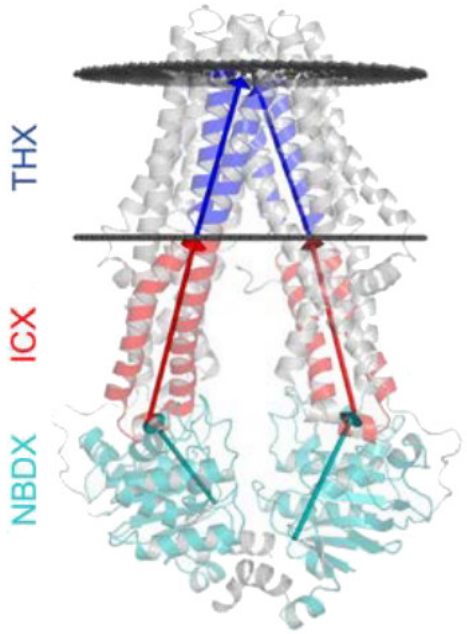
Web alkalmazásunkban nem csak az adott humán ABC fehérje mutációi kereshetők, hanem szekvencia-illesztések használatával, lehetővé tettük a különböző fehérjék szekvenciálisan homológ pozícióiban lévő variánsainak összehasonlítását is. Szintén lehetőség van egyszerű klikkeléssel a mutációk megjelenítésére a fehérjék 3D szerkezetén. Bár kéziratunk idézettségén nem látszik, a Google Analytics web szolgáltatás és személyes visszajelzések alapján sokan használják alkalmazásunkat, ezért 2017-ben jelentős fejlesztéseket hajtottunk végre rajta (pl. DNS-szintű adatok hozzáadása, mutációk hatásának beillesztése SNAP2 és PROVEAN prediktorok alapján[106,107]). Adatbázisunk fejlesztéséhez és újabb funkciókkal való bővítéséhez a web technológiák jelentős átalakulása miatt az egész keretrendszert újra kell írnunk. Tartalmi bővítéséhez az eddig megvalósult legfontosabb rész kísérletes ABC fehérje szerkezetek összegyűjtése és a humán ABC fehérjék szerkezetének generálása AF2-vel (Cikk-9 és http://abc3d.hegelab.org[24]).

4.3.2. Pgp-szerű ABC fehérjeszerkezetek kvantitatív összehasonlítása konftorokkal

(Cikk-7: Quantitative comparison of ABC membrane protein type I exporter structures in a standardized way; http://conftors.hegelab.org)

Egyre több ABC membránfehérje szerkezetét határozzák meg krio-elektronmikroszkópiával. A kapcsolódó tanulmányokban az új szerkezeteket összehasonlítják korábbi szerkezetekkel, de ezek az összehasonlítások általában szemikvantitatív jellegűek és nem teljesek. Míg a

kísérletesen meghatározott szerkezetek validálásához kiszámítanak alapvető fontos mértékeket (pl. φ/ψ szögek, a modell illeszkedése kísérleti adatokhoz)[108,109], addig a szerkezetek geometriai és fizikai-kémiai tulajdonságainak jellemzésére nem határoztak még meg standardizált mérőszámokat. Korábban a rendelkezésre álló nagy felbontású szerkezetek alacsony száma miatt viszonylag egyszerű volt eldönteni, hogy mely ismert konformációkat kell összehasonlítani egy új szerkezet konformációjával. A megoldott szerkezetek számának növekedésével azonban szabványosított, automatikusan generálható mérőszámokra lenne szükség. Saját tanulmányainkhoz is szükségünk volt ilyen mérőszámokra. Egyrészt fontos, hogy ne pusztán a legjobb felbontású szerkezetet válasszuk ki, hanem olyat, amely a megfelelő konformációval rendelkezik. Másrészt számos ABC szerkezet meglepő tulajdonságokat mutatott (pl. nagy membrán-béli dőlés szög), amelyek alapján kísérletet befolyásoló döntést is kellett hoznunk.



7. ábra: Példák Pgp-szerű szerkezetek konformációit leíró dedikált vektorokra. THX: átkereszteződő TM hélixek által definiált konftorok (kék); ICX: ezen TM hélixek intracelluláris részeinek megfeleltethető konftorok (piros); NBDX: csatoló hélix tömegközéppontja és az NBD S9 hélix első aminosava közötti konftor (türkiz). Ezen vektorok által bezárt szögek a befelé nyitottságot jellemzik.

Ezért megvizsgáltuk a Pgp-szerű szerkezetek különféle kvantitatív jellemzőit. Először a fehérjének a lipid kettősréteg normáljához képest mutatott dőlését és a hidrofób kettősréteg fehérje körüli helyzetét tanulmányoztuk. Ezeket az adatokat több adathalmazból és saját szimulációinkból számoltuk. A legtöbb esetben a dölésszög-béli különbségek elhanyagolhatók voltak (<6°). A fehérje körüli kettősréteg elhelyezkedését a transzmembrán hélixek tömegközéppontja és a kettősréteg középpontja közötti z tengely mentén mért távolsággal jellemeztük. Ezek és más (pl. hélixek törése, elfordulása, membránszolvatáció) mérőszámok azonban még mindig nem rendelkeztek megfelelő felbontással különböző szerkezetek pontos összehasonlításához. Ezért az ABC területen szerzett tapasztalatunkra támaszkodva olyan, aminosav párok közötti vektorokat definiáltunk, melyek nagysága vagy egy másik vektorral bezárt szöge jellemző az adott konformációra.

Ezeket a vektorokat konformációs vektoroknak, konftoroknak neveztük el. Például az extracelluláris tér felé való nyitás mértékének összehasonlítására a THX1-2 az egyik

legalkalmasabb vektor-pár (7. ábra), amelyek által bezárt szögek átlaga a "alul-nyitott/befelé-néző", "alul-zárt/befelé-néző", "alul-zárt/felül-zárt", és "alul-zárt/felül-nyitott" konformációkkal rendelkező ABC szerkezetekben 46°, 26°, 39° és 35°.

Összefoglalva, a javasolt metrikák hozzájárulnak az ABC membránfehérjék szerkezeti jellemzőinek mélyebb megértéséhez, a szerkezetek validálásához és molekuladinamikai trajektóriákban megfigyelt mozgások elemzéséhez. Hasonló metrikák más fehérjecsaládokra is kidolgozhatók és alkalmazhatók.

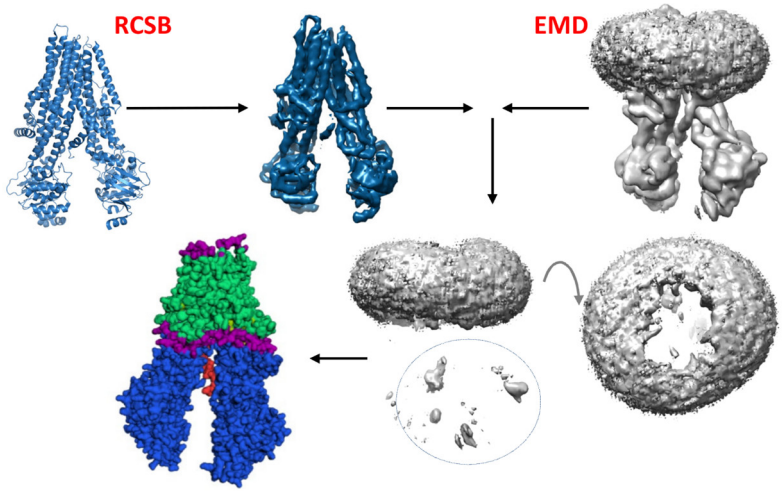
4.3.3. MemBlob szerver TM-régiók krio-EM denzitásból való meghatározására

(Cikk-8: MemBlob database and server for identifying transmembrane regions using cryo-EM maps; http://memblob.hegelab.org)

Membránfehérjék szerkezetének és működésének megértéséhez fontos a transzmembrán, illetve egyéb, lipidekkel kölcsönható régióik azonosítása. Röntgen krisztallográfiával meghatározott szerkezetekben a lipidek ritkán azonosíthatók és nem feltétlen a fiziológiás kötőhelyekhez kapcsolódnak. A lipidekkel kölcsönható helyeket ritkán vizsgálják közvetlenül, azok a kísérletek, amelyekben a feltételezett TM-hélixek köré különböző pozíciókba *tag*-et (címkéket) illesztve azok hozzáférhetőségét vizsgálják, általában alacsony felbontású adatokat szolgáltatnak[110,111]. Ezért számos *in silico* predikciós módszert is fejlesztettek a TM-régiók meghatározására, amelyek közül a legnépszerűbbek a TMDET, a PPM és a MEMPROTMD[112–114].

Az utóbbi évek technikai fejlesztéseinek köszönhetően a membránfehérjék szerkezetét egyre inkább krio-elektronmikroszkópiával oldják meg. Észrevettük, hogy a krio-elektronmikroszkópiát (krio-EM) alkalmazó membránfehérje szerkezetmeghatározás során jelen lévő lipidkörnyezet (pl. micella, bicella és nanodisc) elektronsűrűsége is látható a meghatározott krio-EM denzitásban. Ezért kifejlesztettünk egy olyan munkafolyamatot (MemBlob), ami membránfehérjék kísérletes krio-EM denzitását felhasználva meghatározza a fehérje lipidkörnyezettel körülvett TM részét. Röviden összefoglalva, ha a membránfehérje denzitását kivonjuk a teljes denzitásból, akkor megkapjuk a membránkörnyezet denzitását, amely széleinek a 3D koordinátái meghatározhatók. A denzitás széleit körbe 10 fokonként határoztuk meg, s minden egyes koordinátához megkerestük a legközelebbi fehérje atomot, ami így meghatározta a keresett, lipidbe süllyedő fehérje régiót (8. ábra).

A web alkalmazás egyrészt adatbázisként működik, minden 2018. júliusig krio-EM-mel meghatározott membránfehérje szerkezetre (92 TM fehérje 4 Å vagy annál jobb felbontással) kiszámolt adatot tartalmaz. Lehetővé teszi ugyanakkor a felhasználó számára, hogy egy tetszőleges szerkezetet feltöltsön, s lefuttassa az algoritmust. Az adatbázis összeállítása során kiderült, hogy az elektrondenzitás térképek körülbelül 30%-a nem tartalmazott jól definiált, a membránkörnyezetnek megfelelő részt. Ezeket a szerkezeteket vagy megfelelő lipidkörnyezet hiányában oldották meg, vagy az elektronsűrűség-térképeik alacsony jel/zaj aránya akadályozta a membránpaca határainak meghatározását.



8. ábra: A MemBlob munkafolyamata a CFTR szerkezetén bemutatva. A fehérje atomi szerkezetéből (RCSB, PDB ID: 5uak) denzitást számolunk, amit kivonunk a teljes denzitásból (EMD). A megmaradt membrán-paca széleinek koordinátai definiálják a membrán interfész (lila) és hidrofób (zöld) régióit. Piros: alacsony felbontású szerkezeti rész, aminosav oldallánc hozzárendelés nélkül. Bekarikázott: a rendezetlen R doménből származó denzitások.

Tanulmányunk bemutatta, hogy a krio-EM sűrűségtérképek a fehérjeszerkezeten kívül más értékes információt is tartalmaznak, mint például a lipidkörnyezet helyzete és esetlegesen a rendezetlen régiókhoz tartozó sűrűségek, s ezek a kísérletes adatok bioinformatikai eszközökkel hozzáférhetővé tehetők. A web alkalmazást a megjelenése utáni első évben már sokat használták, s az egyik elsőszerző (G. Csizmadia) előadás formájában egy Gordon konferencián bemutathatta. Tusnády Gáborral (ELKH TTK) együttműködve az általa TM fehérjékre kidolgozott, széles körben használt web alkalmazások közé tervezzük integrálni.

4.3.4. Transzmembrán fehérjék AlphaFold2 predikciója

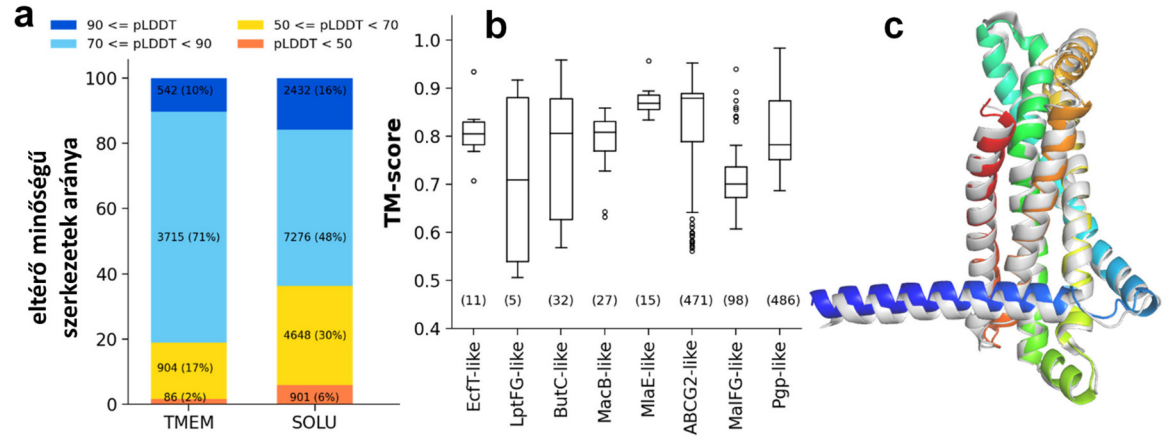
(Cikk-9: Ins and outs of AlphaFold2 transmembrane protein structure predictions; http://alphafold.hegelab.org)

Bár hatalmas erőforrásokat fordítottak a fehérjék szerkezetének szekvenciájuk alapján történő jóslására, ez hosszú évekig kihívást jelentő feladat maradt[115]. A területen komoly áttörés a CASP13 és CASP14 (*Critical Assessment of Protein Structure Prediction*) versenyen történt[116], amikor egy neurális hálózaton alapuló megközelítés, az AlphaFold (AlphaFold2, AF2)

kiemelkedett a többi módszer közül[28]. Az AF2 predikció bemenete egy fehérjelánc szekvenciája, amelyet az algoritmus kiegészít többszörös szekvencia-illesztésből és rokon fehérjék szerkezetéből származó információkkal. Az eredményül kapott szerkezeti modellek minőségét a pLDDT (predicted Local Distance Difference Test) pontszám átlaga jellemzi (amely 0 és 100 közötti értékeket vehet fel; a magasabb érték jobb) és a kimeneti szerkezeti modelleket ez alapján rangsorolják[28]. Azt, hogy az AF2 CASP14-ben megfigyelt nagy pontossága és megbízhatósága bármely fehérjeszerkezet jóslásakor is érvényes[28,117], még nem validálták. Nem ismert, hogy a legmagasabb pLDDT pontszámmal rendelkező szerkezeti modell mindig megfelel-e a natív szerkezetnek. A transzmembrán fehérjék tanulmányozásával foglalkozó kutatók különösen szkeptikusak az AF2 általános megbízhatóságával kapcsolatban, mert ezeknek a fehérjéknek a vizsgálata akár kísérleti, akár számítási módszerekkel különösen nehéz, és az AlphaFold2-t nem hangolták specifikusan TM fehérjékre.

A jósolt transzmembrán (ABC) fehérjeszerkezetek racionalitásának vizsgálata. Annak érdekében, hogy az AF2 által épített TM fehérje szerkezetek jóságát megvizsgáljuk, összehasonlítottuk a TM és a szolubilis fehérjék prediktált szerkezeteinek pLDDT pontszámát. Érdekes módon az oldható fehérjék pontszámai szélesebb eloszlást mutattak, és több alacsonyabb pLDDT értékű szerkezet volt megfigyelhető a TM fehérjékhez képest (9. ábra, a). Ez váratlan volt, mivel az AlphaFold2 tanulási készlete eleve több szolubilis szerkezetet tartalmazott és az algoritmust nem érzékenyítették TM fehérjékre. Ugyanakkor ismert az is, hogy az alacsony pLDDT-értékek szerkezeti rendezetlenséggel is korrelálnak[28], azaz a membránfehérjékre kapott magasabb értékek összhangban vannak azzal, hogy bennük kevesebb rendezetlen régió található a szolubilis fehérjékhez képest. A következő lépésben összehasonlítottuk a TM hélixek AF2 szerkezetekben megfigyelhető térbeli helyzetét a lipid kettősrétegben megfigyelhető és elvárt fiziológiás hélix orientációkkal (Cikk-9, 1. ábra). Azt találtuk, hogy az AF2 szerkezetek többségében a TM régióknak megfelelő hélixek végei kijelölnek két síkot, amelyek távolsága megfeleltethető lipid kettősrétegek vastagságának.

Ezeknél specifikusabban, fehérje-fold szinten is teszteltük az AF2 teljesítményét. Ehhez az ABC fehérje szupercsalád kiváló választás, mivel a jelenleg rendelkezésre álló PDB szerkezeteik TM doménjei igen változatosak, kilenc különböző fold-családba sorolhatók (Cikk-9, S3 ábrán még nyolc szerkezeti család szerepel, azóta már azonosítottunk egy újat – lsd. http://abc3d.hegelab.org)[24,118,119]. Célunk az volt, hogy az AF2 által létrehozott ABC szerkezeteket összehasonlítsuk kísérletesen meghatározott szerkezetekkel. Az összehasonlításhoz TM-pontszám (*TM-score*) értékeket használtunk, amelyet a *TMalign*[120] algoritmussal számoltunk minden egyes AF2 ABC szerkezet és minden egyes ABC fold-család



9. ábra: (a) Transzmembrán (TMEM) és szolubilis (SOLU) AF2-prediktált szerkezetek jósága pLDDT pont alapján. **(b)** AF2-jósolt ABC fehérje szerkezetek hasonlósága kísérletes szerkezetekhez (megegyező *fold*, ha TM-score > 0,5). **(c)** Az MlaE fehérje jósolt szerkezete (kék-piros az N-végtől a C-végig) gyakorlatilag megegyezik a kísérletes szerkezettel (szürke, PDB ID: 7ch0).

referencia szerkezete között. Amennyiben a TM-pontszám nagyobb 0,5-nél, a két fehérje foldja megegyezőnek tekinthető[121]. Eredményeink azt mutatták, hogy az AF2 által jósolt ABC TM domének 99,5%-a megfelel egy kísérletes ABC-foldnak, mert 0,5 feletti értéket mutattak a megfelelő referencia-szerkezettel szemben (9. ábra, b).

Kihívást jelentő és új transzmembrán szerkezetek előrejelzése. Fontos megemlíteni, hogy az EBI adatbázisában elhelyezett[27], AF2-vel jósolt transzmembrán fehérje szerkezetek jelentős részéhez van megfeleltethető kísérletes szerkezet, amely ugyanannak vagy homológ fehérjének a vizsgálatából származik. Ezek nagy része, a 2018-04-30 előtti szerkezetek, benne volt az AF2 tanulási halmazában, a 2020 közepéig publikált szerkezeteket pedig sablonként használták a predikciós futtatások során[28]. Ezért az AF2 szerkezetjóslások megbízhatóságát kétféleképpen teszteltük. Először a CASP14 verseny kihívást jelentő TM célpontját (T1024, LmrP, PDBID: 6t1z, megjelent 2019-10-07[122]) választottuk ki, amelynek voltak ismert homológ szerkezetei, majd olyan új TM szerkezeteket, amelyek szintén 2018-04-30 után kerültek a PDB adatbázisba így nem voltak benne a tanulóhalmazban. Az utóbbi tesztelés során kiterjedt irodalmi, SCOP[123] és PFAM[124] kereséseket végeztünk olyan transzmembrán fehérje szerkezetek vagy homológ szerkezeteik azonosítására, amelyek nem kerültek bele az AF2 tanulóhalmazába. Így azonosítottuk az ABC transzmembrán MlaE-szerű szerkezetet[125], az ER membránfehérje komplex 6-os alegységének szerkezetét, valamint az MprF szerkezetet[126]. Ezeket tudtuk a tanítóhalmaztól teljesen független, AF2 TM fehérje jóslások tesztelésére használni. A sablonok nélküli AF2 futtatások az MlaE (PDBID: 7ch0, RMSD: 1,28Å, TM-score: 0,95; 9c. ábra) és az EMC6 (PDBID: 6ww7, RMSD: 0,96Å, TM-score: 0,93; Cikk-9, 3c. ábra) esetén a kísérleti szerkezetekhez nagyon hasonló modelleket eredményeztek. Ezzel szemben az MprF

transzmembrán domén legjobb pLDDT pontszámú prediktált szerkezete nem egyezett a kísérleti szerkezettel (Cikk-9, 4a. ábra). Ezért ezt a predikciót többször (n=6) elvégeztük a randomszám generátor különböző kiindulásai állapotaival, majd a TM-score segítségével összehasonlítottuk a kimenetet a 7duw transzmembrán doménjével. A pLDDT-pontszámok és a TM-pontszámok ábrázolása (Cikk-9, 4b. ábra) azt mutatta, hogy a 6x5 prediktált szerkezet közül a legjobb pLDDT-pontszámmal rendelkező szerkezet mutatta a legmagasabb TM-pontszámot, azaz ez hasonlított legjobban a célszerkezethez (Cikk-9, 4c. ábra). Fontos megjegyezni, hogy az MprF konformációk közötti különbség két szubdomén (flippáz és szintáz) elkülönülését mutatja[127], azaz az AF2 az LmrP jósláshoz hasonlóan egy, a kísérletes szerkezettől eltérő, de funkcionálisan releváns állapotot ragadhatott meg.

Az AF2 útmutatást adhat ABC szerkezetekhez kapcsolódó kérdések vizsgálatához. Az AF2 is használt sablonokat a szerkezeti modellezéshez, azaz az általa épített szerkezetek egy része magas minőségű homológia modellnek is tekinthető. Az általunk részletesebben vizsgált humán ABC-fehérjék AF2-modellezését kikapcsolt sablonhasználattal is elvégeztük. Eredményeink azt mutatták, hogy az egyláncú fehérjékre betanított AlphaFold2 algoritmus jól működik féltranszporterekre is, mind homodimer, mind heterodimer esetben (Cikk-9, 5. ábra c-e és S5. ábra).

Azt is megvizsgáltuk, hogy az AF2 szerkezeti modellek hogyan egészíthetik ki vagy helyettesíthetik a homológia modelleket MD szimulációkban. A CFTR fehérje esetén, a templát nélküli AF2-predikció egyenes TM8 hélixet eredményezett, ez a szerkezet pedig MD szimulációkban racionálisabb nyitódást mutatott a kísérleteshez képest (Cikk-9, 5. ábra, f-h).

Hozzájárulásunk az AlphaFold2 kódhoz. Két jelentős hibát is felfedeztünk a publikus AF2 kódban (http://alphafold.hegelab.org). Mindkettő memóriafelhasználással kapcsolatban lépett fel abban az esetben, ha a bemenetként kapott többszörös szekvenciaillesztés nagyon nagyméretű. Az első hiba esetén a GPU memória használat növekedett meg annyira jelentősen még rövid fehérjék esetén is (~250 a.a. NBD), hogy egy 24 GB RAM-val rendelkező GPU sem volt elegendő a számolásokhoz. A másik hiba eredete hasonló, csak a rendszer memóriáját terheli meg. Javításainkat a DeepMind már beépítette az AF2 kódba.

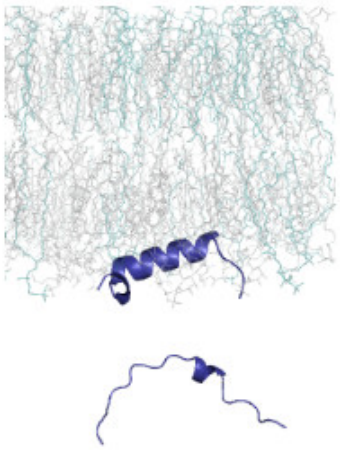
Összefoglalás. Bemutattuk, hogy az AF2 membránfehérjékre is hasonló minőségű szerkezeteket tud jósolni mint szolubilis fehérjékre. Viszont előzetes eredményeink azt mutatják, hogy a 2021 novemberében megjelent AlphaFold-Multimer[128], amelyet fehérjekomplexek jóslására fejlesztettek, membránfehérje komplexeket nem jósol annyira jó hatásfokkal. AF2 felhasználásával megépítettük olyan dimer humán ABC fehérjék szerkezetét, amelyek sem a kísérletes PDB, sem az AF2 EBI adatbázisban nem találhatók meg. Ezeket a

3D-beacon protokol felhasználásával kötöttük be nagyobb adatbázisokba (pl. PDB-KnowledgeBase), hogy mások számára is könnyen hozzáférhetők legyenek (https://3dbeacon.hegelab.org)[24].

4.3.5. MemMoRF, lipidekkel kölcsönható rendezetlen fehérjerégiók adatbázisa

(Cikk-10: The MemMoRF database for recognizing disordered protein regions interacting with cellular membranes; http://memmorf.hegelab.org)

A fehérje-membrán kölcsönhatások alapvető szerepet játszanak számos celluláris folyamatban (pl. jelátvitel, vezikuláris transzport, vírusinvázió). Az ilyen kölcsönhatások kialakításában gyakran vesznek részt rendezetlen régiók (*Intrinsically Disordered Region*, IDR), amelyek nemcsak más fehérjékkel, hanem lipidekkel is képesek specifikus, reverzibilis kölcsönhatásokat kialakítani[129–131]. Míg a fehérjepartnerekkel való kölcsönhatás során rendezetlen-rendezett átmenetet mutató IDR-eket MoRF-nak (*Molecular Recognition Feature*) nevezik[132,133], addig a lipidek által indukált rendezetlen-rendezett átmentet megvalósító szegmenseket MemMoRF-oknak neveztük el (10. ábra). Mivel a MemMoRF-ok kísérletes tanulmányozása és számításos jellemzése is kihívást jelent, s az irodalomban ezekről a régiókról adatok egyelőre csak szórtan találhatók, vizsgálatuk megkönnyítése érdekében létrehoztuk a kísérletileg validált MemMoRF-ok átfogó adatbázisát (https://memmorf.hegelab.org). A MemMoRF azonosítást irodalmi és NMR adatokra alapoztuk. A potenciális MemMoRF-ok dinamikájának jellemzésére a kémiai eltolódásokból számított másodlagos szerkezeti hajlamot[134] és *random coil index*[135] értékeket építettük be adatbázisunkba. Összesen 538 fehérjét vizsgáltunk meg (PDB adatbázis, *Protein Ensemble Database*, és irodalom alapján), s ezek közül 107 fehérjében találtunk membránnal kölcsönható régiókat (n=149). 131 régió mutatott rendezetlen-rendezett átmenetet lipid-kölcsönhatás során, 19 flexibilis maradt kötődés után is, 18 pedig rendezett szerkezettel rendelkezett oldatban is. Ezekhez a fehérjékhez és MemMoRF-okhoz tartozó adatokat cikkekből származó mondatokkal, szerkezeti adatokkal (pl. PFAM-domének[124], rövid lineáris motívumok[136]) és betegségekkel kapcsolatos információkkal (pl. dbSNP[137] és MIM[138]) egészítettük ki. Adatbázisunk (1) szabadon hozzáférhető, könnyen használható, szervezett erőforrás; (2) *gold standard*



10. ábra Molekula dinamika szimulációkban a phospholemman (PDB ID: 2jo1) MemMoRF régiója (a.a. 58-72) membránkörnyezetben megörzi helicitását, míg lipid-kölcsönhatások nélkül elveszti azt.

készletet tartalmaz a rendezetlen régiók lipidkölcsönhatásának szekvencia-függőségének meghatározásához; (3) kiváló minőségű készletet jelent a MemMoRF-azonosításra szolgáló új *in silico* algoritmusok kifejlesztéséhez; és (4) elősegíti a MemMoRF-okat tartalmazó fehérjék további kísérleti vizsgálatát. A MemMoRF adatbázis egy felhasználóbarát felületen keresztül érhető el, így jó eséllyel válhat a széles tudományos közösség számára a transzmembrán és membránhoz kapcsolódó fehérjék rendezetlen régióinak központi forrásává.

5. Kitekintés

A bemutatott cikkekhez kapcsolódó projekteken jelenleg is dolgozunk, a felmerült érdekesebb problémák közül vázolok az alábbiakban néhányat. (1) Az ABC fehérjék összeszerelődését alloszterikus módon erősen befolyásolják a doménjeik közötti kölcsönhatások (készülő kézirat Lukács Gergellyel együttműködésben, McGill University, Kanada). Ez alapján azt is feltételezzük, hogy alloszterikus események a vártnál sokkal jobban meghatározzák a transzport folyamatokat és a szubsztrát-felismerést, amelynek tanulmányozását elkezdtük egy növényi ABCG fehérjén (Markus Geisler kísérletes partnerrel, University of Fribourg, Svájc). (2) Kérdéses, hogy az AlphaFold-Multimer mennyire jól tudja prediktálni membránfehérjék komplexeit. Ezt vizsgáljuk az emberi SERCA kalcium ATPáz és a SARS Cov-2 *Envelope* fehérjére fókuszáltan, kísérletes és elméleti módszerek kombinációjával (NKFIH K137610 pályázat). (3) Az Elixir 3D-bioinfo közösségen keresztül bekapcsolódtunk szerkezeti adatokkal kapcsolatos technológiai fejlesztésekbe (Christine Orengo UCL, Sameer Velankar és Váradi Mihály, EBI, Anglia). Ennek keretében az *in silico* módon épített szerkezeteinket és MD szimulációinkból származó konformációs sokaságainkat a 3D-beacon rendszeren keresztül tervezzük megosztani a tudományos közösséggel (https://3dbeacon.hegelab.org)[24]. Továbbá standardokat, munkafolyamatokat és szoftvereket hozunk létre fehérjecsaládok szerkezetinek összehasonlító elemzésére (http://abc3d.hegelab.org)[24].

Az elmúlt években tudományterületünk eredményei kapcsán és együttműködő partnereinkkel való interakció során számos izgalmas kérdés merült fel kutatási területünk jövőjével kapcsolatban. Molekuladinamikai szimulációink felhasználhatóságát nagyban befolyásolja a lefedhető időskála rövidsége – vajon átütést hozhat-e a területen a kvantumszámítógépek megjelenése? Pillanatnyilag a kvantumkémiai számítások tűnnek legkönnyebben kvantumszámítógépre alkalmazhatónak, ugyanakkor a speciálisan ezekre kifejlesztett numerikus algoritmusok áttörést hozhatnak a klasszikus MD szimulációkban is. Az első praktikusan is használható kvantumszámítógépeket valószínűleg ipari résztvevők fogják létrehozni, ami lassíthatja bizonyos tudományterületeken való használatukat. Azonban az

elmúlt években az ipar és az akadémia kapcsolata is jelentősen átalakult, amelynek egyik kitűnő példája az AlphaFold. A DeepMind vállalat kitűzte egy különösen nehéz probléma megoldását, s erőforrásokat nem kímélve, a célra fókuszálva, rövid időn belül elérte azt. Ráadásul, meglepő módon nem pusztán a prediktált fehérjeszerkezeteket tették közkinccsé, hanem a forráskódot is nyilvánosságra hozták a tudományos közösség számára felhasználható módon. A másik különlegesség az ilyen típusú kutatásban, hogy többé az ipari környezet nem riasztó a tudományos munka iránt érdeklődők számára, hiszen egy ilyen projekt hossza összemérhető egy pályázat futamidejével, a tudományos feladaton kívül nincsen más elvárt teljesítmény (pl. publikációs nyomás), s még az anyagi megbecsülés is kimagasló. A minőségi akadémiai kutatások megmaradása valószínűleg erősen függ attól, hogy ezeket a paramétereket akadémiai környezetben mennyire sikerül közelíteni az ipari paraméterekhez. Az AlphaFold megjelenésével, a nagy megbízhatóságú szekvencia-alapú fehérjeszerkezet építéssel felvetette azt a kérdést is, hogy vajon lezárult-e a fehérjékkel kapcsolatos bioinformatika kora. Erre a válasz egy egyértelmű nem, ugyanis az AlphaFold nem oldott meg minden fehérje-szerkezettel kapcsolatos kérdést. Komplexek szerkezetének AF2-predikciója kimagaslónak tűnik, de nem elégségesnek, különösen membránfehérjék komplexei esetén. Rendezetlen régiók jelenlétét jól tudja jelezni, de más eszközök erre alkalmasabbak lehetnek. Ugyanakkor ezen algoritmusok egyike sem képes a dinamikát és a konformációs sokaságot jellemezni. Továbbá az AF2 a fehérje-feltekeredés folyamatáról semmiféle közvetlen információt nem szolgáltat, amelynek megértése legalább olyan fontos, mint a szekvencia-alapú szerkezet-jóslás, figyelembe véve, hogy számos betegséget fehérje-feltekeredési probléma okoz. Összefoglalva, a fehérje szerkezetek megléte egy alapadathalmaz (mint a genom szekvencia), de az értelmét meg kell találni. Ennek egyik megközelítés módja adatbányászat lehet ebben a nagy mennyiségű szerkezeti adathalmazban, ami pont a szerkezeti bioinformatika *big data* korának kezdetét jelzi. Egy másik fontos megközelítése a fehérjeszerkezetek lényegi megértésének a fehérjedinamika vizsgálata. Bár léteznek és elengedhetetlenül fontosak ehhez kísérletes módszerek (pl. NMR, FRET), az atomi szintű történések (pl. működés, mutációk hatása, gyógyszer-kötődés) megértéséhez az *in silico* vizsgálatok nélkülözhetetlenek. Ezek közül a jelenleg legmegbízhatóbb módszer a molekuláris dinamika, amit méltán nevezett a Nobel-díjas Karplus számításos mikroszkópnak.

6. Új tudományos eredmények

1. Bemutattuk, hogy a CFTR korrektorok többsége az NBD1/CL4 interfészt stabilizálja, ám a terápia hatásosságának növelésére szükség van más helyen ható vegyületekre is, amit interfész stabilizált mutáns fehérjével érdemes keresni.
2. Kimutattuk, hogy a VX-770 potenciátor a CFTR fehérje funkcióját annak destabilizálásán keresztül javítja, ami viszont a korrektorok hatását gyengíti, ezért új potenciátorok fejlesztése szükséges.
3. Molekula dinamika szimulációkkal a klorid-vezetéshez szükséges geometriával rendelkező CFTR-konformációkat tudtunk azonosítani és a klorid-átjutást jellemezni.
4. Elsőként modelleztük az ABCG2 fehérje szerkezetét és molekula dinamika szimulációk alkalmazásával meghatároztuk a köszvénnyel asszociált Q141K és a szubszrátspecifitás változását okozó R482G mutációk lehetséges szerkezeti hatásait is.
5. Az ABCG2 legjelentősebb fiziológiás szubsztrátja, a húgysav jelenlétében végzett MD-szimulációkkal jellemeztük a transzlokációs útvonalat *in silico* dokkolás alkalmazása mellett.
6. Az ABCG2 fehérjével végzett MD-szimulációink alapján megállapíthattuk, hogy a koleszterin a transzportfolyamatot a transzmembrán hélixek intracelluláris oldalon történő zárásának megkönnyítésével segíti elő.
7. Az ABCM2 web alkalmazásunkban lehetővé tettük különböző ABC fehérjék szekvenciálisan homológ pozícióiban lévő variánsainak összehasonlítását.
8. ABC fehérjeszerkezetek összehasonlítására és osztályozására szabványosított, automatikusan generálható mérőszámokat, *konftorokat* hoztunk létre.
9. Kifejlesztettük a MemBlob munkafolyamatot (MemBlob), amely kísérletes adatokból meghatározza a fehérje lipiddel körülvett TM részét.
10. Bemutattuk, hogy az AF2 membránfehérjékre is hasonló minőségű szerkezeteket tud jósolni, mint szolubilis fehérjékre.
11. Adatbázisban tettük elérhetővé az olyan rendezetlen régiókat, amelyek kísérletes adatok alapján lipidekkel is képesek specifikus, reverzibilis kölcsönhatásokat kialakítani (MemMoRF).

7. A dolgozat alapjául szolgáló közlemények

Cikk-1 Okiyoneda, T, G Veit, JF Dekkers, M Bagdany, N Soya, H Xu, A Roldan, AS Verkman, M Kurth, A Simon, T Hegedus, JM Beekman, and GL Lukacs. 2013. "Mechanism-Based Corrector Combination Restores ΔF508-CFTR Folding and Function." *NATURE CHEMICAL BIOLOGY 9* (7): 444–454. doi:10.1038/nchembio.1253.

Cikk-2 Veit, G, R Avramescu, D Perdomo, P Phuan, M Bagdany, P Apaja, F Borot, D Szollosi, Y Wu, W Finkbeiner, T Hegedus, A Verkman, and G Lukacs. 2014. "Some Gating Potentiators, Including VX-770, Diminish ΔF508-CFTR Functional Expression." *SCIENCE TRANSLATIONAL MEDICINE 6* (246). doi:10.1126/scitranslmed.3008889.

Cikk-3 Farkas, Bianka, Hedvig Tordai, Rita Padányi, Attila Tordai, János Gera, Gábor Paragi, and Tamás Hegedűs. 2020. "Discovering the Chloride Pathway in the CFTR Channel." *CELLULAR AND MOLECULAR LIFE SCIENCES 77* (4): 765–778. doi:10.1007/s00018-019-03211-4.

Cikk-4 László, L, B Sarkadi, and T Hegedűs. 2016. "Jump into a New Fold-A Homology Based Model for the ABCG2/BCRP Multidrug Transporter." *PLOS ONE 11* (10). doi:10.1371/journal.pone.0164426.

Cikk-5 Nagy, T, Á Tóth, Á Telbisz, B Sarkadi, H Tordai, A Tordai, and T Hegedűs. 2021. "The Transport Pathway in the ABCG2 Protein and Its Regulation Revealed by Molecular Dynamics Simulations." *CELLULAR AND MOLECULAR LIFE SCIENCES 78* (5): 2329–2339. doi:10.1007/s00018-020-03651-3.

Cikk-6 Gyimesi, G, D Borsodi, H Saranko, H Tordai, B Sarkadi, and T Hegedűs. 2012. "ABCMdb: A Database for the Comparative Analysis of Protein Mutations in ABC Transporters, and a Potential Framework for a General Application." *HUMAN MUTATION 33* (11): 1547–1556. doi:10.1002/humu.22138.

Cikk-7 Csizmadia, Georgina, B Farkas, Z Spagina, H Tordai, and T Hegedűs. 2018. "Quantitative Comparison of ABC Membrane Protein Type I Exporter Structures in a Standardized Way." *COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL 16*: 396–403. doi:10.1016/j.csbj.2018.10.008.

Cikk-8 Farkas, Bianka, G Csizmadia, E Katona, GE Tusnády, and T Hegedűs. 2020. "MemBlob Database and Server for Identifying Transmembrane Regions Using Cryo-EM Maps." *BIOINFORMATICS 36* (8): 2595–2598. doi:10.1093/bioinformatics/btz539.

Cikk-9 Hegedűs, Tamás, M Geisler, GL Lukács, and B Farkas. 2022. "Ins and Outs of AlphaFold2 Transmembrane Protein Structure Predictions." *CELLULAR AND MOLECULAR LIFE SCIENCES 79* (1). doi:10.1007/s00018-021-04112-1.

Cikk-10 Csizmadia, Georgina, G Erdős, H Tordai, R Padányi, S Tosatto, Z Dosztányi, and T Hegedűs. 2020. "The MemMoRF Database for Recognizing Disordered Protein Regions Interacting with Cellular Membranes." *NUCLEIC ACIDS RESEARCH 49* (D1): D355–D360. doi:10.1093/nar/gkaa954.

8. Hivatkozások jegyzéke

1. George, A. M. *ABC Transporters - 40 Years on*. (Springer Cham, 2016).
2. Szakács, G., Váradi, A., Ozvegy-Laczka, C. & Sarkadi, B. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). *Drug Discov Today* **13**, 379–393 (2008).
3. Nikaido, H. & Hall, J. A. Overview of bacterial ABC transporters. *Methods Enzymol* **292**, 3–20 (1998).
4. Klein, I., Sarkadi, B. & Váradi, A. An inventory of the human ABC proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1461**, 237–262 (1999).
5. Borst, P. & Elferink, R. O. Mammalian ABC transporters in health and disease. *Annu Rev Biochem* **71**, 537–592 (2002).
6. Sarkadi, B., Homolya, L., Szakács, G. & Váradi, A. Human multidrug resistance ABCB and ABCG transporters: participation in a chemoimmunity defense system. *Physiol Rev* **86**, 1179–1236 (2006).
7. Borst, P., Váradi, A. & van de Wetering, K. PXE, a Mysterious Inborn Error Clarified. *Trends Biochem Sci* **44**, 125–140 (2019).
8. Riordan, J. R. *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**, 1066–1073 (1989).
9. Csanády, L., Vergani, P. & Gadsby, D. C. STRUCTURE, GATING, AND REGULATION OF THE CFTR ANION CHANNEL. *Physiological reviews* **99**, 707–738 (2019).
10. Tawbeh, A., Gondcaille, C., Trompier, D. & Savary, S. Peroxisomal ABC Transporters: An Update. *Int J Mol Sci* **22**, 6093 (2021).
11. Keppler, D. Multidrug resistance proteins (MRPs, ABCCs): importance for pathophysiology and drug therapy. *Handb Exp Pharmacol* 299–323 (2011) doi:10.1007/978-3-642-14541-4_8.
12. Holló, Z. *et al.* Parallel functional and immunological detection of human multidrug resistance proteins, P-glycoprotein and MRP1. *Anticancer Res* **18**, 2981–2987 (1998).
13. Hegedus, T. *et al.* Interaction of tyrosine kinase inhibitors with the human multidrug transporter proteins, MDR1 and MRP1. *Biochim Biophys Acta* **1587**, 318–325 (2002).
14. Szakács, G., Paterson, J. K., Ludwig, J. A., Booth-Genthe, C. & Gottesman, M. M. Targeting multidrug resistance in cancer. *Nat Rev Drug Discov* **5**, 219–234 (2006).
15. Robey, R. W. *et al.* Revisiting the role of ABC transporters in multidrug-resistant cancer. *Nat Rev Cancer* **18**, 452–464 (2018).
16. Jones, P. M., O'Mara, M. L. & George, A. M. ABC transporters: a riddle wrapped in a mystery inside an enigma. *Trends in Biochemical Sciences* **34**, 520–531 (2009).
17. George, A. M. & Jones, P. M. Perspectives on the structure-function of ABC transporters: the Switch and Constant Contact models. *Prog Biophys Mol Biol* **109**, 95–107 (2012).
18. Locher, K. P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nature structural & molecular biology* **23**, 487–493 (2016).
19. Bryan, J. *et al.* ABCC8 and ABCC9: ABC transporters that regulate K+ channels. *Pflugers Arch* **453**, 703–718 (2007).
20. Yu, Q. *et al.* Structures of ABCG2 under turnover conditions reveal a key step in the drug transport mechanism. *Nat Commun* **12**, 4376 (2021).
21. Prescher, M. *et al.* Evidence for a credit-card-swipe mechanism in the human PC floppase ABCB4. *Structure* **0**, (2021).
22. Hollenstein, K., Frei, D. C. & Locher, K. P. Structure of an ABC transporter in complex with its binding protein. *Nature* **446**, 213–216 (2007).
23. Sharom, F. J. Shedding light on drug transport: structure and function of the P-glycoprotein multidrug transporter (ABCB1). *Biochem Cell Biol* **84**, 979–992 (2006).

24. Tordai, H. *et al.* Comprehensive collection and prediction of ABC transmembrane protein structures in the AI era of structural biology. *bioRxiv* 2022.07.08.499254 (2022) doi:10.1101/2022.07.08.499254.
25. Sarkadi, B., Homolya, L. & Hegedűs, T. The ABCG2/BCRP transporter and its variants - from structure to pathology. *FEBS Lett* (2020) doi:10.1002/1873-3468.13947.
26. Fiser, A. & Sali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* **374**, 461–491 (2003).
27. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439–D444 (2022).
28. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 1–11 (2021) doi:10.1038/s41586-021-03819-2.
29. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem* **30**, 2785–2791 (2009).
30. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455–461 (2010).
31. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *Journal of computational chemistry* **34**, 2135–2145 (2013).
32. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* **14**, 71–73 (2017).
33. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859–1865 (2008).
34. Jo, S. *et al.* CHARMM-GUI 10 years for biomolecular modeling and simulation. *J Comput Chem* **38**, 1114–1124 (2017).
35. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
36. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319–2327 (2011).
37. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering* **9**, 90–95 (2007).
38. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
39. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33–38, 27–28 (1996).
40. Bayer, M. SQLAlchemy. in *The Architecture of Open Source Applications Volume II: Structure, Scale, and a Few More Fearless Hacks* (eds. Brown, A. & Wilson, G.) (aosabook.org, 2012).
41. Hegedus, T. *et al.* F508del CFTR with two altered RXR motifs escapes from ER quality control but its channel activity is thermally sensitive. *Biochim Biophys Acta* **1758**, 565–572 (2006).
42. Veit, G. *et al.* Structure-guided combination therapy to potently improve the function of mutant CFTRs. *Nature medicine* (2018) doi:10.1038/s41591-018-0200-x.
43. Middleton, P. G. *et al.* Elexacaftor-Tezacaftor-Ivacaftor for Cystic Fibrosis with a Single Phe508del Allele. *N Engl J Med* **381**, 1809–1819 (2019).
44. McPhail, G. L. & Clancy, J. P. Ivacaftor: the first therapy acting on the primary cause of cystic fibrosis. *Drugs Today (Barc)* **49**, 253–260 (2013).
45. Heijerman, H. G. M. *et al.* Efficacy and safety of the elexacaftor plus tezacaftor plus ivacaftor combination regimen in people with cystic fibrosis homozygous for the F508del mutation: a double-blind, randomised, phase 3 trial. *Lancet* **394**, 1940–1948 (2019).

46. Qu, B. H., Strickland, E. H. & Thomas, P. J. Localization and suppression of a kinetic defect in cystic fibrosis transmembrane conductance regulator folding. *The Journal of biological chemistry* **272**, 15739–15744 (1997).
47. Vergani, P., Nairn, A. C. & Gadsby, D. C. On the mechanism of MgATP-dependent gating of CFTR Cl- channels. *J Gen Physiol* **121**, 17–36 (2003).
48. Aleksandrov, A. A. & Riordan, J. R. Regulation of CFTR ion channel gating by MgATP. *FEBS Lett* **431**, 97–101 (1998).
49. Seibert, F. S. *et al.* Influence of phosphorylation by protein kinase A on CFTR at the cell surface and endoplasmic reticulum. *Biochimica et biophysica acta* **1461**, 275–283 (1999).
50. Hegedus, T., Serohijos, A. W. R., Dokholyan, N. V., He, L. & Riordan, J. R. Computational studies reveal phosphorylation-dependent changes in the unstructured R domain of CFTR. *Journal of molecular biology* **378**, 1052–1063 (2008).
51. Hegedus, T. *et al.* Role of individual R domain phosphorylation sites in CFTR regulation by protein kinase A. *Biochimica et biophysica acta* **1788**, 1341–1349 (2009).
52. Du, K. & Lukacs, G. L. Cooperative assembly and misfolding of CFTR domains in vivo. *Mol Biol Cell* **20**, 1903–1915 (2009).
53. Du, K., Sharma, M. & Lukacs, G. L. The DeltaF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. *Nat Struct Mol Biol* **12**, 17–25 (2005).
54. Rabeh, W. M. *et al.* Correction of both NBD1 energetics and domain interface is required to restore ΔF508 CFTR folding and function. *Cell* **148**, 150–163 (2012).
55. 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**, 18843–18848 (2011).
56. Clancy, J. P. *et al.* Results of a phase IIa study of VX-809, an investigational CFTR corrector compound, in subjects with cystic fibrosis homozygous for the F508del-CFTR mutation. *Thorax* **67**, 12–18 (2012).
57. Amin, R. & Ratjen, F. Emerging drugs for cystic fibrosis. *Expert Opin Emerg Drugs* **19**, 143–155 (2014).
58. Zhang, Z., Liu, F. & Chen, J. Conformational Changes of CFTR upon Phosphorylation and ATP Binding. *Cell* **170**, 483-491.e8 (2017).
59. El Hiani, Y. & Linsdell, P. Functional Architecture of the Cytoplasmic Entrance to the Cystic Fibrosis Transmembrane Conductance Regulator Chloride Channel Pore. *The Journal of biological chemistry* **290**, 15855–15865 (2015).
60. Linsdell, P. Functional architecture of the CFTR chloride channel. *Molecular membrane biology* **31**, 1–16 (2014).
61. Cui, G., Zhang, Z.-R., O'Brien, A. R. W., Song, B. & McCarty, N. A. Mutations at arginine 352 alter the pore architecture of CFTR. *J Membr Biol* **222**, 91–106 (2008).
62. Norimatsu, Y. *et al.* Cystic fibrosis transmembrane conductance regulator: a molecular model defines the architecture of the anion conduction path and locates a 'bottleneck' in the pore. *Biochemistry* **51**, 2199–2212 (2012).
63. Wang, W., El Hiani, Y. & Linsdell, P. Alignment of transmembrane regions in the cystic fibrosis transmembrane conductance regulator chloride channel pore. *J Gen Physiol* **138**, 165–178 (2011).
64. Sethi, A., Eargle, J., Black, A. A. & Luthey-Schulten, Z. Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A* **106**, 6620–6625 (2009).
65. Branduardi, D., Gervasio, F. L. & Parrinello, M. From A to B in free energy space. *The Journal of chemical physics* **126**, 054103 (2007).
66. Liu, F., Zhang, Z., Csanády, L., Gadsby, D. C. & Chen, J. Molecular Structure of the Human CFTR Ion Channel. *Cell* **169**, 85-95.e8 (2017).

67. Liu, F. *et al.* Structural identification of a hotspot on CFTR for potentiation. *Science* **364**, 1184–1188 (2019).
68. Zhang, Z., Liu, F. & Chen, J. Molecular structure of the ATP-bound, phosphorylated human CFTR. *Proceedings of the National Academy of Sciences of the United States of America* (2018) doi:10.1073/pnas.1815287115.
69. Fay, J. F. *et al.* Cryo-EM Visualization of an Active High Open Probability CFTR Anion Channel. *Biochemistry* **57**, 6234–6246 (2018).
70. Johnson, Z. L. & Chen, J. Structural Basis of Substrate Recognition by the Multidrug Resistance Protein MRP1. *Cell* **168**, 1075-1085.e9 (2017).
71. Carveth, K., Buck, T., Anthony, V. & Skach, W. R. Cooperativity and flexibility of cystic fibrosis transmembrane conductance regulator transmembrane segments participate in membrane localization of a charged residue. *J Biol Chem* **277**, 39507–39514 (2002).
72. Allikmets, R., Schriml, L. M., Hutchinson, A., Romano-Spica, V. & Dean, M. A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance. *Cancer Res* **58**, 5337–5339 (1998).
73. Apáti, A. *et al.* High level functional expression of the ABCG2 multidrug transporter in undifferentiated human embryonic stem cells. *Biochim Biophys Acta* **1778**, 2700–2709 (2008).
74. Woodward, O. M. *et al.* Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc Natl Acad Sci U S A* **106**, 10338–10342 (2009).
75. Dawson, R. J. P. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180–185 (2006).
76. Lee, J.-Y. *et al.* Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature* **533**, 561–564 (2016).
77. Locher, K. P., Lee, A. T. & Rees, D. C. The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**, 1091–1098 (2002).
78. Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L. & Chen, J. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* **450**, 515–521 (2007).
79. Hazai, E. & Bikadi. Homology modeling of breast cancer resistance protein (ABCG2). *J Struct Biol* **162**, 63–74 (2008).
80. Rosenberg, M. F. *et al.* Three-dimensional structure of the human breast cancer resistance protein (BCRP/ABCG2) in an inward-facing conformation. *Acta Crystallogr D Biol Crystallogr* **71**, 1725–1735 (2015).
81. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
82. Taylor, N. M. I. *et al.* Structure of the human multidrug transporter ABCG2. *Nature* **546**, 504–509 (2017).
83. Mózner, O. *et al.* Cellular Processing of the ABCG2 Transporter-Potential Effects on Gout and Drug Metabolism. *CELLS* **8**, (2019).
84. Sarkadi, B., Homolya, L. & Hegedűs, T. The ABCG2/BCRP transporter and its variants - from structure to pathology. *FEBS LETTERS* **594**, 4012–4034 (2020).
85. Zámbó, B. *et al.* Clinically relevant mutations in the ABCG2 transporter uncovered by genetic analysis linked to erythrocyte membrane protein expression. *Scientific reports* **8**, 7487 (2018).
86. Mizuarai, S., Aozasa, N. & Kotani, H. Single nucleotide polymorphisms result in impaired membrane localization and reduced atpase activity in multidrug transporter ABCG2. *Int J Cancer* **109**, 238–246 (2004).
87. Morisaki, K. *et al.* Single nucleotide polymorphisms modify the transporter activity of ABCG2. *Cancer Chemother Pharmacol* **56**, 161–172 (2005).

88. Jackson, S. M. *et al.* Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. *Nature structural & molecular biology* **25**, 333–340 (2018).
89. Orlando, B. J. & Liao, M. ABCG2 transports anticancer drugs via a closed-to-open switch. *Nature communications* **11**, 2264 (2020).
90. Ferreira, R. J., Bonito, C. A., Cordeiro, M. N. D. S., Ferreira, M.-J. U. & Dos Santos, D. J. V. A. Structure-function relationships in ABCG2: insights from molecular dynamics simulations and molecular docking studies. *Scientific reports* **7**, 15534 (2017).
91. Khunweeraphong, N., Szöllősi, D., Stockner, T. & Kuchler, K. The ABCG2 multidrug transporter is a pump gated by a valve and an extracellular lid. *Nature communications* **10**, 5433 (2019).
92. Khunweeraphong, N., Stockner, T. & Kuchler, K. The structure of the human ABC transporter ABCG2 reveals a novel mechanism for drug extrusion. *Sci Rep* **7**, 13767 (2017).
93. Ibrahim, M. A. A. *et al.* In Silico Targeting Human Multidrug Transporter ABCG2 in Breast Cancer: Database Screening, Molecular Docking, and Molecular Dynamics Study. *Mol Inform* **41**, e2060039 (2022).
94. Zattoni, I. F. *et al.* A new porphyrin as selective substrate-based inhibitor of breast cancer resistance protein (BCRP/ABCG2). *Chem Biol Interact* **351**, 109718 (2022).
95. Telbisz, A. *et al.* Membrane cholesterol selectively modulates the activity of the human ABCG2 multidrug transporter. *Biochim Biophys Acta* **1768**, 2698–2713 (2007).
96. Hegedüs, C., Telbisz, Á., Hegedűs, T., Sarkadi, B. & Özvegy-Laczka, C. Lipid regulation of the ABCB1 and ABCG2 multidrug transporters. *Advances in cancer research* **125**, 97–137 (2015).
97. van Meer, G. Lipid traffic in animal cells. *Annu Rev Cell Biol* **5**, 247–275 (1989).
98. Manolaridis, I. *et al.* Cryo-EM structures of a human ABCG2 mutant trapped in ATP-bound and substrate-bound states. *Nature* **563**, 426–430 (2018).
99. Schmidtke, P., Bidon-Chanal, A., Luque, F. J. & Barril, X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **27**, 3276–3285 (2011).
100. Nakayama, A. *et al.* ABCG2 is a high-capacity urate transporter and its genetic impairment increases serum uric acid levels in humans. *Nucleosides Nucleotides Nucleic Acids* **30**, 1091–1097 (2011).
101. Sutto, L., Marsili, S. & Gervasio, F. L. New advances in metadynamics. *WIREs Computational Molecular Science* **2**, 771–779 (2012).
102. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **185**, 604–613 (2014).
103. Orengo, C. A., Todd, A. E. & Thornton, J. M. From protein structure to function. *Curr Opin Struct Biol* **9**, 374–382 (1999).
104. Higgins, D. G., Thompson, J. D. & Gibson, T. J. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* **266**, 383–402 (1996).
105. Caporaso, J. G., Baumgartner, W. A., Randolph, D. A., Cohen, K. B. & Hunter, L. MutationFinder: a high-performance system for extracting point mutation mentions from text. *Bioinformatics* **23**, 1862–1865 (2007).
106. Hecht, M., Bromberg, Y. & Rost, B. Better prediction of functional effects for sequence variants. *BMC Genomics* **16 Suppl 8**, S1 (2015).
107. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).
108. Burley, S. K. *et al.* RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci* **27**, 316–330 (2018).

109. Gore, S. *et al.* Validation of Structures in the Protein Data Bank. *Structure (London, England : 1993)* **25**, 1916–1927 (2017).
110. Chang, X. B., Hou, Y. X., Jensen, T. J. & Riordan, J. R. Mapping of cystic fibrosis transmembrane conductance regulator membrane topology by glycosylation site insertion. *J. Biol. Chem.* **269**, 18572–18575 (1994).
111. Zagotta, W. N., Gordon, M. T., Senning, E. N., Munari, M. A. & Gordon, S. E. Measuring distances between TRPV1 and the plasma membrane using a noncanonical amino acid and transition metal ion FRET. *J. Gen. Physiol.* **147**, 201–216 (2016).
112. Tusnady, G. E., Dosztanyi, Z. & Simon, I. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. *Bioinformatics* **21**, 1276–1277 (2005).
113. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research* **40**, D370–D376 (2012).
114. Stansfeld, P. J. *et al.* MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure (London, England : 1993)* **23**, 1350–1361 (2015).
115. Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. *Nat Rev Mol Cell Biol* **20**, 681–697 (2019).
116. Won, J., Baek, M., Monastyrskyy, B., Kryshtafovych, A. & Seok, C. Assessment of protein model structure accuracy estimation in CASP13: Challenges in the era of deep learning. *Proteins* **87**, 1351–1360 (2019).
117. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* 1–9 (2021) doi:10.1038/s41586-021-03828-1.
118. Srikant, S. Evolutionary history of ATP-binding cassette proteins. *FEBS Lett* **594**, 3882–3897 (2020).
119. Thomas, C. *et al.* Structural and functional diversity calls for a new classification of ABC transporters. *FEBS Lett* **594**, 3767–3775 (2020).
120. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710 (2004).
121. Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).
122. Debruycker, V. *et al.* An embedded lipid in the multidrug transporter LmrP suggests a mechanism for polyspecificity. *Nat Struct Mol Biol* **27**, 829–835 (2020).
123. Andreeva, A., Kulesha, E., Gough, J. & Murzin, A. G. The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Research* **48**, D376–D382 (2020).
124. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412–D419 (2021).
125. Chi, X. *et al.* Structural mechanism of phospholipids translocation by MlaFEDB complex. *Cell Res* **30**, 1127–1135 (2020).
126. Song, D., Jiao, H. & Liu, Z. Phospholipid translocation captured in a bifunctional membrane protein MprF. *Nat Commun* **12**, 2927 (2021).
127. Ernst, C. M. & Peschel, A. MprF-mediated daptomycin resistance. *Int J Med Microbiol* **309**, 359–363 (2019).
128. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2021) doi:10.1101/2021.10.04.463034.
129. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. & Obradović, Z. Intrinsic disorder and protein function. *Biochemistry* **41**, 6573–6582 (2002).

130. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197–208 (2005).
131. Classification of intrinsically disordered regions and proteins - PubMed. https://pubmed.ncbi.nlm.nih.gov/24773235/.
132. Oldfield, C. J. *et al.* Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* **44**, 12454–12470 (2005).
133. Mohan, A. *et al.* Analysis of molecular recognition features (MoRFs). *Journal of molecular biology* **362**, 1043–1059 (2006).
134. Camilloni, C., De Simone, A., Vranken, W. F. & Vendruscolo, M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* **51**, 2224–2231 (2012).
135. Berjanskii, M. V. & Wishart, D. S. A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* **127**, 14970–14971 (2005).
136. Kumar, M. *et al.* ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res* **48**, D296–D306 (2020).
137. dbSNP: the NCBI database of genetic variation - PubMed. https://pubmed.ncbi.nlm.nih.gov/11125122/.
138. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* **33**, D514-517 (2005).

9. A bemutatott közlemények

ARTICLE
dc_2020_22
PUBLISHED ONLINE: 12 MAY 2013 | DOI: 10.1038/NCHEMBIO.1253
nature chemical biology

Mechanism-based corrector combination restores ΔF508-CFTR folding and function

Tsukasa Okiyoneda[1,13], Guido Veit[1], Johanna F Dekkers[2–4], Miklos Bagdany[1], Naoto Soya[1], Haijin Xu[1], Ariel Roldan[1], Alan S Verkman[5,6], Mark Kurth[7], Agnes Simon[8], Tamas Hegedus[9,10], Jeffrey M Beekman[2–4] & Gergely L Lukacs[1,11,12]*

The most common cystic fibrosis mutation, ΔF508 in nucleotide binding domain 1 (NBD1), impairs cystic fibrosis transmembrane conductance regulator (CFTR)-coupled domain folding, plasma membrane expression, function and stability. VX-809, a promising investigational corrector of ΔF508-CFTR misprocessing, has limited clinical benefit and an incompletely understood mechanism, hampering drug development. Given the effect of second-site suppressor mutations, robust ΔF508-CFTR correction most likely requires stabilization of NBD1 energetics and the interface between membrane-spanning domains (MSDs) and NBD1, which are both established primary conformational defects. Here we elucidate the molecular targets of available correctors: class I stabilizes the NBD1-MSD1 and NBD1-MSD2 interfaces, and class II targets NBD2. Only chemical chaperones, surrogates of class III correctors, stabilize human ΔF508-NBD1. Although VX-809 can correct missense mutations primarily destabilizing the NBD1-MSD1/2 interface, functional plasma membrane expression of ΔF508-CFTR also requires compounds that counteract the NBD1 and NBD2 stability defects in cystic fibrosis bronchial epithelial cells and intestinal organoids. Thus, the combination of structure-guided correctors represents an effective approach for cystic fibrosis therapy.

CFTR is an ATP-binding cassette transporter and functions as a cyclic AMP–dependent Cl^- channel at the apical plasma membrane of epithelial cells[1]. CFTR comprises two membrane-spanning domains (MSD1 and MSD2) with four cytosolic loops (CL1–4) and three cytosolic domains: a regulatory domain and two nucleotide-binding domains (NBD1 and NBD2). The coupling helices of CL2 (MSD1) and ICL4 (MSD2) reach across and interact with the NBD of the opposite subunit, providing noncovalent interactions between the MSDs and the NBDs. These domain-domain interfaces seem to be critical in the structural and functional integrity of the channel[2–4] (**Fig. 1a**). Newly synthesized CFTR is N-glycosylated and undergoes cotranslational domain folding and post-translational, coupled-domain assembly in the endoplasmic reticulum[2,5–9], aided by a network of molecular chaperones[10,11] (**Supplementary Fig. 1a**). Upon bypassing the endoplasmic reticulum quality control checkpoints and traversing the Golgi complex, the native CFTR undergoes complex glycosylation[1,11].

Cystic fibrosis, one of the most common inherited diseases in the Caucasian population, is caused by loss-of-function mutations in CFTR that lead to the imbalance of airway surface fluid homeostasis, mucus dehydration, hyperinflammation, bacterial colonization and, consequently, recurrent lung infection, the primary cause of morbidity and mortality in cystic fibrosis patients[1,12,13]. Deletion of Phe508 (ΔF508) in the NBD1, the most prevalent CFTR mutation present in ~90% of cystic fibrosis patients, causes global misfolding and endoplasmic reticulum–associated degradation via the ubiquitin proteasome system, resulting in marginal plasma membrane expression of the partially functional ΔF508-CFTR[1,6,11,14]. The residual plasma membrane channel activity could be enhanced by modulating ΔF508-CFTR's biosynthetic processing, peripheral stability and channel gating through low temperature, chemical chaperones or small molecules[14–19].

Although symptomatic therapies have increased the life expectancy of cystic fibrosis patients, considerable efforts have been devoted to identifying compounds that can increase either the mutant's biosynthetic processing efficiency and cell-surface density ('correctors') or the activity of resident mutant CFTR in the plasma membrane ('potentiators')[19]. VX-770 (Ivacaftor), the only approved potentiator drug, substantially enhances the channel function of several mutants[20], unlike available correctors that are modestly effective and have incompletely understood mechanisms[14,17,21]. In principle, correctors may facilitate ΔF508-CFTR folding via direct binding as pharmacological chaperones[22] or indirectly as chemical chaperones (such as glycerol, trimethylamine *N*-oxide (TMAO) and myo-inositol)[16,23]. Correctors may also enhance the mutant's functional expression as proteostasis regulators[11] by modulating the cellular machineries responsible for folding, degradation and vesicular trafficking of ΔF508-CFTR (for example, 4-phenylbutyrate, HDAC inhibitor, HSF1 inducers and kinase inhibitors; reviewed in refs. 11,19). VX-809, the most promising investigational corrector, restores mutant plasma membrane expression and function to <15% that of wild-type CFTR in immortalized and primary human bronchial epithelia[14]. Although circumstantial evidence suggests that VX-809, like some of the previously identified correctors

[1]Department of Physiology, McGill University, Montréal, Quebec, Canada. [2]Department of Pediatric Pulmonology, University Medical Centre, Utrecht, The Netherlands. [3]Department of Immunology, University Medical Centre, Utrecht, The Netherlands. [4]Centre for Molecular and Cellular Intervention, University Medical Centre, Utrecht, The Netherlands. [5]Department of Medicine, University of California–San Francisco, San Francisco, California, USA. [6]Department of Physiology, University of California–San Francisco, San Francisco, California, USA. [7]Department of Chemistry, University of California–Davis, Davis, California. [8]Department of Molecular Pharmacology, Research Center for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary. [9]MTA-SE Molecular Biophysics Research Group, Hungarian Academy of Sciences, Budapest, Hungary. [10]Department of Biophysics and Radiation Biology, Semmelweis University, Budapest, Hungary. [11]Department of Biochemistry, McGill University, Montréal, Quebec, Canada. [12]GRASP, McGill University, Montréal, Quebec, Canada. [13]Present address: Department of Bioscience, School of Science and Technology, Kwansei Gakuin University, Hyogo, Japan. *e-mail: gergely.lukacs@mcgill.ca

(for example, C3 (VRT-325) and C4 (corr-4a)) and potentiators (for example, P1 (VRT-532) and VX-770), directly targets ΔF508-CFTR during endoplasmic reticulum biogenesis, the molecular basis of correction remains elusive[14,24]. Although VX-809–induced ΔF508-CFTR activity at the plasma membrane is potentiated two-fold with VX-770 in preclinical settings[14], additional improvement seems to be required in corrector efficacy to achieve substantial clinical benefit in cystic fibrosis[21].

Recent studies revealed that ΔF508-NBD1 is thermodynamically and kinetically destabilized at physiological temperature and suggested that NBD1 stabilization would effectively counteract ΔF508-CFTR misprocessing[25,26]. Substantial stabilization of ΔF508-NBD1 by suppressor mutations, however, resulted in a limited increase in folding and plasma membrane expression of ΔF508-CFTR[27,28]. Remarkably, robust rescue (65–80%) of ΔF508-CFTR folding could be achieved by suppressor mutations that simultaneously stabilized ΔF508-NBD1 and the NBD1 interface with the CL4 coupling helix in the MSD2 (refs. 27,28). These observations suggest that efficient correction of both primary structural defects is necessary and sufficient to restore ΔF508-CFTR function to the wild-type level in most cystic fibrosis patients. As a corollary, it is conceivable that correction of one of the primary (NBD1 or the NBD1-MSD2 interface) or secondary (for example, NBD2 misfolding[5]) structural defects accounts for the limited ΔF508-CFTR rescue efficiency of correctors identified to date[11,14,17,19,21,22,29]. Furthermore, combinations of correctors that counteract distinct conformational defects may potentiate each other's effect on ΔF508-CFTR folding and function, in analogy to suppressor mutations of the primary folding defects[27,28] (**Supplementary Fig. 1b**).

On the basis of *in vitro* studies of isolated ΔF508-NBD1 in combination with *in vivo* CFTR processing, functional assays and *in silico* docking analysis as well as published data[22], we propose that correctors can be mechanistically classified into three groups. We also show that the efficacy of VX-809, which targets the NBD1-MSD1/2 interface, can be potentiated with two other classes of chemicals, as determined by ΔF508-CFTR folding, plasma membrane expression, stability and function in cell culture models and intestinal organoids from cystic fibrosis patients. Thus, targeting multiple conformational defects enables us to counteract ΔF508-CFTR misfolding, demonstrating that the structure-guided combination of correctors may provide an effective therapeutic strategy in cystic fibrosis.

RESULTS

Class I corrector plus NBD1 stabilization restores folding

We postulated that small-molecule combinations targeting both primary folding defects would achieve wild type–like folding of ΔF508-CFTR, as documented in the case of second-site suppressor mutations confined to the NBD1 and the NBD1-MSD2 interface[27,28]. To categorize the available corrector molecules (**Supplementary Results**, **Supplementary Table 1**) according to their preferential targets, we determined the rescue efficiency of ΔF508-CFTR containing either genetically stabilized NBD1 (stabilized through the quadruple mutation G550E R553Q R555K F494N (referred to collectively as the R1S mutation; second-site mutations are defined in **Supplementary Table 2**) or the stabilized NBD1-MSD2 interface (stabilized by the R1070W mutation in the CL4; **Fig. 1a**). We confirmed that the R1070W and R1S suppressor mutations modestly increased the plasma membrane density of ΔF508-CFTR from <2% to 7% and 16% that of the wild type, respectively, as determined by cell-surface ELISA in baby hamster kidney (BHK) cells[27] (**Supplementary Fig. 2a**). VX-809, C18 (a VX-809 analog) and C3, the most effective correctors in our screen, increased the plasma membrane density of ΔF508-CFTR^{R1S} (Res^{R1S}) and ΔF508-CFTR^{R1070W} (Res^{R1070W}) to 93–100% and 30–34% that of the wild type, respectively (**Fig. 1b**). This observation suggests that VX-809, C18 and C3 preferentially target the NBD1-MSD2 interface over the NBD1 energetic defect and that they are designated as class I correctors (**Fig. 1b** and **Supplementary Fig. 2a**). The preferential rescue of the NBD1-MSD2 interface and NBD1 stability defects was estimated by the augmented plasma membrane density of Res^{R1S} and Res^{R1070W}, respectively, a measure that indicates the capacity of correctors to synergize with the respective suppressor mutation (**Supplementary Fig. 2a**). The ratio of Res^{R1S} over Res^{R1070W} suggested that VX-809, C18 and C3 have a stabilizing effect on the NBD1-MSD2 interface that is ~3.5-fold stronger than their effect on NBD1 (**Fig. 1c**). This inference was confirmed by using ΔF508-CFTR variants containing the functionally analogous suppressor mutations F494N Q637R F429S (referred to as '3S' mutation), G550E R553Q R555K ('R' mutation) or V510D; the 3S and R mutations stabilize the NBD1, and V510D stabilizes the NBD1-MSD2 interface[7,27,30,31]. VX-809, C18 or C3 increased the plasma membrane density of ΔF508-CFTR^{3S} and ΔF508-CFTR^{R} from 7–25% to 60–120% of the wild type, respectively, and enhanced their complex glycosylation but was less effective in the case of ΔF508-CFTR^{V510D}, which responded similarly to Res^{R1070W} (**Supplementary Fig. 2b**,**c**). A similar rescue efficiency was obtained in the presence of VX-809, C18 or C3 by analyzing the magnitude of fold correction of the cell-surface density relative to that induced by the respective second-site mutation (or mutations) alone. The reduced rescue efficiency of ΔF508-CFTR^{3S} relative to Res^{R1S} by class I correctors may be attributed to the more efficient stabilization of the NBD1-NBD2 interface by the R1S suppressor mutation.

Metabolic pulse-chase experiments confirmed that only the combination of VX-809, C18 or C3 with NBD1 stabilization (R1S) enhanced ΔF508-CFTR folding efficiency near to that of the wild type (22–32%), indicating a four- to eight-fold potentiation between class I correctors and the effect of the R1S mutation (**Fig. 1d** and **Supplementary Fig. 2d**), but ΔF508-CFTR folding was modestly enhanced by this combination of correctors in the presence of mutations conferring interface stabilization (R1070W or V510D). Dual-acting compounds such as C15 (corr-2b[32]) and CoPo-22 (CoPo[33]), which combine both corrector and potentiator effects, only slightly enhanced the plasma membrane density of ΔF508-CFTR variants (**Supplementary Fig. 2a**).

The wild type–like ΔF508-CFTR complex glycosylation, function and plasma membrane density were reproduced by genetic stabilization of NBD1 and the NBD1-MSD2 interface as well as by the combination of class I correctors and NBD1 suppressor mutations in polarized human immortalized cystic fibrosis bronchial respiratory epithelial (CFBE41o-) cells (**Supplementary Fig. 3**). Class I correctors increased the plasma membrane density, complex glycosylation and apical Cl^- current (ICl(apical)) of ΔF508-CFTR to ~100–130% and ~30% that of the wild type in the presence of the R1S and R1070W mutations, respectively, in CFBE41o- cells expressing the mutant channel under the control of a tetracycline-responsive promoter (**Fig. 1e**,**f** and **Supplementary Fig. 3e**–**g**), consistent with preferential stabilization of the NBD1-MSD2 interface by class I correctors in CFBE41o- cells.

To evaluate the efficiency of class I correctors as a function of the ΔF508-NBD1 conformational stability, we established the relationship between NBD1 stability (reflected by the T_m values of the 0S, 1S (F494N), 3S or R4S suppressor mutants of ΔF508-NBD1 or wild-type NBD1, as described in **Supplementary Table 2**) and the plasma membrane density of the respective CFTR variants. The plasma membrane expression of CFTR in the wild type was dependent on NBD1 stability with an eight-fold steeper slope (12.9 ± 3.1% per °C) compared to that of ΔF508-CFTR (1.7 ± 0.2% per °C), confirming the critical role of the Phe508 side chain in CFTR-coupled domain folding[1,6] (**Fig. 1g**). With effects similar to those of the R1070W mutation[27] (9.2 ± 2.3% per °C), VX-809 (9.8 ± 0.7% per °C), C18 (9.6 ± 1.3% per °C) and C3 (10.8 ± 2.1% per °C) restored wild type–like coupling between NBD1 stability

dc_2020_22

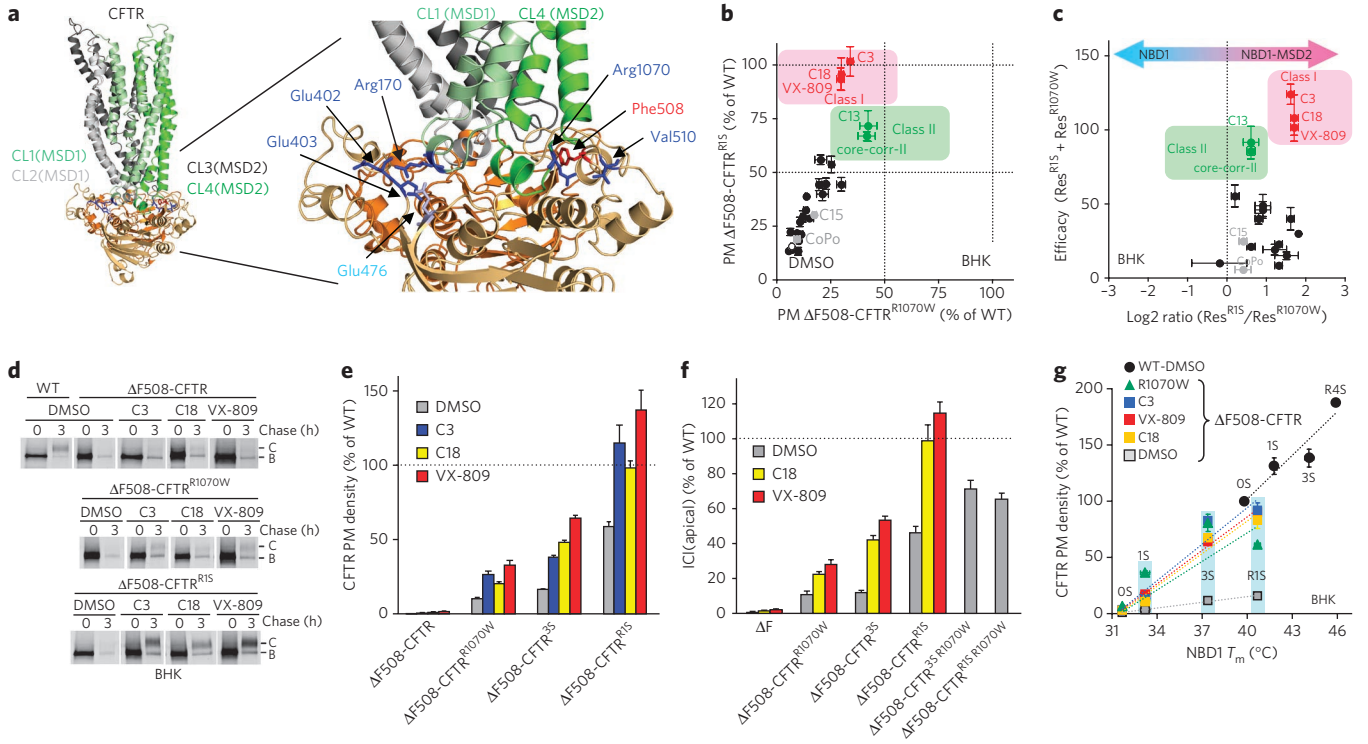


Figure 1 | Combination of corrector and suppressor mutations restores ΔF508-CFTR folding, plasma membrane expression and function. (**a**) NBDs-MSD1/2 interfaces in the CFTR open structural model[36]. Some critical interface residues are indicated. (**b**) Relative plasma membrane (PM) density of Res^{R1S} (*y* axis) or Res^{R1070W} (*x* axis) upon corrector treatment was measured by ELISA in BHK cells (*n* = 6–12). Correctors indicated by black symbols are listed in **Supplementary Figure 2a**. WT, wild type. (**c**) Structural preference of correctors to NBD1-MSD2 interface over NBD1 stability defect was visualized by plotting sum of correction (Res^{R1S} + Res^{R1070W}, *y* axis) as a function of the $\log_2$ ratio of the augmented plasma membrane density of Res^{R1S} and Res^{R1070W} (Res^{R1S}/Res^{R1070W}, *x* axis). (**d**) The maturation efficiency of ΔF508-CFTR was measured by metabolic pulse-chase experiments. B, immature core-glycosylated form; C, mature complex-glycosylated form. (**e**,**f**) Plasma membrane density (**e**; *n* = 8–20) and function (**f**; *n* = 3–4) of ΔF508-CFTR with or without suppressor mutations was measured by ELISA and (ICl(apical), respectively, in CFBE41o- cells. Na^+/K^+-ATPase (ATPase) was used as a loading control. (**g**) Correlation between the T_m of NBD1 variants of 0S, 1S, 3S, R1S and R4S (listed in **Supplementary Table 2**) and plasma membrane density of the respective CFTR variants in BHK cells (*n* = 8–12) in the presence or absence of correctors. The data were fitted by linear regression analysis. Correctors (C3 at 10 μM; C18 and VX-809 at 3 μM) were applied for 24 h at 37 °C. Data represent mean ± s.e.m.

(T_m) and ΔF508-CFTR plasma membrane expression, supporting the possibility that class I correctors stabilize the NBD1-MSD2 interface (**Fig. 1g**). Notably, the improvement of wild-type CFTR plasma membrane expression by NBD1 stabilization reinforces the notion that the inherent conformational fluctuation of NBD1 partly accounts for the limited processing efficiency of the wild-type channel[27] (**Fig. 1g**).

Mechanisms of action of CFTR correction

To assess whether VX-809 or C18 can directly bind ΔF508-CFTR as a pharmacological chaperone, we determined the functional stability of the temperature-rescued mutant channel in an artificial planar phospholipid bilayer[27,34]. These experiments used ΔF508-CFTR with the R29K and R555K mutations ($\Delta F508\text{-}CFTR^{2RK}$) to improve the channel reconstitution success by modestly enhancing plasma membrane expression after the 26 °C rescue while preserving the ΔF508-CFTR channel gating and functional stability defects, as reported before[34] (**Supplementary Fig. 4a–c**). The open probability (P_o) of the phosphorylated $\Delta F508\text{-}CFTR^{2RK}$ channel was progressively decreased from 0.21 to 0.09 upon raising the temperature from 24 °C to 36 °C[34,35] (**Fig. 2a**,**b**). In contrast, the P_o of the wild type was increased from 0.35 to 0.47 (**Fig. 2a**,**b**). VX-809 and C18 prevented the thermal inactivation of the mutant and maintained P_o values of 0.21 and 0.26, respectively, at 36 °C during the course of the measurement (**Fig. 2a**,**b** and **Supplementary Fig. 4d**), providing evidence that these compounds can directly interact with the channel.

To delineate potential VX-809 target sites, we performed *in silico* docking on the NBD1 crystal structure (Protein Data Bank (PDB) code 2BBT) and two CFTR homology models[3,36] following the deletion of Phe508 using AutoDock. VX-809 was docked to the NBD1-CL1(MSD1), NBD1-CL4(MSD2) and NBD1-NBD2 interfaces (**Fig. 2c**, **Supplementary Fig. 5** and **Supplementary Table 3**). Some of these sites have been reported as putative corrector binding sites[37,38]. To test the *in silico* predictions, we disrupted the NBD1-CL4 and NBD1-CL1 interfaces with the R1070W or F508G mutations[5,6,27,39] and the cystic fibrosis–causing R170G mutation (http://www.genet.sickkids.on.ca/), respectively, in wild-type CFTR. The R170G substitution probably disrupts the electrostatic interaction between Arg170 (CL1) and Glu402, Glu403 and/or Glu476 (NBD1) (**Fig. 1a**). The F508G mutation interferes with hydrophobic patch formation at the NBD1-CL4 interface but only marginally compromises NBD1 energetics[3,27]. These events lead to the destabilization of the joint CL1-CL4 interface with NBD1 and the subsequent misfolding of multiple CFTR domains[6]. Remarkably, the severe processing and plasma membrane expression defects of both CL1 and CL4 interface mutants were largely corrected by VX-809, C18 and C3 (**Fig. 3a–c**). Conversely, stabilization of the NBD1-CL4 interface through the V510D or R1070W mutations in the absence of Phe508 attenuated the relative efficiency of

dc_2020_22

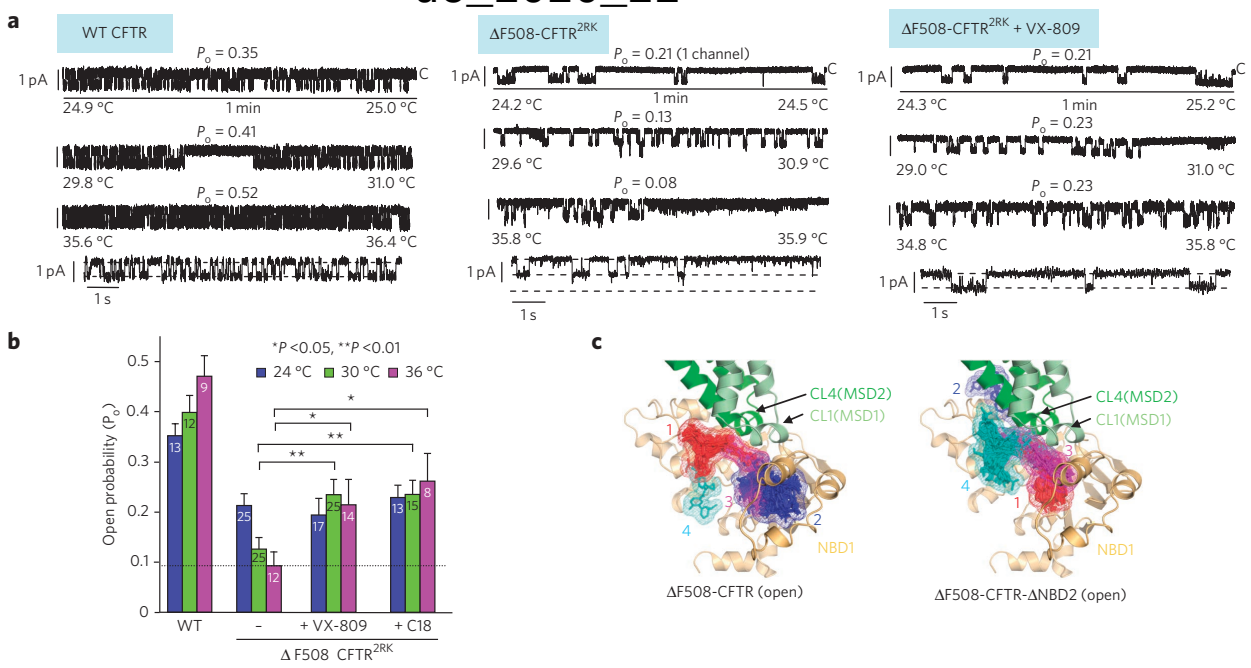


Figure 2 | VX-809 functions as a pharmacological chaperone of CFTR. (**a**,**b**) The effect of VX-809 and C18 correctors on thermal inactivation of ΔF508-CFTR^{2RK} reconstituted into an artificial phospholipid bilayer. In **a**, representative channel activity is shown for the wild type or ΔF508-CFTR^{2RK} at 24 °C, 30 °C or 36 °C during the temperature ramp in the absence or presence of 3 μM VX-809. The processing defect of the ΔF508-CFTR variants was rescued at 26 °C before microsome isolation. Channel activity is also shown at higher time resolution for each condition at 36 °C. The closed (C) state of the channels is indicated. The control gating of ΔF508-CFTR^{2RK} was derived from separate experiments at 24 °C and 27 °C plus 36 °C. Incorporation of two channels was observed at 27–36 °C. The ΔF508-CFTR^{2RK} activity was recorded in separate experiment at 24 °C plus 30 °C and 36 °C in presence of VX-809. (**b**) The P_o of protein kinase A–activated CFTRs was analyzed at the indicated temperature as described in **a**. The cumulative duration of single-channel experiments for the wild type and ΔF508-CFTR^{2RK} was at least 15 min. The number of independent experiments is indicated in the bars (n = 8–25). Data represent mean ± s.e.m. Significance was tested by paired Student's t-test. *P < 0.05, **P < 0.01. (**c**) VX-809 *in silico* docking to open ΔF508-CFTR (top) or ΔF508-CFTR-ΔNBD2 model (bottom) obtained by AutoDock. The first four VX-809 clusters are ranked on the basis of their lowest-binding-energy pose in ascending order and are illustrated on the model using PyMOL. For clarity, NBD2 and the R domain are hidden from the full-length model. Red, blue, magenta and cyan represent clusters of VX-809 with increasing binding free energy.

class I correctors but not that of C4 on ΔF508-CFTR^{3S} (**Fig. 3d** and **Supplementary Fig. 6a**). Moreover, the V510D mutation, which improves wild-type expression conceivably by stabilizing the NBD1-CL4 and NBD1-CL1 interfaces and marginally stabilizing the NBD1 itself[27], also attenuated the effect of class I correctors on wild-type CFTR (**Supplementary Fig. 6b**). To our surprise, the V510D but not the R1S mutation substantially corrected the R170G defect (**Supplementary Fig. 6c**), suggesting that the NBD1-CL1 interface defect is coupled to that of NBD1-CL4. The rescue of CFTR^{R170G} expression by class I correctors, however, was partially diminished by the V510D mutation (**Supplementary Fig. 6d**). In addition to being consistent with *in silico* docking prediction (**Supplementary Fig. 5a**,**c**,**e**), the stronger effect of VX-809 on the CL1 mutants compared to transmembrane helix 1 mutants such as G85E and G91R (**Supplementary Fig. 7a**) was also consistent with the model that VX-809 targets the interdomain interface composed of NBD1 and the CL1 plus CL4, which contains the respective coupling helices[1].

To further dissect the VX-809 target, we assessed its effect on the steady-state expression of wild-type CFTR domain combinations representing stalled biosynthetic folding intermediates[6,40]. This approach is favored by the observation that the conformation of early folding intermediates is similar between the wild type and ΔF508-CFTR[41]. VX-809 increased the amount of wild-type and ΔF508-MSD1-NBD1 domain combination (also known as M1N1) equally as well as it increased the amount of wild-type MSD1 and MSD1-NBD1-R domain combination, but there was no increase for the domain combinations that lacked MSD1 (**Supplementary Fig. 7a–e**), suggesting that VX-809 targeting of MSD1 is a prerequisite for stabilizing the domain interface composed of CL1, CL4 and NBD1.

To determine the possible role of VX-809 at the NBD1-NBD2 interface, we used NBD2 deletion variants (CFTR-ΔNBD2) because they have folding characteristics similar to those of wild-type CFTR (refs. 6–8). Class I correctors and the R1S mutation rescued ΔF508-CFTR-ΔNBD2 plasma membrane expression to 50–75% that of its wild-type counterpart, as in the case of the full-length molecule (**Fig. 3e**,**f**), ruling out any substantial role of NBD1-NBD2 or the NBD2-MSD1 and NBD2-MSD2 interfaces in CFTR correction (**Supplementary Fig. 5c–f**). This result was in agreement with the prediction of preserved VX-809 binding sites in the CFTR-ΔNBD2 model (**Fig. 2c**). In contrast, NBD2 deletion virtually eliminated the ability of C4 (a C13 analog)[17] or core-corr-II[42] to correct ΔF508-CFTR^{R1S} (**Fig. 3e–f**). Therefore, we designated C4 and core-corr-II as members of class II (**Fig. 1b**,**c** and **Supplementary Fig. 2a**). That the class II corrector effect was prevented by the NBD2 deletion is consistent with *in silico* docking predictions, supporting that the putative target of C4 and core-corr-II is the NBD1-NBD2 and/or NBD2-MSD1/2 interface (**Supplementary Figs. 8** and **9**), and with the weaker rescue efficiency of class II correctors on the NBD1-MSD1/2 interface mutants (R1070W, F508G or R170G) compared to that of class I correctors (**Fig. 3a–c**). These results are also in line with the observation that the C4 rescue effect is exerted only after MSD2 translation is completed[40].

dc_2020_22

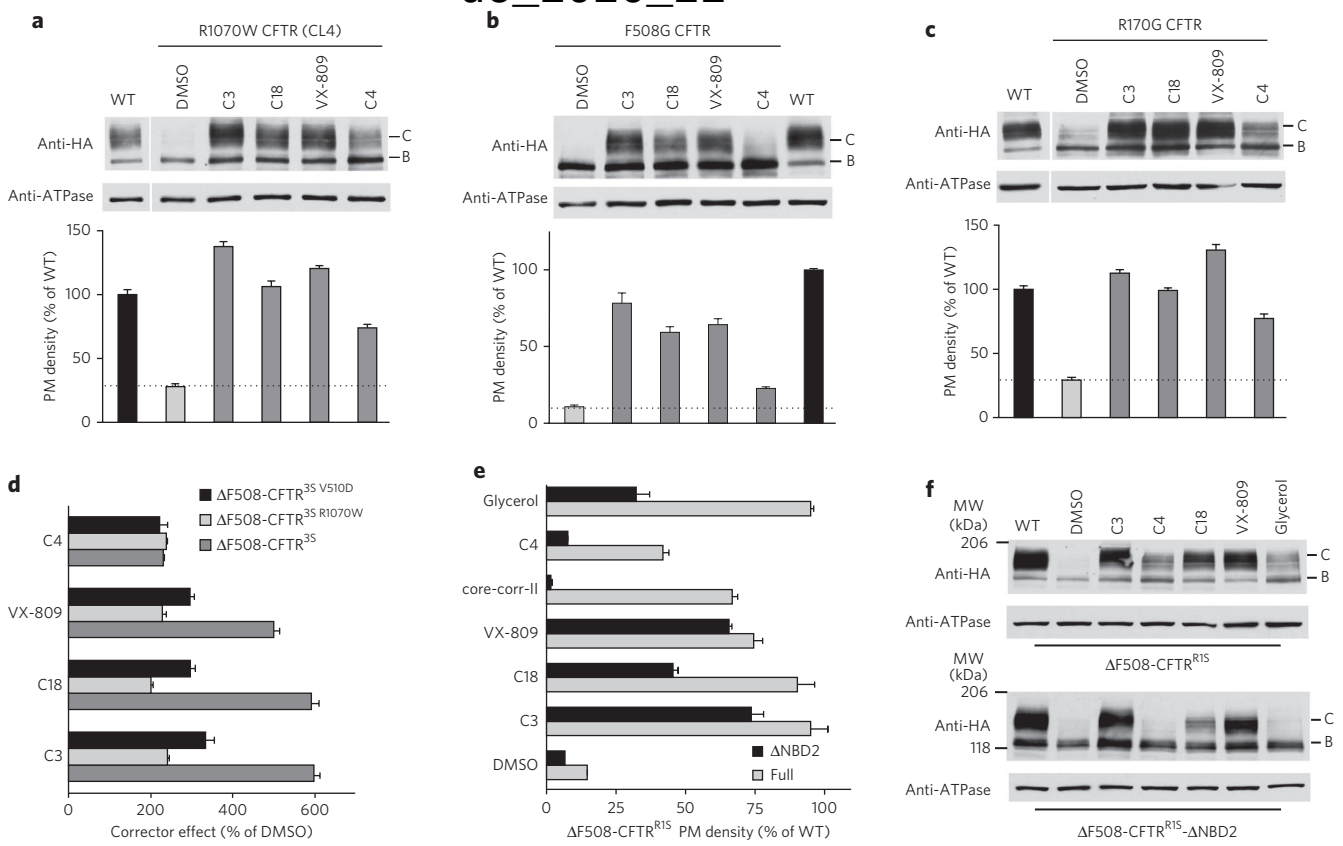


Figure 3 | Evaluating corrector mechanism by using CFTR variants. (**a**–**c**) Effect of correctors (10 μM C3 or C4, 3 μM C18 or VX-809 for 24 h at 37 °C) on cellular expression (top) and plasma membrane density (PM, bottom) of R1070W CFTR (**a**), F508G CFTR (**b**) and R170G CFTR (**c**) in BHK cells, measured by western blotting and cell-surface ELISA (n = 8). WT, wild type. Anti-HA, antibody to hemagglutinin; B, immature core-glycosylated form; C, mature complex-glycosylated form. (**d**) Effect of V510D or R1070W mutation on the plasma membrane density of corrector-treated ΔF508-CFTR^{3S} measured by cell-surface ELISA in BHK cells and expressed as a percentage of the respective DMSO control (n = 8). The same results are also shown in **Supplementary Figure 6a** as a percentage of the wild type. (**e**,**f**) Effect of NBD2 deletion (ΔNBD2) on corrector effect as measured by the plasma membrane density (**e**, n = 6–9) and cellular expression (**f**) of ΔF508-CFTR^{R1S} in BHK cells. Correctors (5 μg ml^{-1} core-corr-II, 10% glycerol; other concentrations are the same as in **a**–**c**) were added for 24 h at 37 °C. Data represent mean ± s.e.m. MW, molecular weight.

Although *in silico* docking predicted the regulatory insertion, a 32-residue segment between the first two β-strands in the NBD1 (ref. 43), as a possible binding site of class II correctors in NBD1 (**Supplementary Figs. 8** and **9**), this scenario was ruled out because C4 remained effective for the ΔF508-CFTR-ΔRI containing energetically stabilized NBD1 (refs. 25,26) (**Supplementary Fig. 9f**).

Chemical chaperones stabilize human ΔF508-NBD1

Although none of the correctors tested seems to preferentially stabilize the NBD1, according to the structural complementation analysis (**Fig. 1b**,**c**), it remains possible that correctors can exert a limited conformational stabilization effect on human ΔF508-NBD1, similarly to RDR1, a compound that directly binds and stabilizes mouse ΔF508-NBD1 (ref. 22). To assess this possibility, we first monitored the thermal unfolding propensity of the domain by differential scanning fluorimetry (DSF)[27]. Initial studies were carried out on human ΔF508-NBD1 containing the single solubilizing mutation F494N (ΔF508-NBD1^{1S}), followed by validation on native ΔF508-NBD1 with protein yield reduced by several factors and increased thermal sensitivity. Most of the available correctors, including RDR1 (ref. 22), failed to counteract the conformational instability of human ΔF508-NBD1^{1S} and ΔF508-NBD1, as reflected by their 8- to 9-°C-lower T_m relative to their wild-type counterparts (**Fig. 4a**–**c** and **Supplementary Fig. 10a**). C6, C11 (dynasore) or C12 weakly increased the T_m of ΔF508-NBD1^{1S}, but this effect was minimal on ΔF508-NBD1 (**Fig. 4b**,**c** and **Supplementary Fig. 10a**). Accordingly, C11 neither attenuated the *in vitro* ubiquitination of unfolded NBD1 by the chaperone-dependent E3 ubiquitin ligase CHIP[27] (**Supplementary Fig. 10b**) nor enhanced the plasma membrane expression of the chimera formed by ΔF508-NBD1^{1S} anchored to truncated CD4 (CD4T-ΔF508-NBD1^{1S}; described below). In contrast, ATP and chemical chaperones (glycerol, TMAO, myo-inositol and D-sorbitol) substantially enhanced the T_m of ΔF508-NBD1^{1S} and ΔF508-NBD1 (**Fig. 4a**–**c** and **Supplementary Fig. 10a**). The DSF results were substantiated by monitoring the *in vivo* folding of NBD1 in the context of the CD4T-ΔF508-NBD1^{1S} chimera. The measurement of the chimera plasma membrane density, as a validated surrogate readout of the biosynthetic processing efficiency[6,27], showed that only chemical chaperones (for example, glycerol and myo-inositol) and solubilizing mutations (R4S) but none of the correctors stabilized cytosolic ΔF508-NBD1^{1S} anchored to the truncated CD4 reporter molecule (**Fig. 4d**).

Considering the enhanced conformational dynamics of residues 507–511 in isolated ΔF508-NBD1^{3S}, as shown by hydrogen deuterium exchange and MS (HDX-MS)[44], we tested whether VX-809 could limit the conformational dynamics of this region. We confirmed the accelerated deuteration of the peptide spanning residues 505–509 in ΔF508-NBD1^{1S} relative to its wild-type counterpart[44] but not that of the remaining >60 peptides, which represent 98% sequence coverage of NBD1^{1S} (**Fig. 4e**–**h** and

dc_2020_22

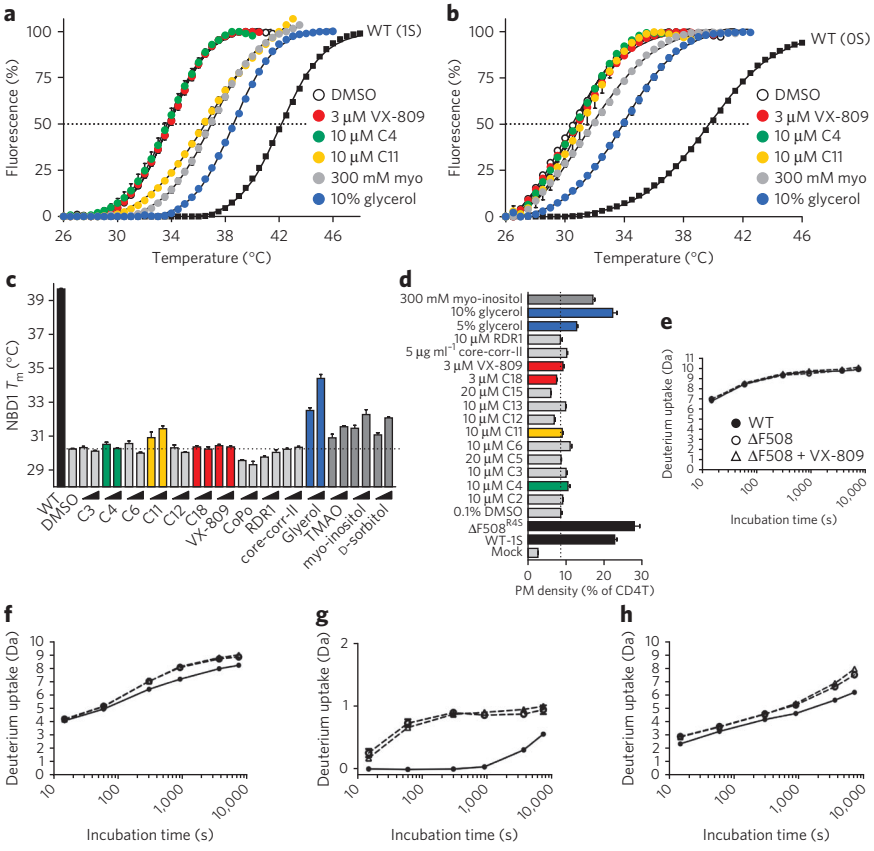


Figure 4 | Effect of correctors on the isolated NBD1 stability *in vitro* and *in vivo*. (**a**–**c**) Melting temperature (T_m) of human ΔF508-NBD1[1S] (**a**) or ΔF508-NBD1 (0S) (**b**,**c**) was determined by DSF[6,27] as described in the Online Methods. The indicated correctors or chemical chaperones (**Supplementary Table 1**) were present during thermal unfolding at the following concentrations: 3 μM and 10 μM C1, C2, C3, C4, C5, C6, C7, C9, C11, C12, C13, C14, C17 and CoPo; 10 μM and 20 μM C8, C15 and C16; 1 μM and 3 μM C18 and VX-809, 5 μM and 15 μM RDR1; 2.5 μg ml^{-1} and 5 μg ml^{-1} core-corr-II; 5% and 10% glycerol; 150 mM and 300 mM TMAO, taurine, myo-inositol (myo) and D-sorbitol. Wild-type NBD1[1S] (WT (1S)) or wild-type NBD1 (WT (0S)) were used as positive controls. In **c**, the T_m of ΔF508-NBD1 was determined by DSF as in **b**. Data represent mean ± s.e.m. (n = 3). The change in T_m was considered statistically significant if it was more ± 3 s.d. (30.2 ± 0.4 °C) of the T_m of DMSO-treated ΔF508-NBD1. (**d**) Plasma membrane (PM) density of CD4T-ΔF508-NBD1[1S] in transiently transfected African green monkey kidney (COS-7) cells was measured by plasma membrane ELISA using anti-CD4. The ELISA signal of mock-transfected cells is also indicated. Cells were treated with corrector for 24 h at 37 °C. CD4T-ΔF508-NBD1[R4S] (ΔF508[R4S]) and CD4T-WT-NBD1[1S] (WT-1S) were used as positive controls. Data represent mean ± s.e.m. (n = 8–12). (**e**–**h**) Conformational dynamics of wild-type and ΔF508-NBD1[1S] was determined by HDX-MS following 5 min of D_2O labeling in the presence or absence of 10 μM VX-809. Data represent mean ± s.e.m. (n = 3). Representative peptides shown span residues 471–490 (**e**), 491–501 (**f**), 505–508 (WT) and 505–509 (ΔF508) (**g**) and 510–525 (**h**) and are from >60 peptides obtained by pepsin digestion.

Supplementary Fig. 10c). VX-809, however, was unable to suppress the HDX kinetics of the 505–509 segment and of the other peptides, suggesting that VX-809 is unable to elicit localized conformational stabilization on ΔF508-NBD1[1S] (**Fig. 4e**–**h**).

Corrector combinations robustly rescue ΔF508 in BHK cells

On the basis of this study and other published data[22,38,40], we propose the categorization of the ΔF508 correctors into three classes (**Fig. 5a**). Class I (C3, C18 and VX-809) primarily stabilizes the NBD1-CL1 and NBD1-CL4 interfaces, class II (core-corr-II and corr-4 analogs (C4, C13)) targets NBD2 and/or its interface, and class III stabilizes ΔF508-NBD1. Because RDR1 stabilizes mouse but not human NBD1 (ref. 22) (**Fig. 4**), we used chemical chaperones such as glycerol as surrogate class III correctors in our proof-of-principle studies.

The combination of chemicals with complementary structural targets is predicted to potentiate the individual correctors via ΔF508-CFTR–coupled domain folding and phenocopy the combination of suppressor mutations[27,28] (**Supplementary Fig. 1b**). Though combining a class I corrector (VX-809, C18 or C3) with a class II corrector (C4 or core-corr-II) only modestly rescued the amount of ΔF508-CFTR in the plasma membrane (<10% of wild-type CFTR), the combination of class I corrector and glycerol enhanced ΔF508-CFTR plasma membrane expression to 30–50% that of the wild type, supporting that correction of both primary defects is required for robust ΔF508-CFTR rescue (**Fig. 5b**). The class I corrector and glycerol each achieved <8% rescue, and this effect was further improved by class II correctors and enhanced ΔF508-CFTR plasma membrane expression to ~60–110% that of the wild type (**Fig. 5b**). Similar phenotypes were confirmed by monitoring the accumulation of complex-glycosylated ΔF508-CFTR with immunoblotting (**Fig. 5c**). Metabolic pulse-chase experiments (**Fig. 5d**,**e**) and ELISA (**Fig. 5f**) revealed that the combined treatment increased the ΔF508-CFTR folding efficiency to ~70% that of the wild type and largely rescued the plasma membrane stability defect of the mutant. Thus, both enhanced endoplasmic reticulum folding efficiency and the plasma membrane stability of ΔF508-CFTR in the presence of two classes of correctors with glycerol account for the wild type–like plasma membrane expression.

ΔF508-CFTR rescue in bronchial epithelia

The rationally selected chemical combination also considerably potentiated the VX-809 rescue of the ΔF508-CFTR in polarized respiratory epithelial cells. The plasma membrane density of ΔF508-CFTR in CFBE41o- cells treated with a class I corrector was enhanced by approximately three-fold in the presence of 5% glycerol and was further augmented by treatment with class II corrector (C4 or core-corr-II), representing an approximately five-fold increase of the VX-809 effect (**Fig. 6a**). The combined treatment of two classes of correctors with glycerol enhanced the plasma membrane density of ΔF508-CFTR to ~10% that of wild-type CFTR, a comparable effect to that achieved by low temperature (**Fig. 6a**). The VX-809 effect was further potentiated by combination with C4 and glycerol and partially restored ΔF508-CFTR complex

dc_2020_22

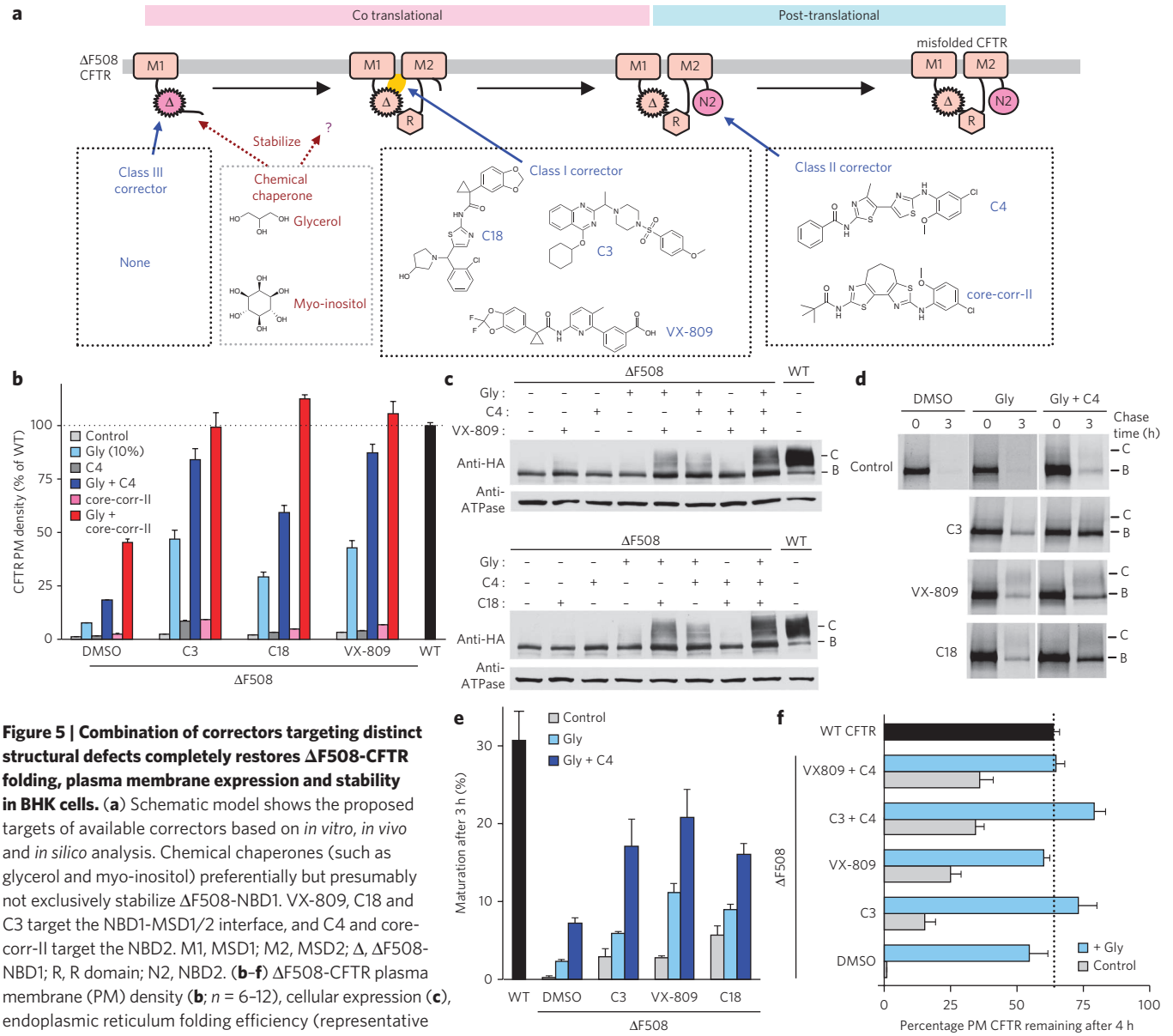


Figure 5 | Combination of correctors targeting distinct structural defects completely restores ΔF508-CFTR folding, plasma membrane expression and stability in BHK cells. (**a**) Schematic model shows the proposed targets of available correctors based on *in vitro*, *in vivo* and *in silico* analysis. Chemical chaperones (such as glycerol and myo-inositol) preferentially but presumably not exclusively stabilize ΔF508-NBD1. VX-809, C18 and C3 target the NBD1-MSD1/2 interface, and C4 and core-corr-II target the NBD2. M1, MSD1; M2, MSD2; Δ, ΔF508-NBD1; R, R domain; N2, NBD2. (**b**–**f**) ΔF508-CFTR plasma membrane (PM) density (**b**; n = 6–12), cellular expression (**c**), endoplasmic reticulum folding efficiency (representative autoradiographs (**d**) and statistical evaluation (**e**) of metabolic pulse-chase results are shown; n = 3–4) and plasma membrane stability after 4 h (**f**; n = 5–12) were determined upon treatment with correctors (10 μM C3 or C4; 3 μM C18 or VX-809, 5 μg ml^{-1} core-corr-II; 10% glycerol (Gly)) or their combinations in BHK cells by cell-surface ELISA, western blotting, metabolic pulse-chase technique and cell-surface ELISA, respectively. All data represent mean ± s.e.m. WT, wild type; anti-HA, antibody specific to hemagglutinin; B, immature core-glycosylated form; C, mature complex-glycosylated form.

glycosylation (**Fig. 6b**), plasma membrane stability (**Fig. 6c**) and function at the apical plasma membrane in CFBE41o- epithelial cells (**Fig. 6d,e**). Similar enhancement was observed in other polarized epithelial cell models. The combined treatment enhanced the plasma membrane density of ΔF508-CFTR to ~40% and ~15% that of the wild type in Madin-Darby Canine Kidney cells and human lung papillary adenocarcinoma epithelial cells (NCI-H441), respectively (**Supplementary Fig. 11**).

Functional correction of ΔF508-CFTR rectal organoids

Finally, the efficacy of the mechanism-based corrector combination was evaluated in human rectal organoids derived from cystic fibrosis patients homozygous for the ΔF508 mutation[45]. CFTR channel activity was measured by monitoring forskolin-induced swelling (FIS) of organoids as the cross-sectional area of organoids was proportional with CFTR-mediated chloride and coupled water transport[45]. Although class I correctors (C18 or VX-809) modestly improved the FIS of *CFTR*$^{\Delta F508/\Delta F508}$ organoids, combining it with myo-inositol[46], a chemical chaperone selected on the basis of its minimal toxicity, enhanced the ΔF508-CFTR–dependent FIS to 30–60% of that observed in organoids from healthy individuals (wild type) (**Fig. 6f–h**). The rescue of ΔF508-CFTR–mediated FIS was completely prevented with CFTR blockers (inhibitor 172 (Inh_{172}) and GlyH101) (**Fig. 6f–h**). Likewise, organoids expressing nonfunctional, truncated CFTR variants harboring a premature truncation (E60X) and the frame shift mutation 4015ATTTdel in the N-terminal tail and NBD2, respectively, also failed to show any FIS (**Fig. 6h**). Remarkably, VX-809, C4 and myo-inositol jointly restored ΔF508-CFTR–dependent FIS to >80% that of the wild type. A comparison of the individual, dual and triple combination effect of correctors on FIS of *CFTR*$^{\Delta F508/\Delta F508}$ organoids indicates a 2- to

dc_2020_22

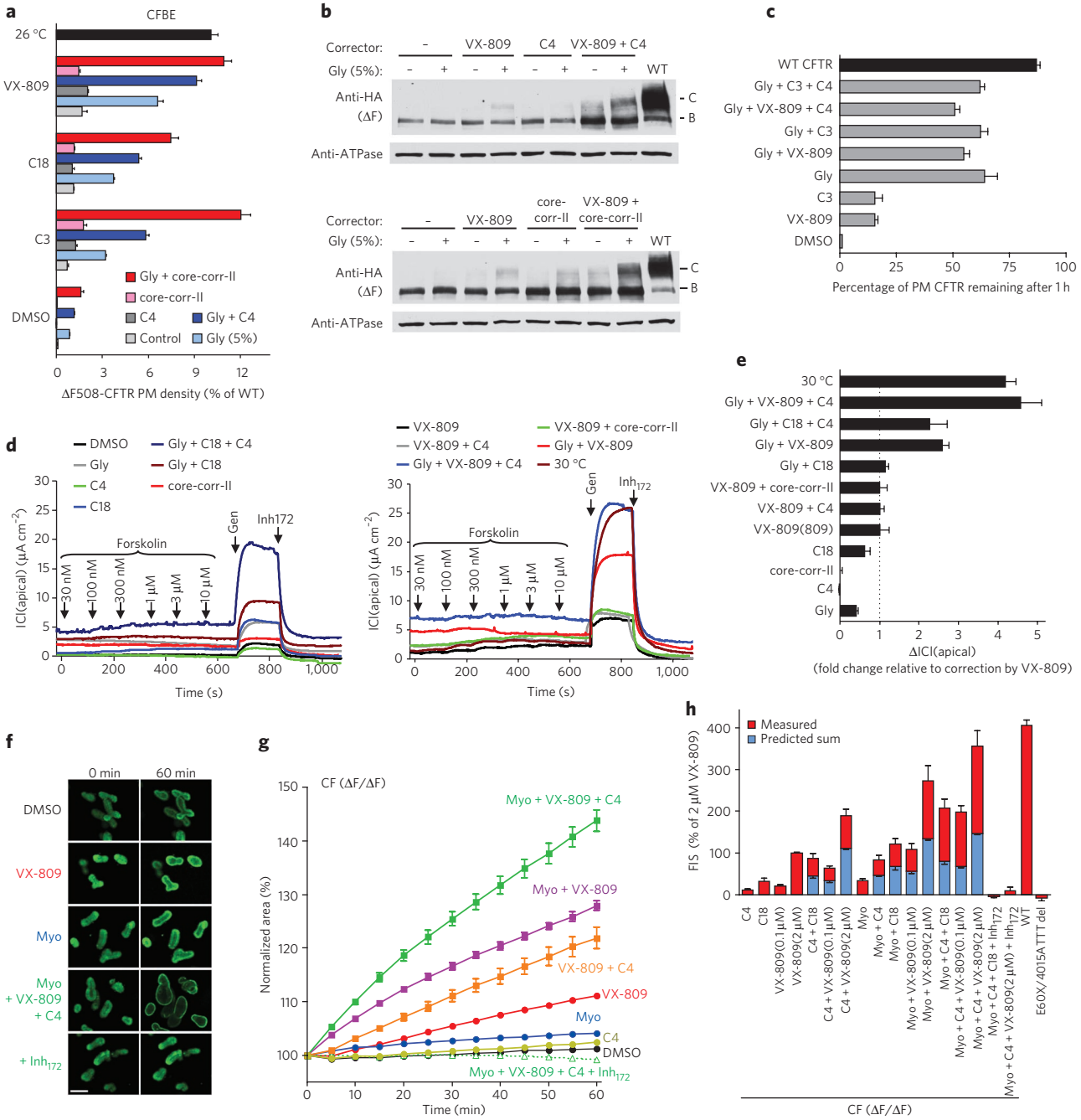


Figure 6 | Corrector combination restores ΔF508-CFTR functional expression in polarized epithelial cell lines and rectal organoids from ΔF508 cystic fibrosis patients. (**a**–**e**) ΔF508-CFTR plasma membrane (PM) density (**a**, *n* = 7–22), cellular expression (**b**), plasma membrane stability (**c**, *n* = 16) and function (**d**, *n* = 3–4) in CFBE41o- cells. Correctors (10 μM C3 or C4, 3 μM C18 or VX-809, 2.5 μg ml^{-1} core-corr-II, 5% glycerol (Gly)) were applied for 24 h at 37 °C. For comparison, ΔF508-CFTR was rescued by low temperature at 26 °C or 30 °C for 36–48 h. WT, wild type; ΔF, ΔF508-CFTR; anti-HA, hemagglutinin-specific antibody; B, immature core-glycosylated form; C, mature complex-glycosylated form. In **d** are representative records of forskolin- and genistein (Gen, 100 μM)-activated ΔF508-CFTR ICl(apical) following corrector treatment are shown. In **e** is a summary of the peak, Inh_{172}-sensitive ICl(apical) of ΔF508-CFTR in CFBE41o- cells after corrector treatment (*n* = 3–4). (**f**) Fluorescence laser confocal images of Calcein green–loaded cystic fibrosis rectal organoids before and after FIS for 60 min in the presence of correctors (2 μM, VX-809, C4 and/or 125 mM myo-inositol (Myo)), and inhibitor (Inh; 50 μM CFTR Inh_{172} and GlyH-101). (**g**) Time course of FIS of cystic fibrosis organoids from a representative experiment. Organoids were treated as described in the Online Methods. FIS is expressed as a percentage of initial cell cross-sectional area before forskolin stimulation. CF (ΔF/ΔF), *CFTR*$^{\Delta F508/\Delta F508}$ organoids. (**h**) FIS is expressed as percentage of the area of VX-809 (2 μM)-treated cells after 60 min of FIS. In addition to the measured FIS (red bar), the predicted sum of the individual corrector effect (2 μM C3, C4, C18; 0.1 μM or 2 μM VX-809; 125 mM myo-inositol) is indicated, assuming additive effect (blue bar). The mean FIS from three healthy controls (WT) is included. No FIS was observed after treatment with myo-inositol, C4 and VX-809 in organoids carrying the E60X and 4015ATTTdel mutations. Data represent mean ± s.e.m. from six independent experiments performed on three patients homozygous for ΔF508.

2.5-fold potentiation (**Fig. 6h**). Together, these results demonstrate that mechanism-based corrector combinations are capable of robustly rescuing ΔF508-CFTR folding and function in cell culture models and primary human cystic fibrosis organoids.

Notably, class I correctors, class II correctors and glycerol also augmented the plasma membrane expression of wild-type CFTR with an intrinsic low folding efficiency (~30%)[1] (**Supplementary Fig. 12a**). The wild-type CFTR rescue efficiency by VX-809 or C18 but not C4 was suppressed in the presence of the V510D mutation and, to a lesser extent, by the 3S or R4S mutation (**Supplementary Fig. 12b**), consistent with the limited conformational instability of the NBD1-MSD1, of NBD1-MSD2 interfaces and of NBD1 in wild-type CFTR[27]. These observations, along with the overlapping VX-809 *in silico* docking data on ΔF508-CFTR and wild-type CFTR (**Supplementary Fig. 12c**) and on *in vitro* data showing MSD1 stabilization are consistent with the possibility that VX-809, similarly to the V510D mutation, targets the conformationally unstable wild-type NBD1-MSD1/2 interface.

DISCUSSION

Here we show that the combination of correctors targeting complementary conformational defects can overcome their individually modest effects and restore ΔF508-CFTR processing and plasma membrane channel function. Using *in vivo*, *in vitro* and *in silico* analyses, we identified subsets of correctors that preferentially target the primary conformational defect at the NBD1-MSD1(CL1) and NBD1-MSD2(CL4) interfaces (class I) or the resulting NBD2 misassembly (class II). None of the tested correctors belong to class III, which targets the human ΔF508-NBD1 energetic defect. Only chemical chaperones, such as glycerol and myo-inositol, can substitute for a class III corrector by stabilizing human ΔF508-NBD1 both *in vitro* and *in vivo*. Although we cannot rule out that glycerol has multiple targets in ΔF508-CFTR, NBD2 deletion diminished glycerol's effect on plasma membrane expression and complex glycosylation of ΔF508-CFTR^{R1S} but left ΔF508-CFTR^{R1070W} expression largely unaffected (**Fig. 3e**,**f** and **Supplementary Fig. 7f**,**g**). In addition, glycerol rescued the amount of ΔF508-CFTR^{R1070W} in the plasma membrane, similar to the effect of the NBD1 stabilizing mutations[24] (**Supplementary Fig. 7f**,**g**). These observations suggest that glycerol may primarily stabilize the NBD1. The proposed classification of correctors also provides a plausible explanation for the additive effect of C3 (class I) and C4 (class II)[47,48] or of VX-809 (class I) and C4 (class II), the attenuated efficiency of C3 and VX-809 (ref. 14) and the potentiation of class I correctors by chemical chaperones (as class III surrogates) (**Figs. 5** and **6**). The limited synergy between class I and class II correctors in the absence of a chemical chaperone may imply that suppression of ΔF508-CFTR primary defects is a prerequisite for correction of the secondary NBD2 defect.

Because of its ability to selectively increase the endoplasmic reticulum folding efficiency and plasma membrane stability in multiple cellular models, VX-809 was predicted to overcome a kinetic folding trap and/or stabilize the native-like conformer by directly interacting with ΔF508-CFTR[14]. The inhibition of ΔF508-CFTR^{2RK} thermal inactivation by VX-809 or C18 suggests that these correctors can directly interact with the mutant at the single-molecule level. Although VX-809 prevented functional inactivation upon raising the temperature from 24 °C to 36 °C, extended exposure to 36 °C attenuated the VX-809 rescue effect, probably because only the NBD1-MSD1/2 interface defect was stabilized. This interpretation is consistent with the modest effect of genetic stabilization of the NBD1-MSD1/2 interface[49] compared to the effect on the NBD1 (ref. 35) and may explain the negligible effect of VX-809 on ΔF508-CFTR^{I539T} thermal inactivation, which was reported recently[38]. Notably, the direct interaction of C3 and potentiators P1 and VX-770 with purified CFTR was also inferred on the basis of their modulation of the channel ATPase and transport activity[24,50,51].

Definitive identification of the VX-809 binding site (or sites) in the absence of a functionally inert, cross-linkable adduct remains a challenge and is further complicated by the coupled folding and misfolding mechanism of CFTR domains[6,8], which may allow allosteric corrections of folding defects from different binding sites. However, our analysis by *in silico* predictions, in concert with *in vitro* assays (DSF and HDX-MS) using isolated human NBD1s and *in vivo* studies using full-length CFTRs with deletion and point mutations, suggest that the NBD1-CL1/4 represents a primary target of VX-809, a conclusion that partly overlaps with a recent proposition[38].

Importantly, mutations confined to transmembrane helix 1 (G91R) and CL2 (M266R and W277R) of MSD1 could also be partially rescued, presumably via targeting the coupled interface defect at NBD1-CL1/4. These results, together with those obtained using CFTR fragments and CL1 mutations, provide a plausible model of VX-809 action in which it binds MSD1 and stabilizes the CL1-CL4 coupling helix, a critical step to form the proper interactions of NBD1, first with MSD1 then with MSD2, and ultimately facilitates cooperative domain assembly upon completion of NBD2 translation[27,28,38]. Counteracting the processing defect of the cystic fibrosis–causing R170G and other missense mutations localized to transmembrane helix 1 and CL1/4 described here and previously[38] by class I correctors also suggests that VX-809 may be successfully used alone or in combination in a variety of rare cystic fibrosis mutations.

Although the additive effect of corrector pairing has been previously observed[14,47,48], the rationale for corrector combination remained elusive. Our most important finding is that the mechanism-based classification of correctors permitted a rationally designed corrector combination approach that achieved a considerable improvement in ΔF508-CFTR rescue efficiency. Chemical correction of NBD1, the NBD1-MSD1/2 interface and NBD2 instability almost completely restored ΔF508-CFTR folding, plasma membrane expression, stability and function in BHK cells. ΔF508-CFTR rescue was weaker in kidney and respiratory epithelial cells (10–30% that of the wild type) than in BHK cells. This result could be attributed to modest NBD1 stabilization by reduced glycerol concentration (5%), which was required to maintain epithelial polarity.

Although the robust combination correction of ΔF508-CFTR transport function in human *CFTR*$^{ΔF508/ΔF508}$ organoids may be an overestimate, owing to the rate-limiting ion transport capacity of the basolateral plasma membrane in wild-type intestinal organoids, VX-809–induced ΔF508-CFTR chloride transport was augmented almost four-fold by treatment with class II compound and chemical chaperone, as shown by the FIS measurements. Considering that VX-809 improves ΔF508-CFTR plasma membrane function to ~15% that of the wild type in primary human bronchial epithelia[14] with marginally translatable clinical benefit[21], it is reasonable to assume that structural defect–targeted corrector combination will eventually confer sufficient CFTR transport capacity in respiratory epithelial cells to achieve substantial clinical improvements in most *CFTR*$^{ΔF508/ΔF508}$ patients.

None of the available correctors but chemical chaperones can counteract the human ΔF508-NBD1 conformational defect, according to studies on *in vitro* and *in vivo* folding. This represents a bottleneck of cystic fibrosis pharmacological therapy, as systemic administration of chemical chaperones is not feasible. To achieve improved pharmacological correction in combination with class I correctors (for example, VX-809), an NBD1 stabilizer must be identified. The initial isolation of pharmacological chaperones stabilizing the NBD1 could be envisioned by high-throughput screening of diverse compounds or fragment libraries *in vitro*, *in vivo* or *in silico* using ΔF508-NBD1 or ΔF508-CFTR cell-based functional or biochemical assays[17,22]. Exploiting NBD1-MSD1/2 interface stabilization by second-site suppressor mutations or class I correctors (for example,VX-809) may bias high-throughput screening efforts toward the isolation of NBD1-stabilizing small molecules.

dc_2020_22

We believe that the successful identification of NBD1 stabilizers would make mechanism-based corrector combination therapy feasible for most cystic fibrosis patients.

Received 12 September 2012; accepted 8 April 2013; published online 12 May 2013

METHODS

Methods and any associated references are available in the online version of the paper.

Referencsses

1. Riordan, J.R. CFTR function and prospects for therapy. *Annu. Rev. Biochem.* **77**, 701–726 (2008).
2. He, L. *et al.* Multiple membrane-cytoplasmic domain contacts in the cystic fibrosis transmembrane conductance regulator (CFTR) mediate regulation of channel gating. *J. Biol. Chem.* **283**, 26383–26390 (2008).
3. Serohijos, A.W. *et al.* 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 (2008).
4. Kim, S.J. & Skach, W.R. Mechanisms of CFTR folding at the endoplasmic reticulum. *Front. Pharmacol.* **3**, 201 (2012).
5. Du, K., Sharma, M. & Lukacs, G.L. The ΔF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. *Nat. Struct. Mol. Biol.* **12**, 17–25 (2005).
6. Du, K. & Lukacs, G.L. Cooperative assembly and misfolding of CFTR domains *in vivo*. *Mol. Biol. Cell* **20**, 1903–1915 (2009).
7. He, L. *et al.* Restoration of domain folding and interdomain assembly by second-site suppressors of the ΔF508 mutation in CFTR. *FASEB J.* **24**, 3103–3112 (2010).
8. Cui, L. *et al.* Domain interdependence in the biosynthetic assembly of CFTR. *J. Mol. Biol.* **365**, 981–994 (2007).
9. Kleizen, B., van Vlijmen, T., de Jonge, H.R. & Braakman, I. Folding of CFTR is predominantly cotranslational. *Mol. Cell* **20**, 277–287 (2005).
10. Rosser, M.F., Grove, D.E., Chen, L. & Cyr, D.M. Assembly and misassembly of cystic fibrosis transmembrane conductance regulator: folding defects caused by deletion of F508 occur before and after the calnexin-dependent association of membrane spanning domain (MSD) 1 and MSD2. *Mol. Biol. Cell* **19**, 4570–4579 (2008).
11. Balch, W.E., Roth, D.M. & Hutt, D.M. Emergent properties of proteostasis in managing cystic fibrosis. *Cold Spring Harb. Perspect. Biol.* **3**, a004499 (2011).
12. Boucher, R.C. New concepts of the pathogenesis of cystic fibrosis lung disease. *Eur. Respir. J.* **23**, 146–158 (2004).
13. Veit, G. *et al.* Proinflammatory cytokine secretion is suppressed by TMEM16A or CFTR channel activity in human cystic fibrosis bronchial epithelia. *Mol. Biol. Cell* **23**, 4188–4202 (2012).
14. 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. USA* **108**, 18843–18848 (2011).
15. Denning, G.M. *et al.* Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-sensitive. *Nature* **358**, 761–764 (1992).
16. Sato, S., Ward, C.L., Krouse, M.E., Wine, J.J. & Kopito, R.R. Glycerol reverses the misfolding phenotype of the most common cystic fibrosis mutation. *J. Biol. Chem.* **271**, 635–638 (1996).
17. Pedemonte, N. *et al.* Small-molecule correctors of defective ΔF508-CFTR cellular processing identified by high-throughput screening. *J. Clin. Invest.* **115**, 2564–2571 (2005).
18. Varga, K. *et al.* Enhanced cell-surface stability of rescued ΔF508 cystic fibrosis transmembrane conductance regulator (CFTR) by pharmacological chaperones. *Biochem. J.* **410**, 555–564 (2008).
19. Hanrahan, J.W., Sampson, H.M. & Thomas, D.Y. Novel pharmacological strategies to treat cystic fibrosis. *Trends Pharmacol. Sci.* **34**, 119–125 (2013).
20. Van Goor, F. *et al.* Rescue of CF airway epithelial cell function *in vitro* by a CFTR potentiator, VX-770. *Proc. Natl. Acad. Sci. USA* **106**, 18825–18830 (2009).
21. Clancy, J.P. *et al.* Results of a phase IIa study of VX-809, an investigational CFTR corrector compound, in subjects with cystic fibrosis homozygous for the F508del-CFTR mutation. *Thorax* **67**, 12–18 (2012).
22. Sampson, H.M. *et al.* Identification of a NBD1-binding pharmacological chaperone that corrects the trafficking defect of F508del-CFTR. *Chem. Biol.* **18**, 231–242 (2011).
23. Howard, M. *et al.* Mammalian osmolytes and *S*-nitrosoglutathione promote ΔF508 cystic fibrosis transmembrane conductance regulator (CFTR) protein maturation and function. *J. Biol. Chem.* **278**, 35159–35167 (2003).
24. Kim Chiaw, P., Wellhauser, L., Huan, L.J., Ramjeesingh, M. & Bear, C.E. A chemical corrector modifies the channel function of F508del-CFTR. *Mol. Pharmacol.* **78**, 411–418 (2010).
25. Protasevich, I. *et al.* Thermal unfolding studies show the disease causing F508del mutation in CFTR thermodynamically destabilizes nucleotide-binding domain 1. *Protein Sci.* **19**, 1917–1931 (2010).
26. Wang, C. *et al.* Integrated biophysical studies implicate partial unfolding of NBD1 of CFTR in the molecular pathogenesis of F508del cystic fibrosis. *Protein Sci.* **19**, 1932–1947 (2010).
27. Rabeh, W.M. *et al.* Correction of both NBD1 energetics and domain interface is required to restore ΔF508 CFTR folding and function. *Cell* **148**, 150–163 (2012).
28. Mendoza, J.L. *et al.* Requirements for efficient correction of ΔF508 CFTR revealed by analyses of evolved sequences. *Cell* **148**, 164–174 (2012).
29. Robert, R. *et al.* Correction of the Δphe508 cystic fibrosis transmembrane conductance regulator trafficking defect by the bioavailable compound glafenine. *Mol. Pharmacol.* **77**, 922–930 (2010).
30. Loo, T.W., Bartlett, M.C. & Clarke, D.M. The V510D suppressor mutation stabilizes ΔF508-CFTR at the cell surface. *Biochemistry* **49**, 6352–6357 (2010).
31. Teem, J.L. *et al.* Identification of revertants for the cystic fibrosis ΔF508 mutation using STE6-CFTR chimeras in yeast. *Cell* **73**, 335–346 (1993).
32. Pedemonte, N. *et al.* Dual activity of aminoarylthiazoles on the trafficking and gating defects of the cystic fibrosis transmembrane conductance regulator chloride channel caused by cystic fibrosis mutations. *J. Biol. Chem.* **286**, 15215–15226 (2011).
33. Phuan, P.W. *et al.* Cyanoquinolines with independent corrector and potentiator activities restore ΔPhe508-cystic fibrosis transmembrane conductance regulator chloride channel function in cystic fibrosis. *Mol. Pharmacol.* **80**, 683–693 (2011).
34. Hegedus, T. *et al.* F508del CFTR with two altered RXR motifs escapes from ER quality control but its channel activity is thermally sensitive. *Biochim. Biophys. Acta* **1758**, 565–572 (2006).
35. Aleksandrov, A.A. *et al.* Allosteric modulation balances thermodynamic stability and restores function of ΔF508 CFTR. *J. Mol. Biol.* **419**, 41–60 (2012).
36. Dalton, J., Kalid, O., Schushan, M., Ben-Tal, N. & Villa-Freixa, J. New model of cystic fibrosis transmembrane conductance regulator proposes active channel-like conformation. *J. Chem. Inf. Model.* **52**, 1842–1853 (2012).
37. Kalid, O. *et al.* Small molecule correctors of F508del-CFTR discovered by structure-based virtual screening. *J. Comput. Aided Mol. Des.* **24**, 971–991 (2010).
38. He, L. *et al.* Correctors of ΔF508 CFTR restore global conformational maturation without thermally stabilizing the mutant protein. *FASEB J.* **27**, 536–545 (2013).
39. Thibodeau, P.H. *et al.* The cystic fibrosis–causing mutation ΔF508 affects multiple steps in cystic fibrosis transmembrane conductance regulator biogenesis. *J. Biol. Chem.* **285**, 35825–35835 (2010).
40. Grove, D.E., Rosser, M.F., Ren, H.Y., Naren, A.P. & Cyr, D.M. Mechanisms for rescue of correctable folding defects in CFTRΔF508. *Mol. Biol. Cell* **20**, 4059–4069 (2009).
41. Zhang, F., Kartner, N. & Lukacs, G.L. Limited proteolysis as a probe for arrested conformational maturation of ΔF508 CFTR. *Nat. Struct. Biol.* **5**, 180–183 (1998).
42. Yu, G.J. *et al.* Potent *S-cis*-locked bithiazole correctors of ΔF508 cystic fibrosis transmembrane conductance regulator cellular processing for cystic fibrosis therapy. *J. Med. Chem.* **51**, 6044–6054 (2008).
43. Aleksandrov, A.A. *et al.* Regulatory insertion removal restores maturation, stability and function of ΔF508 CFTR. *J. Mol. Biol.* **401**, 194–210 (2010).
44. Lewis, H.A. *et al.* Structure and dynamics of NBD1 from CFTR characterized using crystallography and hydrogen/deuterium exchange mass spectrometry. *J. Mol. Biol.* **396**, 406–430 (2010).
45. Dekkers, J. *et al.* A functional CFTR assay using primary cystic fibrosis intestinal organoids. *Nat. Med.* (in the press).
46. Zhang, X.M. *et al.* Organic solutes rescue the functional defect in ΔF508 cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* **278**, 51232–51242 (2003).
47. Wang, Y., Loo, T.W., Bartlett, M.C. & Clarke, D.M. Additive effect of multiple pharmacological chaperones on maturation of CFTR processing mutants. *Biochem. J.* **406**, 257–263 (2007).
48. Pedemonte, N., Tomati, V., Sondo, E. & Galietta, L.J. Influence of cell background on pharmacological rescue of mutant CFTR. *Am. J. Physiol. Cell Physiol.* **298**, C866–C874 (2010).
49. Liu, X., O'Donnell, N., Landstrom, A., Skach, W.R. & Dawson, D.C. Thermal instability of ΔF508 cystic fibrosis transmembrane conductance regulator (CFTR) channel function: protection by single suppressor mutations and inhibiting channel activity. *Biochemistry* **51**, 5113–5124 (2012).
50. Wellhauser, L. *et al.* A small-molecule modulator interacts directly with ΔPhe508-CFTR to modify its ATPase activity and conformational stability. *Mol. Pharmacol.* **75**, 1430–1438 (2009).
51. Eckford, P.D., Li, C., Ramjeesingh, M. & Bear, C.E. Cystic fibrosis transmembrane conductance regulator (CFTR) potentiator VX-770 (Ivacaftor) opens the defective channel gate of mutant CFTR in a phosphorylation-dependent but ATP-independent manner. *J. Biol. Chem.* **287**, 36639–36649 (2012).

Acknowledgments

We thank L. Konerman, T. Wales, J. Engen, M.J. Chalmers and P.R. Griffin for valuable advice on setting up the HDX-MS technique; D. Gruenert (University of California–San Francisco) for the parental CFBE41o- cell line; R.J. Bridges (Rosalind Franklin University of Medicine and Science) and Cystic Fibrosis Foundation Therapeutics (CFFT) Inc. for CFTR modulator panels; D. Thomas (McGill University, Canada) for kindly providing RDR compounds; and W.R. Skach for advice. We are grateful for the financial support of the Dutch Cystic Fibrosis foundation and the Wilhelmina Children's Hospital Research fund to J.M.B., the Hungarian National Science Foundation (MB08C-80039) and the European Union (FP7-IRG 239270) to T.H., the European Research Area (ERA)-Chemistry Hungarian Scientific Research Fund (OTKA) (102166) to A.S., the Tara K. Telford Fund for Cystic Fibrosis Research at the University of California–Davis and the US National Institutes of Health (NIH; grants DK072517 and GM089153) to M.K., NIH–National Institute of Diabetes and Digestive and Kidney Diseases (R01DK75302), CFFT Inc., Cystic Fibrosis Canada, Canadian Institutes of Health Research and Canada Foundation for Innovation to G.L.L. T.H. is a Bolyai Fellow of the Hungarian Academy of Sciences. G.V. was partly supported by a European Molecular Biology Association and Fonds de Recherche Santé Québec Fellowship. G.L.L. is a Canada Research Chair.

Author contributions

T.O. designed and carried out the biochemical studies, including the cell-surface ELISA, pulse-chase and immunoblotting assays, and analyzed the data. G.V. generated the inducible epithelial cell lines and carried out the ICl(apical) measurements. J.F.D. performed the intestinal organoid transport assay under the direction of J.M.B. M.B. carried out the bilayer measurements. N.S. measured the HDX of NBDs. H.X. engineered the CFTR mutants and generated the BHK cell lines. A.R. purified the NBDs and characterized their thermal stability. A.S.V. and M.K. provided reagents. A.S. and T.H. performed the *in silico* docking. G.L.L. conceived and directed the study and wrote the manuscript with T.O.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to G.L.L.

ONLINE METHODS

Cell lines. Full-length and truncated human CFTR variants with the 3HA tag in the fourth extracellular loop[52] were constructed by PCR mutagenesis (**Supplementary Table 2**). Detailed information on PCR mutagenesis is available from the authors on request. BHK cells stably expressing CFTR variants were generated and grown as previously described[27]. MDCK type II cells stably expressing CFTR-3HA variants were generated by lentivirus infection under puromycin selection (1–5 μg/ml) and grown in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen, Carlsbad, CA) supplemented with 10% FBS (FBS). CFBE41o- Tet-on and NCI-H441 Tet-on cells stably expressing CFTR-3HA variants under a tetracycline-responsive promoter were generated by lentivirus transduction using the Lenti-X TetON Advanced Inducible Expression System (Clontech, Mountain View, CA) under puromycin (3 μg/ml) and G418 selection (0.2 mg/ml) and were grown in minimal essential medium (MEM, Invitrogen) supplemented with 10% FBS, 2 mM L-glutamine and 10 mM HEPES and in RPMI-1640 Medium (ATCC) supplemented with 10% FBS, respectively. For propagation, the CFBE41o- cells were cultured in plastic flasks coated with an extracellular matrix (ECM mix) consisting of 10 μg/ml human fibronectin (EMD), 30 μg/ml PureCol collagen preparation (Advanced Biomatrix) and 100 μg/ml bovine serum albumin (Sigma-Aldrich) diluted in LHC basal medium (Invitrogen). The CFTR-3HA expression was induced by 0.5 or 1 μg/ml doxycycline treatment for 4 d.

Antibodies and reagents. Antibodies (Abs) were obtained from the following sources: Monoclonal anti-CFTR L12B4 (no. MAB3484; dilution 1:1,000; recognizing NBD1) and M3A7 (no. MAB3480; dilution 1:1,000; recognizing NBD2) were purchased from Millipore Bioscience Research Reagents (Temecula, CA). Mouse monoclonal anti-HA (no. MMS-101R; dilution 1:1,000) was from Covance Innovative Antibodies (Berkeley, CA). Chemicals were obtained from Sigma-Aldrich (St. Louis, MO) at the highest grade available. Most of the CFTR correctors were obtained from Cystic Fibrosis Foundation Therapeutics (CFFT) Inc., VX-809 was from Selleck (Houston, TX) and class II compounds were synthesized as described[42]. All of the correctors used in this study are listed in **Supplementary Table 1**.

NBD1 protein purification. Recombinant human NBD1 proteins were purified from *E. coli* as previously described[27]. The NBD1 protein was concentrated to 3–5 mg/ml in buffer containing 150 mM NaCl, 1 mM ATP, 2 mM $MgCl_2$, 1 mM TCEP, 10% glycerol and 10 mM HEPES, pH 7.8.

Microsome preparation. Microsomes were isolated by differential centrifugation from BHK cells stably expressing 3HA-tagged WT CFTR or ΔF508-CFTR^{2RK} as described[27]. To enrich for the complex-glycosylated ΔF508-CFTR^{2RK}, cells were incubated at 26 °C for ~36 h followed by treatment with 150 μg/ml cycloheximide (CHX, 26 °C for 12 h) before microsome isolation[52]. WT CFTR–expressing cells were treated with 150 μg/ml CHX for 3 h to eliminate the endoplasmic reticulum–localized core-glycosylated form.

Planar lipid bilayer studies. WT or ΔF508-CFTR reconstitution and channel activity were essentially measured as described previously[27]. A planar phospholipid bilayer was made of a 2:1 (w/w) mixture of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-L-serine (POPS, Avanti Polar Lipids, Alabaster, AL) in an *n*-decane solution at a final lipid concentration of 25 mg/ml and was formed over a 200- or 250-μm-diameter hole in a polystyrene chamber by monitoring the increase in its capacitance to 150–200 pF. The bilayer was bathed in symmetrical 300 mM Tris-HCl, 10 mM HEPES (pH 7.2), 5 mM $MgCl_2$ and 1 mM EGTA. Microsomes (10- to 20-μg proteins) isolated from CFTR-expressing BHK cells were prephosphorylated in the presence of 100 U/ml protein kinase A (PKA) catalytic subunit (Promega) and 2 mM MgATP at room temperature for 10 min, added to the *cis* compartment and fused to the planar lipid bilayer. CFTR channel activity was recorded in the presence of 100 U/ml PKA catalytic subunit and 2 mM MgATP. Currents were measured with a BC-535 amplifier (Warner Instrument, Hamden, CT). Voltage was clamped at $V_m = -60$ mV. In temperature ramp experiments, the chamber was heated from room temperature to 36 °C at a rate of ~1–2 °C/min. Single-channel activities were acquired using pClamp 8.1 (Axon Instruments) with 10 kHz, filtered at 200 Hz with low-pass 8-pole Bessel filter and stored digitally. Records were filtered at 50 Hz and analyzed using Clampfit 10.3 (Axon Instruments). We applied a 50% cutoff between open and closed levels, and events shorter than 10 ms were excluded from the analysis. Single-channel open probability (P_o) was determined as NP_o divided by the number of channels, where NP_o was obtained from event detection features of Clampfit 10.3 and N represents the number of channels.

HDX-MS. Localized conformational dynamics of isolated NBD1s was measured by HDX-MS essentially as described[44]. HDX was initiated by mixing 1–2 μL of NBD1 stock solution to 15 volumes of D_2O-based buffer (pD 7.5, based on pD = pH_{read} + 0.40)[53] containing 150 mM NaCl, 1 mM ATP, 2 mM $MgCl_2$, 1 mM TCEP and glycerol, resulting in a final D_2O concentration of more than 90% (v/v). The final concentration of glycerol was adjusted to 1%. The mixtures were incubated for 15 s, 1 min, 5 min, 15 min, 1 h and 2 h at room temperature and then quenched by adding a 2- to 2.5-μL aliquot of the mixture to 300 mM phosphate buffer containing 8 M urea, pH 2.5 (quenching buffer). Quenched solutions were flash frozen in MeOH containing dry ice, and samples were stored at −80 °C.

Prior to high performance liquid chromatography (HPLC)-MS analysis, the labeled protein solutions were thawed and immediately loaded onto the injection valve. Deuterated NBD1 was digested in an on-line immobilized pepsin column prepared in house. On-line pepsin digestion was carried out at a flow rate of 50 μL/min for ~1.5 min, and resulting peptides were trapped on a C18 trapping column (Optimized Technologies, Oregon City, OR). Following desalting for 1.5 min (at a 150 μl/min flow rate), peptides were loaded onto a C18 column (1 mm i.d. × 50 mm, Thermo Fisher Scientific, Waltham, MA) through a six-port valve. Peptides were separated using a 13–90% linear gradient of acetonitrile containing 0.1% formic acid for 6 min at 50 μl/min. Chromatographic separation was performed using an Agilent 1100 HPLC system. To minimize back-exchange, the column, solvent delivery lines, injector and other accessories were immersed in an ice bath. The C18 column was directly connected to the electrospray ionization source of LTQ Orbitrap XL (Thermo Fisher Scientific). Mass spectra of peptides were acquired in positive-ion mode for m/z 200–2,000. Identification of peptides was carried out in separate experiments by tandem MS (MS/MS) analysis in data-dependent acquisition mode, using collision-induced dissociation. All MS/MS spectra were analyzed using SEQUEST program (Thermo Fisher Scientific). Searching results from SEQUEST were further manually inspected, and only those verifiable were considered in HDX analysis. Triplicate measurements were carried out for each time point. HDExaminer (Sierra Analytics, Modesto, CA) was used to determine the deuteration level as a function of labeling time. The deuterium levels of peptides were not corrected for back-exchange, and therefore presented values reflect the relative exchange levels across the protein samples[54].

Differential scanning fluorimetry (DSF). The melting temperature of recombinant human NBD1 was measured as previously described[27]. DSF scans of NBD1 (7–12 μM) were obtained in 150 mM NaCl, 20 mM $MgCl_2$ and 10 mM HEPES, pH 7.5, and were performed using a Stratagene Mx3005p (Agilent Technologies, La Jolla, CA) qPCR instrument in the presence of 2× Sypro Orange. The medium ATP concentration was kept at 2.5 mM unless otherwise indicated.

Measurement of CFTR cell-surface density and stability. Cell-surface CFTR-3HA density and stability was measured by ELISA using anti-HA as previously described[52]. The plasma membrane density of CD4-NBD1 chimeras was measured by cell-surface ELISA using anti-CD4 (no. 555344; dilution 1:1,000; BD Biosciences; OKT4) in transiently transfected COS-7 cells as described[6,27]. The cell-surface density of CFTR and CD4-NBD1 chimeras was normalized with protein concentrations based on BCA assay. Data are presented as mean ± s.e.m. from at least two independent experiments consisting of multiple (3–8) measurements.

Western blotting and pulse-chase experiments. Western blotting and pulse-chase experiments were performed as previously described[52]. Cells were treated with correctors for 24 h at 37 °C and during the pulse-labeling and chase period.

ICl(apical) measurement. ICl(apical) measurements were essentially performed as described[13]. CFBE41o- cells were plated on ECM mix–coated 12-mm Snapwell filters (Corning, Corning, NY) at a density of 1×10^5 cells/cm^2. Polarized epithelia (≥5 d after confluence) were mounted in Ussing chambers, bathed in Krebs-bicarbonate Ringer and continuously bubbled with 95% O_2 and 5% CO_2. To impose a chloride gradient, Cl^- was replaced by gluconate in the apical compartment. To functionally isolate the apical plasma membrane, the basolateral plasma membrane was permeabilized with 100 μM amphotericin B, and the epithelial sodium channel was inhibited with 100 μM amiloride. CFTR activity

was stimulated by apical forskolin (0.03–10 μM) and genistein (100 μM) followed by the addition of CFTR inhibitor 172 (Inh_{172}, 20 μM) to determine CFTR-specific apical Cl^- current (ICl(apical)). Measurements were performed at 37 °C. The transepithelial resistance (TER) of CFBE was predominantly between 200–1,700 Ω/cm². Although the combination of three chemicals, including glycerol, reduced the TER of permeabilized CFBE cells to <200 Ω/cm², the CFTR-mediated ICl(apical) was measured by monitoring only the CFTR-Inh_{172}–sensitive component of the forskolin- and genistein-stimulated ICl(apical).

***In silico* docking.** *In silico* docking of VX-809, C4 and core-corr-II was performed using *AutoDock* 4.2.

Preparation of the proteins. Proteins were prepared initially in PDB format. The structure of the ΔF508-NBD1 was obtained from the PDB database (PDB code 2BBT)[44]. The F504N and Q646R solubilizing mutations were reverted back to Phe504 and Gln646 followed by energy minimization in SYBYL (Tripos Inc.) using the AMBER force field[55] with AMBERF99 charges[56]. Two full-length models, one representing the closed state (http://dokhlab.unc.edu/research/CFTR/home.html)[3] and the other representing the open state[36], were selected for docking. For comparability, both models were trimmed to contain amino acids 81–365, 391–648, 855–1154 and 1208–1429. Phe508 was deleted followed by the adjustment of the neighboring residues using the Modloop server[57] and by energy minimization using the AMBER force-field with AMBERF99 charges[56]. All of the amino acid numbers mentioned in the paper correspond to the amino acid numbering in the WT CFTR.

Preparation of ligands. Ligands were prepared in mol2 format. VX-809 was downloaded from the PubChem Compound Database in three-dimensional sdf format and converted to mol2 format by SYBYL. Molecules C4 and core-corr-II were drawn in two dimensions using ChemBioDraw Ultra 12.0 (CambridgeSoft) and saved in mol2 format. Their coordinates were copied to the PRODRG server[58] to obtain an initial three-dimensional structure. The resulting mol2 files of C4 and core-corr-II were corrected to contain appropriate atom types and were subjected to energy minimization in SYBYL. The Tripos force field with Gasteiger-Hückel charges were used until the RMS gradient of the energy derivative reached 0.01 kcal/mol/Å.

Running the docking simulations. The docking was performed using AutoDock[59]. Input files were prepared using AutoDock 4.2 with AutoDockTools (MGLTools 1.5.4). Ligands were loaded with Gasteiger charges and were saved in pdbqt format. Polar hydrogen atoms and Kollmann charges were added to the proteins. The grid box was centered on coordinates 95 Å, 70 Å, 170 Å, and its size was set to 70 Å × 70 Å × 70 Å with 0.5-Å spacing to include the two NBDs and the coupling helices (**Supplementary Fig. 5b**). To achieve this box size, AutoDock was compiled with a value of MAX_GRID_POINTS higher than the default 126, which allowed us to set 140 grid points in each direction. Lamarckian Genetic Algorithm was used for docking with default settings, except for the parameters ga_pop_size (300) and ga_num_evals (30,000,000) to perform exhaustive sampling. Three docking simulations with different random seeds for each drug and structure were run on the Hungarian HPC infrastructure (NIIF Institute, Hungary). From each run, 250 hits were collected.

Analysis of the docking simulations. All the 750 poses for each docking were used for analysis. These conformations were analyzed using SciPy (http://www.scipy.org/) and in-house written Python scripts. A Python package for SciPy called hcluster was used to perform hierarchic clustering based on r.m.s. deviation using the centroid distance measure. The threshold for forming flat clusters was set to include conformations with r.m.s. deviation values smaller than 3 Å into each cluster. Binding poses were clustered by r.m.s. deviation and ordered by the binding pose with the lowest binding free energy in each cluster. The four clusters, in which the pose with the lowest binding free energy exhibits one of the lowest binding free energies, are shown in figures, which were prepared in the PyMOL Molecular Graphics System (Schrödinger, LLC). The CFTR amino acid binding sites of correctors were determined using PyMOL by selecting residues with less than a 4 Å distance from every molecule in each cluster.

***In vitro* ubiquitination of NBD1.** *In vitro* ubiquitination of purified ΔF508-1S-NBD1 was performed as previously[27]. NBD1-1S (1 μM) was incubated for 5 min at 34 °C with 0.1% DMSO (as control) or C11 in reaction buffer (20 mM HEPES, pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT, 20 μM MG-132, 5 μg/ml leupeptine, 5 μg/ml pepstatin A) in the presence of 2 μM Hsc70. After the heat denaturation, the CHIP-mediated NBD1 ubiquitination was initiated by incubation at 26 °C for 2 h with 0.1 μM E1, 4 μM UbcH5c, 3 μM CHIP and 10 μM ubiquitin (Sigma). NBD1was analyzed by immunoblotting with anti-NBD1 (L12B4).

Crypt isolation and organoid culture from human rectal biopsies. Crypt isolation and culture of human intestinal organoids have been described previously[60]. Rectal suction biopsies were obtained for intestinal current measurements (ICM) during standard cystic fibrosis care by a procedure described previously[45] that has been approved by the Ethics Committee of the University Medical Centre Utrecht and the Erasmus Medical Centre of Rotterdam. In short, rectal biopsies were washed with cold complete chelation solution and incubated with 10 mM EDTA for 60–120 min at 4 °C. Supernatant was harvested, and EDTA was washed away. Crypts were isolated by centrifugation and embedded in matrigel (growth factor reduced, phenol-free, BD bioscience) and seeded (50–100 crypts per 50 μl matrigel per well) in 24-well plates. The matrigel was polymerized for 10 min at 37 °C and immersed in complete culture medium: advanced DMEM/F12 supplemented with penicillin-streptomycin, 10 mM HEPES, Glutamax, N2, B27 (all from Invitrogen), 1 mM *N*-acetylcysteine (Sigma) and the following growth factors: 50 ng/ml mEGF, 50% Wnt3a-conditioned medium and 10% Noggin-conditioned medium, 20% Rspo1-conditioned medium, 10 μM nicotinamide (Sigma), 10 nM gastrin (Sigma), 500 nM A83-01 (Tocris) and 10 μM SB202190 (Sigma). The medium was refreshed every 2–3 d, and organoids were passaged 1:4 every 7–10 days. Organoids from passages 1–10 were used for confocal live cell imaging.

CFTR transport activity of human rectal organoids. Organoids were seeded from a 7-d-old culture in a 4-μl matrigel placed into a flat-bottom 96-well culture plate (Nunc) and commonly contained 40–80 organoids and 100 μl culture medium. To visualize volume changes, 1 d after seeding, organoids were loaded with 10 μM Calcein green for 60 min (Invitrogen). After Calcein green treatment (with or without CFTR inhibition), 5 μM forskolin was added, and the organoids' morphology was monitored by time-lapse fluorescence laser confocal microscopy (LSM710, Zeiss, 5× objective). To inhibit CFTR, organoids were preincubated with 50 μM CFTR-Inh_{172} and 50 μM GlyH-101 (Cystic Fibrosis Foundation Therapeutics, Inc.) for 3 h. Images were collected every 10 min for 90 min in a top-stage incubator (5% CO_2 at 37 °C). Each condition was monitored in triplicate wells. For ΔF508-CFTR correction, organoids were preincubated for 20–24 h with 2 μM C4, 2 μM C18, 100 nM or 2 μM VX-809 or 125 mM myo-inositol (Sigma-Aldrich) as single treatment or in combinations. DMSO concentration was identical under all conditions and did not exceed 0.2% (w/v). The organoid surface area was quantified using Volocity (Improvision) imaging software. The normalized total organoid surface area was calculated and averaged from three individual wells per condition. The area under the curve was calculated using Prism (GraphPad Software).

Statistical analysis. Results are presented as mean ± s.e.m. for the number (n) of experiments. Statistical analysis was performed by two-tailed Student's t-test with the means of at least three independent experiments, and a 95% confidence level was considered significant.

52. Okiyoneda, T. *et al.* Peripheral protein quality control removes unfolded CFTR from the plasma membrane. *Science* **329**, 805–810 (2010).
53. Glasoe, P.K. & Long, F.A. Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.* **64**, 188–190 (1960).
54. Wales, T.E. & Engen, J.R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev.* **25**, 158–170 (2006).
55. Cornell, W.D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
56. Wang, J., Cieplak, P. & Kollmann, P.A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074 (2000).
57. Fiser, A. & Sali, A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **19**, 2500–2501 (2003).
58. Schüttelkopf, A.W. & van Aalten, D.M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363 (2004).
59. Morris, G.M. *et al.* AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
60. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).

doi:10.1038/nchembio.1253

CYSTIC FIBROSIS

Some gating potentiators, including VX-770, diminish ΔF508-CFTR functional expression

Guido Veit,[1] Radu G. Avramescu,[1] Doranda Perdomo,[1] Puay-Wah Phuan,[2] Miklos Bagdany,[1] Pirjo M. Apaja,[1] Florence Borot,[1] Daniel Szollosi,[3,4] Yu-Sheng Wu,[1] Walter E. Finkbeiner,[5] Tamas Hegedus,[3,4] Alan S. Verkman,[2] Gergely L. Lukacs[1,6,7]*

Cystic fibrosis (CF) is caused by mutations in the CF transmembrane regulator (*CFTR*) that result in reduced anion conductance at the apical membrane of secretory epithelia. Treatment of CF patients carrying the G551D gating mutation with the potentiator VX-770 (ivacaftor) largely restores channel activity and has shown substantial clinical benefit. However, most CF patients carry the ΔF508 mutation, which impairs CFTR folding, processing, function, and stability. Studies in homozygous ΔF508 CF patients indicated little clinical benefit of monotherapy with the investigational corrector VX-809 (lumacaftor) or VX-770, whereas combination clinical trials show limited but significant improvements in lung function. We show that VX-770, as well as most other potentiators, reduces the correction efficacy of VX-809 and another investigational corrector, VX-661. To mimic the administration of VX-770 alone or in combination with VX-809, we examined its long-term effect in immortalized and primary human respiratory epithelia. VX-770 diminished the folding efficiency and the metabolic stability of ΔF508-CFTR at the endoplasmic reticulum (ER) and post-ER compartments, respectively, causing reduced cell surface ΔF508-CFTR density and function. VX-770–induced destabilization of ΔF508-CFTR was influenced by second-site suppressor mutations of the folding defect and was prevented by stabilization of the nucleotide-binding domain 1 (NBD1)–NBD2 interface. The reduced correction efficiency of ΔF508-CFTR, as well as of two other processing mutations in the presence of VX-770, suggests the need for further optimization of potentiators to maximize the clinical benefit of corrector-potentiator combination therapy in CF.

INTRODUCTION

Cystic fibrosis (CF), one of the most common inherited diseases in the Caucasian population, is caused by mutations in the CF transmembrane regulator (*CFTR*) gene that lead to loss of CFTR channel function and impaired epithelial anion transport in the lung, intestine, pancreas, and other organs (*1*, *2*). The nearly 2000 different mutations identified in the *CFTR* gene (http://www.genet.sickkids.on.ca) have been categorized into six different classes according to the resulting molecular aberration (*3*, *4*). The most prevalent class II mutation, deletion of phenylalanine 508 (ΔF508), results in misfolded CFTR channels that are predominantly recognized and degraded by the endoplasmic reticulum (ER) quality control machinery (*2*, *5*). ΔF508-CFTR molecules that escape from the ER are functionally impaired (class III mutation) and conformationally unstable, with rapid removal from the plasma membrane (PM) by the peripheral quality control and targeting for endolysosomal degradation (*6*). G551D, the third most common CF-causing mutation that affects ~4% of CF patients, belongs to class III and displays normal processing and cell surface expression but severe functional impairment (*7*).

The CFTR protein is an ATP (adenosine 5′-triphosphate)–binding cassette transporter family member that comprises two membrane-spanning domains (MSD1 and MSD2) and three cytosolic domains, two nucleotide-binding domains (NBD1 and NBD2) and a regulatory domain (*8*). The ΔF508 mutation in the NBD1 produces multiple structural defects in CFTR. At least two of those, NBD1 misfolding and NBD1-MSD1/2 interfacial instability, have to be reversed genetically and/or pharmacologically to achieve near wild type–like PM expression (*9*–*13*).

Mechanistically, the available investigational small-molecule CFTR modulators fall into three classes: (i) suppressor molecules that prevent premature termination of protein synthesis; (ii) correctors that partially revert the folding and processing defects; and (iii) potentiators that increase channel gating and conductance (*14*–*16*). The potentiator ivacaftor (VX-770, Kalydeco) has been approved for therapy of CF patients with one copy of G551D (*17*) or some other rare gating mutations (*18*, *19*). VX-770 treatment of patients with G551D and other class III mutations demonstrated marked clinical benefit, including ~10 to 14% increase in the forced expiratory volume in 1 s (FEV_1), decrease in pulmonary exacerbations, and weight gain relative to placebo treatment (*20*–*22*).

Nasal potential difference and short-circuit current (I_{sc}) measurements in rectal biopsies of CF patients as well as in primary human bronchial epithelial (HBE) cells indicate that a subset of homozygous ΔF508 patients have residual ΔF508-CFTR PM function (*23*–*25*). The PM expression and activity of ΔF508-CFTR inversely correlate with CF disease severity (*23*, *25*, *26*). Acute addition of VX-770 in HBE cell cultures from some patients homozygous for the ΔF508 mutation increased the residual forskolin-stimulated channel activity from ~4 to 16% of that in HBE cultures from non-CF individuals, whereas other cultures were not responsive (*24*). A phase 2 trial in ΔF508 homozygous patients, however, showed no improvement in FEV_1,

[1]Department of Physiology, McGill University, Montréal, Quebec H3G 1Y6, Canada. [2]Departments of Medicine and Physiology, University of California, San Francisco, San Francisco, CA 94143–0521, USA. [3]MTA-SE Molecular Biophysics Research Group, Hungarian Academy of Sciences, 1444 Budapest, Hungary. [4]Department of Biophysics and Radiation Biology, Semmelweis University, 1444 Budapest P.O. Box 263, Hungary. [5]Department of Pathology, University of California, San Francisco, San Francisco, CA 94143–0511, USA. [6]Department of Biochemistry, McGill University, Montréal, Quebec H3G 1Y6, Canada. [7]Groupe de Recherche Axé sur la Structure des Protéines (GRASP), McGill University, Montréal, Quebec H3G 1Y6, Canada.
*Corresponding author. E-mail: gergely.lukacs@mcgill.ca

although a small reduction in sweat chloride concentration upon VX-770 treatment occurred (*27*).

Likewise, treatment with VX-809 (lumacaftor) alone, a promising investigational corrector drug that restores the ΔF508-CFTR PM expression and function to ~15% of wild-type CFTR activity in non-CF HBE cells (*28*), failed to show robust improvement in lung function of *ΔF508/ΔF508* patients (*29*). In cell cultures, a combination of chronic VX-809 and acute VX-770, together with a cAMP (cyclic adenosine 3′,5′-monophosphate) agonist, increased ΔF508-CFTR conductance to ~25% of that in non-CF HBE (*28*). These preclinical results motivated the ongoing phase 2-3 clinical trials of combination treatment with VX-770 and VX-809, or VX-661, another investigational corrector (*14*) (http://www.clinicaltrials.gov NCT01225211 and NCT01531673). The results of a phase 2 trial in homozygous ΔF508 patients receiving VX-809 and VX-770 combination treatment suggested an improvement in FEV_1 of 8.6% compared to placebo ($P < 0.001$) and a marginal decrease in sweat chloride concentration (*30*), although sustained clinical benefit awaits verification. The recent news release of the first phase 3 trials reported a mean absolute improvement in FEV_1 compared to placebo in the range of 2.6 to 4.0% ($P \leq 0.0004$) (*31*). The limited clinical efficacy of combination therapy based on these data may be accounted for by insufficient tissue concentration of the drugs, decreased susceptibility to correction in the inflamed lung, and/or conformational destabilization of the mutant upon chronic exposure to VX-770. To evaluate the latter possibility, we determined the effect of prolonged exposure to VX-770, and to other investigational potentiators, on the biochemical and functional expression of ΔF508-CFTR. The results indicate that VX-770 and some, but not all, other potentiators cause ΔF508-CFTR destabilization at multiple cellular sites in model systems and primary CF HBE, with consequent reduced functional expression of ΔF508-CFTR at the cell surface.

RESULTS

Prolonged exposure to VX-770 reduces the PM and cellular expression of ΔF508-CFTR

To investigate the effect of prolonged exposure to VX-770 on ΔF508-CFTR PM expression, we first used the human CF bronchial epithelial cell line CFBE41o− (referred to as CFBE), a widely validated model system with $CFTR^{\Delta F508/\Delta F508}$ genetic background but no detectable CFTR protein expression (*32*). CFBE cells were engineered for inducible expression of CFTR variants as described (*10*, *33*). To facilitate the PM detection of ΔF508-CFTR, horseradish peroxidase isoenzyme C (HRP-C) was genetically engineered into its fourth extracellular loop. The functional and biochemical properties of ΔF508-CFTR-HRP are similar to those of the 3HA-tagged variant (*13*, *34*) (fig. S1, A to D).

Acute addition of VX-770 to low temperature–rescued ΔF508-CFTR (rΔF508) in CFBE cells increased the cAMP-dependent protein kinase (PKA)–activated current by up to sixfold with EC_{50} of 12.8 ± 1.0 nM (fig. S2, A and B), similar to that reported in ΔF508/ΔF508 HBE cells (22 ± 10 nM) (*24*). Prolonged exposure (24 hours) to VX-770, however, caused a concentration-dependent decrease in the PM density of ΔF508-CFTR, regardless of whether the preincubation with VX-770 was done at physiological temperature or at 26 to 30°C, which facilitated ΔF508 CFTR biosynthetic processing (Fig. 1, A and B). The maximal reduction in ΔF508-CFTR PM density was attained at ~30 nM VX-770, well below the plasma concentration of ~3.5 μM in VX-770–treated CF patients (*35*). Although increasing the concentration of human serum (0 to 100%) raised the EC_{50} of VX-770 from 2.5 ± 0.2 nM to 23.1 ± 4.6 nM in the presence of VX-809, it did not affect the reduced PM density achieved by long-term treatment with ≥100 nM VX-770 (fig. S2, G and H). In contrast, the PM density of wild-type CFTR or G551D-CFTR was not reduced by prolonged VX-770 exposure (Fig. 1, A and C).

VX-809 partially restored ΔF508-CFTR biogenesis, function, and PM expression by about three- to fourfold in CFBE and primary HBE monolayers (fig. S2C) (*10*, *13*, *28*). VX-809 alone or in combination with low-temperature rescue, however, failed to prevent the VX-770–dependent reduction in ΔF508-CFTR PM density (Fig. 1, A and B, and fig. S2C). Similar results were obtained for ΔF508-CFTR rescued with the corrector VX-661 (fig. S2D). The VX-770–induced reduction in ΔF508-CFTR PM density was independent of channel gating because neither activation of adenyl cyclase by forskolin nor blocking the channel with BPO-27 (*36*) influenced the VX-770 effect (Fig. 1B and fig. S2C). Extended exposure to VX-770 did not affect cell viability (fig. S2E). PM down-regulation of 3HA-tagged ΔF508-CFTR by VX-770 in low temperature–rescued CFBE, NCI-H441 (a lung adenocarcinoma cell line exhibiting some Clara cell features), and MDCK II (Madin-Darby canine kidney) epithelial cells suggested that the VX-770 effect is not CFBE-specific or related to the HRP-tag insertion (Fig. 1D and fig. S2F).

To evaluate whether VX-770 causes the redistribution of PM resident ΔF508-CFTR to intracellular pools or exerts a global down-regulation of mature ΔF508-CFTR in post-ER compartments, we determined the cellular expression of ΔF508-CFTR by immunoblot analysis. VX-770 treatment for 24 hours decreased the amount of the complex-glycosylated ΔF508-CFTR (C-band) in CFBE lysates in a dose-dependent manner (Fig. 1E). The VX-770 effect was attenuated in VX-809– or VX-661–treated cells, probably due to partial stabilization of the mature ΔF508-CFTR pool by VX-809, as reported previously (Fig. 1, F and G) (*10*, *28*, *37*). In contrast, the complex-glycosylated form of wild-type CFTR and G551D-CFTR was not affected by prolonged VX-770 exposure (Fig. 1H). The modest, albeit significant ($P = 0.02$), decrease in the steady-state level of core-glycosylated ΔF508-CFTR (B-band) may be due to reduced biogenesis and/or accelerated ER degradation upon exposure to 100 nM VX-770 (Fig. 1, E to G). These observations suggest that the VX-770 effect cannot be explained merely by accelerated internalization or impeded recycling of rΔF508-CFTR.

ΔF508-CFTR chloride conductance decreases after long-term VX-770 exposure in CFBE and primary respiratory epithelia

To assess the functional consequence of prolonged VX-770 exposure of CFBE and primary HBE cells, we performed short-circuit current (I_{sc}) measurements after 24 hours of incubation with 100 nM VX-770. Forskolin-stimulated I_{sc} was measured after inhibition of ENaC (epithelial sodium channel) by amiloride and maximal acute potentiation of cell surface ΔF508-CFTR function with 10 μM VX-770 (Fig. 2A). Forskolin-stimulated I_{sc} (1.7 ± 0.3 μA/cm^2) was reduced to 1.1 ± 0.2 μA/cm after incubation of ΔF508-CFTR–expressing CFBE cells with VX-770 for 24 hours. A comparable reduction in I_{sc} was observed in VX-809– and VX-661–corrected cells (Fig. 2, A and B).

To confirm the relevance of these results to human tissues, we assessed the VX-770 effect in primary HBE cell cultures, isolated from the lungs of six $CFTR^{\Delta F508/\Delta F508}$ patients and four $CFTR^{WT/WT}$ donors. The HBE cells were differentiated on Snapwell filter inserts under air-liquid

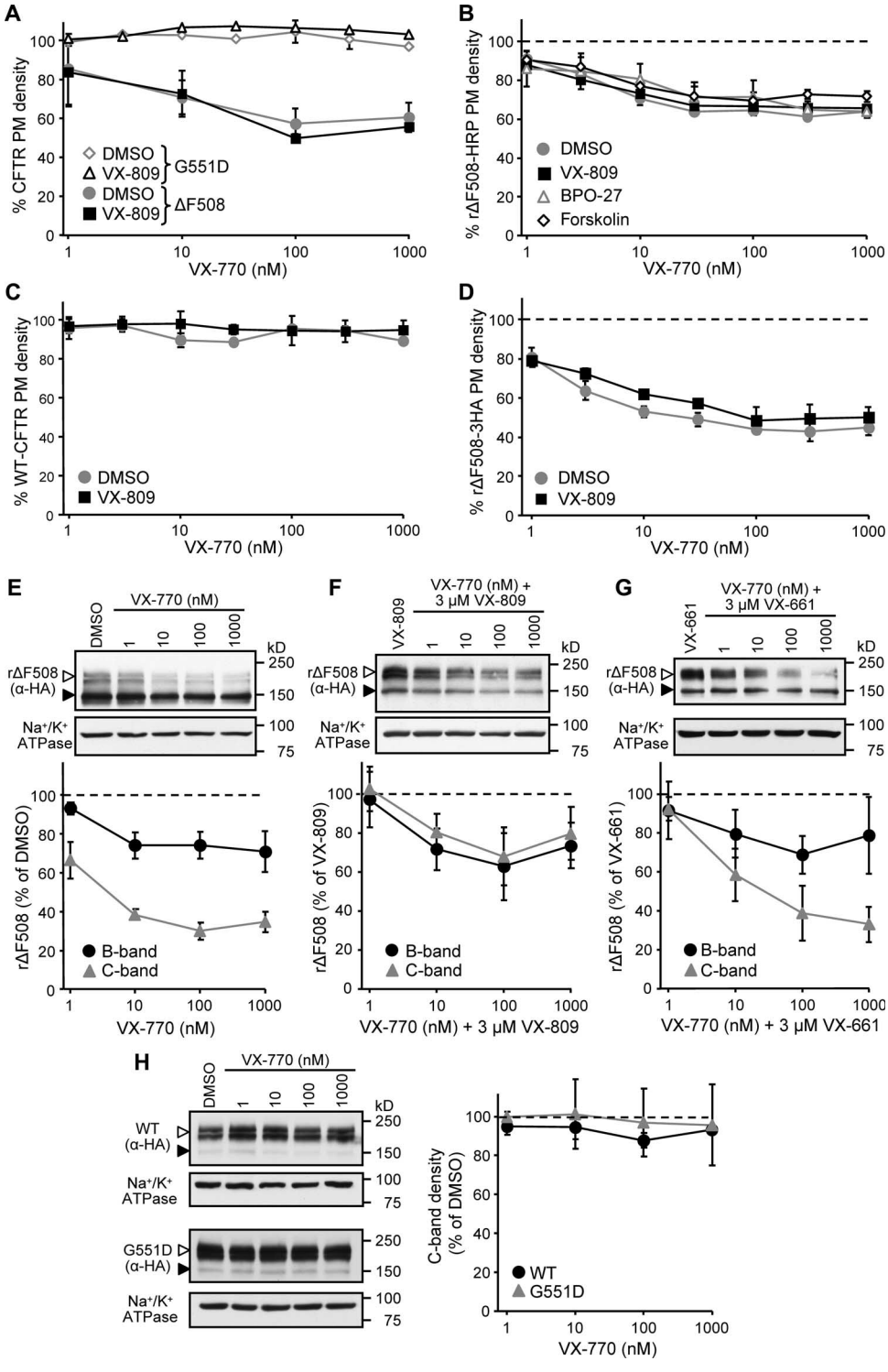


Fig. 1. Prolonged VX-770 exposure reduces the PM expression of ΔF508-CFTR but not wild-type (WT) or G551D-CFTR in human CFBE. (**A** and **C**) PM density of ΔF508-CFTR-HRP (ΔF508) (A), G551D-CFTR-3HA (G551D) (A), and WT-CFTR-3HA (C). Cells were treated with VX-770 for 24 hours in the presence or absence of 3 μM VX-809 at 37°C, and the values are expressed as percentage of non–VX-770-treated controls (n = 3). (**B** and **D**) PM density of low temperature (48 hours, 26°C)–rescued ΔF508-CFTR-HRP (rΔF508-HRP) (B) or ΔF508-CFTR-3HA (rΔF508-3HA) (D). Cells were treated with VX-770 in the presence or absence of VX-809 (3 μM), BPO-27 (25 μM), or forskolin (1 μM) for 24 hours at 26°C followed by a 1-hour chase at 37°C (n = 3). (**E** to **G**) Effect of VX-770 on the expression pattern of low temperature–rescued ΔF508-CFTR-3HA determined by immunoblot. Cells were treated with VX-770 alone (E) or in combination with VX-809 (3 μM) (F) or VX-661 (3 μM) (G) for 24 hours at 26°C. CFTR was visualized with anti-HA antibody, and anti–Na^+/K^+-ATPase antibody served as loading control. Densitometric analysis of the core-glycosylated (B-band, filled arrowhead) or complex-glycosylated (C-band, empty arrowhead) ΔF508-CFTR is expressed as percentage of control (lower panels, n = 3 to 4). (**H**) Effect of VX-770 on WT- and G551D-CFTR expression measured by immunoblot (left panel) and quantification of the C-band density (right panel, n = 3). Error bars indicate SEM of three or four independent experiments.

interface (ALI) conditions for at least 4 weeks either in Ultroser G medium (i), which increases the ENaC- and CFTR-mediated currents (*38*), or in ALI medium (ii) (*39*) (Fig. 2, C and D). The residual CFTR-mediated I_{sc} in the ΔF508-CFTR HBE was augmented by treatment with the correctors VX-809 or VX-661 (3 μM, 24 hours) (*13*). To further increase CFTR-mediated I_{sc} and isolate the apical anion conductance, some cells were differentiated in Ultroser G medium and analyzed after basolateral permeabilization and in the presence of a basolateral-to-apical Cl^- gradient. Independent of the differentiation method and presence of a chloride gradient, exposure to VX-770 for 24 hours decreased the VX-809– or VX-661–corrected ΔF508-CFTR current by 33 ± 6% and 47 ± 8% (mean ± SEM, n = 6), respectively (Fig. 2, C and D, and Table 1). In contrast, VX-770 pretreatment did not affect the PKA-activated wild-type CFTR current in HBE (Fig. 2E and table S1), in line with the absence of changes in PM and C-band density in wild-type CFTR (Fig. 1H).

Potentiator P5 does not impair the PM density and function of ΔF508-CFTR

To determine whether down-regulation of rΔF508-CFTR is a universal phenomenon of long-term potentiator exposure, we tested a panel of CFTR potentiators with distinct chemical structures. These investigational small molecules, abbreviated as P1 to P10, were made available

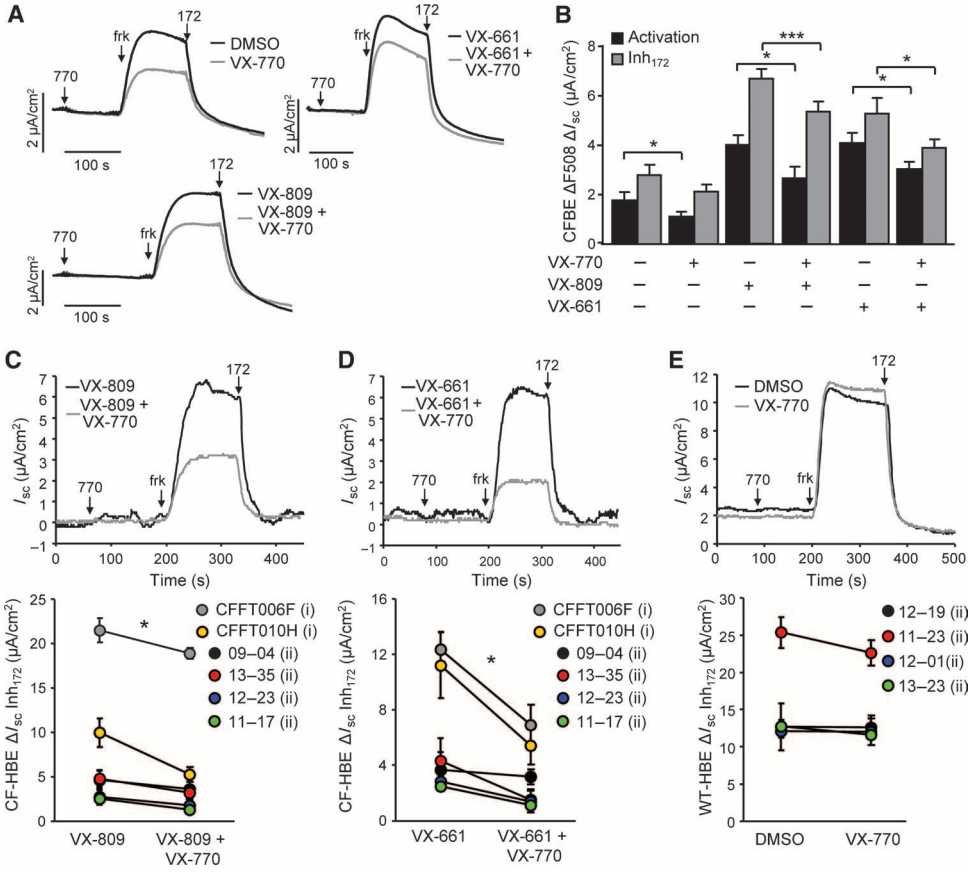


Fig. 2. Prolonged incubation with VX-770 reduces the maximal CFTR-mediated anion current in CFBE and primary cultures of ΔF508-CFTR but not WT-CFTR HBE. (**A** and **B**) Representative I_{sc} recordings (A) and quantification of the changes in I_{sc} (n = 3) (B) in CFBE monolayer expressing ΔF508-CFTR with or without 24-hour VX-770 (100 nM), VX-809 (3 μM), or VX-661 (3 μM) pretreatment. CFTR-mediated currents were induced by sequential acute addition of VX-770 (770, 10 μM) and forskolin (frk, 20 μM) followed by CFTR inhibition with Inh_{172} (172, 20 μM) in the presence of a basolateral-to-apical chloride gradient after basolateral permeabilization with amphotericin B. (**C** to **E**) Representative I_{sc} traces (upper panels) and quantification of the Inh_{172} inhibited current (ΔI_{sc} Inh_{172}, lower panels) in HBE isolated from six different homozygous ΔF508 CF patients (C and D) or four WT-CFTR donors (E) with or without VX-770 treatment (24 hours, 100 nM) alone (E) or in combination with VX-809 (3 μM) (C) and VX-661 (3 μM) (D). The HBE cells were either polarized in Ultroser G medium (i) followed by measurement in the presence of a basolateral-to-apical chloride gradient and basolateral permeabilization with amphotericin B, or polarized in ALI medium (ii) and measured as an intact monolayer with equimolar chloride concentrations in both chambers. Error bars indicate SEM of three independent experiments (B) or SD of triplicate measurements (C to E). *P < 0.05, ***P < 0.001 (exact P values are listed in table S3).

by the Cystic Fibrosis Foundation Therapeutics Inc. (CFFT) for the research community (fig. S3A). The potency and efficacy of P1 to P10 on the activity of low-temperature rΔF508-CFTR were demonstrated in CFBE cells using the halide-sensitive yellow fluorescent protein (YFP) quenching assay. Acute addition of P1 to P8 confirmed the potentiation of the rΔF508-CFTR activity, whereas P9 and P10 had only small effects (Fig. 3, A to H, and fig. S3, B to D). The dose-response curve of genistein (P6), a flavone widely used for acute potentiation of CFTR activity, did not reach saturation activity at 100 μM, suggesting that rΔF508-CFTR has lower affinity for genistein than wild-type CFTR (*40*, *41*) (Fig. 3F).

Prolonged treatment (24 hours) with most potentiators produced a concentration-dependent decrease in ΔF508-CFTR PM density in CFBE partially rescued with low temperature alone or in combination with VX-809 (Fig. 3, A to H). This was especially prominent for genistein (P6), with ~60 and ~75% decrease in rΔF508-CFTR PM density and conductance, respectively (Fig. 3, F, I, and J). P5 did not reduce the ΔF508-CFTR PM density and potentiated the rΔF508-CFTR activity by up to about sevenfold in CFBE cells (Fig. 3E). This result was confirmed by immunoblot analysis of low-temperature and VX-809 rΔF508-CFTR. Increasing concentrations of P5 did not alter the relative abundance of core- and complex-glycosylated ΔF508-CFTR, suggesting that P5 does not affect ΔF508-CFTR ER processing and stability (Fig. 4A).

I_{sc} measurements in CFBE expressing ΔF508-CFTR and primary HBE cells isolated from four $CFTR^{\Delta F508/\Delta F508}$ patients also confirmed the lack of effect of prolonged P5 exposure on the maximal activation of ΔF508-CFTR current at physiological temperature (Fig. 4, B and C, and Table 1). P5 may thus be a useful investigational compound to potentiate ΔF508-CFTR function without impairing its PM expression.

VX-770 impairs biogenesis, stability, and endocytic trafficking of the ΔF508-CFTR

Because VX-770 may impair both the biogenesis and the metabolic stability of mature ΔF508-CFTR, according to the immunoblot analysis, we measured conformational maturation of newly formed ΔF508 CFTR by the metabolic pulse-chase technique (*11*). Phosphorimage analysis was used to quantify the conversion efficiency of core-glycosylated ΔF508-CFTR (B-band), labeled with [^{35}S]methionine and [^{35}S]cysteine, into complex-glycosylated ΔF508-CFTR (C-band) upon traversing the cis/medial Golgi in CFBE cells. The folding efficiency of ΔF508-CFTR in the presence of VX-809 decreased from 1.8 ± 0.2% to 1.3 ± 0.1% after VX-770 treatment for 24 hours, representing a ~25% (P = 0.026) reduction (Fig. 5A, left panel). Similarly, the folding efficiency of VX-770–treated ΔF508-CFTR-3S, carrying NBD1-stabilizing second-site mutations, decreased from 4.0 ± 0.2% to 2.6 ± 0.6% (P = 0.049) (Fig. 5A, right panel). The reduced ER maturation efficiency cannot be attributed to decreased transcription or profoundly increased degradation of the core-glycosylated ΔF508-CFTR because neither the mRNA level nor the B-band stability was affected by VX-770 (fig. S4, A to C). The decreased incorporation of radioactivity during the 30-min pulse is consistent with increased cotranslational degradation and/or partial translational inhibition (fig. S4D).

Table 1. Effect of 24 hours of potentiator treatment on I_{sc} measurements of primary *CFTR*$^{\Delta F508/\Delta F508}$ HBE. n.d., not determined.

Patient	Differentiation medium	VX-809					VX-661		
		DMSO Mean ± SD (μA/cm^2)	VX-770 Mean ± SD (μA/cm^2)	Change %	P5 Mean ± SD (μA/cm^2)	Change %	DMSO Mean ± SD (μA/cm^2)	VX-770 Mean ± SD (μA/cm^2)	Change %
09-04	ALI (*39*)	4.61 ± 1.13	3.63 ± 1.17	−21	3.30 ± 0.44	−28	3.67 ± 1.27	3.17 ± 0.55	−14
13-35	ALI	4.77 ± 0.88	3.18 ± 0.17	−33	5.17 ± 0.23	8	4.33 ± 1.63	1.53 ± 0.74	−65
12-23	ALI	2.72 ± 0.94	1.78 ± 0.94	−35	2.57 ± 0.55	−6	2.85 ± 0.35	1.38 ± 0.79	−52
11-17	ALI	2.47 ± 0.40	2.47 ± 0.40	−49	3.03 ± 0.49	23	2.47 ± 0.15	1.13 ± 0.23	−54
Mean	ALI	3.64	2.46	−34	3.52	−1	3.33	1.80	−46
CFFT006F	Ultroser G (*38*)	21.52 ± 1.38	18.89 ± 0.59	−12	n.d.	n.d.	12.34 ± 1.27	6.90 ± 1.48	−44
CFFT010H	Ultroser G	9.95 ± 1.62	5.26 ± 0.80	−47	n.d.	n.d.	11.21 ± 2.36	5.44 ± 1.37	−51
Mean	Ultroser G	15.74	12.07	−30			11.78	6.17	−48
Combined (mean ± SEM)				−33 ± 6		−1 ± 11			−47 ± 8

The peripheral stability of rΔF508-CFTR was determined both at the PM and in post-ER compartments in CFBE cells. After the accumulation of low temperature–rescued ΔF508-CFTR at the PM, it was rapidly removed with a $T_{1/2}$ ~2.5 hours at 37°C (Fig. 5B), probably as a result of accelerated internalization, lysosomal targeting, and attenuated recycling, as reported in HeLa cells (*6*). The $T_{1/2}$ of rΔF508-CFTR at the PM was decreased by VX-770 to ~1.75 hours, regardless of whether VX-770 was present for 24 or 3 hours (Fig. 5B). VX-770 also accelerated the PM turnover of rΔF508-CFTR that was modestly stabilized by VX-809 (*28*), as reflected by the reduction of $T_{1/2}$ from ~3 to ~2.25 hours (Fig. 5C). In addition, VX-770 destabilized the complex-glycosylated rΔF508-CFTR pool both in the presence and in the absence of VX-809, as determined by cycloheximide (CHX) chase and immunoblot (Fig. 5, D and E). Similar results were obtained by metabolic pulse-chase for rΔF508-CFTR containing the 3S NBD1 stabilizing mutation (fig. S4E). Together, these observations suggest that VX-770 interferes with both the biogenesis and the peripheral stability of mature ΔF508-CFTR in the presence or absence of VX-809, which likely accounts for reduced ΔF508-CFTR PM function. As found for VX-770, prolonged treatment with potentiators P1, P2, P4, P6, and P7 led to a destabilization of PM-localized rΔF508-CFTR (Fig. 5F). Notably, destabilization was not seen in CFBE treated with P3 or P5 (Fig. 5F).

The conformational destabilization of rΔF508-CFTR by VX-770 may be recognized by the peripheral quality control machinery that targets non-native PM proteins for lysosomal degradation (*42*). This possibility was assessed by determining the post-endocytic fate of rΔF508-CFTR by measuring the pH of CFTR-containing endocytic vesicles with fluorescence ratio image analysis (FRIA) (*43*). rΔF508-CFTR was accumulated in polarized, filter-grown CFBE at reduced temperature and exposed for 24 hours to DMSO, VX-809, VX-770, or VX-809 + VX-770. The rΔF508-CFTR-3HA was unfolded at 37°C for 1.5 hours and then labeled with pH-sensitive FITC (fluorescein isothiocyanate) at 0°C, using the antibody capture technique as described in Materials and Methods. Internalization of labeled PM ΔF508-CFTR was initiated by shifting the temperature to 37°C.

Endocytosed rΔF508-CFTR was delivered after 30 min into multivesicular bodies (MVB)/lysosomes (pH 5.25 ± 0.01), whereas wild-type CFTR largely remained in early endosomes (Fig. 5, G and H, and fig. S4F). MVB/lysosomal delivery was inhibited by VX-809, as seen by preferential confinement to early endosomes (pH 6.31 ± 0.1) during a 2-hour chase (Fig. 5, G and H). VX-770 partially reversed the VX-809 trafficking effect by facilitating rΔF508-CFTR transfer to late endosomes, as indicated by the pH (5.84 ± 0.1) of the rΔF508-CFTR–containing vesicular compartment after a 2-hour chase (Fig. 5, G and H), suggesting that VX-770 increases ΔF508-CFTR susceptibility to recognition by the peripheral quality control machinery.

The effect of VX-770 at the single-molecule level was measured by determining the channel function of rΔF508-CFTR with or without second-site suppressor mutations in reconstituted planar phospholipid bilayer (Fig. 6, A to E). R29K and R555K mutations were introduced (ΔF508-CFTR-2RK) to increase channel reconstitution efficiency (*44*). The open probability (P_o) of phosphorylated ΔF508-CFTR-2RK channel decreased from 0.19 to 0.09 upon increasing the temperature from 24 to 36°C (Fig. 6B). VX-770 enhanced ΔF508-CFTR-2RK function at low temperature (P_o = 0.43 at 24°C), but accelerated its inactivation rate, as indicated by the ~3.5-fold faster loss of channel activity between 32 and 36°C (−18.1 ± 1.4%/°C in the presence of VX-770 versus −4.9 ± 2.8%/°C inactivation rate in the control) (Fig. 6, A to C). These results suggest a direct interaction of VX-770 with ΔF508-CFTR-2RK, resulting in its destabilization.

Second-site mutations modulate ΔF508-CFTR susceptibility to VX-770–mediated down-regulation

Second-site suppressor mutations in ΔF508-CFTR have been used to investigate mechanisms of small-molecule CFTR modulators (*10*) (fig. S5A). An increasing number of solubilizing mutations (1S, 2S, and 3S) progressively stabilize the isolated ΔF508-NBD1 domain energetically (see table S2 for list of all mutants) (*11*, *45*–*47*). Similarly, NBD1-MSD2 interface–stabilizing mutants (for example, R1070W or V510D) (*8*, *48*) enhance the PM expression of ΔF508-CFTR PM to ~5 to 10% of wild-type CFTR. Combining the two classes of mutations increased expression to ~50% of wild-type CFTR by aiding coupled domain folding (*10*, *11*). Unexpectedly, neither solubilizing nor interface-stabilizing mutations alone or in combination prevented the reduction in ΔF508-

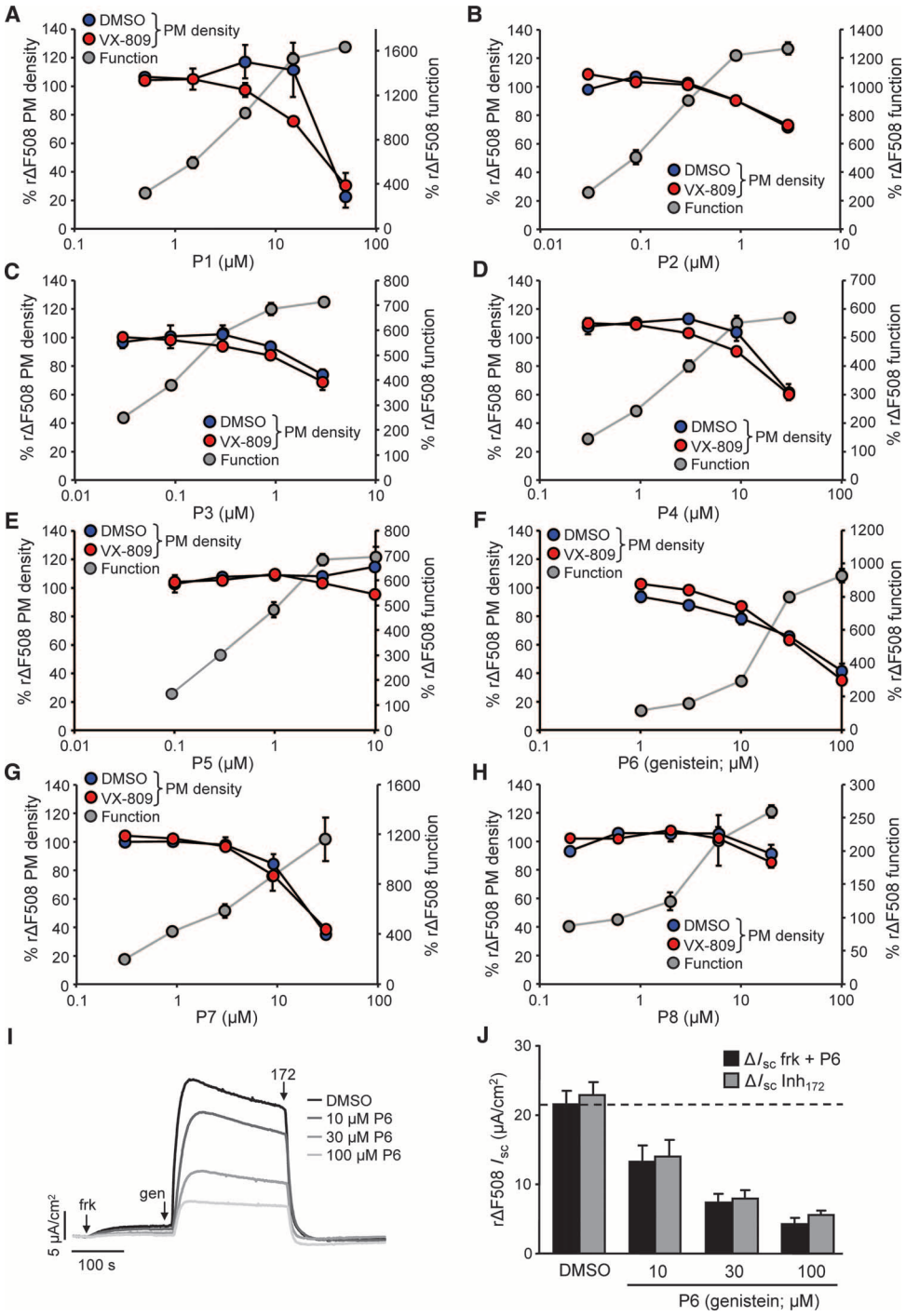


Fig. 3. Prolonged treatment with most potentiators reduces the PM density of rescued ΔF508-CFTR. (**A** to **H**) Effect of potentiators P1 (A), P2 (B), P3 (C), P4 (D), P5 (E), P6 (F), P7 (G), and P8 (H) on rΔF508-CFTR PM density in the presence or absence of 3 μM VX-809 (24-hour exposure, 26°C + 1-hour chase at 37°C, left axis, blue and red circles, $n = 3$) and function (acute addition, 32°C, right axis, gray circles, $n = 3$) in CFBE cells. (**I** and **J**) Representative I_{sc} recordings (I) and quantification of the changes in I_{sc} ($n = 3$) (J) in CFBE monolayer expressing rΔF508 with or without 24-hour P6 (genistein, 10 to 100 μM) pretreatment. CFTR was activated by sequential acute addition of forskolin (20 μM) and genistein (10 to 100 μM) followed by CFTR inhibition with Inh_{172} (20 μM) in the presence of a basolateral-to-apical chloride gradient after basolateral permeabilization with amphotericin B. Error bars indicate SEM of three independent experiments.

CFTR expression after prolonged VX-770 exposure, regardless of the presence of VX-809 (Fig. 7, A and B, and fig. S5, B to F). Solubilizing mutations augmented the loss of ΔF508-CFTR PM expression by VX-770 from ~45 to ~80% and decreased the IC_{50} of VX-770 from ~10 nM to 1 to 2 nM ($P = 0.0020$ to 0.0153) (Fig. 7, A to C, and fig. S5, B and C). In contrast, the C-terminal 70–amino acid truncation (Δ70 CFTR), which reduced the PM stability and expression by ~90% relative to wild-type CFTR (*49*), was resistant to prolonged VX-770 treatment (Fig. 7A and fig. S5G). The lack of apparent correlation between global down-regulation of ΔF508-CFTR PM density and the extent of VX-770–induced destabilization is further supported by the phenotype of revertant mutations [3R; G550E, R553Q, and R555K (*50*)] alone or in combination with 1S (R1S). These mutations energetically stabilize the ΔF508-NBD1 to an extent comparable to 3S, increase the PM density to ~40 to 50% of wild-type CFTR (*11*), and either directly or indirectly delay NBD1-NBD2 dimer dissociation (*11*, *51*) and channel closing, manifesting in about twofold increased P_o (Fig. 6D and fig. S6A). 3R or R1S mutations prevented the VX-770–induced down-regulation of ΔF508-CFTR from the PM (Fig. 7, A and D, and fig. S5H), and the R1S mutation eliminated the thermal inactivation in the bilayer and attenuated the potentiating effect of VX-770 (Fig. 6D and fig. S6A). Similarly, the ΔF508-E1371S (E1371S) mutation, which increases NBD1-NBD2 dimer stability by preventing the hydrolysis of bound ATP to the Walker A and B motifs in NBD2 and thereby the dissociation of the NBD1-NBD2 dimer (*52*, *53*), protected ΔF508-CFTR from thermal inactivation in reconstituted planar phospholipid bilayer experiments regardless of the presence of VX-770 (Fig. 6E and fig. S6B).

To offer a possible explanation for the VX-770 interaction with ΔF508-CFTR, we propose the formation of multiple binding pockets in the partially unfolded ΔF508-NBD1/2, based on molecular dynamics simulations and docking studies (fig. S7A). The putative binding sites of VX-770 are labeled in red in the ΔF508-CFTR structure (fig. S7B). High-affinity/low binding energy (less than −6.5 kcal/mol) interactions of VX-770 with amino acids in the NBD1/2 interface and the coupling

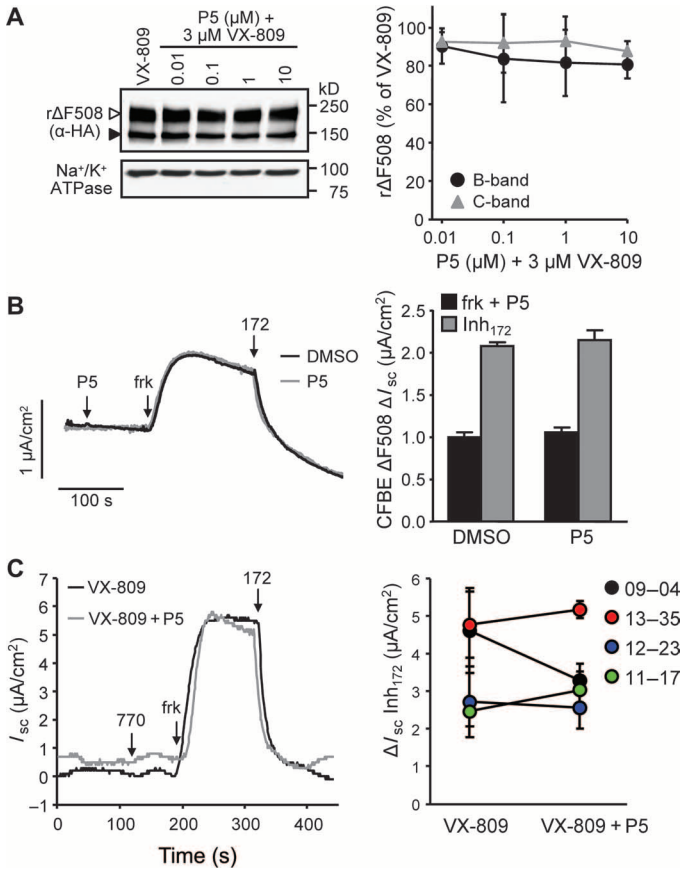


Fig. 4. Prolonged exposure to potentiator P5 does not impair the expression and function of ΔF508-CFTR. (**A**) Effect of P5 on the expression pattern of rΔF508, determined by immunoblot. Cells were treated with P5 in combination with VX-809, and CFTR was visualized using anti-HA antibody. Anti–Na^+/K^+-ATPase antibody served as loading control. Densitometric analysis of the core-glycosylated (B-band, filled arrowhead) or complex-glycosylated (C-band, empty arrowhead) rΔF508 is expressed as percentage of non–P5-treated controls (right panel, $n = 3$). (**B**) Representative I_{sc} recordings (left panel) and quantification of the changes in I_{sc} ($n = 3$, right panel) in CFBE monolayer expressing ΔF508-CFTR with or without 24-hour P5 (3 μM) pretreatment. Measurements were performed in the presence of a basolateral-to-apical chloride gradient after basolateral permeabilization. (**C**) Representative I_{sc} traces (left panel) and quantification of the Inh_{172} inhibited current (ΔI_{sc} Inh_{172}, right panel) in HBE isolated from four different homozygous ΔF508 CF patients with or without P5 treatment (3 μM, 24 hours, 37°C) in combination with VX-809 (3 μM). The HBE cells were polarized in ALI medium and measured without permeabilization with symmetrical chloride-containing solutions. Error bars indicate SEM of three independent experiments (A and B) or SD of triplicate measurements (C).

helix of CL1 in ΔF508-CFTR only partially overlap with those in wild-type CFTR (fig. S7C), consistent with the absence of the destabilizing effect of VX-770 in wild-type CFTR. The lack of VX-770 destabilizing effect on the revertant and NBD2 ATPase (adenosine triphosphatase) mutants is in line with the notion that prolonged NBD1-NBD2 dimerization (*51*, *54*) either directly or allosterically hinders the VX-770 association with destabilizing sites. Although the dockings studies suggest partially overlapping putative binding sites for P5 and VX-770, P5 also has unique interactions in NBD1 (amino acids 621 to 623) that are not observed for VX-770 (fig. S7D).

Rare CF mutations sensitize CFTR to VX-770–induced down-regulation

Gating potentiation of class III mutations by acute VX-770 exposure in preclinical settings rationalized the approval of ivacaftor in patients with eight rare CF mutations (*19*). To evaluate whether prolonged VX-770 exposure may interfere with the expression of other class III and class II mutations, we determined the PM density and function of R347H-, R170G-, and P67L-CFTR (fig. S8A) in CFBE. Whereas R347H-CFTR, a class III mutation (*55*), was resistant to VX-770, PM expression was reduced by 30 to 43% and forskolin-stimulated I_{sc} was reduced by 32 to 38% in VX-809–rescued R170G-CFTR and P67L-CFTR by VX-770 (Fig. 7, E to H, and fig. S8, B and C), raising the possibility that multiple class II mutations are susceptible to VX-770–mediated destabilization.

DISCUSSION

Here, we provide evidence that basal and VX-809– or VX-661–rescued PM densities of ΔF508-CFTR are reduced in parallel with the loss of chloride conductance upon prolonged exposure to VX-770. The reduction in channel density was not prevented by partial correction of the ΔF508-CFTR biogenesis defect by low temperature. In contrast, the PM expression of wild-type CFTR and G551D-CFTR was not reduced by prolonged VX-770 treatment.

Although the cellular basis of the destabilizing action of VX-770 on ΔF508-CFTR is well documented both in this work and in an accompanying publication (*56*), its molecular details remain to be elucidated. Although indirect effect cannot be ruled out, we postulate that direct association of VX-770 with ΔF508-CFTR increases its unfolding propensity, an inference supported by the accelerated rate of functional inactivation of temperature-rescued ΔF508-CFTR-2RK in reconstituted planar lipid bilayers. The conformational destabilization of ΔF508-CFTR by VX-770 is also in line with its reduced ER folding efficiency, accelerated PM, and post-ER pool turnover, as well as lysosomal targeting from early endosomes, phenomena that have been described as hallmarks of the peripheral quality control of non-native membrane proteins (*6*, *42*).

Conformational stabilization of NBD1 or the NBD1-MSD2 interface with second-site mutations (1S, 2S, or 3S and R1070W or V510D, respectively) sensitized ΔF508-CFTR to VX-770, as reflected by the increased fractional down-regulation and decreased IC_{50} (from ~10 to ~2 nM) despite partial rescue of misprocessing of these mutant variants (*11*). Despite conferring similar energetic stabilization as NBD1-3S, the revertant mutations that are clustered at residues 550 to 555 (*11*) bestowed complete resistance to PM down-regulation and thermal functional inactivation in the bilayer by VX-770. Because the revertant mutations probably stabilize the NBD1-NBD2 dimer (*51*), this observation suggests that an unstable NBD1-NBD2 interface is a prerequisite for the VX-770 destabilizing action, an inference that is supported by the resistance of ΔF508-CFTR-E1371S to VX-770. The E1371S substitution retards channel closing by preventing the hydrolysis of bound ATP to the Walker A and B motifs in NBD2 and thereby the dissociation of the NBD1-NBD2 dimer (*52*, *53*).

Monotherapy with VX-809, which increased ΔF508-CFTR–mediated I_{sc} to ~14% of that in wild-type CFTR in primary HBE (*28*), did not confer substantial clinical benefit (*29*). The maximally stimulated I_{sc} in VX-809–rescued ΔF508/ΔF508-HBE in combination with acute

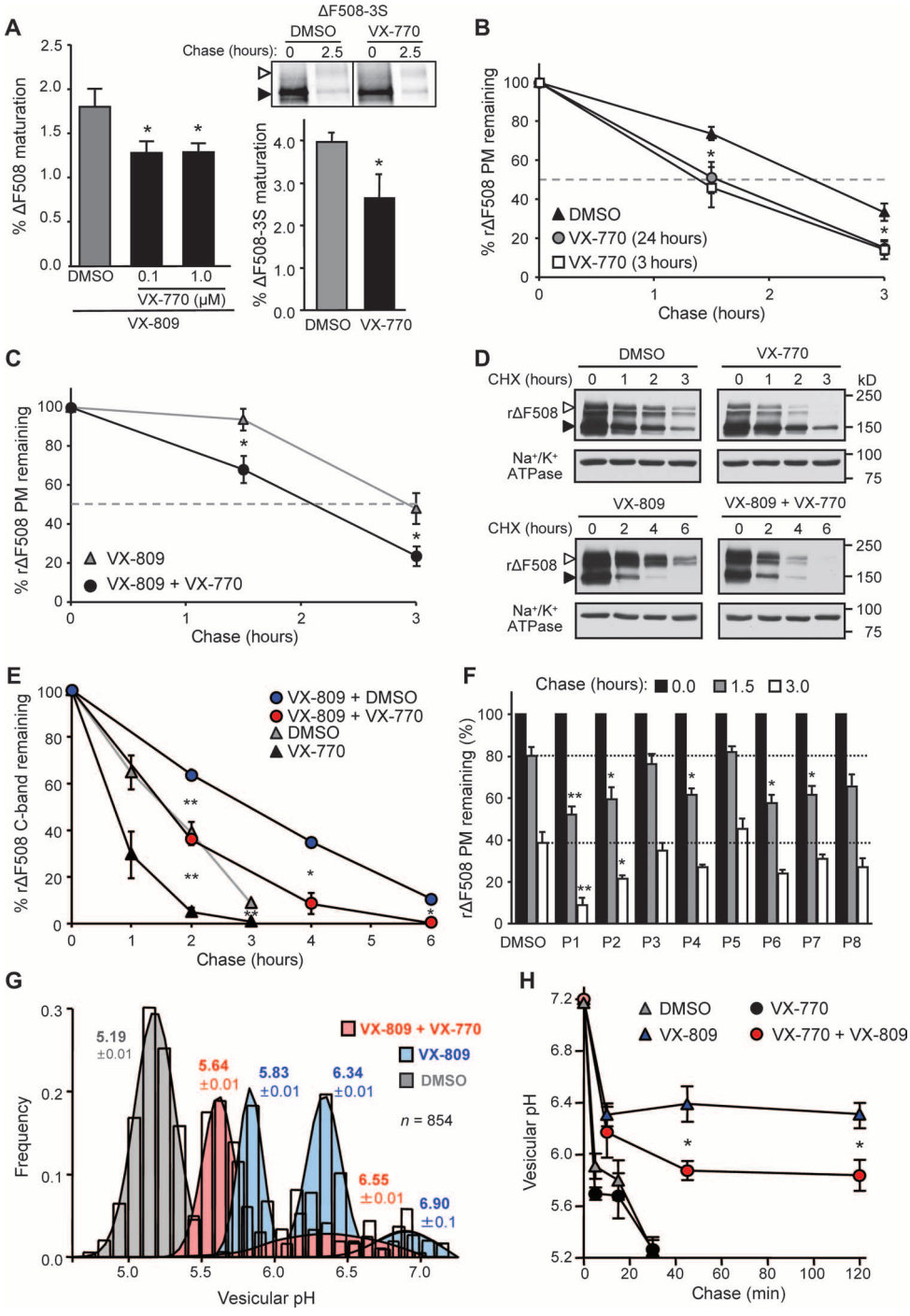


Fig. 5. VX-770 attenuates the maturation and accelerates the PM removal of ΔF508-CFTR. (**A**) Determination of ER folding efficiency of ΔF508-CFTR in the presence of VX-809 (3 μM, 24 hours, left panel, n = 5) or of ΔF508-3S in the absence of VX-809 (right panel, n = 3) by metabolic pulse-chase in CFBE cells with or without VX-770 (1.0 or 0.1 μM, 1-hour pretreatment). The folding efficiency was calculated as the percentage of pulse-labeled, immature core-glycosylated ΔF508-CFTR (B-band, filled arrowhead) conversion into the mature complex-glycosylated form (C-band, open arrowhead). Labeling was performed for 30 min followed by chase for 2.5 hours at 37°C. (**B** and **C**) Effect of VX-770 on the PM stability of low temperature–rescued (48 hours, 26°C) ΔF508-CFTR (rΔF508). CFBE monolayer was treated with VX-770 (1 μM, 3 or 24 hours, 26°C) alone (B) or in combination with VX-809 (3 μM) (C) followed by chase at 37°C for 1.5 or 3 hours (n = 4). (**D**) Stability of rΔF508 in CFBE cells treated with VX-770 (100 nM, 24 hours) and VX-809 (3 μM, 24 hours) was determined by immunoblot with CHX chase. (**E**) Complex-glycosylated CFTR [open arrowhead in (D)] disappearance was quantified by densitometry and is expressed as percentage of initial amount (n = 3). (**F**) Effect of each potentiator (P1, P4, P7, and P8, 30 μM; P2 and P3, 3 μM; P5, 10 μM; and P6, 100 μM; 24 hours, 26°C) on the PM stability of rΔF508 (n = 3). (**G**) Representative histogram of rΔF508-containing vesicular pH measured by FRIA. The cells were treated with dimethyl sulfoxide (DMSO), VX-809 (3 μM), or VX-809 and 0.1 μM VX-770 for 24 hours and during the 120-min chase. The Gaussian distribution of the vesicular pH of 280 vesicles is indicated for each condition. (**H**) Influence of VX-770 on the endolysosomal transfer kinetics of rΔF508. The graph shows the mean vesicular pH at each chase point (n = 3). Error bars indicate SEM of three to five independent experiments. *P < 0.05, **P < 0.01 (exact P values are listed in table S3).

VX-770 potentiation reached ~23% of that in non-CF HBE in our studies (Table 1 and table S1). These results are consistent with published data suggesting an I_{sc} equivalent to ~25% of that in non-CF HBE (twofold increase over VX-809 alone) (*28*). Prolonged VX-770 exposure reduced the ΔF508 I_{sc} by 33 ± 6% (mean ± SEM, n = 6) in VX-809–corrected HBE, albeit with a considerable range of values in cells from different CF patients (12 to 49%, Table 1), which is likely due to the influence of the genetic or epigenetic variability between individuals on the proteostasis network activity (*57*, *58*). In VX-661–rescued ΔF508/ΔF508-HBE, the VX-770–induced functional attenuation ranged between 14 and 65% with a mean of 47 ± 8% (± SEM, n = 6) (Table 1). If these results translate to the clinical setting, combination of corrector therapy with VX-770 could be beneficial for a subpopulation of ΔF508 CF patients, but may reduce the overall rescue efficiency below the threshold required for clinical benefit in poor VX-809 responders and/or individuals susceptible to VX-770–mediated channel destabilization. The negative

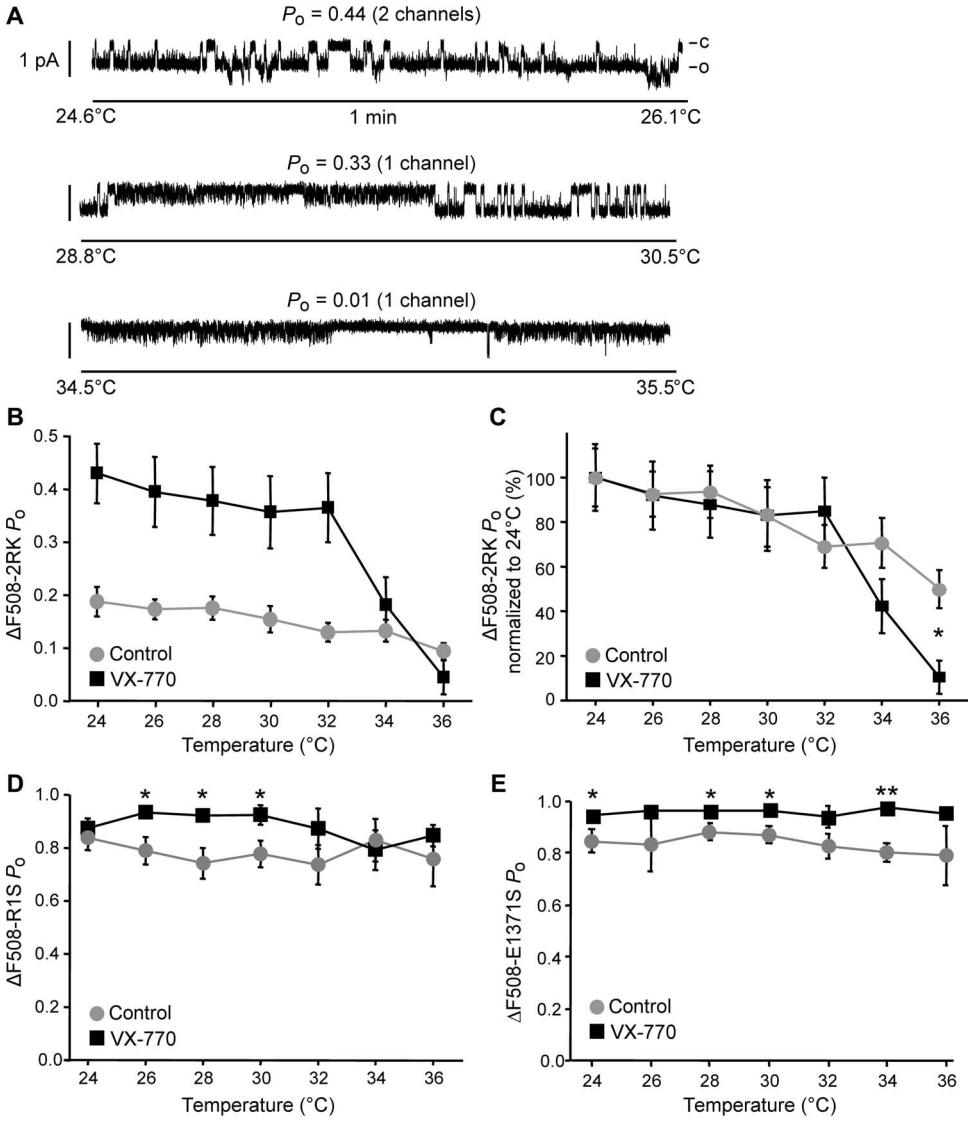


Fig. 6. VX-770 directly interacts with and destabilizes the ΔF508-CFTR-2RK channel. (**A**) Representative records show ΔF508-R29K-R555K-CFTR (ΔF-2RK) channel function in the presence of 1 μM VX-770 from two separate experiments at ~25°C (two channels incorporated), ~30°C, and ~35°C (one channel incorporated). The closed (c) and open (o) states of the channels are indicated. (**B**) Single-channel open probabilities (P_o) of protein kinase A–activated ΔF508-CFTR-2RK in the presence or absence of VX-770 (1 μM) (n = 7 to 46). The cumulative duration of single-channel measurements for any given temperature exceeded 8.5 min. (**C**) Temperature-dependent inactivation of ΔF508-2RK in the presence or absence of VX-770 (1 μM) derived by normalization with 24°C values of the results depicted in (B). (**D** and **E**) Single-channel open probabilities (P_o) of protein kinase A–activated ΔF508-CFTR-R1S (ΔF508-R1S, n = 4 to 13) (D) or ΔF508-CFTR-E1371S (ΔF-E1371S, n = 4 to 6) (E) determined by artificial phospholipid bilayer measurements in the presence or absence of VX-770 [1 μM in (D), 0.1 μM in (E)]. The cumulative measurement time for any given temperature was >4.5 min for ΔF508-R1S and >3 min for ΔF508-E1371S. Error bars indicate SEM of 4 to 46 independent experiments. *P < 0.05, **P < 0.01 (exact P values are listed in table S3).

action of VX-770 would be more pronounced in patients having a single copy of ΔF508-CFTR.

A limitation of our study is that the concentration of free VX-770 in lung cells of CF patients treated with VX-770 is not known, and hence, the VX-770–induced down-regulation of ΔF508-CFTR can only be extrapolated. VX-770 treatment with the recommended dose of 150 mg every 12 hours produced a peak plasma concentration of 3.5 μM after 5 days (*35*). Because ~97% of VX-770 is bound to plasma protein (*35*, *59*), the free drug concentration is estimated to be ~100 nM in vivo, which is in line with its clinical benefit in patients carrying at least one G551D allele (*21*, *22*) despite its relatively high EC_{50} for activation of G551D-CFTR (236 nM VX-770) in primary HBE (*24*). In our studies, maximal reduction of ΔF508-CFTR PM density and function was seen at ~30 nM VX-770 in CFBE. The effective in vivo intracellular concentration of VX-770, however, could be even higher than predicted due to the accumulation of VX-770 in primary HBE cells (*56*) and its eightfold enrichment in the epithelial lining fluid of the rat lung relative to plasma (*59*).

Because CFTR-mediated transepithelial transport is closely correlated with CF disease severity (*23*, *25*, *60*), evaluation of sustained exposure to modulators should be valuable in identifying individuals who would benefit from corrector-potentiator combination therapy with, for example, patient-derived primary or conditionally reprogrammed respiratory cells (*61*), or intestinal organoids (*62*). As a complementary strategy, new potentiators might be identified that lack the ΔF508-CFTR destabilizing action, as exemplified by the P5 (*63*). P5 efficiently stimulated the activity of phosphorylated rΔF508-CFTR, but not G551D (*63*). The latter suggests that the mechanism of action of P5 is distinct from that of VX-770 and therefore could be valuable to correct the gating defect of ΔF508-CFTR while preserving its PM density.

In summary, our results suggest that VX-770, as well as most of the available investigational potentiators, impairs the biochemical stability of ΔF508-CFTR and other class II processing mutations (such as P67L and R170G). These findings in cell cultures may translate to reduced efficacy of corrector-potentiator combination therapy in the clinical setting. Further structure-activity studies of existing potentiators, as well as identification of potentiators that do not destabilize mutant CFTRs, are warranted to enhance the therapeutic benefit of corrector-potentiator combination therapy in CF.

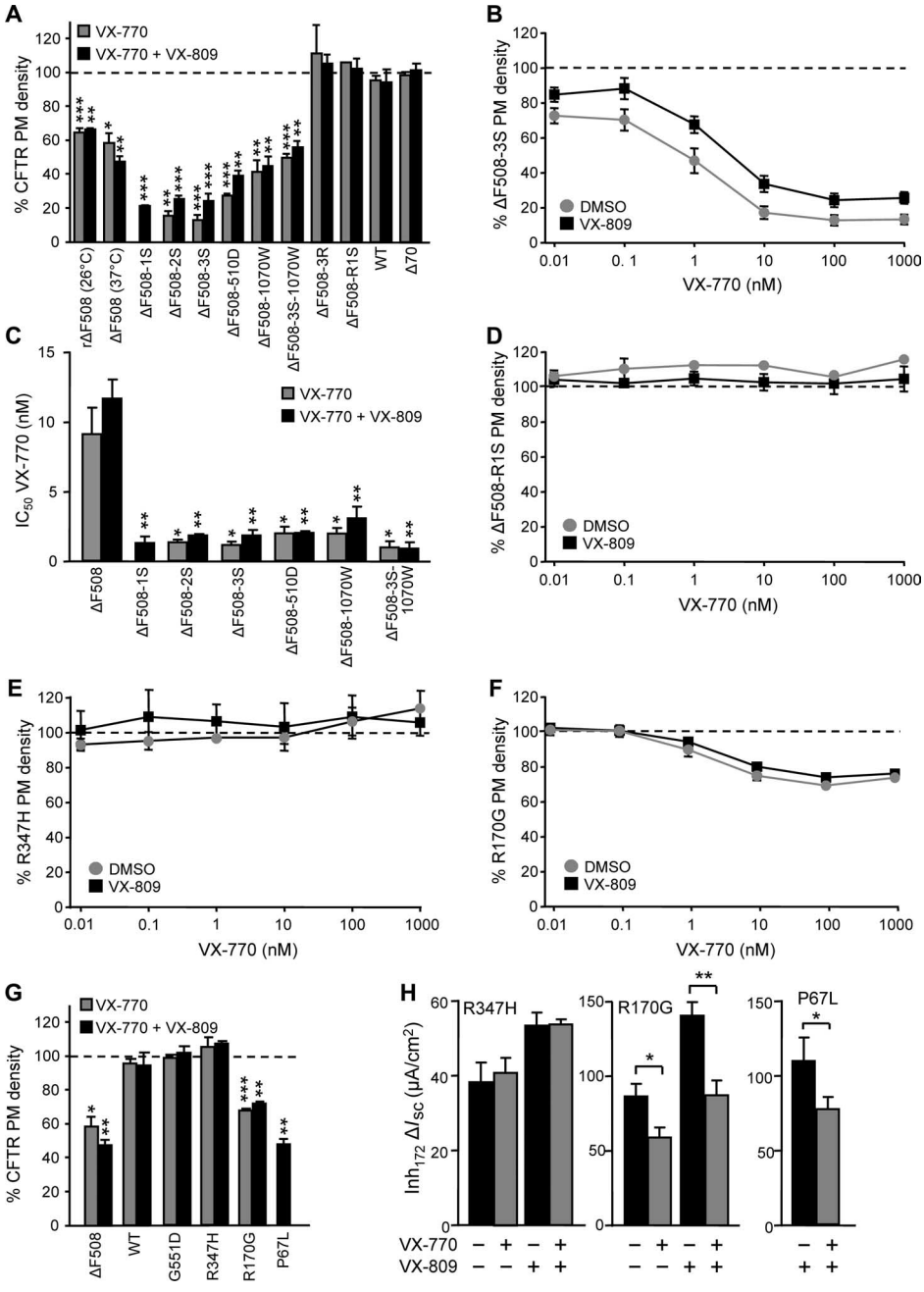


Fig. 7. Suppressor mutations of ΔF508-CFTR and other CF-causing CFTR mutations modulate the susceptibility to VX-770–mediated PM down-regulation. (**A**) PM density of ΔF508-CFTR with or without second-site suppressor mutations and of non–ΔF508-CFTR variants was measured after VX-770 (100 nM, 24 hours) incubation (n = 3). (**B** and **D**) PM density of ΔF508-3S (B) (n = 3) and ΔF508-R1S (D) (n = 3) after 24 hours of treatment with increasing concentrations of VX-770 in the presence or absence of VX-809 (3 μM). (**C**) 50% inhibitory concentration (IC_{50}) of VX-770 on CFTR variants' PM expression, calculated on the basis of the measurements shown in (B), (D), and fig. S5 (B to F). (**E** and **F**) PM density of R347H-CFTR (E) and R170G-CFTR (F) after 24 hours of treatment with increasing concentrations of VX-770 in the presence of DMSO or VX-809 (3 μM). (**G**) CFTR PM density after VX-770 (100 nM, 24 hours) incubation, measured for CF-causing mutants G551D-, R347H-, R170G-, and P67L-CFTR (n = 3). The values for ΔF508- and WT-CFTR are shown for comparison. (**H**) Quantification of the changes in I_{sc} (n = 3) (B) in CFBE monolayers expressing R347H-, R170G-, or P67L-CFTR with or without 24-hour VX-770 (100 nM) and VX-809 (3 μM) pretreatment (n = 3 to 4). Error bars indicate SEM of three to four independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001 (exact P values are listed in table S3).

MATERIALS AND METHODS

Study design

The goal of the study was to measure the effect of long-term administration of VX-770 and other investigational potentiators, either alone or in combination with small-molecule correctors, on the biochemical and functional expression of ΔF508-CFTR in immortalized and primary human respiratory epithelia. CFBE heterologously expressing wild-type, G551D-, or ΔF508-CFTR and primary HBE with a $CFTR^{WT/WT}$ (from four donors) or a $CFTR^{\Delta F508/\Delta F508}$ (from six patients) genotype were subjected to chronic treatment with VX-770 (24 hours) in the presence or absence of the correctors VX-809 or VX-661, and CFTR function was determined by short-circuit current measurement or halide-sensitive YFP fluorescence quenching. The PM density, PM stability, cellular expression, conformational maturation, metabolic stability, and lysosomal targeting of ΔF508-CFTR were determined in CFBE to examine the cellular phenotype of the VX-770 destabilizing action. To elucidate the molecular mechanism of the VX-770 effect, we examined the functional destabilization of the mutant in a planar lipid bilayer and investigated the PM density and function of ΔF508-CFTR containing second-site mutations. Putative binding sites of potentiators were identified by molecular dynamic simulation and in silico docking. Finally, to assess the VX-770 effect specificity, the PM density and function of other class III and class II mutations (R347H-, R170G-, and P67L-CFTR) were determined in CFBE.

Antibodies and reagents

Mouse monoclonal anti–hemagglutinin (HA) anitbody was purchased from Covance Innovative Antibodies. VX-770, VX-809, and VX-661 were acquired from Selleckchem. The CFTR potentiators P1 to P10 were made available by R. J. Bridges (Rosalind Franklin University of Medicine and Science) and CFFT. All other chemicals were purchased from Sigma-Aldrich at the highest grade available.

Cell lines

Full-length human CFTR variants with the 3HA-tag in the fourth extracellular loop have been described before (*10*).

The HRP-C was introduced into the fourth extracellular loop replacing the 3HA-tag by using the Eco RV/Avr II restriction sites with a 5′ linker (ctcgaatcaggaggtagtggtggcggaagt). The CFTR variants used in this study are listed in table S2. Maintenance of the human CF bronchial epithelial cell line CFBE41o– (*32*), with a $CFTR^{\Delta F508/\Delta F508}$ genotype [a gift from D. Gruenert, University of California, San Francisco (UCSF)], and stable cell line generation of CFTR-3HA variants under the control of a tetracycline-responsive transactivator were described before (*33*). NCI-H441 and MDCK II cells expressing inducible ΔF508-CFTR have been described (*10*).

Primary HBE and short-circuit current measurements

Primary cultures of HBE cells from four $CFTR^{\Delta F508/\Delta F508}$ CF patients and four $CFTR^{WT/WT}$ donors were isolated and grown at ALI in ALI differentiation medium, as described (*39*, *64*). Primary cultures differentiated in Ultorser G medium (*38*) from two $CFTR^{\Delta F508/\Delta F508}$ CF patients were purchased from ChanTest. I_{sc} measurements of primary HBE and CFBE epithelia were performed as described (*33*, *65*).

PM density measurement

The PM density of 3HA-tagged CFTR variants was determined by cell surface enzyme-linked immunosorbent assay (ELISA) (*6*). HRP-tagged CFTR PM density was measured in a VICTOR Light Plate Reader (PerkinElmer) after addition of HRP-substrate (50 μl per well; SuperSignal West Pico, Thermo Fisher Scientific). PM density measurements were normalized with cell viability determined by alamarBlue Assay (Invitrogen).

Halide-sensitive YFP quenching assay

Assay of ΔF508-CFTR function by halide-sensitive YFP fluorescence quenching was performed as described (*33*). CFBE cells expressing inducible ΔF508-CFTR were transduced with lentiviral particles encoding the halide sensor YFP-F46L/H148Q/I152L (*66*) followed by isolation of double-expressing clones. YFP-expressing cells or controls were seeded onto 96-well microplates at a density of 2×10^4 cells per well, induced for ΔF508-CFTR expression for 2 days at 37°C, and low temperature–rescued for an additional 48 hours at 26°C. During the assay, cells were incubated in phosphate-buffered saline (PBS)–chloride (50 μl per well) (140 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 1.1 mM $MgCl_2$, 0.7 mM $CaCl_2$, and 5 mM glucose, pH 7.4) containing the indicated potentiator concentrations, followed by well-wise injection of activator solution (50 μl per well) [20 μM forskolin, 0.5 mM 3-isobutyl-1-methyl-xanthine (IBMX), 0.5 mM 8-(4-chlorophenylthio)-adenosine-3′,5′-cyclic monophosphate (cpt-cAMP)] and 100 μl of PBS-iodide, in which NaCl was replaced with NaI. The fluorescence was monitored for 36 s with a 5-Hz acquisition rate at 485-nm excitation and 520-nm emission wavelengths using a POLARstar OPTIMA (BMG LABTECH) fluorescence plate reader. After background subtraction and normalization to YFP signals before NaI injection, the I^- influx rate was calculated by linear fitting to the initial slope.

Pulse-chase labeling

Experiments were performed as described (*11*). Briefly, CFBE cells expressing ΔF508-CFTR were pretreated with 3 μM VX-809 for 24 hours followed by 1-hour exposure to 1 or 0.1 μM VX-770. CFBE cells expressing ΔF508-CFTR-3S were exposed to 1 μM VX-770 for 1 hour. CFTR variants were pulse-labeled with [^{35}S]methionine and [^{35}S]cysteine (0.2 mCi/ml) (EasyTag EXPRESS Protein Labeling Mix, PerkinElmer) in cysteine- and methionine-free medium for 30 min and chased in full medium for 2.5 or 4.5 hours at 37°C in the presence of the indicated compounds. Radioactivity incorporated into the core- and complex-glycosylated CFTR was visualized by fluorography and quantified by phosphorimage analysis with a Typhoon imaging platform (GE Healthcare).

Fluorescence ratiometric image analysis

The methodology for FRIA of endocytic vesicles containing CFTR as a cargo has been described in detail (*43*). Briefly, filter-grown CFBEi-ΔF508-3HA was allowed to polarize for 5 days and temperature-rescued for 48 hours at 30°C. VX-809 or VX-809 (3 μM) with 0.1 μM VX-770 was added for the last 24 hours and kept during the experiment. Before labeling, the cells were shifted to 37°C for 1.5 hours. Subsequently, rΔF508-CFTR-3HA was labeled with anti-HA antibody and FITC-conjugated goat anti-mouse secondary Fab (Jackson ImmunoResearch) on ice. Synchronized internalization was performed at the indicated times at 37°C. FRIA was performed on a Zeiss AxioObserver Z1 inverted fluorescence microscope (Carl Zeiss MicroImaging) equipped with an X-Cite 120Q fluorescence illumination system (Lumen Dynamics Group Inc.) and Evolve 512 EMCCD (electron-multiplying charge-coupled device) camera (Photometrics Technology). The acquisition was carried out at 495 ± 5–nm and 440 ± 10–nm excitation wavelengths with a 535 ± 25–nm emission filter and analyzed with MetaFluor software (Molecular Devices).

Planar lipid bilayer studies

Isolated CFTR-containing microsomes (containing 20 to 40 μg of total protein) were fused to planar lipid bilayers, and currents were analyzed as described previously (*10*, *11*). Briefly, voltage was clamped at –60 mV, and currents were measured with a BC-535 amplifier (Warner Instrument) and pCLAMP 9 or 10 software (Axon Instruments), filtered at 200 Hz, and sampled at 10 kHz with an eight-pole Bessel filter and Digidata 1320 or 1440 digitizer (Axon Instruments). The chamber was gradually warmed from 23 to 37°C, reaching the maximum temperature within 8 to 9 min. Open probability (P_o) values were calculated for 2°C temperature intervals between 23 and 37°C, and the midpoints are depicted on the *x* axis.

Statistical analysis

Results are presented as means ± SEM, with the number of experiments indicated. Statistical analysis was performed by two-tailed Student's *t* test with the means of at least three independent experiments. The 95% confidence level was considered significant. For I_{sc} measurements, we performed paired *t* test analysis of patient/donor cells measured under different conditions. We used the Hill equation to calculate IC_{50} values in the GraphPad Prism software package. Original data and exact *P* values are provided in table S3.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/6/246/246ra97/DC1

Materials and Methods

Fig. S1. Comparison between 3HA- and HRP-tagged ΔF508-CFTR.

Fig. S2. Acute VX-770 potentiation in CFBE and the effect of extended VX-770 exposure on the PM density of rΔF508-CFTR in other epithelial cell models.

Fig. S3. Structure and effect of CFFT potentiator panel on the rΔF508-CFTR function and PM density.

Fig. S4. The effect of prolonged exposure to VX-770 on the biogenesis and stability of CFTR variants.

Fig. S5. The effect of VX-770 on the PM density of CFTR variants.
Fig. S6. Representative records of ΔF508-R1S and ΔF508-E1371S activities in artificial phospholipid bilayer.
Fig. S7. In silico modeling of VX-770 and P5 binding to wild-type and ΔF508-CFTR cytosolic regions.
Fig. S8. The effect of prolonged VX-770 exposure on the CF-causing mutants R347H-, R170G-, and P67L-CFTR.
Table S1. Effect of 24 hours of potentiator treatment on I_{sc} measurements in primary *CFTR*$^{WT/WT}$ HBE.
Table S2. CFTR mutants used in this study.
Table S3. Data and derived *P* values used in composite graphs (provided as an Excel file).
References (*67*–*70*)

REFERENCES AND NOTES

1. F. S. Collins, Cystic fibrosis: Molecular biology and therapeutic implications. *Science* **256**, 774–779 (1992).
2. J. R. Riordan, CFTR function and prospects for therapy. *Annu. Rev. Biochem.* **77**, 701–726 (2008).
3. M. P. Rogan, D. A. Stoltz, D. B. Hornick, Cystic fibrosis transmembrane conductance regulator intracellular processing, trafficking, and opportunities for mutation-specific treatment. *Chest* **139**, 1480–1490 (2011).
4. M. J. Welsh, A. E. Smith, Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell* **73**, 1251–1254 (1993).
5. K. Du, M. Sharma, G. L. Lukacs, The ΔF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. *Nat. Struct. Mol. Biol.* **12**, 17–25 (2005).
6. T. Okiyoneda, H. Barrière, M. Bagdány, W. M. Rabeh, K. Du, J. Höhfeld, J. C. Young, G. L. Lukacs, Peripheral protein quality control removes unfolded CFTR from the plasma membrane. *Science* **329**, 805–810 (2010).
7. S. M. Rowe, S. Miller, E. J. Sorscher, Cystic fibrosis. *N. Engl. J. Med.* **352**, 1992–2001 (2005).
8. A. W. Serohijos, T. Hegedus, A. A. Aleksandrov, L. He, L. Cui, N. V. Dokholyan, J. R. Riordan, Phenylalanine-508 mediates a cytoplasmic–membrane domain contact in the CFTR 3D structure crucial to assembly and channel function. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3256–3261 (2008).
9. J. L. Mendoza, A. Schmidt, Q. Li, E. Nuvaga, T. Barrett, R. J. Bridges, A. P. Feranchak, C. A. Brautigam, P. J. Thomas, Requirements for efficient correction of ΔF508 CFTR revealed by analyses of evolved sequences. *Cell* **148**, 164–174 (2012).
10. T. Okiyoneda, G. Veit, J. F. Dekkers, M. Bagdany, N. Soya, H. Xu, A. Roldan, A. S. Verkman, M. Kurth, A. Simon, T. Hegedus, J. M. Beekman, G. L. Lukacs, Mechanism-based corrector combination restores ΔF508-CFTR folding and function. *Nat. Chem. Biol.* **9**, 444–454 (2013).
11. W. M. Rabeh, F. Bossard, H. Xu, T. Okiyoneda, M. Bagdany, C. M. Mulvihill, K. Du, S. di Bernardo, Y. Liu, L. Konermann, A. Roldan, G. L. Lukacs, Correction of both NBD1 energetics and domain interface is required to restore ΔF508 CFTR folding and function. *Cell* **148**, 150–163 (2012).
12. H. Y. Ren, D. E. Grove, O. De La Rosa, S. A. Houck, P. Sopha, F. Van Goor, B. J. Hoffman, D. M. Cyr, VX-809 corrects folding defects in cystic fibrosis transmembrane conductance regulator protein through action on membrane-spanning domain 1. *Mol. Biol. Cell* **24**, 3016–3024 (2013).
13. P. W. Phuan, G. Veit, J. Tan, A. Roldan, W. E. Finkbeiner, G. L. Lukacs, A. S. Verkman, Synergy-based small-molecule screen using a human lung epithelial cell line yields ΔF508-CFTR correctors that augment VX-809 maximal efficacy. *Mol. Pharmacol.* **86**, 42–51 (2014).
14. R. Amin, F. Ratjen, Emerging drugs for cystic fibrosis. *Expert Opin. Emerg. Drugs* **19**, 143–155 (2014).
15. L. J. Galietta, Managing the underlying cause of cystic fibrosis: A future role for potentiators and correctors. *Paediatr. Drugs* **15**, 393–402 (2013).
16. S. M. Rowe, A. S. Verkman, Cystic fibrosis transmembrane regulator correctors and potentiators. *Cold Spring Harb. Perspect. Med.* **3**, a009761 (2013).
17. P. B. Davis, U. Yasothan, P. Kirkpatrick, Ivacaftor. *Nat. Rev. Drug Discov.* **11**, 349–350 (2012).
18. U.S. Food and Drug Administration approves KALYDECO™ (ivacaftor) for use in eight additional mutations that cause cystic fibrosis, press release http://investors.vrtx.com/releasedetail.cfm?ReleaseID=827435 (21 February 2014).
19. H. Yu, B. Burton, C. J. Huang, J. Worley, D. Cao, J. P. Johnson Jr., A. Urrutia, J. Joubran, S. Seepersaud, K. Sussky, B. J. Hoffman, F. Van Goor, Ivacaftor potentiation of multiple CFTR channels with gating mutations. *J. Cyst. Fibros.* **11**, 237–245 (2012).
20. Results from phase 3 study of ivacaftor monotherapy showed statistically significant improvements in lung function in people with non-G551D gating mutations, press release http://investors.vrtx.com/releasedetail.cfm?ReleaseID=781005 (29 July 2013).
21. F. J. Accurso, S. M. Rowe, J. P. Clancy, M. P. Boyle, J. M. Dunitz, P. R. Durie, S. D. Sagel, D. B. Hornick, M. W. Konstan, S. H. Donaldson, R. B. Moss, J. M. Pilewski, R. C. Rubenstein, A. Z. Uluer, M. L. Aitken, S. D. Freedman, L. M. Rose, N. Mayer-Hamblett, Q. Dong, J. Zha, A. J. Stone, E. R. Olson, C. L. Ordoñez, P. W. Campbell, M. A. Ashlock, B. W. Ramsey, Effect of VX-770 in persons with cystic fibrosis and the G551D-*CFTR* mutation. *N. Engl. J. Med.* **363**, 1991–2003 (2010).
22. B. W. Ramsey, J. Davies, N. G. McElvaney, E. Tullis, S. C. Bell, P. Dřevínek, M. Griese, E. F. McKone, C. E. Wainwright, M. W. Konstan, R. Moss, F. Ratjen, I. Sermet-Gaudelus, S. M. Rowe, Q. Dong, S. Rodriguez, K. Yen, C. Ordoñez, J. S. Elborn, A CFTR potentiator in patients with cystic fibrosis and the *G551D* mutation. *N. Engl. J. Med.* **365**, 1663–1672 (2011).
23. I. Bronsveld, F. Mekus, J. Bijman, M. Ballmann, H. R. de Jonge, U. Laabs, D. J. Halley, H. Ellemunter, G. Mastella, S. Thomas, H. J. Veeze, B. Tümmler, Chloride conductance and genetic background modulate the cystic fibrosis phenotype of ΔF508 homozygous twins and siblings. *J. Clin. Invest.* **108**, 1705–1715 (2001).
24. F. Van Goor, S. Hadida, P. D. Grootenhuis, B. Burton, D. Cao, T. Neuberger, A. Turnbull, A. Singh, J. Joubran, A. Hazlewood, J. Zhou, J. McCartney, V. Arumugam, C. Decker, J. Yang, C. Young, E. R. Olson, J. J. Wine, R. A. Frizzell, M. Ashlock, P. Negulescu, Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18825–18830 (2009).
25. H. J. Veeze, D. J. Halley, J. Bijman, J. C. de Jongste, H. R. de Jonge, M. Sinaasappel, Determinants of mild clinical symptoms in cystic fibrosis patients. Residual chloride secretion measured in rectal biopsies in relation to the genotype. *J. Clin. Invest.* **93**, 461–466 (1994).
26. N. Kälin, A. Claass, M. Sommer, E. Puchelle, B. Tümmler, ΔF508 CFTR protein expression in tissues from patients with cystic fibrosis. *J. Clin. Invest.* **103**, 1379–1389 (1999).
27. P. A. Flume, T. G. Liou, D. S. Borowitz, H. Li, K. Yen, C. L. Ordoñez, D. E. Geller; VX 08-770-104 Study Group, Ivacaftor in subjects with cystic fibrosis who are homozygous for the F508del-CFTR mutation. *Chest* **142**, 718–724 (2012).
28. F. Van Goor, S. Hadida, P. D. Grootenhuis, B. Burton, J. H. Stack, K. S. Straley, C. J. Decker, M. Miller, J. McCartney, E. R. Olson, J. J. Wine, R. A. Frizzell, M. Ashlock, P. A. Negulescu, Correction of the F508del-CFTR protein processing defect in vitro by the investigational drug VX-809. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18843–18848 (2011).
29. J. P. Clancy, S. M. Rowe, F. J. Accurso, M. L. Aitken, R. S. Amin, M. A. Ashlock, M. Ballmann, M. P. Boyle, I. Bronsveld, P. W. Campbell, K. De Boeck, S. H. Donaldson, H. L. Dorkin, J. M. Dunitz, P. R. Durie, M. Jain, A. Leonard, K. S. McCoy, R. B. Moss, J. M. Pilewski, D. B. Rosenbluth, R. C. Rubenstein, M. S. Schechter, M. Botfield, C. L. Ordoñez, G. T. Spencer-Green, L. Vernillet, S. Wisseh, K. Yen, M. W. Konstan, Results of a phase IIa study of VX-809, an investigational CFTR corrector compound, in subjects with cystic fibrosis homozygous for the *F508del-CFTR* mutation. *Thorax* **67**, 12–18 (2012).
30. Final data from phase 2 combination study of VX-809 and KALYDECO™ (ivacaftor) showed statistically significant improvements in lung function in people with cystic fibrosis who have two copies of the F508del mutation, press release http://investors.vrtx.com/releasedetail.cfm?ReleaseID=687394 (28 June 2012).
31. Two 24-week phase 3 studies of lumacaftor in combination with ivacaftor met primary endpoint with statistically significant improvements in lung function (FEV1) in people with cystic fibrosis who have two copies of the F508del mutation, press release http://investors.vrtx.com/releasedetail.cfm?releaseid=856185 (24 June 2014).
32. C. Ehrhardt, E. M. Collnot, C. Baldes, U. Becker, M. Laue, K. J. Kim, C. M. Lehr, Towards an in vitro model of cystic fibrosis small airway epithelium: Characterisation of the human bronchial epithelial cell line CFBE41o-. *Cell Tissue Res.* **323**, 405–415 (2006).
33. G. Veit, F. Bossard, J. Goepp, A. S. Verkman, L. J. Galietta, J. W. Hanrahan, G. L. Lukacs, Proinflammatory cytokine secretion is suppressed by TMEM16A or CFTR channel activity in human cystic fibrosis bronchial epithelia. *Mol. Biol. Cell* **23**, 4188–4202 (2012).
34. N. Pedemonte, G. L. Lukacs, K. Du, E. Caci, O. Zegarra-Moran, L. J. Galietta, A. S. Verkman, Small-molecule correctors of defective ΔF508-CFTR cellular processing identified by high-throughput screening. *J. Clin. Invest.* **115**, 2564–2571 (2005).
35. Access data FDA: 203188Orig1s000. http://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/203188Orig1s000OtherRedt.pdf.
36. D. S. Snyder, L. Tradtrantip, C. Yao, M. J. Kurth, A. S. Verkman, Potent, metabolically stable benzopyrimido-pyrrolo-oxazine-dione (BPO) CFTR inhibitors for polycystic kidney disease. *J. Med. Chem.* **54**, 5468–5477 (2011).
37. T. W. Loo, M. C. Bartlett, D. M. Clarke, Corrector VX-809 stabilizes the first transmembrane domain of CFTR. *Biochem. Pharmacol.* **86**, 612–619 (2013).
38. T. Neuberger, B. Burton, H. Clark, F. Van Goor, Use of primary cultures of human bronchial epithelial cells isolated from cystic fibrosis patients for the pre-clinical testing of CFTR modulators. *Methods Mol. Biol.* **741**, 39–54 (2011).
39. M. L. Fulcher, S. Gabriel, K. A. Burns, J. R. Yankaskas, S. H. Randell, Well-differentiated human airway epithelial cell cultures. *Methods Mol. Med.* **107**, 183–206 (2005).
40. K. A. Lansdell, Z. Cai, J. F. Kidd, D. N. Sheppard, Two mechanisms of genistein inhibition of cystic fibrosis transmembrane conductance regulator Cl^- channels expressed in murine cell line. *J. Physiol.* **524** (Pt. 2), 317–330 (2000).
41. F. Wang, S. Zeltwanger, I. C. Yang, A. C. Nairn, T. C. Hwang, Actions of genistein on cystic fibrosis transmembrane conductance regulator channel gating. Evidence for two binding sites with opposite effects. *J. Gen. Physiol.* **111**, 477–490 (1998).
42. P. M. Apaja, H. Xu, G. L. Lukacs, Quality control for unfolded proteins at the plasma membrane. *J. Cell Biol.* **191**, 553–570 (2010).
43. H. Barrière, P. Apaja, T. Okiyoneda, G. L. Lukacs, Endocytic sorting of CFTR variants monitored by single-cell fluorescence ratiometric image analysis (FRIA) in living cells. *Methods Mol. Biol.* **741**, 301–317 (2011).

44. T. Hegedus, A. Aleksandrov, L. Cui, M. Gentzsch, X. B. Chang, J. R. Riordan, F508del CFTR with two altered RXR motifs escapes from ER quality control but its channel activity is thermally sensitive. *Biochim. Biophys. Acta* **1758**, 565–572 (2006).
45. H. Hoelen, B. Kleizen, A. Schmidt, J. Richardson, P. Charitou, P. J. Thomas, I. Braakman, The primary folding defect and rescue of ΔF508 CFTR emerge during translation of the mutant domain. *PLOS One* **5**, e15458 (2010).
46. H. A. Lewis, X. Zhao, C. Wang, J. M. Sauder, I. Rooney, B. W. Noland, D. Lorimer, M. C. Kearins, K. Conners, B. Condon, P. C. Maloney, W. B. Guggino, J. F. Hunt, S. Emtage, Impact of the ΔF508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. *J. Biol. Chem.* **280**, 1346–1353 (2005).
47. I. Protasevich, Z. Yang, C. Wang, S. Atwell, X. Zhao, S. Emtage, D. Wetmore, J. F. Hunt, C. G. Brouillette, Thermal unfolding studies show the disease causing F508del mutation in CFTR thermodynamically destabilizes nucleotide-binding domain 1. *Protein Sci.* **19**, 1917–1931 (2010).
48. P. H. Thibodeau, J. M. Richardson III, W. Wang, L. Millen, J. Watson, J. L. Mendoza, K. Du, S. Fischman, H. Senderowitz, G. L. Lukacs, K. Kirk, P. J. Thomas, The cystic fibrosis-causing mutation ΔF508 affects multiple steps in cystic fibrosis transmembrane conductance regulator biogenesis. *J. Biol. Chem.* **285**, 35825–35835 (2010).
49. M. Haardt, M. Benharouga, D. Lechardeur, N. Kartner, G. L. Lukacs, C-terminal truncations destabilize the cystic fibrosis transmembrane conductance regulator without impairing its biogenesis. A novel class of mutation.. *J. Biol. Chem.* **274**, 21873–21877 (1999).
50. A. C. DeCarvalho, L. J. Gansheroff, J. L. Teem, Mutations in the nucleotide binding domain 1 signature motif region rescue processing and functional defects of cystic fibrosis transmembrane conductance regulator ΔF508. *J. Biol. Chem.* **277**, 35896–35905 (2002).
51. C. M. Farinha, J. King-Underwood, M. Sousa, A. R. Correia, B. J. Henriques, M. Roxo-Rosa, A. C. Da Paula, J. Williams, S. Hirst, C. M. Gomes, M. D. Amaral, Revertants, low temperature, and correctors reveal the mechanism of F508del-CFTR rescue by VX-809 and suggest multiple agents for full correction. *Chem. Biol.* **20**, 943–955 (2013).
52. S. G. Bompadre, J. H. Cho, X. Wang, X. Zou, Y. Sohma, M. Li, T. C. Hwang, CFTR gating II: Effects of nucleotide binding on the stability of open states. *J. Gen. Physiol.* **125**, 377–394 (2005).
53. L. Cui, L. Aleksandrov, Y. X. Hou, M. Gentzsch, J. H. Chen, J. R. Riordan, A. A. Aleksandrov, The role of cystic fibrosis transmembrane conductance regulator phenylalanine 508 side chain in ion channel gating. *J. Physiol.* **572**, 347–358 (2006).
54. Z. Xu, L. S. Pissarra, C. M. Farinha, J. Liu, Z. Cai, P. H. Thibodeau, M. D. Amaral, D. N. Sheppard, Revertant mutants modify, but do not rescue, the gating defect of the cystic fibrosis mutant G551D-CFTR. *J. Physiol.* **592**, 1931–1947 (2014).
55. F. Van Goor, H. Yu, B. Burton, B. J. Hoffman, Effect of ivacaftor on CFTR forms with missense mutations associated with defects in protein processing or function. *J. Cyst. Fibros.* **13**, 29–36 (2014).
56. D. M. Cholon, N. L. Quinney, M. L. Fulcher, C. R. Esther, J. Das, N. V. Dokholyan, S. H. Randell, R. C. Boucher, M. Gentzsch, Potentiator ivacaftor abrogates pharmacological correction of ΔF508 CFTR in cystic fibrosis. *Sci. Transl. Med.* **6**, 246ra96 (2014).
57. L. Guillot, J. Beucher, O. Tabary, P. Le Rouzic, A. Clement, H. Corvol, Lung disease modifier genes in cystic fibrosis. *Int. J. Biochem. Cell Biol.* **52**, 83–93 (2014).
58. W. Xu, C. Hui, S. S. Yu, C. Jing, H. C. Chan, MicroRNAs and cystic fibrosis—An epigenetic perspective. *Cell Biol. Int.* **35**, 463–466 (2011).
59. European medicines agency, assessment report: Kalydeco EMA/473279/2012. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002494/WC500130766.pdf.
60. M. Wilschanski, A. Dupuis, L. Ellis, K. Jarvi, J. Zielenski, E. Tullis, S. Martin, M. Corey, L. C. Tsui, P. Durie, Mutations in the cystic fibrosis transmembrane regulator gene and in vivo transepithelial potentials. *Am. J. Respir. Crit. Care Med.* **174**, 787–794 (2006).
61. F. A. Suprynowicz, G. Upadhyay, E. Krawczyk, S. C. Kramer, J. D. Hebert, X. Liu, H. Yuan, C. Cheluvaraju, P. W. Clapp, R. C. Boucher Jr., C. M. Kamonjoh, S. H. Randell, R. Schlegel, Conditionally reprogrammed cells represent a stem-like state of adult epithelial cells. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20035–20040 (2012).
62. J. F. Dekkers, C. L. Wiegerinck, H. R. de Jonge, I. Bronsveld, H. M. Janssens, K. M. de Winter-de Groot, A. M. Brandsma, N. W. de Jong, M. J. Bijvelds, B. J. Scholte, E. E. Nieuwenhuis, S. van den Brink, H. Clevers, C. K. van der Ent, S. Middendorp, J. M. Beekman, A functional CFTR assay using primary cystic fibrosis intestinal organoids. *Nat. Med.* **19**, 939–945 (2013).
63. H. Yang, A. A. Shelat, R. K. Guy, V. S. Gopinath, T. Ma, K. Du, G. L. Lukacs, A. Taddei, C. Folli, N. Pedemonte, L. J. Galietta, A. S. Verkman, Nanomolar affinity small molecule correctors of defective ΔF508-CFTR chloride channel gating. *J. Biol. Chem.* **278**, 35079–35085 (2003).
64. M. Yamaya, W. E. Finkbeiner, S. Y. Chun, J. H. Widdicombe, Differentiated structure and function of cultures from human tracheal epithelium. *Am. J. Physiol.* **262**, L713–L724 (1992).
65. W. Namkung, W. E. Finkbeiner, A. S. Verkman, CFTR-adenylyl cyclase I association responsible for UTP activation of CFTR in well-differentiated primary human bronchial cell cultures. *Mol. Biol. Cell* **21**, 2639–2648 (2010).
66. W. Namkung, J. R. Thiagarajah, P. W. Phuan, A. S. Verkman, Inhibition of Ca^{2+}-activated Cl^- channels by gallotannins as a possible molecular basis for health benefits of red wine and green tea. *FASEB J.* **24**, 4178–4186 (2010).
67. A. Fiser, A. Sali, ModLoop: Automated modeling of loops in protein structures. *Bioinformatics* **19**, 2500–2501 (2003).
68. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
69. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
70. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).

Acknowledgments: We thank D. Gruenert (UCSF) for the parental CFBE41o– cell line, P. Wolters (UCSF) for assistance in collecting lung transplant tissues, R. J. Bridges (Rosalind Franklin University of Medicine and Science) and CFFT Inc. for CFTR modulator panels, P. J. Thomas (University of Texas Southwestern Medical Center Dallas, TX, USA) for the pBI-CMV2-CFTR-P67L and -R347H plasmids, and V. Malhotra (Center for Genomic Regulation, Barcelona, Spain) for the HRP-C complementary DNA. **Funding:** This work was supported by NIH DK72517 as well as the Research Development Program from the Cystic Fibrosis Foundation (to W.E.F. and A.S.V.), KTIA-AIK-12-2012-0025 (to T.H.), NIH–The National Institute of Diabetes and Digestive and Kidney Diseases R01DK75302, and Cystic Fibrosis Canada and Canadian Institutes of Health Research to (G.L.L). G.V. was partly supported by Fonds de Recherche Santé Québec Fellowship, and R.G.A. partly by the Solvay Fellowship and R.I. Birks Award. T.H. is a Bolyai Fellow of the Hungarian Academy of Sciences. G.L.L. is a Canada Research Chair. **Author contributions:** The overall design of the study was by G.V. and G.L.L.; G.V., R.G.A., D.P., P.-W.P., M.B., and P.M.A. performed experiments and analyzed the results. D.S. and T.H. performed the molecular dynamics simulations and in silico docking. W.E.F. contributed the primary HBE cells. F.B. and Y.-S.W. cloned the CFTR constructs. The manuscript was primarily written by G.V., G.L.L., and A.S.V. with input from all authors. **Competing interests:** GLL has consulted for Genzyme in 2011–2012 and has also served as an unpaid advisor to CFFT Inc. All other authors declare that they have no competing interests. **Data and materials availability:** CFBE41o– cell lines were obtained under material transfer agreement (MTA) from D. Gruenert (UCSF). CFTR expression constructs are available under MTA from McGill University.

Submitted 24 February 2014
Accepted 30 June 2014
Published 23 July 2014
10.1126/scitranslmed.3008889

Citation: G. Veit, R. G. Avramescu, D. Perdomo, P.-W. Phuan, M. Bagdany, P. M. Apaja, F. Borot, D. Szollosi, Y.-S. Wu, W. E. Finkbeiner, T. Hegedus, A. S. Verkman, G. L. Lukacs, Some gating potentiators, including VX-770, diminish ΔF508-CFTR functional expression. *Sci. Transl. Med.* **6**, 246ra97 (2014).

Science Translational Medicine

Some gating potentiators, including VX-770, diminish ΔF508-CFTR functional expression

Guido Veit, Radu G. Avramescu, Doranda Perdomo, Puay-Wah Phuan, Miklos Bagdany, Pirjo M. Apaja, Florence Borot, Daniel Szollosi, Yu-Sheng Wu, Walter E. Finkbeiner, Tamas Hegedus, Alan S. Verkman and Gergely L. Lukacs

Sci Transl Med **6**, 246ra97246ra97.
DOI: 10.1126/scitranslmed.3008889

Potentiating Trouble

Cystic fibrosis is a genetic disease caused by mutations of the CFTR ion channel, resulting in pulmonary and other complications. Ivacaftor is the only targeted drug approved for cystic fibrosis, but it is not effective enough to treat the severest and most common form of this disease. Ivacaftor is a "potentiator," which means that it improves the activity of mutant CFTR, but cannot work if there is no CFTR on the cell surface. Other drugs, called "correctors," help bring mutant CFTR to the cell surface, but two manuscripts by Cholon and Veit and co-authors now show that combining the two types of drugs does not work effectively because potentiators make CFTR less stable, accelerating the removal of this channel from the cell membrane.

Cellular and Molecular Life Sciences
https://doi.org/10.1007/s00018-019-03211-4

Cellular and Molecular Life Sciences

ORIGINAL ARTICLE

Discovering the chloride pathway in the CFTR channel

Bianka Farkas[1,2] · Hedvig Tordai[1] · Rita Padányi[1,3] · Attila Tordai[4] · János Gera[5] · Gábor Paragi[6,7] · Tamás Hegedűs[1,3]

Received: 22 February 2019 / Revised: 22 June 2019 / Accepted: 26 June 2019

Abstract

Cystic fibrosis (CF), a lethal monogenic disease, is caused by pathogenic variants of the CFTR chloride channel. The majority of CF mutations affect protein folding and stability leading overall to diminished apical anion conductance of epithelial cells. The recently published cryo-EM structures of full-length human and zebrafish CFTR provide a good model to gain insight into structure–function relationships of CFTR variants. Although, some of the structures were determined in the phosphorylated and ATP-bound active state, none of the static structures showed an open pathway for chloride permeation. Therefore, we performed molecular dynamics simulations to generate a conformational ensemble of the protein and used channel detecting algorithms to identify conformations with an opened channel. Our simulations indicate a main intracellular entry at TM4/6, a secondary pore at TM10/12, and a bottleneck region involving numerous amino acids from TM1, TM6, and TM12 in accordance with experiments. Since chloride ions entered the pathway in our equilibrium simulations, but did not traverse the bottleneck region, we performed metadynamics simulations, which revealed two possible exits. One of the chloride ions exits includes hydrophobic lipid tails that may explain the lipid-dependency of CFTR function. In summary, our in silico study provides a detailed description of a potential chloride channel pathway based on a recent cryo-EM structure and may help to understand the gating of the CFTR chloride channel, thus contributing to novel strategies to rescue dysfunctional mutants.

Keywords Cystic fibrosis · ABCC7 · Chloride channel · Structure · Molecular dynamics

Introduction

Cystic fibrosis is a monogenic disease associated with high mortality and is caused by mutations affecting the CFTR (cystic fibrosis transmembrane conductance regulator also known as ABCC7) protein, a chloride channel in the plasma membrane of epithelial cells. In the presence of mutations a lack of functional CFTR expression in the apical membrane is observed resulting in imbalanced salt and water homeostasis [1]. The most frequent mutation in patients with cystic fibrosis (CF) is the deletion of F508 (F508del) impairing both protein folding and function [2]. Recent clinical trials are promising, and combinations of several drugs have been reported to act on the affected steps of CFTR biogenesis, trafficking and function [3]. For understanding the effects of mutations and drugs on CFTR, the knowledge of high-resolution structure and dynamics of the channel in the context of gating is essential. This information is also crucial for the development of more potent treatment modalities correcting the effect of F508del and other mutations.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00018-019-03211-4) contains supplementary material, which is available to authorized users.

✉ Tamás Hegedűs
hegedus@hegelab.org

[1] Department of Biophysics and Radiation Biology, Semmelweis University, Budapest, Hungary

[2] Faculty of Information Technology, Pázmány Péter Catholic University, Budapest, Hungary

[3] MTA-SE Molecular Biophysics Research Group, Hungarian Academy of Sciences, Budapest, Hungary

[4] Department of Pathophysiology, Semmelweis University, Budapest, Hungary

[5] Department of Medical Chemistry, University of Szeged, Szeged, Hungary

[6] MTA-SZTE Biomimetic System Research Group, Hungarian Academy of Sciences, Szeged, Hungary

[7] Institute of Physics, University of Pécs, Pecs, Hungary

Published online: 20 July 2019

CFTR (ABCC7) is the member of the ATP binding cassette (ABC) superfamily. It consists of two transmembrane domains (TMD1 and TMD2), each consisting of 6 TM helices, two nucleotide binding domains (NBD1 and NBD2), and a disordered segment, called regulatory (R) domain [4, 5]. The R domain phosphorylation is a prerequisite for channel gating. The N-terminus, also called Lasso/L0 region, plays a role in the functional expression of CFTR through its intramolecular and intermolecular protein interactions [6, 7]. The ATP binding sites are formed by Walker A and B motifs in one of the NBDs and the ABC signature motif in the other NBD. The site with the Walker motifs from NBD1 is non-hydrolytic. An additional intriguing property of NBD1 is the presence of a roughly 40 amino acid long regulatory insertion between the β1 and β2 strands [8]. This insertion destabilizes NBD1 and is highly flexible, thus it is not visible in structures [9]. The conformation changes in the NBDs are transmitted to the TMDs via the so-called intracellular loops that are not loops in a structural sense, since they are formed by the intracellular continuation of two neighboring TM helices (TM2/3, TM4/5, TM8/9, and TM10/11) connected by a small helix (Fig. S1). These four short helices are called coupling helixes (CH1-4), as they interact with the NBDs and couple the ATP dependent conformational changes to the transmembrane domains [10]. Importantly, two TM helices from the two TM domains cross over to the opposite NBD in a domain swapping-like conformation [10]. For many years, only the structure of NBD1 had been known at atomic resolution [11]. After publishing the first full length ABC protein structure, several CFTR homology models have been built and some of their properties have also been experimentally confirmed [10, 12–15]. Major aspects of the homology models have been reinforced by recent cryo-EM structures, which also exhibit features unpredictable by homology modeling, for example the kink in TM8 [16–19]. In addition, the cryo-EM structures by Chen et al. and Fay et al. provide interesting, albeit low resolution information on R domain localization and on extracellular loop 4 (EL4) conformation [16, 20]. EL4 is the longest among the extracellular loops, is glycosylated, and is not visible at high resolution in any of the structural models. The CFTR structure has been determined in the absence of ATP and also in its active state, when ATP molecules were bound to the phosphorylated protein.

Although cryo-EM structures were determined in various conformations, understanding of channel function is still lacking. To overcome the limitations of static structures, several molecular dynamics studies have been performed. Tordai et al. performed simulations with the apo zebrafish CFTR structure, which was determined in the absence of ATP and exhibited separated NBDs [21]. In this study, it was suggested that this conformation is not highly populated under physiological condition, since closure of the NBDs could be observed even in short simulations. Corradi et al. have also detected this closure and membrane defects in the bilayer around TM8 [22]. They found that the position of the kinked TM8 is stable in long simulations using either the apo or the ATP-bound conformation. In a recent paper, Hoffmann et al. combined their ATP-bound CFTR homology model with metadynamics simulation to describe conformational changes associated to gating, since even the cryo-EM structure determined in the active form did not exhibit an open pathway for chloride ions [6]. Chin et al. also performed long molecular dynamics (MD) simulations with an ATP-bound human CFTR homology model, based on the ATP-bound zebrafish structure and focused on the lipid interactions of the CFTR protein [23]. This and other studies showed that lipid interactions are essential for high CFTR ATPase activity, which hydrolytic activity had been demonstrated and thought to be very low compared to other ABC transporters [1, 16, 23, 24].

Since none of the CFTR structures determined in active state demonstrate an open chloride channel and none of the in silico studies characterized in detail the chloride passage through the protein, in the present study we performed MD simulations to generate open conformations with the intention of identifying chloride pathways, and characterizing the interaction of CFTR protein with chloride ions.

Methods

Structure preparation

We used the cryo-EM structure (PDBID: 5W81) of phosphorylated, ATP-bound zebrafish CFTR (zCFTR) [17]. Residue indices in the figures are renumbered according to the human CFTR indexing. TM helix numbering is presented in Fig. S1.

Molecular dynamics (MD)

The starting structure was oriented along the membrane normal with the help of the OPM database and server [25]. MD simulations were performed using the GROMACS 2018 program package in which the CHARMM36m force field and the TIP3P water model were selected [26, 27]. The CHARMM-GUI web interface was employed to build a lipid bilayer environment around the protein model, as well as to generate input files for energy minimization, equilibration steps (NVT, NPT), and production run [28–30]. Input MDP files can be downloaded from http://cftr.hegelab.org. During system preparation, the following options were applied: a homogenous lipid bilayer was built from POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) molecules and 150 mM KCl was set for physiological liquid environment;

grid information for PME (Particle-Mesh Ewald) electrostatics was generated automatically, and NPT ensemble was selected with constant number of particles (N), pressure (P) of 1 bar, and temperature of 310 K. Parameters for ATP were provided by Komuro et al. [31]. All the structures were energy minimized in the first step using the steepest descent integrator (converged when force < 1000 kJ/mol/nm). The fast smooth PME algorithm was used to calculate electrostatic interactions and LINCS to constrain bonds. The equilibration procedure was performed 22 times to generate inputs for independent simulations with different starting velocities and this was followed by 35 ns long production simulations. A simulation with open conformations and five other randomly selected ones were extended to 100 ns long simulations to increase the possibility of observing open conformations. RMSD (root mean square deviation from the initial structure) and R_g (radius of gyration) plots indicated that the simulations with the near-atomic resolution zCFTR structural model (3.37 Å) were stable (Fig. S2). A brief summary of simulations can be found in Table S1. During the analysis (e.g. for RMSD and contact map calculations), GROMACS tools, the MDAnalysis package [32] and in-house Python scripts were applied.

Metadynamics

A snapshot from the equilibrium trajectory with open conformations was selected as a starting structure for further studies. According to the MD simulation, two chloride ions (chloride #1 and #2) entered the interior region, and our starting structure selection was based on the position of chloride #2, which approached the bottleneck region closer than chloride #1. This starting geometry was equilibrated for 25 ns with a position restraint on chloride #2 and the final state was used as input for a well-tempered metadynamics simulation. A distance based collective variable (CV) was biased in the metadynamical running, namely the distance between chloride #2 and the center of mass (COM) of four Cα atoms (residues 96, 348, 932, and 1149; human indexing: 95, 347, 924 and 1141). Focusing on the interior region, an extra constraint was introduced for the CV variable to restrict its value between $z = 5$ Å and $z = 20$ Å. This approach prevented the escape of the chloride ion towards the intracellular or the more distant extracellular regions. Finally, an angle restraint provided further control of the ion movement to keep it in the region of interest. Namely, we considered three points including the COM of the POPC P atoms in the extracellular leaflet, the COM of the POPC P atoms in the intracellular leaflet, and the actual position of the chloride ion. Connecting these three points in this order, an angle was defined and it was kept smaller than 80° during the simulation (Fig. S3). The well-tempered metadynamics simulation was running for 600 ns using the GROMACS augmented with the PLUMED 2.3 package [33]. The free energy surface (FES) for chloride #2 was projected onto the x/y and x/z planes using PLUMED tools. Convergence of the metadynamics simulation is demonstrated in Fig. S4 and input files can be accessed at http://cftr.hegelab.org.

Channel detection

As the first step, the Hole2 program [34] was used to identify conformations with pathways sufficient for chloride passage in the transmembrane region. All frames from every simulation were screened. A conformation was considered as open in the TM region if the radius of the detected pore was larger than 1.8 Å (radius of the chloride ion) in the membrane region, defined by the COM of two groups of Cα atoms located approximately at the level of the intracellular lipid head groups (a.a. 190, 250, 992, and 1047; human indexing: 189, 223, 1018, 1023) and at the level of the extracellular head groups (a.a. 218, 223, 1018, and 1023 human indexing: 217, 222, 1010, and 1015). In the second step, the conformations open in the TM region were subjected to the Caver program [35], which allows a more sophisticated tunnel search at the expense of speed. The starting point of this search was defined by the COM of two Cα atoms (190 and 990; human indexing: 189 and 982). The residues of the transmembrane domains (a.a. 1–384 and 844–1181; human indexing: 1–383, 846–1173) were exclusively used in the calculation. The cutoff radius was set to 1.8 Å. The parameters of shell radius and shell depth were set to 6 Å and 3 Å, respectively. The clustering threshold was set to 4.5 Å. The "one tunnel" setting in the snapshot option was turned off. The tunnel residues were calculated using the default contact distance of Caver (3.0 Å). All other parameters were set to default values. Several tunnels were detected from the starting point ending at either the intracellular or extracellular sites. Conformations containing at least two tunnels connecting the intracellular and extracellular spaces were selected for further analysis. The pair of tunnels with the lowest scores (tunnel cost parameter) were set as the most probable open pathway. Hole2 and Caver input parameters can be downloaded from http://cftr.hegelab.org and their output was analyzed using in-house Python scripts. The interaction of amino acids with Caver spheres and chloride ions were calculated using MDAnalysis. Clustering the channels in the open conformations was performed based on atom positions of the channel-lining residues (criterion: Ward's method). The frames were aligned to the initial structure for analysis and the $z = 0$ corresponds to the bottom of the simulation box.

Network analysis

The NetworkView plugin of VMD was applied to build, analyze, and display the network based on the pairwise correlated motions of amino acids [36–38]. MDTraj [39] was used to convert a part of the trajectory (from 80 to 100 ns) in a suitable format for Carma [40], which calculated the correlation in Cα motions. The *subopt* code from the authors of NetworkView plugin was run to calculate optimal and suboptimal paths between specific residues. The configuration files and parameters can be found at http://cftr.hegelab.org.

Visualization

Structures are visualized using PyMOL (The PyMOL Molecular Graphics System, Version 1.8.4 Schrödinger, LLC) or VMD [36]. Figures were generated by Matplotlib [41].

Results

Identification of open channels suited for chloride conductance

The determination of CFTR structure in its active state with the phosphorylated R domain and the ATP-bound nucleotide binding domains (NBD) enabled us to examine the structural background of chloride conductance. However, the cryo-EM structure of the phosphorylated and ATP-bound zebrafish CFTR structure (PDBID: 5W81) discloses no pathway suitable for chloride passage [17]. To find CFTR conformations with open channels, we performed equilibrium molecular dynamics (MD) simulations ($n = 22$) with different initial velocities using this zCFTR structure (Table S1). We used the "open conformation" expression for structures with a radius larger than 1.8 Å along the entire pathway, which geometry can ensure chloride conduction. Firstly, we identified conformations with an open transmembrane region using the Hole2 program [34]. Subsequently, the resulting structures were filtered for continuous channels between the intra- and extracellular spaces by the Caver code [35]. Among the 22 equilibrium simulations, only a single trajectory (see detailed below) exhibited conformations with open channels.

Concerning the full trajectory, a small portion of the total frames (54/10,000) contained channels having simultaneously open geometry at both sides. Importantly, opening events were not grouped around a given time point, but distributed evenly over the 100 ns simulation time (see below). In order to compare the open conformations, we performed clustering on the channel-lining residues. An amino acid was selected as a pore facing residue if the distance between any of its atoms and the Caver channel spheres was smaller than 3 Å (Cavers's default value). The tunnel clustering was performed using the pairwise RMSD of the channel-lining residues as a distance measure. Tunnels in all the resulted clusters were highly similar in the TM region and diverged only in the intracellular parts (Fig. S5). While differences in the intracellular ends could be observed, these tunnels were all open between TM3, TM4 and TM6 around amino acids at position K190, R248 (K in zCFTR), G366 (R in zCFTR), K370 (human CFTR residue indexing is used throughout the paper) (Fig. 1). A second, symmetrical opening with lower Caver scores was also apparent in some of the simulations, between TM9, TM10 and TM12 around amino acids G971 (R in zCFTR), R975, Q1035 (R in zCFTR), K1041, R1048, R1158, R1162 and K1165 (Fig. S5). The physiological relevance of these two intracellular gates, which are surrounded by positively charged residues, is supported by experiments [42].

Interestingly, some of the conformations exhibited a large opening to the hydrophobic part of the membrane bilayer in the TM8 region (Fig. 1) indicating that lipid molecules can interact with the pore residues. In MD simulations, we observed POPC molecules that could even enter the channel. As a consequence, the channel became closed by the hydrophobic lipid tail that can prevent the chloride passage.

In order to validate our simulations, we related the identified channel-lining residues to locations, which have been experimentally identified as potential residues facing the pathway or playing a role in the chloride conductance [43, 44]. We counted the residues in the 54 open conformations, which were closer than 3 Å to any points of the detected pathways (Caver spheres). The normalized values are plotted in Fig. 2 and experimentally determined channel positions are labeled with a letter "e". There are numerous residues, which are equally indicated to face the tunnel by both experiments and our simulation. Some of the experimentally identified residues, such as those in TM9, do not participate in the channel formation in our in silico studies. These residues most likely act through allosteric mechanisms, since TM9 is located laterally distant from the putative tunnels. There are also amino acids in TM2, TM5 or TM8, which face the channel in our simulations, but we were unable to find experimental evidences if they are similarly located in vivo.

To assess the structural characteristics of the pathway and the possible presence of bottleneck regions, we plotted the average radius of the channel along the z-coordinate of the simulation box (the z axis is closely perpendicular to the membrane bilayer and the protein) (Fig. 3). The intracellular entry ($z = 50$–75 Å) displayed several critical low values, while in the next region ($z = 75$–95 Å), considered as a large vestibule based on experiments [43–47], the profile showed large radius values. The region between 95 and 110 Å provided bottleneck characteristics with low values close to

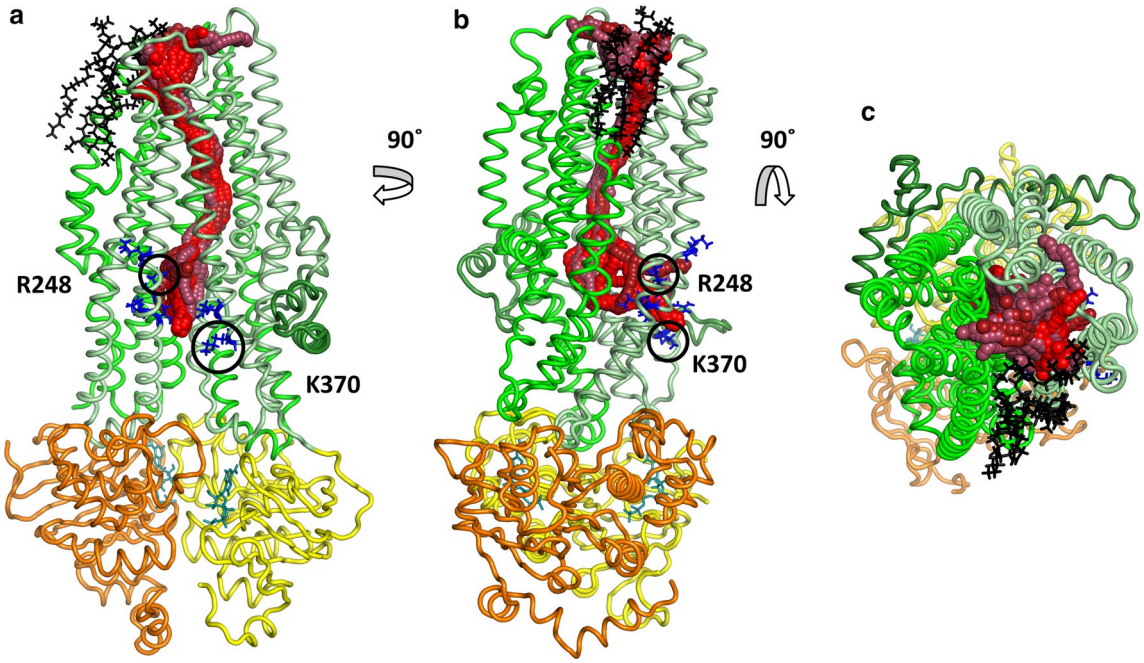


Fig. 1 Channels with extracellular and intracellular ends. Tunnels suitable for chloride conductance were identified in one case out of 22 equilibrium simulations with the ATP-bound CFTR structure (PDBID: 5W81) using Caver. The channels with top scores (spheres with red colors) were clustered based on the atom positions of the channel-lining residues. Intracellular residues with positive charges at the TM4/6 entry site are labeled with blue sticks. Lasso/L0 region, TMD1, TMD2, NBD1, and NBD2 domains are marked with dark green, pale green, green, yellow, and orange, respectively. Lipid molecules lining the pore in the region of TM8 are shown by black sticks. **a**, **b** Side views from the bilayer hydrophobic core. **c** A top view from the extracellular space

the radius of chloride ion. Residues at positions Q98, P99, and L102 in TM1, I344 (M in zCFTR) and V345 in TM6, and N1138 (Q in zCFTR) in TM12 had previously been experimentally identified as channel forming residues. In our simulations, these residues were located in the inner vestibule [43, 45–49] and were found in pore-facing positions. In addition, I336, F337, T338, T339, I340 (L in zCFTR) and S341, which were positioned in the bottleneck region we identified in silico, have been implicated as the selectivity filter by several experimental studies [15, 43, 48].

Interaction of chloride ions with the channel during simulation

Ions show extensive movements in MD simulations compared to large molecules, such as proteins. Therefore, the contacts of chloride ions with amino acids were expected to provide valuable information on chloride interaction sites. It is important to note that the chloride concentration was 150 mM on both sides of the membrane because of the periodic boundary condition in MD simulations.

First, we calculated and subsequently normalized the contact map of chloride ions with the protein in all our 100 ns long simulations ($n=6$). The contact of each amino acid with chloride ions ($d<4$ Å) was mapped to the structure and colored according to contact frequency (Fig. 4). Intensive interactions could be observed at the entry site defined by TM10 and TM12 close to the N-terminal L0/Lasso region. The frequency of interactions was more pronounced here than at the other entry site defined by TM4 and TM6 near to K370. These interaction sites corresponded to the intracellular pores identified by Caver in our simulations (Figs. 1, S5).

During the equilibration process, we observed that chloride ions and water molecules filled up the cavity. In the trajectory with open frames, we monitored the position of two chloride ions (labeled as #1 and #2), which entered the inner space of the protein along the membrane normal (Fig. 5). As expected, chloride ions repulsed, and did not remain close to each other. However, these ions did not pass the above described bottleneck region, starting around $z=95$ Å. Figure 5 also shows that the infrequent open conformations (54 out of 10,000 frames) were distributed over the 100 ns trajectory and were not grouped around a single time point. This observation suggests that we could notice independent opening events in spite of their perceptibility in a single simulation. Because of the low probability of open conformations, most likely none of the chloride ions was in an appropriate position at the moment of an opening, thus we were not able to observe an ion passage. We also calculated the contact frequency of these chloride ions with the pore

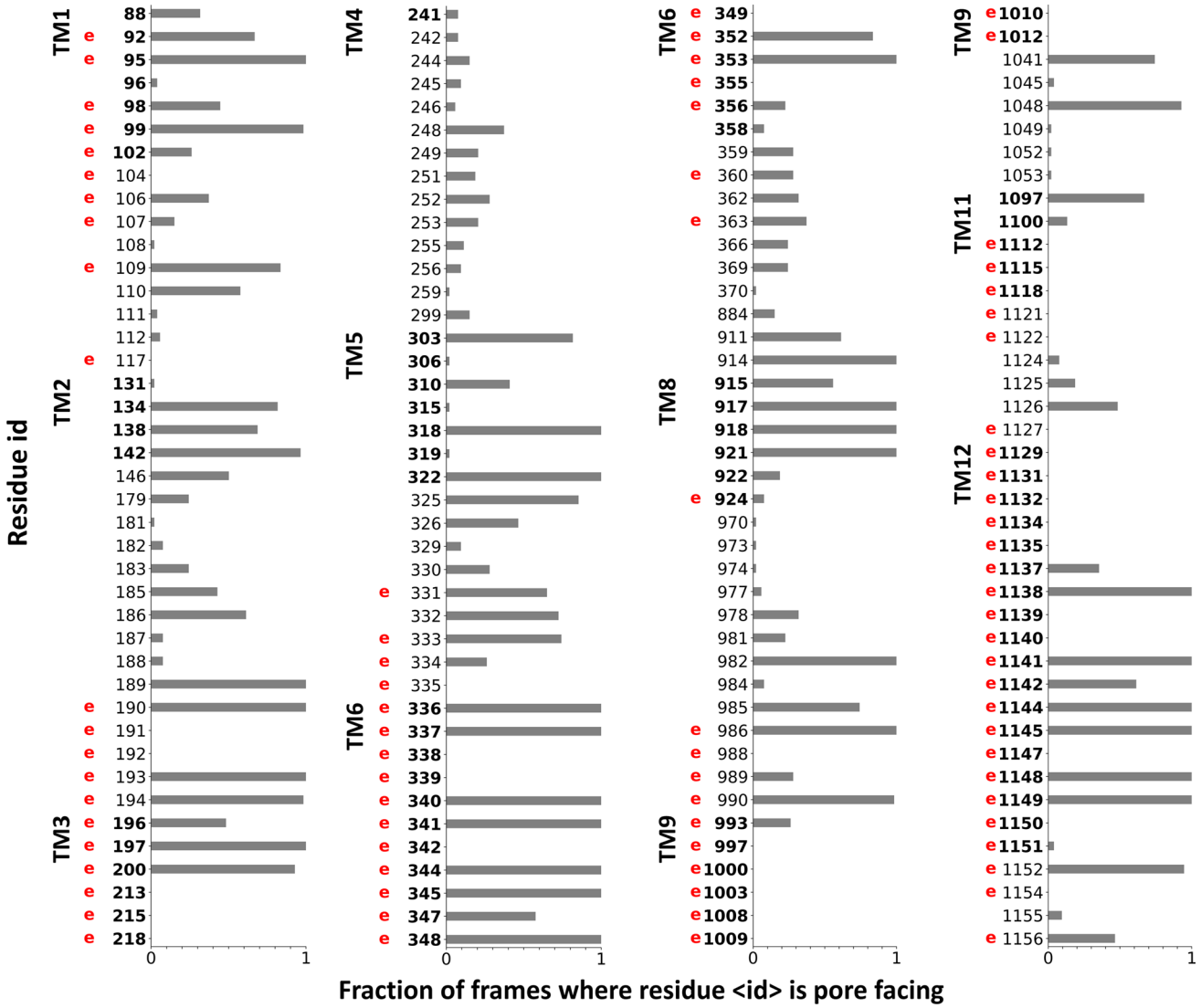


Fig. 2 Comparing channel-lining residues identified in simulations to experimental data. Residues interacting with the Caver spheres were counted from all open frames ($n = 54$) and normalized. Only residues lining the channel based on in silico or in vitro data, are indicated in the plot. A letter "e" in red was placed at the residue number (human numbering) to indicate that the residue had been shown by laboratory experiments to influence the CFTR chloride conductance. Bold numbers indicate amino acids located in the TM region

forming residues (Fig. 5). We observed that chloride ions spent more time close to positively charged amino acids, namely K95, R134, K190, R248 (K in zCFTR), R303, R352 and R1097, facing the large inner vestibule and lining the intracellular opening. Some of these residues had been indicated to be functionally important by experiments, as they affected the chloride conductance in studies using mutant CFTR forms [42, 43, 47, 50–52]. The chloride interaction was less frequent with non-positively charged residues. The two chlorides entering the TM region could reach the bottleneck region but stopped around residues Q98, M1137 and N1138 (L in zCFTR) without passing through the TM region. We also analyzed the interaction and movement of positively charged potassium ions as a negative control. No entry of any potassium into the protein could be detected (Fig. S6).

When we compared the contact maps of open and closed conformations (Fig. S7), residues I344 (M in zCFTR) and N1138 (L in zCFTR) were detected at the extracellular opening of the channel (between TM 1, 6 and 12), participating in closing the chloride pathway. Other residues, namely F337, S341, and L102, surrounding this extracellular pore opening have been suggested to affect channel gating by mutagenesis experiments [15, 43, 45, 48]. Some of these amino acids were located around the observed small opening-up of the ATP-bound human CFTR structure (PDBID: 6MSM), which conformation was suggested to be closer to an open state [20].

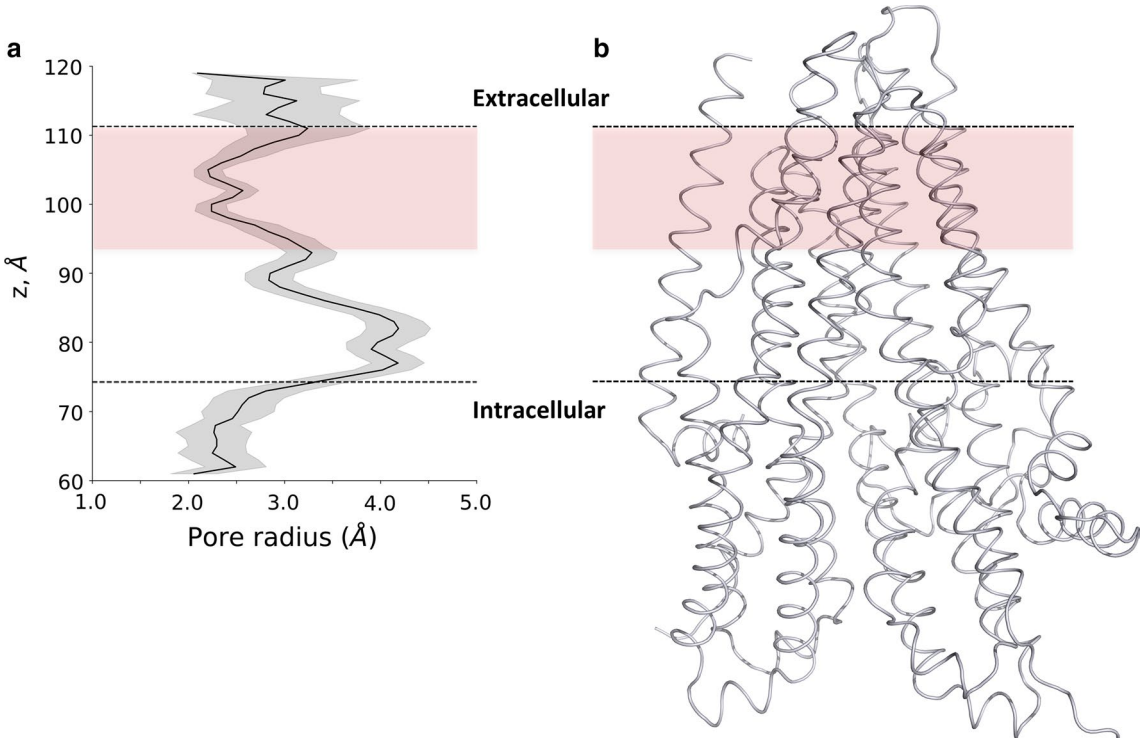


Fig. 3 The chloride channel profile. **a** The Caver sphere radius from all open channels was averaged over z coordinates. The large inner vestibule is indicated by large values (up to 4 Å), while bottleneck regions in the extracellular leaflet and in the intracellular region are characterized by radius values between 2 and 2.5 Å. The bottleneck in the extracellular membrane leaflet is depicted by the light coral box area. **b** The 3D structures of transmembrane domains are shown in the context of the channel profile

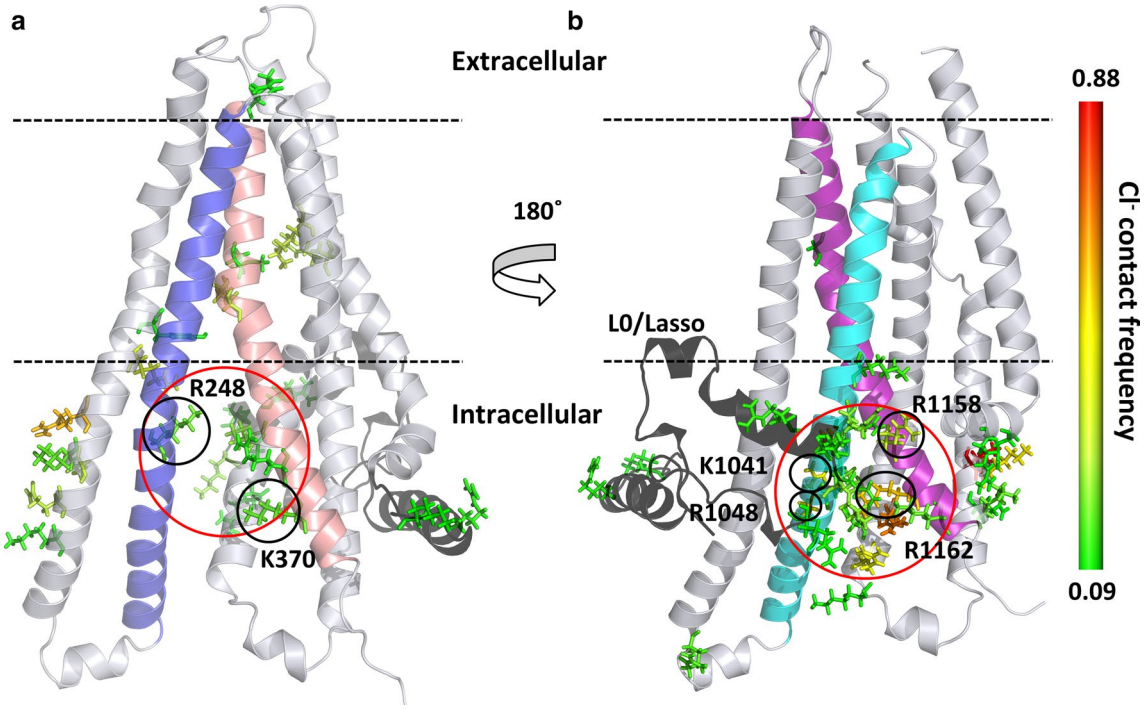


Fig. 4 Interaction sites of chloride ions with the CFTR protein. Contact sites are depicted by stick representation of amino acids and color-coded according to the contact frequency from green (low) to red (high). **a** TMD1 and the chloride entry pore at K370, between TM4 and TM6 are shown. **b** TMD2 and the pore close to the L0/Lasso motif (black), between TM10 and TM12 are depicted. TM4, TM6, TM10 and TM12 are colored by blue, salmon, cyan, and purple, respectively. The entry pores are circled (red)

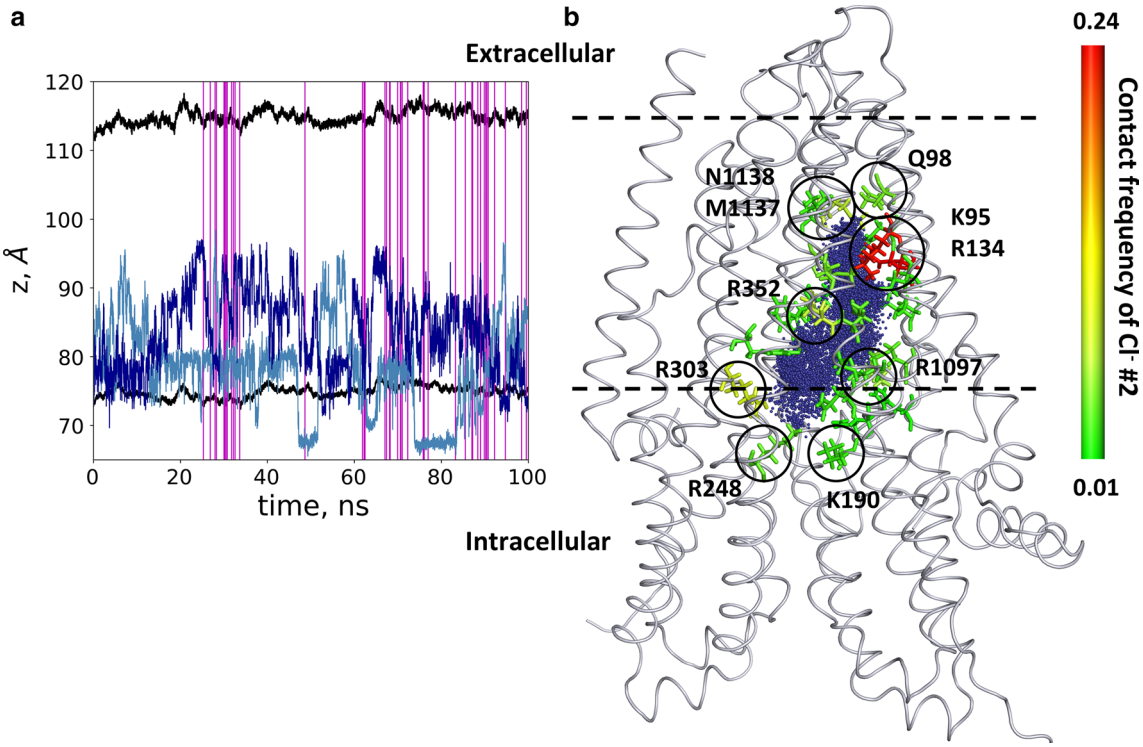


Fig. 5 Chloride ions entered the pathway but did not pass. **a** The positions of two chloride ions (#1 light blue and #2 blue), entering the protein, are shown by blue lines. Black lines indicate the boundaries of the membrane bilayer determined by the center of mass of P atoms in POPC molecules. Vertical magenta lines mark frames with an open pathway. **b** The locations of Cl^- #2 from every frame are shown with blue dots in the context of the initial structure. The contact frequency of this chloride ion with the protein was projected onto the structure and color coded from green (low) to red (high). Residues with the highest interaction frequencies are K95 and R134 (red)

The comparison of the contact maps of open and closed conformations (Fig. S7) revealed regions which have different contacts in open and closed states. These regions are located between the intracellular and extracellular ends of the TM domains including amino acids mainly from TM6 and TM12 and could represent allosteric communication sites for transmitting conformational changes from the nucleotide binding domains to the bottleneck region. In order to characterize the allosteric communication between NBDs and the extracellular ends of the TM helices, we performed a network analysis based on correlations in pairwise residue motion using the VMD NetworkView pipeline [37, 38]. A network was built from amino acids as nodes and their connections were weighted by the level of correlation in their motion. We determined the optimal and suboptimal paths between various "source" and "sink" residues (Fig. 6). The central amino acids (V171, V272, T963, and L1065) of the four coupling helices (CH), which are the main interaction sites between the nucleotide binding and transmembrane domains, were set as "source" nodes. Two amino acids (R334 in TM6 and Y914 in TM8) from the bottleneck residues were selected as "sinks". We determined the optimal and suboptimal paths in all eight combinations of sources and sinks, since the number of paths correlates with the strength of dynamic coupling. The highest number of suboptimal paths, thus the strongest coupling was observed between the CH4 and sink residues (Table S3). For visual comparison, we plotted L1065/R334 and V272/R334 pathways in Fig. 6. In the first case, several helices (TM1, TM2, TM3, TM6, TM11, and TM12) are included in the paths between the two nodes, while in the latter case only one helix, TM4 and a small region of TM5 and TM6 are involved in the allosteric communication between the intra- and extracellular parts of the CFTR.

Identification of the barrier of chloride passage using metadynamics simulations

Since the direct passage of chloride ions was not observed during our conventional MD simulations, we performed metadynamics computations to reveal the pathway through the bottleneck region and describe the potential surface of the transition. Metadynamics facilitates the escape from local minima by accumulating history-dependent Gaussian potential on specific reaction coordinates (collective variables, CV) [53]. The frame, in which chloride #2 was most proximally located to the bottleneck region, was selected from the trajectory. This frame was applied in a long,

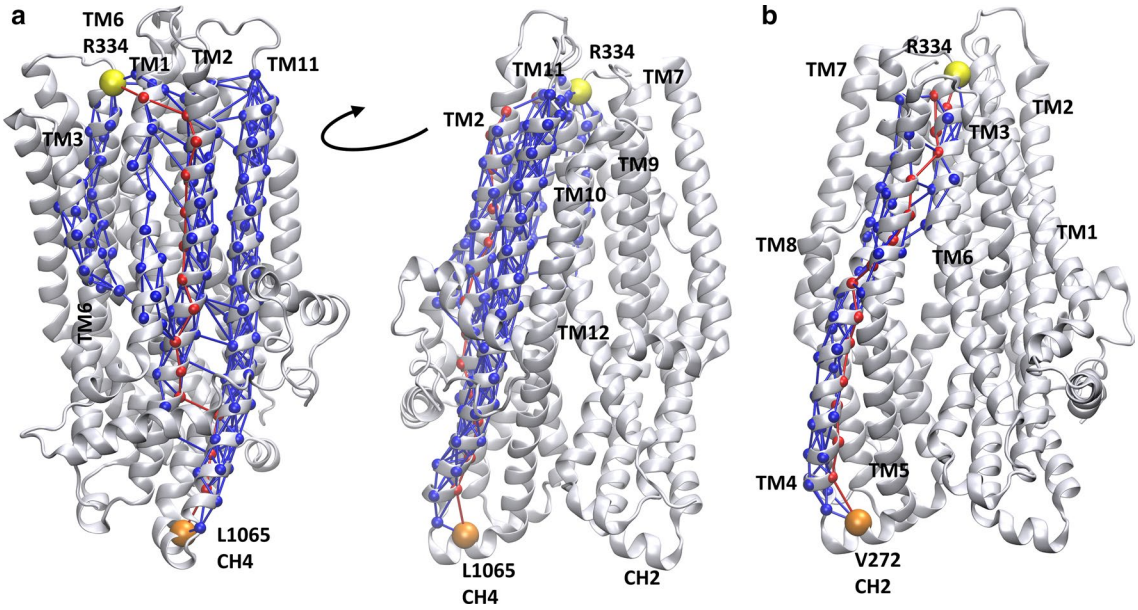


Fig. 6 Coupling helix 4 dynamics is strongly coupled to the bottleneck region. Optimal and suboptimal paths in the network of amino acids (spheres), connected by edges (lines) corresponding to correlation in motions, were determined. Number of suboptimal paths between coupling helix 4 (CH4, L1065) and R334 (**a**) is significantly higher than paths from any other coupling helix, such as from coupling helix 2 (CH2, V272) to R334 (**b**). Red: optimal paths; blue: suboptimal paths; orange: source residue; yellow: sink residue

well-tempered metadynamics simulation, where the bias was based on the distance between chloride #2 and the center of mass of four Cα atoms (residues 96, 348, 932, and 1149; human indexing: 95, 347, 924, and 1141). Constraints were applied to prevent the escape of the chloride from the protein towards either the intracellular or the far extracellular directions as described in the Methods section and demonstrated in Fig. S3.

The well-tempered metadynamics simulation was running for 600 ns allowing convergence (Fig. S4). 2D free energy surfaces were calculated along *x*/*y* and *x*/*z* (Fig. 7a, b) using the *x*, *y*, and *z* components of the distance CV. The *x*/*y* projection, corresponding to a top view from the extracellular space, indicated two deep minima, which were revealed as two extracellular exit sites by the *x*/*z* projection (a side view). This *x*/*z* projection showed that the path split after the bottleneck region and no significant barrier was present along these routes. Amino acids located towards the extracellular side of the bottleneck region and interacting with the chloride ion, include I106, A107, Y109, D110 from TM1, I331, I332 (N in zCFTR), L333 R334 from TM6 and Y914 from TM8 (Fig. 7c, d).

Discussion

Our computational studies revealed several novel aspects of CFTR chloride channel function, which may prove important for future studies, since to date no open chloride pathway has been identified at the atomic level in spite of recently published cryo-EM structures. Although Das et al. have generated structural models combining experimental data with extensive molecular modeling, their structures were highly different from any known ABC protein structures [54]. Hoffman et al. performed metadynamics simulations biasing the extracellular region of the central helices, applying their homology model that exhibited differences from recent experimental CFTR conformations [6]. In spite of the differences, their results rationally suggest several implications regarding the mechanism of CFTR gating (see below).

Although the cryo-EM structure (PDBID: 5W81) was determined under activating conditions, it did not exhibit an open pathway and it was regarded as a transiently closed conformation. We performed equilibrium simulations to observe the channel opening process with the intention of providing conformations sufficient for chloride conductance. Both, the number of trajectories with open conformations and the probability of channel opening along a trajectory were found to be low, namely 1 out of 22 trajectories and 54 out of 10,000 frames, respectively. This low number of events can be explained by recent electrophysiological data, since Zhang et al. demonstrated that zCFTR exhibited more than tenfold lower open probability when compared to the human CFTR [55]. Importantly, the sampling of the pore forming residues in our simulations corresponded properly to experimental data (see details in "Results"). Most of the observed discrepancies between our simulations and

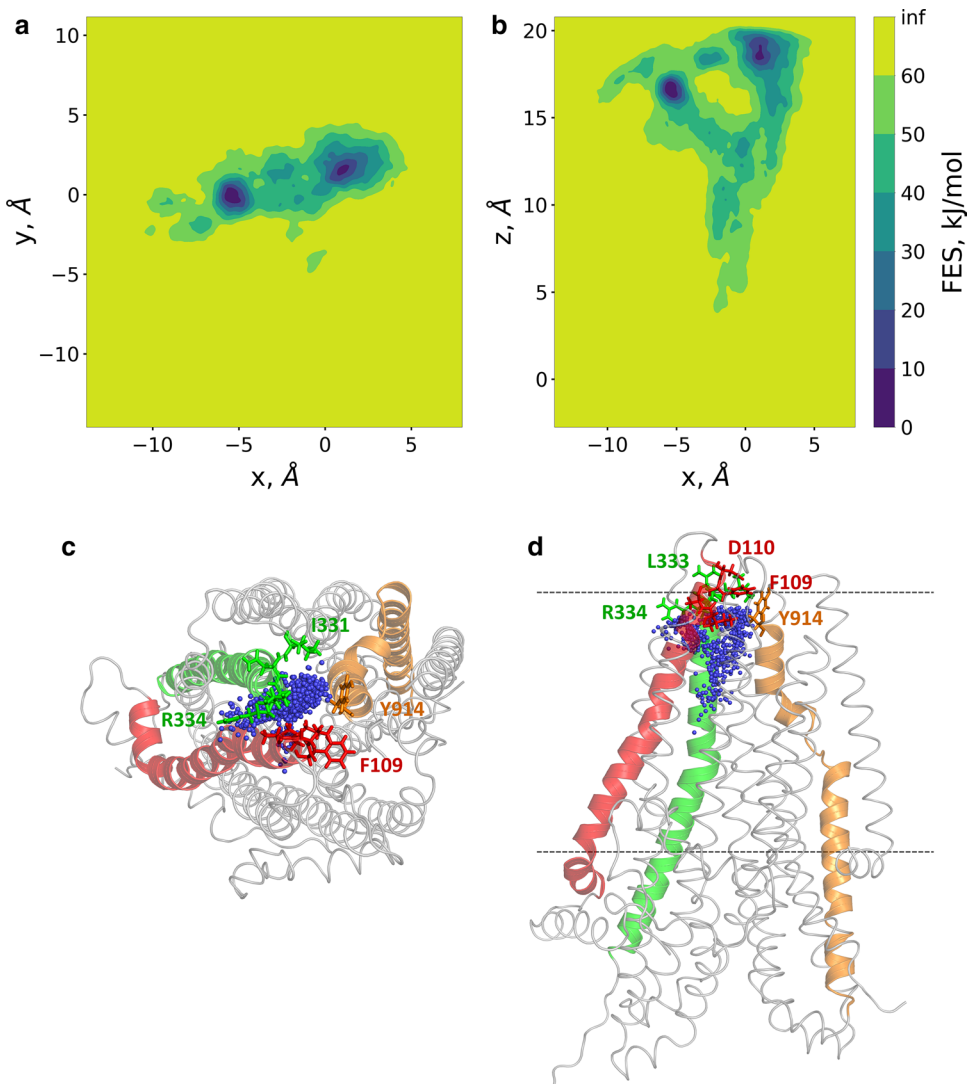


Fig. 7 Two routes towards the extracellular regions were detected by metadyanamics simulations. Well-tempered metadynamics simulations were performed to identify potential exit routes of chloride ions and the amino acids narrowing this region of the pathway. The distance between the center of mass of four Cα atoms (a.a. 95, 347, 924, and 1141) and Cl^- #2 was used as a reaction coordinate (collective variable, CV). 2D free energy surfaces were calculated along the *x*, *y*, and *z* components of this distance CV and plotted (panels **a**, **b**). Top (panel **c**) and side (panel **d**) views of chloride ions from every 100th frames are shown by blue spheres in relation to the initial structure. TM1, TM6, and TM8 are colored by red, green, and orange, respectively. Sticks represent the following amino acids in the bottleneck region: I106, A107, Y109 (F in zCFTR), D110 (TM1), I331, I332 (N in zCFTR), L333, R334 (TM6), Y914 (TM8)

experimental results may originate from allosteric effects. A trivial example of allosteric effects is TM9, which is located laterally distant from the central pore. Therefore, the residues in this helix cannot participate in the chloride pathway, although some of the residues have been indicated by experiments to influence conductance. Similarly, allosteric sites in ABCG2 have been identified by interpreting functional experiments combined with structural data [56]. Such examples strongly suggest that careful analysis and discussion of contradictory biochemical and structural data are important and can become a rich source for identifying allosteric interactions at a high resolution, promoting drug target selection. The other type of discrepancies, where in silico modeling revealed new components of the ion pathway (e.g. TM2, TM5 or TM8), could prompt further mutation experiments. The intriguing concordant data of experiments with hCFTR and in silico simulations with zCFTR suggest that the main conformational changes of these two related proteins associated with channel function may be similar in spite of the evident differences in channel opening probability and conductance [55]. This similarity is also supported by the 54% amino acid identity level with an additional 29% similarity of the zebrafish and human sequences. In addition, the structure of zCFTR determined under activating conditions (PDBID: 5W81 [17]) is highly similar to the more recently published structure of hCFTR (PDBID: 6MSM [20]) solved under similar conditions (RMSD = 1.8 Å). The difference in sequences may result in a stronger interaction of residues in the closed conformation of zCFTR leading to decreased probability of open conformation when compared to hCFTR.

Despite the similarities between experimental and in silico data, several questions remain open concerning channel gating. While several intracellular portals have been indicated by both experimental and computational studies [6, 42], in the light of the structure only one entry site around K370 was favored [17] and the other one at TM10/TM12, in the vicinity of the L0/Lasso motif was neglected. This shift in attention most likely occurred since a cysteine at

the position of one of the many positively charged amino acids around this TM10/TM12 pore did not react with methanethiosulfonate (MTS) reagents [42]. Importantly, K1041 and R1048 at this lateral pore are indicated by the same study to play an important role in the electrostatic attraction of chloride ions [42]. We determined low radius values for the channel profile at the intracellular pore region suggesting that the entry point may be selective not only for charge (Fig. S6), but also for size. We were unable to detect chloride entry into the CFTR channel from the extracellular side that could have been attributed to the extracellularly located bottleneck region. The low interaction frequency of chloride ions with amino acids at the extracellular portal and the lack of a well-defined positively charged patch of residues in this extracellular region suggest an important role for EL4. This loop, which was not present in the simulations as it was non-visible in the cryo-EM structures, involves K892 and R899 residues. These positively charged amino acids may attract chloride ions in the area of the extracellular mouth of the protein in case of flow from the extracellular to the intracellular direction.

The intracellular pores are indicated as the first selective part of the channel, since they are narrow and likely provide selectivity for negatively charged ions because of the surrounding positively charged amino acids (Fig. S6). The second level of selectivity and potentially the limiting factor of chloride conductance were present at the bottleneck region in the extracellular membrane leaflet (Fig. 3). The diameter of the pore at this region in the open conformations during our simulations only slightly exceeded the size of the chloride and the diameter in the closed conformations (Fig. S7). This small diameter difference may not be sufficient to allow the passage of the chloride ion surrounded by a hydrate shell. Under physiological conditions, this shell can be weakened by the negative membrane potential, which cannot be modeled in our MD simulations [57]. Although the protein structure was not constrained in an open conformation during our metadynamics simulation, the chloride ion visited both sides of the bottleneck region. Biasing the location of the chloride ion led to a gentle pushing of amino acid side chains and helices by the biased ion. These events were difficult to track because of the subtle conformational differences that were sufficient to provide an open tunnel (Fig. S7). The open state of one of the routes seems to be controlled by the distance between TM1 and TM6, which helices have been shown crucial for gating [15, 43, 45–49].

An intriguing question in the gating mechanism of any channel is, how the conformational changes in the regulatory regions drive the opening of the TM helices. Despite the low number of open conformations, our contact map analysis revealed two hot spots, which were located asymmetrically in the two TM domains, between the NBDs and the bottleneck region along a putative allosteric pathway (Fig. S7). This asymmetry is not unexpected, as the CFTR structure itself is not symmetric and the TM helices are arranged differently compared to helices of other ABC transporters [17, 54]. However, the highly uneven role of coupling helices and pairs of TM helices in dynamic coupling between the NBDs and the extracellular pore region is astonishing (Fig. 6). Coupling helix 4, which is in contact with F508, exhibits an order of magnitude larger number of suboptimal paths towards the bottleneck region compared to all other three coupling helices. This high level of coupling is especially striking, as coupling helix 4 and F508 are in the vicinity of the degenerate, non-hydrolytic ATP site with a tightly bound ATP molecule [58–60]. We propose that the importance of this degenerate site with the stably bound ATP lies in the stabilization of the above mentioned crucial allosteric pathway. The ATP molecule bound to this degenerate site may be structurally indispensable for gating, thus should not be hydrolyzed at a high rate, in order to maintain a stable conformation of the interface between NBD1 and coupling helix 4.

Interestingly, lipid molecules participated in forming the wall of the chloride channel pathway close to the bottleneck region which feature was clearly caused by the helix break in TM8. Although this seems peculiar, the influence of various lipid molecules on CFTR function has been experimentally demonstrated and it is under active investigations. The observed ATPase activity of CFTR has been very low compared to ABC transporters [61, 62]. This phenomenon was anticipated since the regulation of a channel function is not expected to consume comparable amounts of ATP to those required by active transport processes. However, a recent study described that phosphatidylserine promoted the ATPase activity of CFTR to a level comparable to other ABC transporters [24]. Chin et al. detected both, increased CFTR activity and potentiation in the presence of various lipids [23]. Some of these studies also employed in silico methods to reveal binding sites and to characterize the mode of lipid action. Microsecond MD simulations revealed lipid binding sites in functionally important regions, which may also play a role in conformational transitions. These authors [23] observed preferential binding of phosphatidylserine at helices (e.g. TM4, TM6, and TM10), which participate in forming the intracellular pores. According to our results, at least one of the pathways at the bottleneck region can likely be the target of lipid molecule interactions (Figs. 1, 7). POPC interactions in this region, around the TM8 kink were also detected by Chin et al. [23] and lipid perturbation in the same area was observed in simulations by Corradi et al. [22]. These suggest a hot spot around the TM8 kink, whose conformational transition towards an open state may be influenced or mediated by lipids.

In summary, we described here the chloride pathway of CFTR in atomic resolution for the first time. Our results

suggest that several intra- and extracellular entry sites may exist, no large conformational changes of the closed structures are required for opening and lipids may influence the channel path directly not only in an allosteric manner. Our study and methodology may help to understand the gating mechanism of wild type and mutant CFTR proteins.

Acknowledgements Open access funding provided by Semmelweis University (SE). We appreciate the support provided by Jiří Damborský and his group members (https://www.caver.cz). This work was supported by NKFIH K111678, NKFIH K127961, an award from the Cystic Fibrosis Foundation (CFF HEGEDU18I0), and the Semmelweis Science and Innovation Fund (Grant no. Sci_Innov18). We acknowledge the computational resources made available on the GPU cluster of the Governmental Information-Technology Development Agency and Wigner GPU Laboratory (http://gpu.wigner.mta.hu), and the support of their stuff is gratefully acknowledged. Donation of a Quadro P6000 GPU by NVIDIA Corporation is greatly appreciated.

References

1. Riordan JR, Rommens JM, Kerem B et al (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245:1066–1073
2. Veit G, Avramescu RG, Chiang AN et al (2016) From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. Mol Biol Cell 27:424–433. https://doi.org/10.1091/mbc.E14-04-0935
3. Davies JC, Moskowitz SM, Brown C et al (2018) VX-659–tezacaftor–ivacaftor in patients with cystic fibrosis and one or two Phe508del alleles. N Engl J Med 379:1599–1611. https://doi.org/10.1056/NEJMoa1807119
4. Meng X, Clews J, Martin ER et al (2018) The structural basis of cystic fibrosis. Biochem Soc Trans 46:1093–1098. https://doi.org/10.1042/BST20180296
5. Csanády L, Vergani P, Gadsby DC (2019) Structure, gating, and regulation of the CFTR anion channel. Physiol Rev 99:707–738. https://doi.org/10.1152/physrev.00007.2018
6. Hoffmann B, Elbahnsi A, Lehn P et al (2018) Combining theoretical and experimental data to decipher CFTR 3D structures and functions. Cell Mol Life Sci 75:3829–3855. https://doi.org/10.1007/s00018-018-2835-7
7. Cormet-Boyaka E, Jablonsky M, Naren AP et al (2004) Rescuing cystic fibrosis transmembrane conductance regulator (CFTR)-processing mutants by transcomplementation. Proc Natl Acad Sci USA 101:8221–8226. https://doi.org/10.1073/pnas.0400459101
8. Csanády L, Chan KW, Nairn AC, Gadsby DC (2005) Functional roles of nonconserved structural segments in CFTR's NH2-terminal nucleotide binding domain. J Gen Physiol 125:43–55. https://doi.org/10.1085/jgp.200409174
9. Aleksandrov AA, Kota P, Aleksandrov LA et al (2010) Regulatory insertion removal restores maturation, stability and function of DeltaF508 CFTR. J Mol Biol 401:194–210. https://doi.org/10.1016/j.jmb.2010.06.019
10. Serohijos AWR, Hegedus T, Aleksandrov AA et al (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. https://doi.org/10.1073/pnas.0800254105
11. Lewis HA, Zhao X, Wang C et al (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. https://doi.org/10.1074/jbc.M410968200
12. Mornon J-P, Lehn P, Callebaut I (2009) Molecular models of the open and closed states of the whole human CFTR protein. Cell Mol Life Sci 66:3469–3486. https://doi.org/10.1007/s00018-009-0133-0
13. Corradi V, Vergani P, Tieleman DP (2015) Cystic fibrosis transmembrane conductance regulator (CFTR): CLOSED AND OPEN STATE CHANNEL MODELS. J Biol Chem 290:22891–22906. https://doi.org/10.1074/jbc.M115.665125
14. Mornon J-P, Hoffmann B, Jonic S et al (2015) 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 72:1377–1403. https://doi.org/10.1007/s00018-014-1749-2
15. Alexander C, Ivetac A, Liu X et al (2009) Cystic fibrosis transmembrane conductance regulator: using differential reactivity toward channel-permeant and channel-impermeant thiol-reactive probes to test a molecular model for the pore. Biochemistry 48:10078–10088. https://doi.org/10.1021/bi901314c
16. Fay JF, Aleksandrov LA, Jensen TJ et al (2018) Cryo-EM visualization of an active high open probability CFTR ion channel. Biochemistry. https://doi.org/10.1101/274316
17. Zhang Z, Liu F, Chen J (2017) Conformational changes of CFTR upon phosphorylation and ATP binding. Cell 170:483–491.e8. https://doi.org/10.1016/j.cell.2017.06.041
18. Liu F, Zhang Z, Csanády L et al (2017) Molecular structure of the human CFTR ion channel. Cell 169:85–95.e8. https://doi.org/10.1016/j.cell.2017.02.024
19. Zhang Z, Chen J (2016) Atomic structure of the cystic fibrosis transmembrane conductance regulator. Cell 167:1586–1597.e9. https://doi.org/10.1016/j.cell.2016.11.014
20. Zhang Z, Liu F, Chen J (2018) Molecular structure of the ATP-bound, phosphorylated human CFTR. Proc Natl Acad Sci USA 115:12757–12762. https://doi.org/10.1073/pnas.1815287115
21. Tordai H, Leveles I, Hegedűs T (2017) Molecular dynamics of the cryo-EM CFTR structure. Biochem Biophys Res Commun 491:986–993. https://doi.org/10.1016/j.bbrc.2017.07.165
22. Corradi V, Gu R-X, Vergani P, Tieleman DP (2018) Structure of transmembrane helix 8 and possible membrane defects in CFTR. Biophys J 114:1751–1754. https://doi.org/10.1016/j.bpj.2018.03.003
23. Chin S, Ramjeesingh M, Hung M et al (2018) Lipid interactions enhance activation and potentiation of cystic fibrosis transmembrane conductance regulator (CFTR). BioRxiv. https://doi.org/10.1101/495010
24. Hildebrandt E, Khazanov N, Kappes JC et al (2017) Specific stabilization of CFTR by phosphatidylserine. Biochim Biophys Acta Biomembr 1859:289–293. https://doi.org/10.1016/j.bbamem.2016.11.013
25. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. Bioinformatics 22:623–625. https://doi.org/10.1093/bioinformatics/btk023
26. Pronk S, Páll S, Schulz R et al (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845–854. https://doi.org/10.1093/bioinformatics/btt055

27. Best RB, Zhu X, Shim J et al (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ_1 and χ_2 dihedral angles. J Chem Theory Comput 8:3257–3273. https://doi.org/10.1021/ct300400x
28. Lee J, Cheng X, Swails JM et al (2016) CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. J Chem Theory Comput 12:405–413. https://doi.org/10.1021/acs.jctc.5b00935
29. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29:1859–1865. https://doi.org/10.1002/jcc.20945
30. Brooks BR, Brooks CL, Mackerell AD et al (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30:1545–1614. https://doi.org/10.1002/jcc.21287
31. Komuro Y, Re S, Kobayashi C et al (2014) CHARMM force-fields with modified polyphosphate parameters allow stable simulation of the ATP-bound structure of Ca^{2+}-ATPase. J Chem Theory Comput 10:4133–4142. https://doi.org/10.1021/ct5004143
32. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem 32:2319–2327. https://doi.org/10.1002/jcc.21787
33. Tribello GA, Bonomi M, Branduardi D et al (2014) PLUMED 2: new feathers for an old bird. Comput Phys Commun 185:604–613. https://doi.org/10.1016/j.cpc.2013.09.018
34. Smart OS, Neduvelil JG, Wang X et al (1996) HOLE: a program for the analysis of the pore dimensions of ion channel structural models. J Mol Gr 14:354–360. https://doi.org/10.1016/S0263-7855(97)00009-X
35. Petřek M, Otyepka M, Banáš P et al (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities. BMC Bioinform 7:316. https://doi.org/10.1186/1471-2105-7-316
36. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Gr 14(33–38):27–28
37. Sethi A, Eargle J, Black AA, Luthey-Schulten Z (2009) Dynamical networks in tRNA:protein complexes. Proc Natl Acad Sci USA 106:6620–6625. https://doi.org/10.1073/pnas.0810961106
38. Eargle J, Luthey-Schulten Z (2012) NetworkView: 3D display and analysis of protein·RNA interaction networks. Bioinformatics 28:3000–3001. https://doi.org/10.1093/bioinformatics/bts546
39. McGibbon RT, Beauchamp KA, Harrigan MP et al (2015) MDTraj: a modern open library for the analysis of molecular dynamics trajectories. Biophys J 109:1528–1532. https://doi.org/10.1016/j.bpj.2015.08.015
40. Glykos NM (2006) Software news and updates carma: a molecular dynamics analysis program. J Comput Chem 27:1765–1768. https://doi.org/10.1002/jcc.20482
41. Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55
42. El Hiani Y, Linsdell P (2015) Functional architecture of the cytoplasmic entrance to the cystic fibrosis transmembrane conductance regulator chloride channel pore. J Biol Chem 290:15855–15865. https://doi.org/10.1074/jbc.M115.656181
43. Norimatsu Y, Ivetac A, Alexander C et al (2012) Cystic fibrosis transmembrane conductance regulator: a molecular model defines the architecture of the anion conduction path and locates a "bottleneck" in the pore. Biochemistry 51:2199–2212. https://doi.org/10.1021/bi201888a
44. Linsdell P (2014) Functional architecture of the CFTR chloride channel. Mol Membr Biol 31:1–16. https://doi.org/10.3109/09687688.2013.868055
45. Wang W, El Hiani Y, Linsdell P (2011) Alignment of transmembrane regions in the cystic fibrosis transmembrane conductance regulator chloride channel pore. J Gen Physiol 138:165–178. https://doi.org/10.1085/jgp.201110605
46. Gao X, Bai Y, Hwang T-C (2013) Cysteine scanning of CFTR's first transmembrane segment reveals its plausible roles in gating and permeation. Biophys J 104:786–797. https://doi.org/10.1016/j.bpj.2012.12.048
47. El Hiani Y, Linsdell P (2012) Tuning of CFTR chloride channel function by location of positive charges within the pore. Biophys J 103:1719–1726. https://doi.org/10.1016/j.bpj.2012.09.020
48. 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. https://doi.org/10.1074/jbc.M411935200
49. Sheppard DN, Travis SM, Ishihara H, Welsh MJ (1996) Contribution of proline residues in the membrane-spanning domains of cystic fibrosis transmembrane conductance regulator to chloride channel function. J Biol Chem 271:14995–15001
50. Zhou J-J, Li M-S, Qi J, Linsdell P (2010) Regulation of conductance by the number of fixed positive charges in the intracellular vestibule of the CFTR chloride channel pore. J Gen Physiol 135:229–245. https://doi.org/10.1085/jgp.200910327
51. Aubin CNS, 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. https://doi.org/10.1085/jgp.200609516
52. Cui G, Zhang Z-R, O'Brien ARW et al (2008) Mutations at arginine 352 alter the pore architecture of CFTR. J Membr Biol 222:91–106. https://doi.org/10.1007/s00232-008-9105-9
53. Branduardi D, Gervasio FL, Parrinello M (2007) From A to B in free energy space. J Chem Phys 126:054103. https://doi.org/10.1063/1.2432340
54. Csizmadia G, Farkas B, Spagina Z et al (2018) Quantitative comparison of ABC membrane protein type I exporter structures in a standardized way. Comput Struct Biotechnol J 16:396–403. https://doi.org/10.1016/j.csbj.2018.10.008
55. Zhang J, Yu Y-C, Yeh J-T, Hwang T-C (2018) Functional characterization reveals that zebrafish CFTR prefers to occupy closed channel conformations. PLoS One 13:e0209862. https://doi.org/10.1371/journal.pone.0209862
56. Taylor NMI, Manolaridis I, Jackson SM et al (2017) Structure of the human multidrug transporter ABCG2. Nature. https://doi.org/10.1038/nature22345
57. He Z, Cui H, Hao S et al (2018) Electric-field effects on ionic hydration: a molecular dynamics study. J Phys Chem B 122:5991–5998. https://doi.org/10.1021/acs.jpcb.8b02773
58. Szabó K, Szakács G, Hegeds T, Sarkadi B (1999) Nucleotide occlusion in the human cystic fibrosis transmembrane conductance regulator. Different patterns in the two nucleotide binding domains. J Biol Chem 274:12209–12212. https://doi.org/10.1074/jbc.274.18.12209
59. Aleksandrov L, Aleksandrov AA, Chang X-B, Riordan JR (2002) The first nucleotide binding domain of cystic fibrosis transmembrane conductance regulator is a site of stable nucleotide interaction, whereas the second is a site of rapid turnover. J Biol Chem 277:15419–15425. https://doi.org/10.1074/jbc.M111713200
60. Basso C, Vergani P, Nairn AC, Gadsby DC (2003) Prolonged nonhydrolytic interaction of nucleotide with CFTR's NH2-terminal nucleotide binding domain and its role in channel gating. J Gen Physiol 122:333–348. https://doi.org/10.1085/jgp.200308798
61. Li C, Ramjeesingh M, Wang W et al (1996) ATPase activity of the cystic fibrosis transmembrane conductance regulator. J Biol Chem 271:28463–28468. https://doi.org/10.1074/jbc.271.45.28463

62. Schultz BD, Bridges RJ, Frizzell RA (1996) Lack of conventional ATPase properties in CFTR chloride channel gating. J Membr Biol 151:63–75

RESEARCH ARTICLE

Jump into a New Fold—A Homology Based Model for the ABCG2/BCRP Multidrug Transporter

Laura László[1], Balázs Sarkadi[1,2], Tamás Hegedűs[1]*

1 Molecular Biophysics Research Group and Department of Biophysics and Radiation Biology, Semmelweis University, Budapest, Hungary, **2** Research Centre for Natural Sciences, Institute of Enzymology, Hungarian Academy of Sciences, Budapest, Hungary

* tamas.hegedus@hegelab.org

OPEN ACCESS

Citation: László L, Sarkadi B, Hegedűs T (2016) Jump into a New Fold—A Homology Based Model for the ABCG2/BCRP Multidrug Transporter. PLoS ONE 11(10): e0164426. doi:10.1371/journal.pone.0164426

Editor: Eugene A. Permyakov, Russian Academy of Medical Sciences, RUSSIAN FEDERATION

Received: August 11, 2016

Accepted: September 23, 2016

Published: October 14, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. Additional information will be available at http://abcg.hegelab.org.

Funding: This work was supported by National Research, Development and Innovation Office, http://nkfih.gov.hu, K111678 (T.H.) and K115375 (B.S.); and Hungarian Academy of Sciences, http://mta.hu, Bolyai Fellowship (T.H.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

ABCG2/BCRP is a membrane protein, involved in xenobiotic and endobiotic transport in key pharmacological barriers and drug metabolizing organs, in the protection of stem cells, and in multidrug resistance of cancer. Pharmacogenetic studies implicated the role of ABCG2 in response to widely used medicines and anticancer agents, as well as in gout. Its Q141K variant exhibits decreased functional expression thus increased drug accumulation and decreased urate secretion. Still, there has been no reliable molecular model available for this protein, as the published structures of other ABC transporters could not be properly fitted to the ABCG2 topology and experimental data. The recently published high resolution structure of a close homologue, the ABCG5-ABCG8 heterodimer, revealed a new ABC transporter fold, unique for ABCG proteins. Here we present a structural model of the ABCG2 homodimer based on this fold and detail the experimental results supporting this model. In order to describe the effect of mutations on structure and dynamics, and characterize substrate recognition and cholesterol regulation we performed molecular dynamics simulations using full length ABCG2 protein embedded in a membrane bilayer and *in silico* docking simulations. Our results show that in the Q141K variant the introduced positive charge diminishes the interaction between the nucleotide binding and transmembrane domains and the R482G variation alters the orientation of transmembrane helices. Moreover, the R482 position, which plays a role the substrate specificity of the transporter, is located in one of the substrate binding pockets identified by the *in silico* docking calculations. In summary, the ABCG2 model and *in silico* simulations presented here may have significant impact on understanding drug distribution and toxicity, as well as drug development against cancer chemotherapy resistance or gout.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

ATP Binding Cassette (ABC) transporters form one of the largest families of membrane proteins and are involved in numerous physiological and pharmacological functions [1]. These proteins are present from bacteria to human, and while in bacteria they may work both as importers and exporters, in eukaryotes only the exporter function has been preserved. The human ABC protein family includes 48 members, from which the members of the ABCB, ABCC and ABCG proteins are key cellular exporters for xenobiotics and endobiotics [1, 2]. These promiscuous transporters provide the basis of protecting cells and tissue barriers against hydrophobic toxic materials, regulate the ADME-Tox (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties of numerous clinically applied drugs, and are involved in cancer chemotherapy resistance.

In spite of the huge amount of data about the mechanism of action, cellular localization, and pharmacogenetically important polymorphisms and mutations of the human ABC drug transporters, the structural features of these large membrane proteins are hardly known [3–5]. There are several crystallization-based experimental data for bacterial importers and exporters, while practically only one single set of such data are available for the mammalian (mouse) ABCB1 drug transporter [6]. Still, numerous homology models have been built for other human ABC drug transporters, either based on this ABCB1, or the bacterial ABC transporter structures [7–9].

The ABCG2 protein is a multifunctional human membrane transporter–it is expressed at a high level in the gut, in the blood-brain and the feto-maternal barriers, and its function is especially relevant in stem cell protection [10–12]. Overexpression of ABCG2 has been shown to cause cancer multidrug resistance [13–15], while a reduced expression or function is an established cause of hyperuricemia and gout [16–18]. Still, it has been especially challenging to model the atomic level structure of ABCG2, as the nucleotide binding domain (NBD) and the transmembrane domain (TMD) arrangements are in an inverse order than in the ABCB or ABCC families, and membrane topology studies indicated a completely different transmembrane helix arrangement for this protein.

Very recently the crystal structure of the ABCG5-ABCG8 heterodimer membrane protein, the key human transporter for cholesterol, has been published [19]. These proteins, as members of the human ABCG subfamily, show close homology to the homodimeric ABCG2 transporter. Since ABCG2 has major medical importance, here we provide a homology model for this protein, based on the ABCG5-ABCG8 structure. We also provide explanations for the polymorphism and mutations with experimental or clinical relevance [20–23]. Moreover, by utilizing molecular dynamics simulations and *in silico* docking calculations, we describe the potential drug binding and transport regions, as well as the residues responsible for the special cholesterol sensitivity of this promiscuous drug transporter. Clearly, these structural data may provide important clues to decipher the effects of the ABCG2 variants, design drugs to rescue their expression, or drugs to modulate their function in stem cell development, protection of the fetus or cancer drug resistance.

Methods

Homology modeling

Sequence alignment of ABCG2_HUMAN, ABCG5_HUMAN, and ABCG8_HUMAN (UniProt) was generated using ClustalW [24]. Modeller 9.12 was employed to generate the homology models [25]. Cysteines 592 and 608 were constrained to form intramolecular disulfide bonds, while cysteines at position 603 were forced to participate in an intermolecular S-S

bridge. One hundred models were prepared and the best model was selected by Modeller's DOPE score. Since the extracellular loops contained knots, the loops between C592 and C608 were refined also by Modeller, employing the same approach: 100 loop models were created and evaluated by DOPE score.

Mutant constructs could be generated simply by PyMOL (The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC.), since either the side chain of the new residue was small or it was on the protein surface. PyMOL was used to generate all the molecular graphics. Sequence alignments, structural models, and additional information on molecular dynamics and in silico docking (see below) can be downloaded from http://abcg.hegelab.org to facilitate further studies.

Molecular Dynamics (MD) simulations

The structural models were oriented along the membrane normal based on the ABCG5-ABCG8 orientation in the OPM (Orientations of Proteins in Membranes) database [26]. MD simulations were performed using GROMACS 5.1 with CHARMM36 force field [27, 28]. The input files for energy minimization, several equilibration steps (NVT, NPT), and production run were generated via the CHARMM-GUI web interface [29, 30]. The following options were selected: terminal residues were patched by ACE (acetylated N-terminus) and CT3 (N-methylamide C-terminus), the extracellular cysteines were set to form disulfide bridges; homogenous POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) lipid bilayer were selected with default parameters and 150 mM NaCl was inserted; grid information for PME (Particle-Mesh Ewald) electrostatics was generated automatically, NPγT ensemble was selected with constant number of particles (N), pressure (P) of bar, surface tension zero (γ), and temperature of 310 K. All the wild type and mutant structures were energy minimized in the first step using the steepest descent integrator (maximum number to integrate: 5,000 or converged when force is $<$1,000 kJ/mol/nm). From each energy minimized structure we forked six parallel simulations containing consecutive sets of equilibration steps, when decreasing force constants (from 4,000 to 50 kJ/mol/nm^2) in these steps were applied for protein and lipids. Production runs were performed without restraints. Berendsen thermostat and barostat were used in the equilibration steps, while Nose-Hoover thermostat and Parrinello-Rahman barostat with semiisotropic coupling were employed in the production run. Time constants were 1 ps and 5 ps for the thermostats and barostats, respectively. Electrostatic interactions were calculated using the fast smooth PME algorithm and LINCS algorithm was used to constrain bonds. Simulations were carried out in constant particle number, pressure, and temperature ensembles with a time step of 2 fs. In summary, all the parameters provided by the CHARMM-GUI interface were unchanged and used except the simulation time in the production step, that was set to 50 ns. Thus six parallel runs resulted in 300 ns total simulation time for each constructs including WT, Q141K, R486G, and Y413S. Simulations were executed on a GPU cluster of the NIIF National Information Infrastructure Development Institute (http://www.niif.hu/en). Simulations were analyzed by the MDAnalysis Python package [31] and in-house Python scripts on our local small HPC cluster.

In silico docking

Molecules are listed in the Supplementary material (S6 Fig). The OpenBabel package and MGLTools scripts were used to convert between molecule file formats and to prepare the molecules for docking using AutoDock Vina [32, 33], respectively. The six conformations of ABCG2 wild type were taken from the last frame of the equilibrium simulations and prepared for docking using MGLTools (Gasteiger charges were added). The docking space was defined

by a box around the whole transmembrane domain including also some parts of the NBD and the extracellular loops (S5 Fig). Because of the large volume of the box, exhaustiveness was set to 128 instead of the default value 8 and the number of required poses in the output (num_-modes) was set to 20. Analysis was performed by PyMOL and in-house Python scripts.

Results and Discussion

Homology modeling based on the new ABC transporter fold provides a plausible and stable ABCG2 model

In contrast to previous ABC transporter structures, the recent ABCG5-ABCG8 high resolution structure provides an excellent template for modeling ABCG2 [19]. ABCG2 exhibits 27% and 26% identities and 48% and 44% similarities when compared to ABCG5 and ABCG8, respectively. Although these values seem to be low for general homology modeling, for longer sequences (> 100 a.a.) and especially for membrane ABC proteins they are sufficient (see Modeller's tutorial and [34]). Even in a worse scenario, when the N- and C-terminal halves of CFTR/ABCC7 (Cystic Fibrosis Transmembrane Conductance Regulator) exhibit only 18% and 21% identities compared to the *Staphylococcus aureus* Sav1866 protein, using this distantly related protein as a template resulted in a high quality CFTR homology model that could be confirmed by experiments [9]. Bacterial homolog based MDR1 (P-gp) models also have been generated and widely used to guide experimental and computational studies [35–37]. Although the ABCG5-ABCG8 template is a heterodimer, the structural differences between the two halves are subtle (S1 Fig). Based on these considerations, we generated a homology model of ABCG2 based on the ABCG5-ABCG8 structure (Fig 1) employing the sequence alignment shown in S2 Fig. Although the alignment generation was relatively straightforward, some parts of the ABCG2 protein were not modeled, either because they are mobile and thus invisible in the template structure, or their sequence and length differ from the corresponding regions in the template (e.g. the loop between the β1 and β2 strands of NBD and the linker region between the NBD and TMD; see details below and S2 Fig).

Large deviations from the starting structure are expected even in short (<20 ns) MD simulations when the initial structure is wrong or inaccurate [6, 38–40]. Therefore we performed molecular dynamics simulations employing the ABCG2 homodimer embedded in a membrane bilayer to confirm the stability of the homology model. RMSD values of frames compared to the initial structure indicated the stability of our model (S3 Fig). Although it would be interesting to perform experiments testing specific aspects of the ABCG2 structural model employing devised mutations, as in the case of the CFTR homology model [9], the ABCG2 transporter is highly sensitive to mutations and its cysteine-less form cannot be functionally expressed [41]. However, there are several experimental observations, which we discuss in the next section, supporting our model.

The nucleotide binding domains (NBDs) are the most conserved regions in all ABC proteins from bacteria to human, and consist of a RecA-like core domain present in all P-loop ATPases, and an α-subdomain characteristic exclusively for ABC proteins [4, 5, 42]. The Walker A and B motives responsible for ATP binding are located in the core domain, while the ABC signature sequence (LSGGQ), which provides the catalytic base toward the γ-phosphate, is situated in the α-subdomain (S2 Fig). Since ATP binds to the Walker A sequence in one NBD, and the signature sequence is provided from the other NBD, for ATP hydrolysis an intimate interaction of the two NBDs is required [4, 5, 42]. It is also important to mention that the ABCG5-ABCG8 heterodimer exhibits a functional asymmetry in the NBDs, as ABCG5 possesses a degenerate signature sequence thus unable to cleave ATP. In the ABCG2 dimer both ATP sites are active. The ABCG5-ABCG8 structure does not contain bound ATP, thus the NBDs were separated

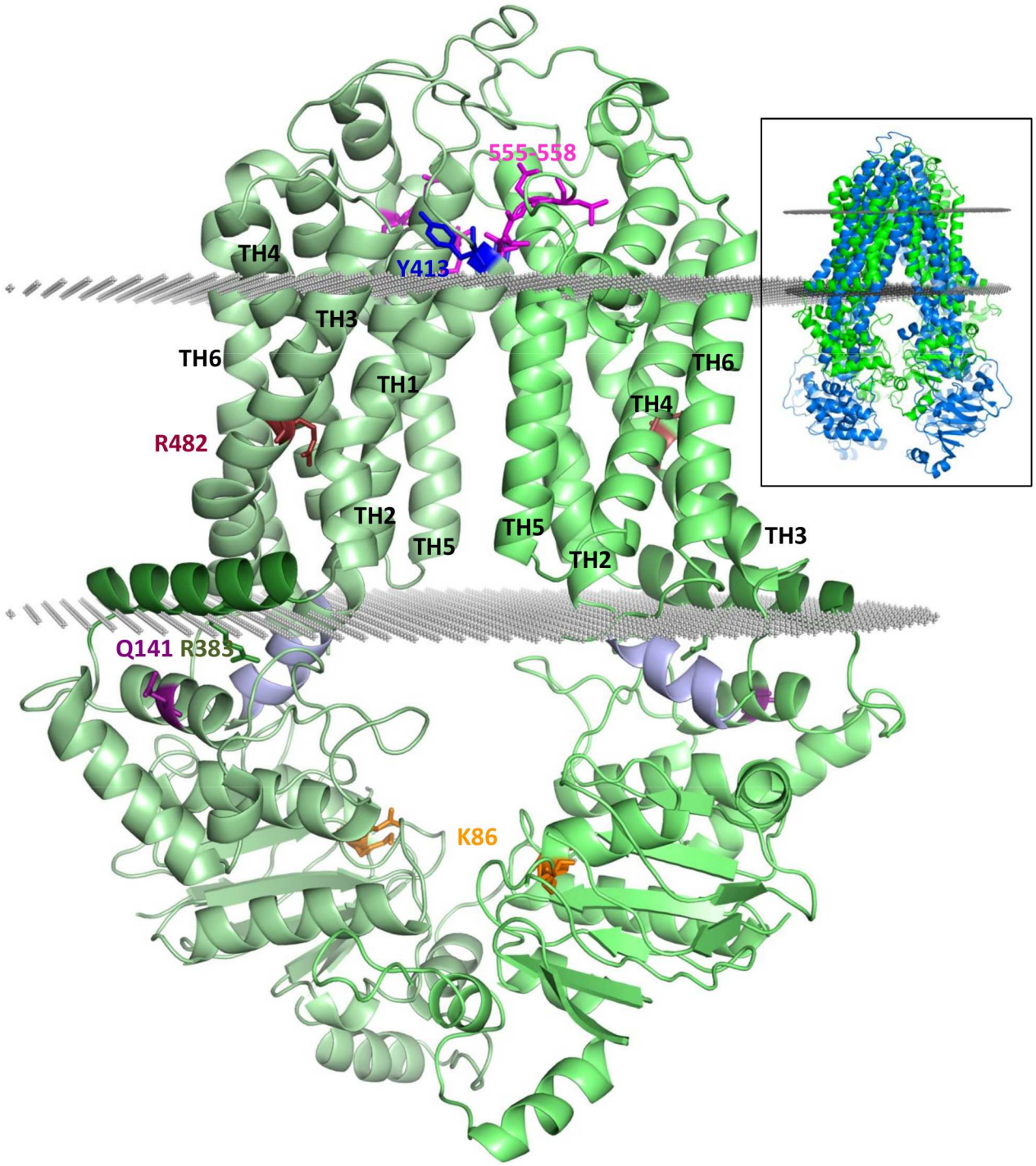


Fig 1. The general structural properties of the ABCG2 homology model. The two monomers are colored by different light green colors. The most important parts, providing the interface between the TMD and NBD are the coupling helix (light blue) and the connecting helix (dark green). The functionally important R482 is colored ruby. The site of the most frequent polymorphism, Q141 is deep purple. The location of important mutations affecting biogenesis and function are labeled by dark green (R383) and orange (K86), respectively. Residues, which are probably significant in cholesterol modulation, are blue (Y413) and magenta (a.a. 555–558). Gray dots represent the boundaries of the hydrophobic region of the bilayer, defined by the OPM webserver. **Insert:** ABCG2 (green) and mouse ABCB1/Pgp (blue, PDBID: 4M1M) are overlaid. The mouse ABCB1 NBD is much further from the membrane bilayer and the distance between the NBDs of ABCB1 and ABCG2 (the Cα atoms of the Walker A Lys residues; K433 and K86, respectively) is 26 Å.

doi:10.1371/journal.pone.0164426.g001

[19], and so are they now in the ABCG2 structure presented. However, in contrast to that in the mouse MDR1 structure [6], the NBDs are not fully separated, and a connection is provided by helices located at the C-terminus of NBDs. Till now this conformation could have been observed only in lower eukaryotes [43, 44].

In the ABCG type proteins the loop between the first and second β-strands of the core domain are longer than in most ABC transporters (approx. 20 a.a. in ABCG2, and 20 and 40 in ABCG5 and ABCG8, respectively; S2 Fig). This loop is invisible in both the ABCG5 and ABCG8 crystal structures, thus most likely highly mobile. The only loop, which has been described at the same location, is the so called "regulatory insertion" in the CFTR protein [45]. However, this region of the NBD does not play any role in phosphorylation or nucleotide dependent regulation of CFTR function. When this "regulatory insertion" was deleted in CFTR, the thermostability of this channel was increased significantly, but no physiological function for this segment could be identified [46]. Until now the existence of this long loop in ABCG proteins has not been known because of the lower sequence conservation of the β1-strand and its function is unknown. Since it includes the short A-loop motif [47], one of its functions supposed to be ATP stabilization. However, its length suggests additional roles, and we propose that it may serve as a filter at the entry to the substrate binding cavity (see below), and contribute to the first step of allosteric communication of signaling drug biding to the ATP binding site.

The transmembrane domain in the ABCG proteins exhibits a completely new ABC transporter fold. The intracellular loops are shorter compared to known ABC exporter structures, that results in a small distance of the NBDs from the inner layer of the membrane (Fig 1 insert). In this respect the structures of the ABCG proteins resemble the bacterial importers, but their transmembrane fold is completely different. In addition, no intracellular loops do cross over from one TMD to the opposite NBD, as observed in previous ABC exporter structures [9, 48]. Moreover, the arrangement of the short, so called coupling helices, which are located at the NBD/TMD interface and couple the motions resulting from ATP binding and hydrolysis in the NBDs to conformational changes in the TMD, are completely different from those found in the ABCB or ABCC type proteins. One of the intracellular loops between TH4 and TH5 (ICL2), in contrast to other ABC exporter folds, is so short that it does not leave the bilayer, thus does not reach the NBD and cannot function as a coupling helix, providing a molecular coupling. The only potential coupling helix in ICL1 (a.a. 452–461), located between TH2 and TH3, binds to the NBDs similarly to that observed in other ABC structures [9, 42] (Fig 1), but in a slightly different conformation. Interestingly, while there are bacterial importers that possess only one coupling helix in one half of the transporter, ABCGs still possess more interactions between the TMDs and NBDs. An amphipathic helix (a.a. 373–390; Figs 1 and 2 and S2 and S3 Figs) similar to a coupling helix provides additional interactions and is located in the linker region directly before TH1. The coupling helix like conformation of this region, which is named "connecting helix" [19] and forms an alpha helix perpendicular to the TM helices, is established by its amphipathic nature. However, this connecting helix does not penetrate as deeply into the NBDs as the coupling helices in Type I ABC exporters. In addition, this connecting helix exhibits similar structural and functional roles as those coupling helices which cross over from one TMD to the opposite NBD in the other half of the molecule, in type I ABC exporter structures.

An interesting and ABCG specific region of the new fold is an extracellular loop between TH5 and TH6 (ECL3; a.a. 552–621) with a special conformation. The residues located between 562 and 586 form two consecutive helices, exhibiting a V-shape, immersed partially into the bilayer (S2 and S4 Figs). This re-entering into the membrane resembles the P-loop of ion channels, and thus we propose to name it a G-loop. The extracellular loop following the

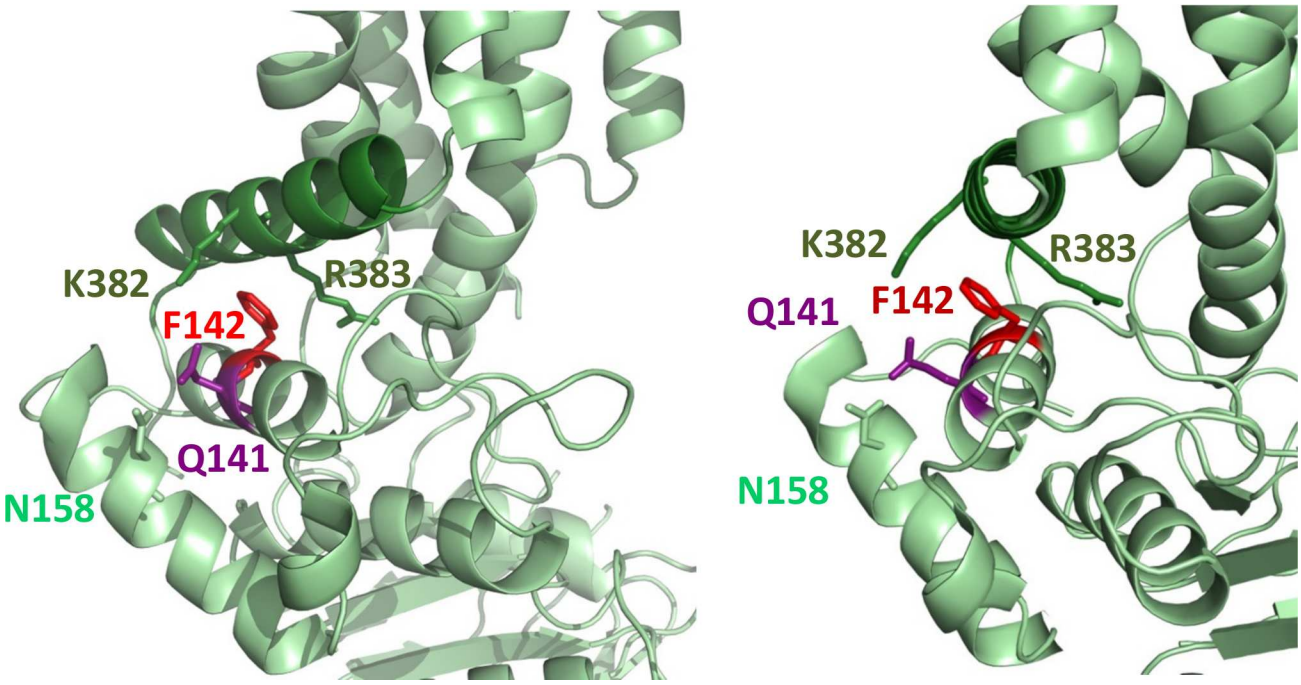


Fig 2. The Q141K variant interferes with the coupling between the NBD and the connecting helix. The side chain of F142, which is at a homologous position as the CFTR F508 (S2 Fig), is clamped by the positively charged K382 and R383. The positive charge of 141K destabilizes this interaction by repulsion with K382, as shown by molecular dynamics simulations (S5 Fig).

doi:10.1371/journal.pone.0164426.g002

G-loop in the ABCG proteins is not highly conserved and is slightly different in its length. ABCG8 does not possess any cysteines in this region, while the two cysteines in ABCG5 G-loop are at similar positions as C592 and C608 in the ABCG2. It has been experimentally documented that these cysteines form an intramolecular disulfide bond in the ABCG2 protein [49], while this covalent bond is not observable in the ABCG5-ABCG8 structure, because purification and crystallization were done under reducing conditions. In addition, in ABCG2 it has been shown that the two C603 residues in the two halves of ABCG2 form an intermolecular disulfide bond [49, 50]. The experimentally verified N-glycosylation site in ABCG2 at position N596 [51] in our ABCG2 model is found in the flexible extracellular loop 3 and accessible for glycosylation.

As described above, the ABCG family members have entirely different conformational arrangements in the transmembrane helices and connecting regions than the other known ABC exporters. Therefore it is not surprising that until now the structure of ABCG-like proteins could not have been properly predicted—all attempts were based on the presumption that the NBDs are located far from the bilayer and are connected to the TMD with long intracellular loops [8, 52–54]. In addition, experimental findings that supported the long ABC fold in ABCG type proteins may raise serious concerns (see Fig 1 insert). Moreover, studies which attempted to determine the ABCG2 transmembrane topology by HA-insertion experiments, seriously failed [54, 55]. This was most likely caused by a misinterpretation of the expression and function of the mutant variants, biased by the misconception of presuming long intracellular loops in TMDs. In contrast, the *in silico* prediction of the ABCG2 membrane topology performed much better, and resulted in TM helix boundaries similar to those deducted from the ABCG5-ABCG8 crystal structure and the present ABCG2 model (S1 Table) [56, 57].

The new model contributes to the understanding of the effects of ABCG2 variations

Structural models, even homology models with their limitations, are important tools to assess the effects of both natural variations, and mutations generated for structure/function studies. A prominent example in the field is the CFTR homology model based on the Sav1866 structure [9]. While these proteins share very low homology in the TMDs, major structural features could have been confirmed by experiments and the model still serves as a fundamental basis from basic studies to drug development. Below we detail the experimentally examined specific residues and regions in the ABCG2 protein, in order to assess the suitability of the new model to understand their effects on ABCG2 structure and function.

The K86M mutation, within the Walker A ATP binding motif has been used in numerous studies to generate a non-functional ABCG2 transporter, lacking both ATPase and transport activities [58]. Similarly to other ABC transporters, the NBD/NBD domain interface is highly sensitive for mutations, irrespectively whether the actual residue plays a role in the catalysis. The E211Q mutation, causing loss of function, is localized in this region [59].

Two major polymorphic variants of ABCG2 are V12M and Q141K [22]. V12M, present in 5–10% frequency in human populations, has been shown not to have a measurable effect on the processing or function of the protein. The lack of effect can be explained by the position of this variation located in the short and flexible N-terminal region of the protein (a.a. 1–30). Tagging experiments are also in line with the spatial arrangement of the N-terminus: either a 6–10 histidine tag, or even a large GFP tag, attached to this end, is well tolerated in ABCG2 processing and function [60].

Q141K, present in about 15–30 percent of people in various ethnic groups, causes a significant reduction in intracellular trafficking and plasma membrane localization of ABCG2 [20, 21]. Due to the reduced expression at the site of action, this variant contributes to the development of gout and enhances the side effects and toxicity of various drugs [16–18, 61–63]. Interestingly, Q141 within the NBD is located next to F142, homologous to the CFTR F508 (Fig 2) [64], in a helix interacting with the amphipathic "connecting helix". This site is analogous to the crossed-over coupling helix of cytoplasmic loop 3 in CFTR, and may similarly have a role in stabilizing the NBD/TMD interface [9, 65].

The side chain of Q141 is directed towards N158, thus the substitution of glutamine by the larger lysine with a positive charge, may displace the α-helix of N158. This helix is on the external side of the NBD, thus expected to cause only a minor effect on the transporter function. Molecular dynamics simulations indicate that these helices do not move differently in the Q141K mutant as compared to the wild type protein. On the other hand, the distances between the connecting helix and the helix of Q141 and F142 exhibit differences in dynamics (S5 Fig). The interaction at this NBD/TMD interface is stabilized by a special arrangement of three amino acids (Fig 2). The side chains of K382 and R383 in the connecting helix form a V-shape, and clamp the residue F142 located in the NBD. In the Q141K variant the positive charge interferes with K382, the interaction of the two helices is destabilized, and exhibits an increased probability of divergence for the two interfacing helices. In the light of this observation, the crucial role of R383 in stabilizing the NBD/TMD interface is highlighted and the deleterious impact of R383 mutations on ABCG2 biogenesis is interpretable [66].

Experimental studies have shown that while the Q141K variant causes only a partial impairment in ABCG2 processing, the ΔF142 mutant has a more severe effect than the ΔF508 mutation in the CFTR protein [64]. Misprocessing of ΔF142 ABCG2 cannot be rescued either by temperature or correctors, most likely because in the case of ABCG2 every single mutation behaves as a double mutation in the homodimer transporter. The molecular modeling and

experimental studies may significantly contribute to explore the already available CFTR correctors to rescue the ABCG2 Q141K variant, e.g. in the therapy of gout [65]. Still, the local molecular environment is dissimilar to that observed in CFTR, namely the connecting helix is less embedded into the NBD, and the interacting residues are not hydrophobic in ABCG2, thus any CFTR corrector compounds tested for rescuing ABCG2 Q141K [65] should be tuned for the structure of ABCG2. In addition, CFTR F508 resides in a loop following a helix, while ABCG2 F142 is located in an α-helix.

In ABCG2 a historically and functionally important residue is the arginine in position 482 [58, 67]. In the first ABCG2 cloning experiments drug resistant cell lines expressed the R482G variant, which has a different substrate and inhibitor specificity and apparently a higher drug transport turnover that the wild-type protein. Still, this variant has not been found *in vivo*, probably because it cannot transport negatively charged substrates, including uric acid or conjugated hydrophobic drugs [23]. In the new ABCG2 model R482 resides in TH3, very close (2 a.a.) to a kink generated by P480. This proline, and most likely also the kink, are conserved in ABCGs (S2 Fig). As discussed below, the role of R482 in the substrate handling of ABCG2 is strongly supported by the current model.

The short cytoplasmic C terminus of the ABCG2 protein has been shown to be very sensitive to any experimental modification or tagging. This is well explained by the localization of these amino acids, facing the inner "cavity" of the dimer in very close proximity to the coupling helix and most probably interacting with it.

Docking calculations reveal drug binding sites along a potential transport pathway

In silico docking studies on multidrug transporters have not been a great success. Since it is challenging to handle the plasticity of a binding site capable of interacting with chemically different compounds and *in silico* docking to a homology model raises the concern regarding side chain orientations, we performed docking to several conformations generated by the equilibration steps of the molecular dynamics simulations. This process also provided a more physiological orientation of the side chains in the lipid regions, as the protein in the simulations was embedded in a lipid bilayer. These conformations possess similar backbones, since the protein is position-restrained during the main part of the equilibration process. To these conformations we docked various ABCG2 substrates including sulfasalazine, methotrexate, rhodamine 123, flavopiridol, and also molecules that do not interact with ABCG2 such as verapamil and calcein (S6–S8 Figs) [2, 12, 15], employing AutoDock Vina [33].

Interestingly, the conformations exhibited various characteristic locations for substrates (Fig 3): some of the conformations indicated potential binding sites around the entry pore from the cytosol (Sites 1 and 2), others delineated the entry to the interface of the two transmembrane domains (Site 3), and some exhibited a partially extracellular location (Site 4). Site 1 is situated below the connector and coupling helix, and above the loop connecting the core and α-helical subdomains of NBD. Site 2 is a more defined binding pocket, intercalated between TH1 and TH4, which also includes the R482 and the P480 kink. Site 3 is located between the two monomeric subunits, surrounded by TH2 and TH5, provided by both monomers. Amino acids of Site 4 are part of the tip of the TH helices and extracellular loops. Binding to this site was a rare event, indicating that the conformations of this region captured in the MD simulations do not form a real binding pocket, as expected for an off-site.

Interestingly, while both substrates and non-substrates can bind to Site 1, the entry into Site 2 of non-substrates is limited (e.g. for verapamil or calcein; S7 and S8 Figs and S2 Table). Thus substrate selection, e.g. differentiation of toxic molecules from natural metabolites, may happen

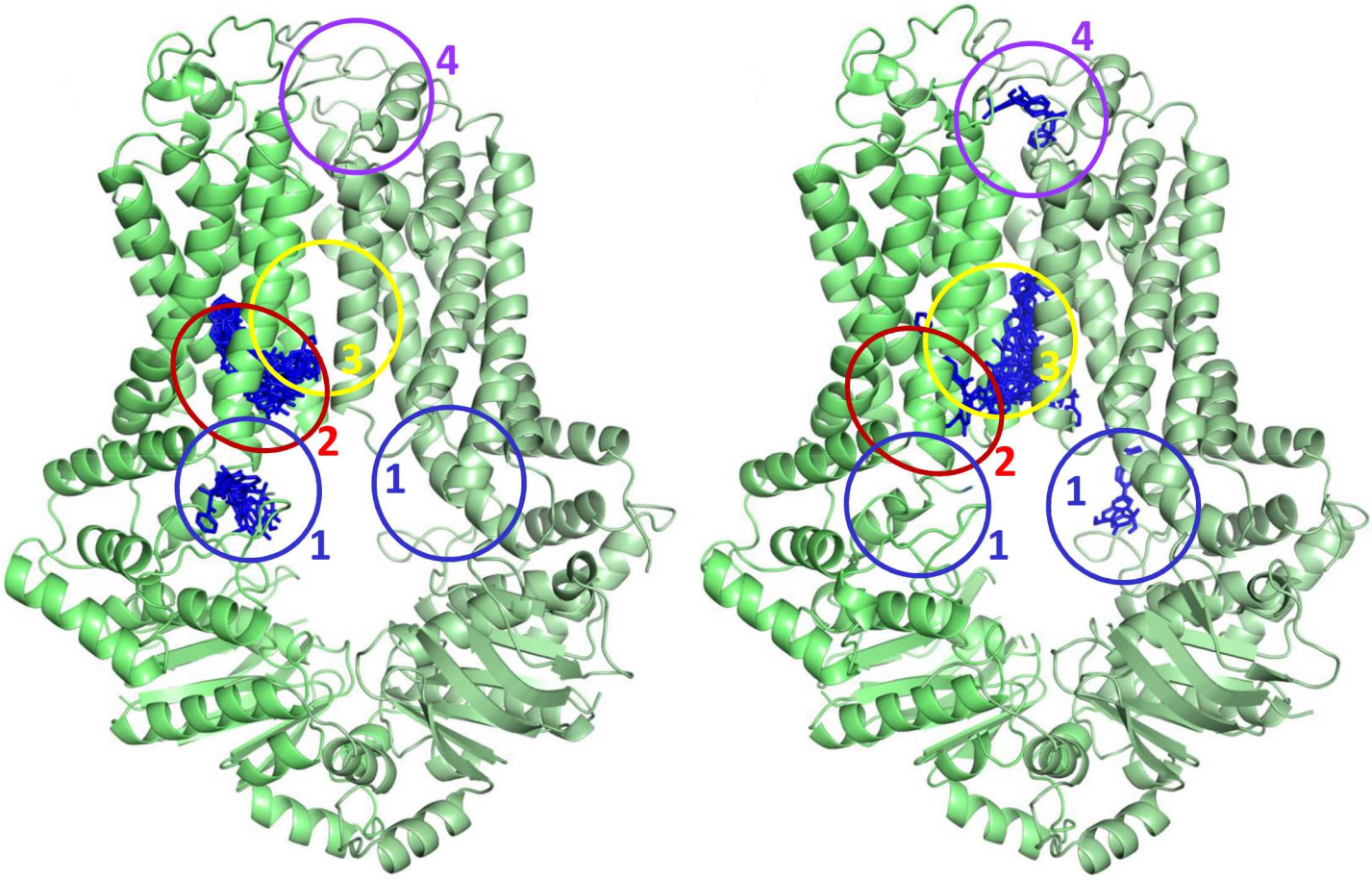


Fig 3. *In silico* docking shows binding sites along a potential substrate pathway. Substrates and non-substrates were docked to six ABCG2 conformations. Both types of molecules could dock at Site 1 (blue), while only the binding of substrates could be observed at Site 2 (red). The central Site 3 (yellow) resides between the two monomers. A potential off-site at the extracellular part is also revealed (Site 4, magenta). Here, docking poses of sulfasalazine are shown in the case of two ABCG2 conformations (two out of the last frames of the six equilibrations are shown).

doi:10.1371/journal.pone.0164426.g003

at this site. This proposal is also strengthened by the fact that R482, which exhibits a strong effect on substrate selectivity, is part of Site 2. It is also important to note that mutations of T402 and P485 in this pocket (S8 Fig) have been reported to reduce the transport of many substrates [8, 68], further supporting the existence of Site 2. It seems that all the molecules examined can bind into the central pocket (Site 3), which might be the entry point into the pathway between the two TMDs. In addition, for substrates, binding regions overlapping between Sites 2 and 3 also can be observed, delineating a potential transition spot from Site 2 to Site 3. Site 1 and Site 2 are present at both monomers, although their presence is not so pronounced in one of the monomers in our conformations because of the inherent asymmetry of the ABCG5-ABCG8 template. Site 3 is located between the two monomers, as a part of the main translocation path, and substrates can enter this pocket from Site 2 of either monomer in an alternating fashion (Fig 3). The central large cavity in the apo structures of ABCG5-ABCG8 and ABCG2 dimers exposes both hydrophobic and hydrophilic residues, and their pattern also may play a role in discriminating substrates and non-substrates. In addition, the loop between the β1 and β2 strands of NBDs may also limit the access to the entry sites and participate in substrate selection.

In silico modeling facilitates to uncover the atomic details of cholesterol dependence of ABCG2 function

Membrane cholesterol is a major modulator of ABCG2 function, as documented by several experimental studies [69–71]. In fact, purified ABCG2 is practically non-functional without the addition of cholesterol to the reconstituting lipids [60]. Experimental studies indicated that the R482G mutation also influences the cholesterol sensitivity of ABCG2, that is less cholesterol is required for full transport function [69]. Since this amino acid is in the hydrophobic region, it is questionable how could cholesterol interact with this residue. Most probably R482 alters the conformation and dynamics of TM helices, resulting in altered cholesterol sensitivity and substrate specificity (see above).

In order to test this hypothesis we executed MD simulations using the R482G mutant embedded in a lipid bilayer, and compared the conformation of the TM helices close to this residue, located in TH3. Even in the case of a limited accessible time scale, large conformational changes could be observed (Fig 4 and S9 Fig). In the absence of the large R482 side chain, TH3 moved closer to TH4 and at the same time drifting away from TH1. These conformational changes caused by R482 substitution indeed have a significant effect on regions, which have been proposed as cholesterol binding sites (e.g. TH1), and also on drug binding site 2, thus

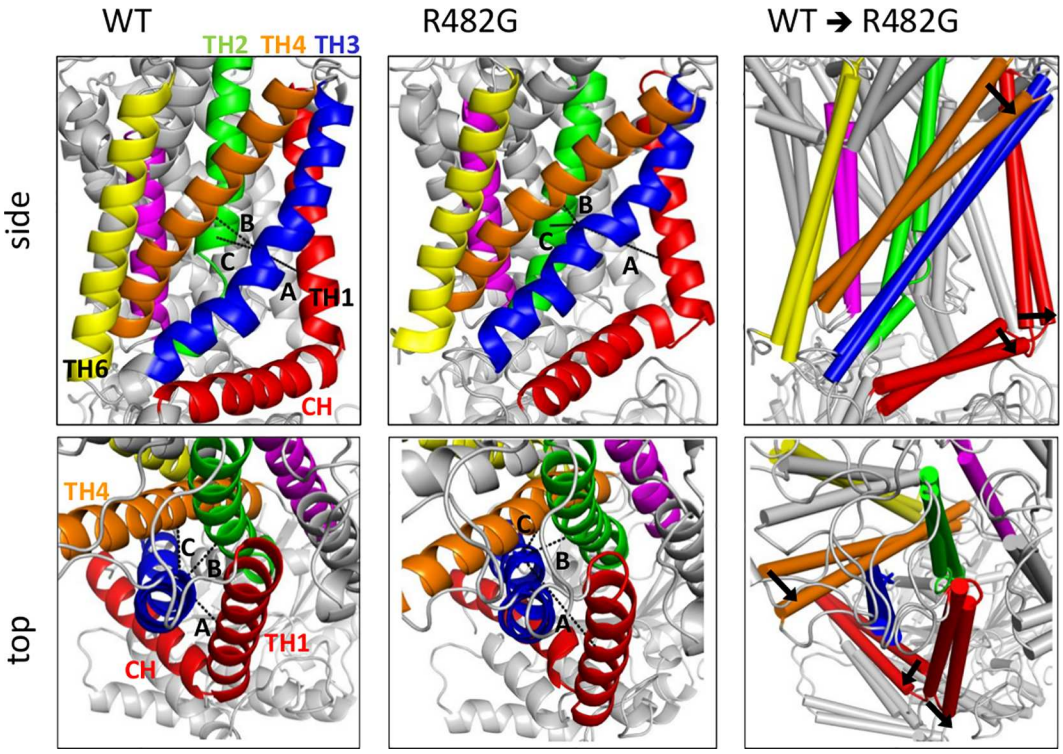


Fig 4. Structural effects of the R482 variations. The R482G mutation is able to alter the positioning of TM helices and the conformation of the P480 kink. Two structures were taken from the end of two simulations, which exhibited the largest changes, to decipher and demonstrate the effect of R482. The distances between Cα of R482 (TH3) and that of Q398 (TH1), S441 (TH2), and A517 (TH4) were measured throughout the simulation trajectory, and in the last frame exhibited the following values: distances of A, B, and C in WT are 8.4 Å, 7.2 Å, and 7.4 Å, while in the R482G variant are 15.1 Å, 8.1 Å, and 4.9 Å, respectively. The right panels contain both the WT and R482 structures in cylindrical representation. Arrows are placed at spots, which exhibit the most pronounced differences between the two constructs, and point from the wild type to the mutant conformation. TH1-6 are colored by red, green, blue, orange, magenta, and yellow, respectively.

doi:10.1371/journal.pone.0164426.g004

provide suitable explanations for both altered cholesterol sensitivity and substrate specificity of this variant.

Two recent publications explored the putative cholesterol binding sites in ABCG2 by mutagenesis. Gal *et al.* [72] investigated the effects of mutations of the key tyrosine residues in the putative cholesterol recognition amino acid consensus (CRAC) motives, located in ABCG2 at positions Y413, Y459, Y469, Y570 and Y645 (S4 Fig). The Y459S mutation prevented protein expression, the Y469S and Y645S mutants lost their transport and ATPase activities, while the only significant effect on cholesterol modulation of ABCG2 function was caused by the Y413S mutation. In the second related study [73] a leucine based potential cholesterol binding motif (a.a. 555–558) was found to play a significant role in the cholesterol dependence of ABCG2.

The structural model of ABCG2 presented here revealed that Y459 is located in the coupling helix and its mutation understandably caused a major detrimental effect. The Y413S CRAC motif mutation, that significantly altered the cholesterol sensitivity of ABCG2 without major functional effects, is located in the extracellular tip of TH1, in the area of the external lipid head groups, as supported by molecular dynamics simulations, performed with the transporter embedded in a POPC lipid bilayer (Fig 5). Most interestingly, the leucine based motif (a.a. 555–558), affecting the cholesterol sensitivity of ABCG2 is located just before the G-loop, in the outer, charged leaflet of the bilayer, close and in a potentially interacting position with the Y413 in the CRAC motif in the opposite monomer. These observations strongly suggest that the cholesterol binding site is located in this region. It may be supposed that an intermolecular interaction between Y413, L555, and V556 provide a specific, potentially dynamic conformation for the TM

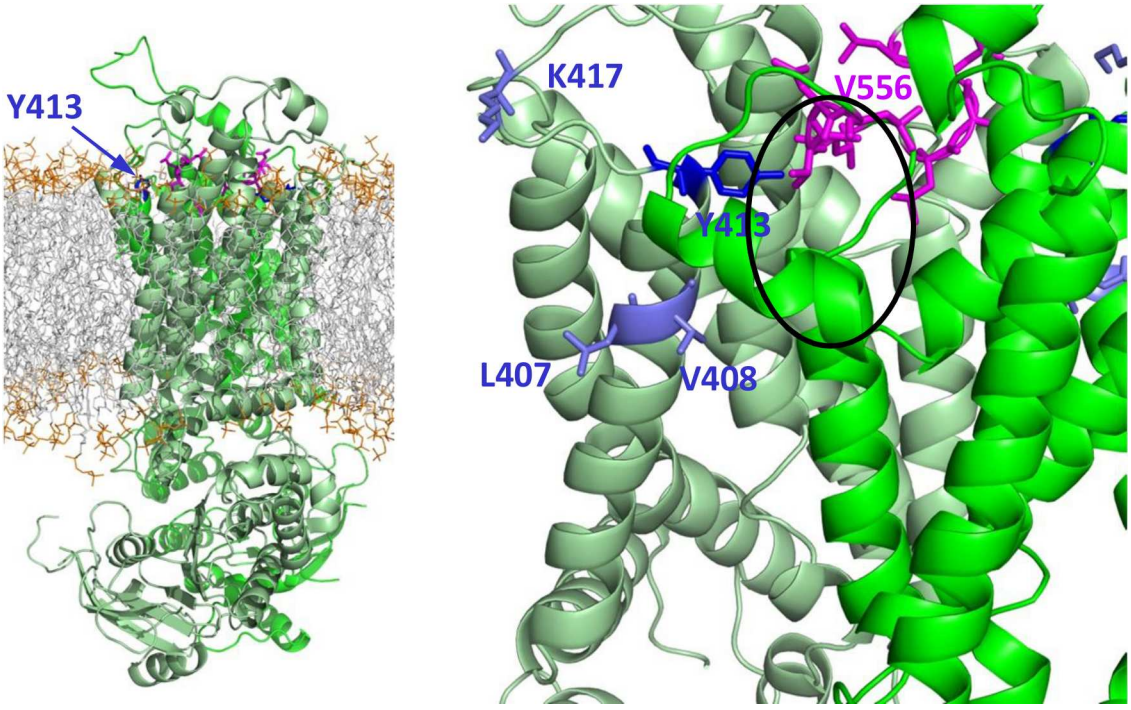


Fig 5. The structural background of cholesterol regulation. The last frame of a 50 ns long MD simulation with ABCG2 embedded in a POPC bilayer shows that the CRAC motif, containing Y413, is located in the charged area of POPC head groups (orange), as a rational location for cholesterol biding. Also, the leucine based cholesterol binding motif (magenta) is situated in this layer. Right panel: zoomed area reveals a close contact between the CRAC and the leucine based motives (e.g. Y413 and V556 are closer than 5 Å) and may provide a cholesterol binding site (black circle). Gray: POPC hydrophobic tails; orange: charged head groups of lipids.

doi:10.1371/journal.pone.0164426.g005

helices in cholesterol binding. The special localization of these two motives also suggests that cholesterol may contribute to the stabilization of the dimer in a specific conformation.

Since the Y413S mutation may affect cholesterol sensitivity of ABCG2 allosterically, and experimental studies to identify the direct binding of such a hydrophobic molecule are extremely difficult, we performed MD simulations with the Y413S mutant construct. We could not observe any significant changes in the dynamics in TH1 (S10 Fig), that also suggests that the altered cholesterol sensitivity observed in this mutant is caused directly by this functional CRAC motif.

Concluding remarks

The ABCG5-ABCG8 based homology model of ABCG2 presented here is capable to shed light on the effects of mutations, substrate handling, and also on the regulation of this transporter. The altered dynamics of the R482G variant provides explanation for both altered substrate specificity *via* affecting the drug binding pocket (Site 2), and to the altered cholesterol regulation through allosteric communication via TH1 to CRAC motif (Y413). As this ABCG5-ABCG8 based homology model behaved unexpectedly well in the simple computational approaches we applied, without any sophisticated additional methods, most likely the ABCG5-ABCG8 structure is a physiologically relevant conformation. Importantly, this conformation exhibits only slightly separated NBDs without bound ATP and exposes drug binding sites without widely separated TM helices at the cytoplasmic membrane leaflet.

The existence of the observed binding sites in the six conformations used in *in silico* docking calculations leads to important implications. Namely, subtle conformational changes (the maximum RMSD between the six equilibrated structures is 1.3 Å) are sufficient to provide binding sites at different regions of the protein, without the need of large movements or largely separated NBDs. Most likely the alternating access mechanism, which has been proposed to involve a drug binding conformation with widely separated intracellular parts and a release conformation widely opened to the extracellular space, is realized differently in the case of ABCG exporters. Based on the ABCG5-ABCG8 structure and our results with the ABCG2 model we hypothesize that the conformational changes needed for transport involve subtle repositioning and rotation of the transmembrane helices and these play a more pronounced role in the transport than have thought before. Supported by our results we are convinced that the ABCG2 model presented here may pave the road for drug design and understanding multidrug recognition and transport.

Supporting Information

S1 Fig. Structural alignment of ABCG5 and ABCG8. Chain A (ABCG5, green) and chain B (ABCG8, blue) from the heterodimer structure (PDBID: 5DO7) are highly similar (RMSD 2.1 Å) except in the linker region.
(TIF)

S2 Fig. Sequence alignments of ABCG subfamily members and CFTR NBD1. (A) Sequences of the ABCG proteins downloaded from the UniProt database (ABCG1_HUMAN, ABCG2_HUMAN, ABCG4_HUMAN, ABCG5_HUMAN, and ABCG8_HUMAN) were aligned employing ClustalW. ABCG2 N-terminus including NBD (a.a. 1–300) and CFTR NBD1 (CFTR_HUMAN, a.a. 381–645) were aligned separately, since CFTR NBD1 sequence has unique features, which result in suboptimal alignments even for demonstration purposes when aligned together with all ABCG proteins. The CFTR NBD1 was merged manually from this pairwise alignment into the multiple alignment. Since the CFTR transmembrane domain is a

Type I exporter fold and not similar in either length or sequence to ABCG proteins, this TMD is not included in the alignment. Important regions and amino acids positions are labeled above the alignment and their numbering refers to ABCG2 positions. Since ABCG2 exhibit somewhat higher similarity to ABCG1 and ABCG4, the alignment of ABCG2 to ABCG5 and ABCG8 is not optimal at a few minor positions (labeled with red arrowheads). **(B)** We also generated pairwise alignments of the ABCG2 to the templates that was used for homology modeling and contains concatenated dimers of ABCG2 (labeled as ABCG22) and ABCG5-ABCG8 (labeled as ABCG58). The loop between the NBD β1 and β2 strands are highly different in length resulting in highly misaligned sequences of this region, thus the β1 sequence of ABCG8 was manually aligned to the β1 sequence of ABCG2. The linker region was not modeled because of low sequence similarities and its flexible missing parts in ABCG5-ABCG8. The ABCG2 sequence exhibits sufficiently high similarity to the templates (identity ~25% and similarity over 40%) that allows generating high quality homology models in the case of these ABC transmembrane proteins (Modeller tutorial and [34]). The alignment was drawn using Jalview, colored according to the ClustalX color scheme, and can be downloaded from http://abcg.hegelab.org. The monomer's boundaries are indicated by red arrowheads, while locations of breaks in the structure are indicated by black arrowheads.
(TIF)

S3 Fig. MD simulations show the stability of the homology model and the mutant constructs. We performed six parallel 50 ns long MD simulations with every constructs embedded in a lipid bilayer. All of them exhibited a stable structure, with sufficiently stabilized RMSD and energy values that can be considered acceptable for such a large and stable system. Major distortions, which were published for MD simulations with crystal structures of other ABC transporters, could not be observed in the case of our structural model [38,39].
(TIFF)

S4 Fig. Important regions and amino acid locations are highlighted on the ABCG2 model. Cysteines (red) were constrained to form intramolecular (C592-C608) or intermolecular (C603-C603) disulfide bonds. The glycosylation site N596 (black) resides in a mobile loop indicated by molecular dynamics simulations. Amino acids between 562 and 586 form two consecutive helices, which are immersed partially into the bilayer and resembles to the P-loop of ion channels, thus we propose to name it as a G-loop (yellow). CRAC motives are labeled blue and the coupling helix light blue. The tyrosines in the potential CRAC motifs (Y413, Y459, Y469, Y570 and Y645, see text) are colored blue. R482 and P480, which proline creates a kink in the helix, are colored ruby and pale green, respectively.
(TIFF)

S5 Fig. The Q141K mutation may cause repulsion with K382 and weaken the NBD/TMD interface. Although the Q141 side chain points towards N158 (Fig 2), the Q141K mutations does not exhibit a pronounced effect of the distances of the helices, in which these amino acids are located (not shown). However, the positively charged 141K interferes with the side chain of K382, which restrains F142 with R383 (see Fig 2). This is revealed by MD simulations showing increased distances and higher variability of distances between the Cα atoms of restudies 141 and 382 in the Q141K variant, as compared to the wild type protein. All the 5,000 frames were analyzed and every consecutive ten distance values were averaged to smooth the plots of all distance measurements in this study.
(TIFF)

S6 Fig. Compounds selected for *in silico* docking. 3D and 2D structures were downloaded from PubChem and ChemSpider. ABCG2 substrates: Acetaminophen-sulfate (CID: 83939),

Acyclovir (CID: 2022), Afatinib (CID: 10184653), Albendazole-sulphoxide (CID: 83969), Arry-334543 (CID: 42642648), Benzo(a)pyrene (CID: 2336), Benzoylphenylurea (CID: 74566), Ciprofloxacin (CID: 2764), D-Luciferin (CID: 5484207), Danusertib (CID: 11442891), Eltrombopag (CID: 9846180), Flavopiridol (CID: 5287969), Icotinib (CID: 22024915), Masitinib (CID: 10074640), Methotrexate (CID: 126941), N-Acetyl-amonafide (CID: 10064887), Pheophorbide A (CID: 5323510), Protoporphyrin IX (CID: 4971), Purpurin18 (CID: 5489047), Rhodamine123 (CID: 65218), Rosuvastatin (CID: 446157), Sulfasalazine (CID: 5359476), Tandutinib (CID: 3038522), Telatinib (CID: 9808844), and Uric-Acid (CID: 1175). Non-substrates: Calcein (CID: 65079), Calcein AM (ChemSpiderID: 346571), Colchicine (CID: 6167), Digoxin (ChemSpiderID: 206532), Doxycycline (CID: 54671203), Fluo-3 (ChemSpiderID: 94730), LTC4 (ChemSpiderID: 4444133), NEM-GS (CID: 443150), Ketoconazole (CID: 456201), Loperamide (CID: 3955), Quinidine (CID: 441074), Reserpine (CID: 5770), Verapamil (CID: 2520), and Vinblastine (ChemSpiderID: 12773).
(TIF)

S7 Fig. Non-substrates are limited in accessing binding Site 2. We performed *in silico* docking calculations for flavopiridol and methotrexate, two established ABCG2 substrates (green labels) as well as verapamil and calcein, which are not transported substrates of ABCG2 (red labels), to the six equilibrated conformation of the transporter. Although we could observe binding to Site 1 (blue) for all compounds, access of verapamil to Site 2 (red), which is the binding pocket including R482, was decreased compared to that of substrates. Calcein also exhibited a limited access to this site, without deep penetration into it. These observations suggest that these ABCG2 conformations can be employed in future studies for developing *in silico* methods to distinguish substrates and non-substrates. Moreover, Site 2 may be the gate that differentiates between toxic molecules and tolerated metabolites. The black box indicates the search space defined in all of our docking calculations. Docking to one of the six equilibrated conformations is shown.
(TIFF)

S8 Fig. Quantitative analysis of sites of a potential translocation pathway. After visual inspection of all the binding poses (20 poses x 6 protein conformations) of falvopiridol and methotrexate, we assigned amino acids semi-manually to binding **Site 1** (Q126, D127, D128, D128, V129, V129, V130, V130, M131, M131, G132, G132, T133, T133, L134, V178, V178, G179, G179, T180, T180, Q181, F182, F182, I183, I183, R191, R191, N387, N387, L388, G390, N391, N391, P392, Q393, Q393, A394, A394, A397, Q398, S443, A444, E446, L447, L447, F448, V449, V449, V450, V450, E451, K452, K453, K453, L454, I456, K473, D477), **Site 2** (L388, A394, A397, Q398, I399, V401, T402, L405, Q437, C438, S440, S441, V442, S443, A444, V445, E446, L447, F448, V450, K473, D477, L478, M481, R482, P485, S486, A517, A520, S521, A524), **Site 3** (A397, V401, F439, F439, S440, V442, V442, S443, S443, V445, V445, E446, E446, L447, V449, V533, V534, V534, S535, S535, V536, A537, T538, T538, L539, L540, M541, M541, T542, I543, F545) and **Site 4** (Y413, I423, Q424, Q424, N425, A427, A427, G428, G553, G553, L554, L554, L555, L555, V556, N557, N557, F578, F586, P602, C603, N604, Y605, A606, T607, L614, Q617, G618, I619, L621). **(A)** Colors red, blue, yellow, magenta, deep purple, orange, and green label Sites 1, 2, 3, and 4, overlap of Sites 1–2, 2–3, and 1–3. Although there is an overlap in the amino acids of sites, a docked pose can be assigned to a binding site based on the numbers of interacting amino acids of two sites with the small molecule. The overlap of binding sites indicates a pathway and a mechanism of transport by binding of a molecule somewhere to the pathway followed by its moving forward to a next site closer to the extracellular space. For example binding to Site 1, which seems to be the most accessible site, can be followed by the movement to Site 2. Site 2 includes R482, which position has an effect on

substrate specificity, and may participate in substrate selection. Non-substrates may not penetrate deeply into this site to induce conformational changes necessary for moving forward on the pathway. Although the initial set of molecules (4 substrates and 2 non-substrates) provided a hint for binding site identification, their low number may not be sufficient for representative results. Therefore we investigated the docking of additional 21 substrates and 12 non-substrates. All poses ((25+14 molecules) x 20 poses x 6 protein conformations) were automatically assigned to binding sites. **(B)** Although a density plot of binding affinities indicates lower binding energies for substrates compared to non-substrates, there are also poses of non-substrates with low binding energies, thus *in silico* values cannot be used to characterize the binding of molecules. Therefore we determined the frequency of substrates and non-substrates in each binding site. **(C)** The number of substrates and non-substrates bound to each binding site was counted and normalized to the number of molecules in the two categories (substrate or not) and to the number of poses. The values are shown in the table. Binding Site 1 was highly populated and more non-substrate poses can be found in this site as compared to substrate poses. Although many poses of non-substrate molecules can be found in Site 2, still this is the site showing increased binding of substrates as compared to non-substrates. These results strongly indicate a distinguished role of Site 2 in substrate selection. When generating sophisticated *in silico* methods for predicting substrates of ABCG2, this Site 2 should be considered with higher weights in model building. Site 3, which is located in the hydrophobic transmembrane region between the two ABCG2 monomers, accommodates both substrates and non-substrates equally well. However, it is important to note that the access to this site might be limited *in vivo* (e.g. by the loop between the β1 and β2 strands of NBD), while *in silico* methods can place any fitting molecule into this volume.
(TIFF)

S9 Fig. The R482G mutation has a pronounced effect on TM helix orientation and dynamics. In spite of the relatively short time periods we can cover in MD simulations, two interesting changes can be observed in this mutant, as compared to the wild type protein. In one of the simulations in chain B the distance between R482 (TH3) and Q398 (TH1) was largely increased in the mutant (top right panel). Moreover, in the other chain and in all of the simulations, the distance between R482 and Q398 increased faster, as compared to the wild type (top left panel). These observations strongly suggest that while these relative movements of TH1 and TH3 are observable in both the wild type and the mutant proteins, the interaction between TH1 and TH3 is more dynamic in the R482G variant as compared to the wild type. In addition, the distance between R482 and A517 significantly shortened in R482G (bottom panels), thus TH3 and TH4 can get into a more intimate contact, because of the lack of a long side chain in TH3. We could not observe significant differences in distances between TH3 and TH2 (middle panels) caused by this mutation.
(TIFF)

S10 Fig. The CRAC mutant Y413S does not affect the dynamics of TH1. In order to characterize the relative orientation and movement of TH1, in which the mutated residue is located, the interaction of amino acids located at the two ends of TH1 were measured throughout the six trajectories for both the wild type and the Y413S mutant ABCG2. C_α distances of I399 (turquoise) and P485 (red) were calculated in addition to that of Y413 (blue) and V556 (magenta). We could observe a larger change only in one Y413S simulations, when the distance between these two residues were altered. However, similar changes could also be observed for a WT simulation. This most likely happens because the Leu-based motif resides in a loop region that may exhibit propensity for higher dynamics.
(TIFF)

S1 Table. The *in silico* method (HMMTOP) [56,57] predicted better location of TM helices compared to experimentally concluded boundaries based on HA-insertion [54,55].
(TIF)

S2 Table. Vina docking scores are better (lower values) for ABCG2 transported substrates, as compared to the values for non-substrates in most of the cases. For clarity only values for the initial set of 4 substrates and 2 non-substrates are listed.
(TIF)

Acknowledgments

We appreciate the help of Ágnes Telbisz, Csilla Özvegy-Laczka, Gergely Szakács, András Váradi, Hedvig Tordai, Boglárka Zámbó and Gergő Gógl in discussing the model. We acknowledge NIIF for awarding us access to HPC resources based in Hungary and the technical help of Ferenc Bartha.

Author Contributions

Conceptualization: BS TH.

Formal analysis: LL TH.

Funding acquisition: BS TH.

Investigation: TH.

Methodology: LL TH.

Project administration: TH.

Resources: TH.

Software: LL TH.

Supervision: TH.

Validation: LL TH.

Visualization: LL TH.

Writing – original draft: BS TH.

Writing – review & editing: LL BS TH.

References

1. ABC Transporters—40 Years on: Springer International Publishing; 2016 2016.
2. Sarkadi B, Homolya L, Szakacs G, Varadi A. Human multidrug resistance ABCB and ABCG transporters: participation in a chemoimmunity defense system. Physiological reviews. 2006; 86(4):1179–236. Epub 2006/10/04. doi: 10.1152/physrev.00037.2005 PMID: 17015488.
3. Kerr ID, Jones PM, George AM. Multidrug efflux pumps: the structures of prokaryotic ATP-binding cassette transporter efflux pumps and implications for our understanding of eukaryotic P-glycoproteins and homologues. The FEBS journal. 2010; 277(3):550–63. Epub 2009/12/08. doi: 10.1111/j.1742-4658.2009.07486.x PMID: 19961540.
4. Rees DC, Johnson E, Lewinson O. ABC transporters: the power to change. Nature reviews Molecular cell biology. 2009; 10(3):218–27. Epub 2009/02/24. doi: 10.1038/nrm2646 PMID: 19234479; PubMed Central PMCID: PMC2830722.
5. Locher KP. Review. Structure and mechanism of ATP-binding cassette transporters. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2009; 364(1514):239–45.

Epub 2008/10/30. doi: 10.1098/rstb.2008.0125 PMID: 18957379; PubMed Central PMCID: PMC2674090.

6. Li J, Jaimes KF, Aller SG. Refined structures of mouse P-glycoprotein. Protein science: a publication of the Protein Society. 2014; 23(1):34–46. Epub 2013/10/25. doi: 10.1002/pro.2387 PMID: 24155053; PubMed Central PMCID: PMC3892297.
7. Corradi V, Vergani P, Tieleman DP. Cystic Fibrosis Transmembrane Conductance Regulator (CFTR): CLOSED AND OPEN STATE CHANNEL MODELS. The Journal of biological chemistry. 2015; 290 (38):22891–906. Epub 2015/08/01. doi: 10.1074/jbc.M115.665125 PMID: 26229102; PubMed Central PMCID: PMC4645605.
8. Ni Z, Bikadi Z, Cai X, Rosenberg MF, Mao Q. Transmembrane helices 1 and 6 of the human breast cancer resistance protein (BCRP/ABCG2): identification of polar residues important for drug transport. American journal of physiology Cell physiology. 2010; 299(5):C1100–9. Epub 2010/08/27. doi: 10.1152/ajpcell.00160.2010 PMID: 20739628; PubMed Central PMCID: PMC2980320.
9. Serohijos AW, Hegedus T, Aleksandrov AA, He L, Cui L, Dokholyan NV, et al. Phenylalanine-508 mediates a cytoplasmic-membrane domain contact in the CFTR 3D structure crucial to assembly and channel function. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(9):3256–61. Epub 2008/02/29. doi: 10.1073/pnas.0800254105 PMID: 18305154; PubMed Central PMCID: PMC2265173.
10. Zhou S, Schuetz JD, Bunting KD, Colapietro AM, Sampath J, Morris JJ, et al. The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. Nature medicine. 2001; 7(9):1028–34. Epub 2001/09/05. doi: 10.1038/nm0901-1028 PMID: 11533706.
11. Gottesman MM, Fojo T, Bates SE. Multidrug resistance in cancer: role of ATP-dependent transporters. Nature reviews Cancer. 2002; 2(1):48–58. Epub 2002/03/21. doi: 10.1038/nrc706 PMID: 11902585.
12. Szakacs G, Varadi A, Ozvegy-Laczka C, Sarkadi B. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). Drug discovery today. 2008; 13(9–10):379–93. Epub 2008/05/13. doi: 10.1016/j.drudis.2007.12.010 PMID: 18468555.
13. Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. Nature reviews Cancer. 2005; 5 (4):275–84. Epub 2005/04/02. doi: 10.1038/nrc1590 PMID: 15803154.
14. Litman T, Brangi M, Hudson E, Fetsch P, Abati A, Ross DD, et al. The multidrug-resistant phenotype associated with overexpression of the new ABC half-transporter, MXR (ABCG2). Journal of cell science. 2000; 113 (Pt 11):2011–21. Epub 2000/05/12. PMID: 10806112.
15. Natarajan K, Xie Y, Baer MR, Ross DD. Role of breast cancer resistance protein (BCRP/ABCG2) in cancer drug resistance. Biochemical pharmacology. 2012; 83(8):1084–103. Epub 2012/01/18. doi: 10.1016/j.bcp.2012.01.002 PMID: 22248732; PubMed Central PMCID: PMC3307098.
16. Dehghan A, Kottgen A, Yang Q, Hwang SJ, Kao WL, Rivadeneira F, et al. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. Lancet. 2008; 372 (9654):1953–61. Epub 2008/10/07. doi: 10.1016/S0140-6736(08)61343-4 PMID: 18834626; PubMed Central PMCID: PMC2803340.
17. Woodward OM, Kottgen A, Coresh J, Boerwinkle E, Guggino WB, Kottgen M. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(25):10338–42. Epub 2009/06/10. doi: 10.1073/pnas.0901249106 PMID: 19506252; PubMed Central PMCID: PMC2700910.
18. Matsuo H, Takada T, Ichida K, Nakamura T, Nakayama A, Ikebuchi Y, et al. Common defects of ABCG2, a high-capacity urate exporter, cause gout: a function-based genetic analysis in a Japanese population. Science translational medicine. 2009; 1(5):5ra11. Epub 2010/04/07. doi: 10.1126/scitranslmed.3000237 PMID: 20368174.
19. Lee JY, Kinch LN, Borek DM, Wang J, Urbatsch IL, Xie XS, et al. Crystal structure of the human sterol transporter ABCG5/ABCG8. Nature. 2016; 533(7604):561–4. Epub 2016/05/05. doi: 10.1038/nature17666 PMID: 27144356.
20. Mizuarai S, Aozasa N, Kotani H. Single nucleotide polymorphisms result in impaired membrane localization and reduced atpase activity in multidrug transporter ABCG2. International journal of cancer. 2004; 109(2):238–46. Epub 2004/01/30. doi: 10.1002/ijc.11669 PMID: 14750175.
21. Morisaki K, Robey RW, Ozvegy-Laczka C, Honjo Y, Polgar O, Steadman K, et al. Single nucleotide polymorphisms modify the transporter activity of ABCG2. Cancer chemotherapy and pharmacology. 2005; 56(2):161–72. Epub 2005/04/20. doi: 10.1007/s00280-004-0931-x PMID: 15838659.
22. Kondo C, Suzuki H, Itoda M, Ozawa S, Sawada J, Kobayashi D, et al. Functional analysis of SNPs variants of BCRP/ABCG2. Pharmaceutical research. 2004; 21(10):1895–903. Epub 2004/11/24. PMID: 15553238.

23. Ozvegy-Laczka C, Koblos G, Sarkadi B, Varadi A. Single amino acid (482) variants of the ABCG2 multidrug transporter: major differences in transport capacity and substrate recognition. Biochimica et biophysica acta. 2005; 1668(1):53–63. Epub 2005/01/27. doi: 10.1016/j.bbamem.2004.11.005 PMID: 15670731.

24. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23(21):2947–8. Epub 2007/09/12. doi: 10.1093/bioinformatics/btm404 PMID: 17846036.

25. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annual review of biophysics and biomolecular structure. 2000; 29:291–325. Epub 2000/08/15. doi: 10.1146/annurev.biophys.29.1.291 PMID: 10940251.

26. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI. OPM: orientations of proteins in membranes database. Bioinformatics. 2006; 22(5):623–5. Epub 2006/01/07. doi: 10.1093/bioinformatics/btk023 PMID: 16397007.

27. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013; 29(7):845–54. Epub 2013/02/15. doi: 10.1093/bioinformatics/btt055 PMID: 23407358; PubMed Central PMCID: PMC3605599.

28. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. Journal of chemical theory and computation. 2012; 8(9):3257–73. doi: 10.1021/ct3004000x PMID: ISI:000308830700029.

29. Wu EL, Cheng X, Jo S, Rui H, Song KC, Davila-Contreras EM, et al. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. Journal of computational chemistry. 2014; 35 (27):1997–2004. Epub 2014/08/19. doi: 10.1002/jcc.23702 PMID: 25130509; PubMed Central PMCID: PMC4165794.

30. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. Journal of chemical theory and computation. 2016; 12(1):405–13. Epub 2015/12/04. doi: 10.1021/acs.jctc.5b00935 PMID: 26631602; PubMed Central PMCID: PMC4712441.

31. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. Journal of computational chemistry. 2011; 32(10):2319–27. Epub 2011/04/19. doi: 10.1002/jcc.21787 PMID: 21500218; PubMed Central PMCID: PMC3144279.

32. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. Journal of cheminformatics. 2011; 3:33. Epub 2011/10/11. doi: 10.1186/1758-2946-3-33 PMID: 21982300; PubMed Central PMCID: PMC3198950.

33. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry. 2009; 30(16):2785–91. Epub 2009/04/29. doi: 10.1002/jcc.21256 PMID: 19399780; PubMed Central PMCID: PMC2760638.

34. Kelly L, Pieper U, Eswar N, Hays FA, Li M, Roe-Zurz Z, et al. A survey of integral alpha-helical membrane proteins. Journal of structural and functional genomics. 2009; 10(4):269–80. Epub 2009/09/18. doi: 10.1007/s10969-009-9069-8 PMID: 19760129; PubMed Central PMCID: PMC2780624.

35. Wen PC, Verhalen B, Wilkens S, McHaourab HS, Tajkhorshid E. On the origin of large flexibility of P-glycoprotein in the inward-facing state. The Journal of biological chemistry. 2013; 288(26):19211–20. Epub 2013/05/10. doi: 10.1074/jbc.M113.450114 PMID: 23658020; PubMed Central PMCID: PMC3696692.

36. Pajeva IK, Globisch C, Wiese M. Comparison of the inward- and outward-open homology models and ligand binding of human P-glycoprotein. The FEBS journal. 2009; 276(23):7016–26. Epub 2009/11/03. doi: 10.1111/j.1742-4658.2009.07415.x PMID: 19878299.

37. Bessadok A, Garcia E, Jacquet H, Martin S, Garrigues A, Loiseau N, et al. Recognition of sulfonylurea receptor (ABCC8/9) ligands by the multidrug resistance transporter P-glycoprotein (ABCB1): functional similarities based on common structural features between two multispecific ABC proteins. The Journal of biological chemistry. 2011; 286(5):3552–69. Epub 2010/11/26. doi: 10.1074/jbc.M110.155200 PMID: 21098040; PubMed Central PMCID: PMC3030360.

38. Gyimesi G, Ramachandran S, Kota P, Dokholyan NV, Sarkadi B, Hegedus T. ATP hydrolysis at one of the two sites in ABC transporters initiates transport related conformational transitions. Biochimica et biophysica acta. 2011; 1808(12):2954–64. Epub 2011/08/16. doi: 10.1016/j.bbamem.2011.07.038 PMID: 21840296.

39. Ivetac A, Sansom MS. Molecular dynamics simulations and membrane protein structure quality. European biophysics journal: EBJ. 2008; 37(4):403–9. Epub 2007/10/26. doi: 10.1007/s00249-007-0225-4 PMID: 17960373.

40. Chang G, Roth CB, Reyes CL, Pornillos O, Chen YJ, Chen AP. Retraction. Science. 2006; 314 (5807):1875. Epub 2006/12/23. doi: 10.1126/science.314.5807.1875b PMID: 17185584.

41. Liu Y, Yang Y, Qi J, Peng H, Zhang JT. Effect of cysteine mutagenesis on the function and disulfide bond formation of human ABCG2. The Journal of pharmacology and experimental therapeutics. 2008; 326(1):33–40. Epub 2008/04/24. doi: 10.1124/jpet.108.138115 PMID: 18430864; PubMed Central PMCID: PMC2632310.

42. Locher KP, Lee AT, Rees DC. The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. Science. 2002; 296(5570):1091–8. Epub 2002/05/11. doi: 10.1126/science.1071142 PMID: 12004122.

43. Hohl M, Briand C, Grutter MG, Seeger MA. Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nature structural & molecular biology. 2012; 19(4):395–402. Epub 2012/03/27. doi: 10.1038/nsmb.2267 PMID: 22447242.

44. Ward A, Reyes CL, Yu J, Roth CB, Chang G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(48):19005–10. Epub 2007/11/21. doi: 10.1073/pnas.0709388104 PMID: 18024585; PubMed Central PMCID: PMC2141898.

45. Csanady L, Chan KW, Nairn AC, Gadsby DC. Functional roles of nonconserved structural segments in CFTR's NH2-terminal nucleotide binding domain. The Journal of general physiology. 2005; 125(1):43–55. Epub 2004/12/15. doi: 10.1085/jgp.200409174 PMID: 15596536; PubMed Central PMCID: PMC2217481.

46. Aleksandrov AA, Kota P, Aleksandrov LA, He L, Jensen T, Cui L, et al. Regulatory insertion removal restores maturation, stability and function of DeltaF508 CFTR. Journal of molecular biology. 2010; 401 (2):194–210. Epub 2010/06/22. doi: 10.1016/j.jmb.2010.06.019 PMID: 20561529; PubMed Central PMCID: PMC4361937.

47. Ambudkar SV, Kim IW, Xia D, Sauna ZE. The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding. FEBS letters. 2006; 580 (4):1049–55. Epub 2006/01/18. doi: 10.1016/j.febslet.2005.12.051 PMID: 16412422.

48. Dawson RJ, Locher KP. Structure of a bacterial multidrug ABC transporter. Nature. 2006; 443 (7108):180–5. Epub 2006/09/01. doi: 10.1038/nature05155 PMID: 16943773.

49. Henriksen U, Fog JU, Litman T, Gether U. Identification of intra- and intermolecular disulfide bridges in the multidrug resistance transporter ABCG2. The Journal of biological chemistry. 2005; 280 (44):36926–34. Epub 2005/08/19. doi: 10.1074/jbc.M502937200 PMID: 16107343.

50. Ozvegy-Laczka C, Laczko R, Hegedus C, Litman T, Varady G, Goda K, et al. Interaction with the 5D3 monoclonal antibody is regulated by intramolecular rearrangements but not by covalent dimer formation of the human ABCG2 multidrug transporter. The Journal of biological chemistry. 2008; 283 (38):26059–70. Epub 2008/07/23. doi: 10.1074/jbc.M803230200 PMID: 18644784; PubMed Central PMCID: PMC3258862.

51. Diop NK, Hrycyna CA. N-Linked glycosylation of the human ABC transporter ABCG2 on asparagine 596 is not essential for expression, transport activity, or trafficking to the plasma membrane. Biochemistry. 2005; 44(14):5420–9. Epub 2005/04/06. doi: 10.1021/bi0479858 PMID: 15807535.

52. Li YF, Polgar O, Okada M, Esser L, Bates SE, Xia D. Towards understanding the mechanism of action of the multidrug resistance-linked half-ABC transporter ABCG2: a molecular modeling study. Journal of molecular graphics & modelling. 2007; 25(6):837–51. Epub 2006/10/10. doi: 10.1016/j.jmgm.2006.08.005 PMID: 17027309.

53. Rutledge RM, Esser L, Ma J, Xia D. Toward understanding the mechanism of action of the yeast multidrug resistance transporter Pdr5p: a molecular modeling study. Journal of structural biology. 2011; 173(2):333–44. Epub 2010/11/03. doi: 10.1016/j.jsb.2010.10.012 PMID: 21034832; PubMed Central PMCID: PMC3026082.

54. Rosenberg MF, Bikadi Z, Chan J, Liu X, Ni Z, Cai X, et al. The human breast cancer resistance protein (BCRP/ABCG2) shows conformational changes with mitoxantrone. Structure. 2010; 18(4):482–93. Epub 2010/04/20. doi: 10.1016/j.str.2010.01.017 PMID: 20399185; PubMed Central PMCID: PMC2858361.

55. Wang H, Lee EW, Cai X, Ni Z, Zhou L, Mao Q. Membrane topology of the human breast cancer resistance protein (BCRP/ABCG2) determined by epitope insertion and immunofluorescence. Biochemistry. 2008; 47(52):13778–87. Epub 2008/12/10. doi: 10.1021/bi801644v PMID: 19063604; PubMed Central PMCID: PMC2649121.

56. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics. 2001; 17(9):849–50. Epub 2001/10/09. PMID: 11590105.

57. Dobson L, Remenyi I, Tusnady GE. CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic acids research. 2015; 43(W1):W408–12. Epub 2015/05/07. doi: 10.1093/nar/gkv451 PMID: 25943549; PubMed Central PMCID: PMC4489262.

58. Ozvegy C, Varadi A, Sarkadi B. Characterization of drug transport, ATP hydrolysis, and nucleotide trapping by the human ABCG2 multidrug transporter. Modulation of substrate specificity by a point mutation. The Journal of biological chemistry. 2002; 277(50):47980–90. Epub 2002/10/11. doi: 10.1074/jbc.M207857200 PMID: 12374800.

59. Hou YX, Li CZ, Palaniyandi K, Magtibay PM, Homolya L, Sarkadi B, et al. Effects of putative catalytic base mutation E211Q on ABCG2-mediated methotrexate transport. Biochemistry. 2009; 48(38):9122–31. Epub 2009/08/21. doi: 10.1021/bi900675v PMID: 19691360; PubMed Central PMCID: PMC2770347.

60. Telbisz A, Ozvegy-Laczka C, Hegedus T, Varadi A, Sarkadi B. Effects of the lipid environment, cholesterol and bile acids on the function of the purified and reconstituted human ABCG2 protein. The Biochemical journal. 2013; 450(2):387–95. Epub 2012/12/05. doi: 10.1042/BJ20121485 PMID: 23205634.

61. Cusatis G, Gregorc V, Li J, Spreafico A, Ingersoll RG, Verweij J, et al. Pharmacogenetics of ABCG2 and adverse reactions to gefitinib. Journal of the National Cancer Institute. 2006; 98(23):1739–42. Epub 2006/12/07. doi: 10.1093/jnci/djj469 PMID: 17148776.

62. Ieiri I. Functional significance of genetic polymorphisms in P-glycoprotein (MDR1, ABCB1) and breast cancer resistance protein (BCRP, ABCG2). Drug metabolism and pharmacokinetics. 2012; 27(1):85–105. Epub 2011/11/30. PMID: 22123128.

63. Noguchi K, Katayama K, Sugimoto Y. Human ABC transporter ABCG2/BCRP expression in chemoresistance: basic and clinical perspectives for molecular cancer therapeutics. Pharmacogenomics and personalized medicine. 2014; 7:53–64. Epub 2014/02/14. doi: 10.2147/PGPM.S38295 PMID: 24523596; PubMed Central PMCID: PMC3921828.

64. Saranko H, Tordai H, Telbisz A, Ozvegy-Laczka C, Erdos G, Sarkadi B, et al. Effects of the gout-causing Q141K polymorphism and a CFTR DeltaF508 mimicking mutation on the processing and stability of the ABCG2 protein. Biochemical and biophysical research communications. 2013; 437(1):140–5. Epub 2013/06/27. doi: 10.1016/j.bbrc.2013.06.054 PMID: 23800412.

65. Woodward OM, Tukaye DN, Cui J, Greenwell P, Constantoulakis LM, Parker BS, et al. Gout-causing Q141K mutation in ABCG2 leads to instability of the nucleotide-binding domain and can be corrected with small molecules. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110(13):5223–8. Epub 2013/03/16. doi: 10.1073/pnas.1214530110 PMID: 23493553; PubMed Central PMCID: PMC3612674.

66. Polgar O, Ediriwickrema LS, Robey RW, Sharma A, Hegde RS, Li Y, et al. Arginine 383 is a crucial residue in ABCG2 biogenesis. Biochimica et biophysica acta. 2009; 1788(7):1434–43. Epub 2009/05/02. doi: 10.1016/j.bbamem.2009.04.016 PMID: 19406100; PubMed Central PMCID: PMC4163909.

67. Honjo Y, Hrycyna CA, Yan QW, Medina-Perez WY, Robey RW, van de Laar A, et al. Acquired mutations in the MXR/BCRP/ABCP gene alter substrate specificity in MXR/BCRP/ABCP-overexpressing cells. Cancer research. 2001; 61(18):6635–9. Epub 2001/09/18. PMID: 11559526.

68. Ni Z, Bikadi Z, Shuster DL, Zhao C, Rosenberg MF, Mao Q. Identification of proline residues in or near the transmembrane helices of the human breast cancer resistance protein (BCRP/ABCG2) that are important for transport activity and substrate specificity. Biochemistry. 2011; 50(37):8057–66. Epub 2011/08/23. doi: 10.1021/bi200573t PMID: 21854076; PubMed Central PMCID: PMC3172388.

69. Telbisz A, Muller M, Ozvegy-Laczka C, Homolya L, Szente L, Varadi A, et al. Membrane cholesterol selectively modulates the activity of the human ABCG2 multidrug transporter. Biochimica et biophysica acta. 2007; 1768(11):2698–713. Epub 2007/07/31. doi: 10.1016/j.bbamem.2007.06.026 PMID: 17662239.

70. Storch CH, Ehehalt R, Haefeli WE, Weiss J. Localization of the human breast cancer resistance protein (BCRP/ABCG2) in lipid rafts/caveolae and modulation of its activity by cholesterol in vitro. The Journal of pharmacology and experimental therapeutics. 2007; 323(1):257–64. Epub 2007/07/27. doi: 10.1124/jpet.107.122994 PMID: 17652262.

71. Velamakanni S, Janvilisri T, Shahi S, van Veen HW. A functional steroid-binding element in an ATP-binding cassette multidrug transporter. Molecular pharmacology. 2008; 73(1):12–7. Epub 2007/12/21. doi: 10.1124/mol.108.038299 PMID: 18094074.

72. Gal Z, Hegedus C, Szakacs G, Varadi A, Sarkadi B, Ozvegy-Laczka C. Mutations of the central tyrosines of putative cholesterol recognition amino acid consensus (CRAC) sequences modify folding,

activity, and sterol-sensing of the human ABCG2 multidrug transporter. Biochimica et biophysica acta. 2015; 1848(2):477–87. Epub 2014/12/03. doi: 10.1016/j.bbamem.2014.11.006 PMID: 25445676.

73. Hegedus C, Telbisz A, Hegedus T, Sarkadi B, Ozvegy-Laczka C. Lipid regulation of the ABCB1 and ABCG2 multidrug transporters. Advances in cancer research. 2015; 125:97–137. Epub 2015/02/03. doi: 10.1016/bs.acr.2014.10.004 PMID: 25640268.

Cellular and Molecular Life Sciences
https://doi.org/10.1007/s00018-020-03651-3

Cellular and Molecular Life Sciences

ORIGINAL ARTICLE

The transport pathway in the ABCG2 protein and its regulation revealed by molecular dynamics simulations

Tamás Nagy[1] · Ágota Tóth[1] · Ágnes Telbisz[2] · Balázs Sarkadi[1,2] · Hedvig Tordai[1] · Attila Tordai[3] · Tamás Hegedűs[1]

Received: 26 May 2020 / Revised: 1 September 2020 / Accepted: 15 September 2020

Abstract

Atomic-level structural insight on the human ABCG2 membrane protein, a pharmacologically important transporter, has been recently revealed by several key papers. In spite of the wealth of structural data, the pathway of transmembrane movement for the large variety of structurally different ABCG2 substrates and the physiological lipid regulation of the transporter has not been elucidated. The complex molecular dynamics simulations presented here may provide a breakthrough in understanding the steps of the substrate transport process and its regulation by cholesterol. Our analysis revealed drug binding cavities other than the central binding site and delineated a putative dynamic transport pathway for substrates with variable structures. We found that membrane cholesterol accelerated drug transport by promoting the closure of cytoplasmic protein regions. Since ABCG2 is present in all major biological barriers and drug-metabolizing organs, influences the pharmacokinetics of numerous clinically applied drugs, and plays a key role in uric acid extrusion, this information may significantly promote a reliable prediction of clinically important substrate characteristics and drug-drug interactions.

Keywords ABCG2 · Multidrug transport · Molecular dynamics · Cholesterol regulation

Introduction

ABCG2 mediated membrane transport is an important mechanism for the elimination of several toxic substrates from the cell. Localization of this important ABC transporter in all-important tissue barriers (apical surface of hepatocytes, renal cells, blood–brain barrier etc.) and stem cells renders ABCG2 a key factor in drug resistance and uric acid elimination. ABCG2 is also crucial in the pharmacokinetics of several compounds [1–3] which is highlighted by the fact that the US Food and Drug Administration and the European Medicines Agency list ABCG2 among the transporters to be investigated for pharmacokinetics and drug-drug interactions [4, 5]. Experimental data showed that ABCG2 can handle substrates with a wide variety in size and polarity [2, 6, 7]. Uric acid is a small and amphiphilic physiological substrate [8, 9], Hoechts 33342 [10] is a large hydrophobic molecule, and methotrexate [11] or various sulfated and glucuronide conjugates [12] are more polar substrates. Several large molecules, such as Ko143 or elacridar can inhibit substrate transport by a putative competitive inhibition mechanism [13, 14]. A conformation sensitive antibody (5D3), which binds to the extracellular surface of the ABCG2 dimer, also influences the transport cycle and inhibits ATPase activity. Interestingly, this antibody exhibits an increased binding to an inhibitor-bound state of ABCG2 [14–16]. ABCG2 mediated transport is significantly modulated by the lipid composition of the membrane environment [17–19]. ABCG2 transport activity requires cholesterol, but it is not known whether this lipid exerts its effects through affecting the membrane bilayer structure, by binding specifically to the ABCG2 protein as an allosteric modulator, or by both mechanisms [20].

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00018-020-03651-3) contains supplementary material, which is available to authorized users.

✉ Tamás Hegedűs
hegedus@hegelab.org

[1] Department of Biophysics and Radiation Biology, Semmelweis University, Tuzolto u. 37-47, 1094 Budapest, Hungary

[2] Institute of Enzymology, Research Centre for Natural Sciences, Magyar Tudosok krt. 2, 1117 Budapest, Hungary

[3] Department of Transfusion Medicine, Semmelweis University, Nagyvarad ter 4, 1089 Budapest, Hungary

Published online: 26 September 2020

The ABCG2 protein includes an N-terminal intracellular nucleotide-binding domain and a C-terminal transmembrane domain built from six TM helices. The functional form of this half transporter is a homodimer [21, 22]. Cryo-electron microscopy (cryo-EM) studies revealed that the TM helices in ABCG2 are short, bringing the nucleotide-binding domains (NBDs) in close proximity to the intracellular leaflet of the membrane bilayer (Fig. S1). Therefore, the structure of ABCG2 has been classified as a type II exporter fold [18, 23, 24]. The NBD and TMD are connected by a so-called linker region, which involves a V-shaped α-helical region surrounded by highly dynamic segments not resolved in any of the cryo-EM structures [25, 26]. The amphipathic elbow helix, called connecting helix or TH1a in ABCG proteins and the short intracellular loops called coupling helices, play an important role in coupling the conformational changes between the NBDs and TMDs. The large extracellular loop (EL3), which contain intra- and intermolecular disulfide bridges, has been demonstrated to interact with all other extracellular loops and the reentrant, creek-forming G-loop (Fig. S1) [27]. The potential role of the unresolved, highly dynamic segments is currently unknown.

The EL3 loops create a lid-like structure contributing to an exit site to the extracellular space. This lid covers the extracellular ends of TH5 helices involving a so-called Leu-valve (residues 554 and 555) site, which is the boundary of the main drug-binding pocket in the dimer between the TH2, TH2′, TH5, and TH5′. In a cryo-EM study, the estron-3-sulfate substrate was localized in this central, hydrophobic, and cytoplasm-facing cavity [23]. This study demonstrated that N436 and F439 were important residues for substrate recognition and transport. Another cryo-EM study revealed the localization and interaction of MZ29, a Ko143 derivative inhibitor with ABCG2 [18]. The inhibitory molecule was found to fill the central cavity with high affinity and thereby trapping the conformation with separated NBDs. Interestingly, the TMDs were closed in a newly published cryo-EM ABCG2 structure in the absence of ATP, and substrates were shown to open the translocation pathway [28]. Based on an ABCG2 homology model, recently we identified the exit site (Site 4 or cavity 2) and the central binding pocket (Site 3 or cavity 1) using in silico docking of substrates and non-substrates [29]. Moreover, we were able to detect two additional pockets towards the intracellular space. We proposed that Site 2, which involves the position R482, influencing substrate specificity, maybe the site responsible for substrate selection. Site 1 equally interacted with all docked molecules and it was expected to form the entry site of the translocation pathway.

While the highly valuable experimental structures are the fundamentals of structural studies, they inherently lack information on dynamics. Therefore, molecular dynamics simulations have increasingly been used to dissect the effects of mutations on protein structure, expression, and function, as well as to characterize transport mechanism [25, 27, 30]. However, all experimental and the majority of in silico studies focused on a smaller part of the translocation pathway and molecular dynamics simulations with ABCG2 have not involved a substrate molecule. Therefore, in the present study, we performed MD simulations with ABCG2 embedded in a lipid bilayer with the additional presence of a physiological substrate, uric acid. We aimed to characterize the full translocation pathway, highlighting the dynamic alteration of the central drug-binding pocket and the important role of access routes to this binding site.

Methods

Structural models

Human ABCG2 structures were used in our simulations in the absence of ATP (PDB IDs: 6HIJ and 6HCO) [18, 24] and in an ATP-bound conformation (PDB ID: 6HZM) [23]. The latter structure was determined in the presence of E211Q mutation to facilitate structure determination by lowering the rate of ATP hydrolysis. We reverted this mutation back to wild type using VMD [31]. Non-protein molecules (e.g. inhibitors and water) were removed. All the structures were trimmed similarly to remove the resolved parts of the linker region and are available for download (https://www.hegelab.org/resources.html).

Molecular dynamics simulations

Classical molecular dynamics (MD) simulations were used to investigate the effect of cholesterol on ABCG2 dynamics and to characterize the intracellular parts of the substrate translocation pathway. GROMACS 2018 with the CHARMM36m force field was used to run molecular dynamics simulations [32, 33]. Simulations are summarized in Table S1.

Most of the simulation systems were prepared using CHARMM-GUI [34, 35]. First, the structural models were oriented according to the OPM (Orientations of Proteins in Membranes) database [36]. Then all N- and C-termini were patched with ACE (acetyl) and CT3 (N-Methylamide) groups, respectively, and disulfide bridges were set between C592-C608 in chain A, C592′-C608′ in chain B, and C603-C603′ between chains A and B. The 6HIJ structure with associated cholesterol and DOPE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine) molecules were inserted into a POPC bilayer (1-palmitoyl-2-oleoyl-sn- glycero-3-phosphocholine, $n = 232$). The 6HCO structure was inserted into a pure POPC bilayer ($n = 228$), 1:1 POPC:cholesterol ($n = 144$, $m = 144$), and 1:1 POPC:sitosterol ($n = 144$,

$m = 144$). KCl at a concentration of 150 mM was used. Grid information for PME (Particle-Mesh Ewald) electrostatics was generated automatically, and NPT ensemble was selected with a constant number of particles (N), pressure (P) of 1 bar, and temperature of 310 K. Uric acid molecules were inserted into the system by replacing water molecules, using GROMACS tools (e.g. gmx insert-molecules). Parameters for uric acid were generated by the CHARMM general force field (CGenFF) at CHARMM-GUI. In the case of simulations with the ATP-bound 6HZM structure, the protein was inserted into a 1:1 POPC:cholesterol bilayer ($n = 163$, $m = 163$), similarly to simulations with inward-facing ABCG2 structures.

Each system was energy minimized using the steepest descent integrator, which stopped when the largest force in the system becomes less than 500 kJ/mol/nm. To increase sampling, several simulations were forked using the energy minimized system, with different velocities. Six equilibration steps with decreasing position restraints were performed. The corresponding parameter (mdp) files are also available for download. Nose–Hoover thermostat and Parrinello-Rahman barostat with semi-isotropic coupling was employed in the production run. Time constants for the thermostat and the barostat were set to 1 picosecond and 5 picosecond, respectively. Electrostatic interactions were calculated using the fast smooth PME algorithm and LINCS algorithm was used to constrain bonds.

Targeted molecular dynamics

We used the MOVINGRESTRAINT procedure implemented in PLUMED for transforming the inward-facing apo conformation to an inward-closed, ATP- and substrate-bound conformation. Uric acid was docked to the central binding pocket of the apo structure (PDB ID: 6HCO) using Autodock Vina [37, 38]. Two ATP molecules were placed onto the nucleotide-binding domains in the same conformation as in the ATP-bound structure (PDB ID: 6HZM). To achieve this, isolated Mg-ATP·NBD complexes from the 6HZM structure were aligned to the full-length 6HCO structure and the two Mg-ATP molecules and the apo ABCG2 structure were merged. The Cα and Mg-ATP coordinates were extracted from the ATP-bound 6HZM structure and were used for targeting. The reaction coordinate (collective variable, CV) was the root mean square deviation (RMSD) between the 6HCO and the target 6HZM structures, calculated after an optimal alignment (TYPE = OPTIMAL) of the evolving 6HCO and the static target structures. In the first 5 ns of the simulation, the force constant (κ) was increased from zero to 40,000 kJ/mol/nm, by which the closure of such a large system was achievable as determined by testing several values. In the second 5 ns of the simulations, the force constant was decreased to zero. After this simulation, performed in POPC bilayer, the protein with ATP and uric acid was inserted into a POPC:cholesterol membrane, equilibrated, and used in equilibrium and metadynamics simulations.

Metadynamics

The translocation through the Leu-valve was accelerated by performing well-tempered metadynamics simulations [39], using GROMACS 2018 in combination with PLUMED [40]. A distance-based collective variable (CV) was biased during metadynamics simulations, namely the distance between the center of masses (COM) of uric acid and four protein Cα atoms (residues 439 and 542) in TH2, TH2′, TH5, and TH5′. To inhibit the backward movement of the substrate towards the intracellular space, a lower wall was defined between the protein COM and the initial position of uric acid (5 Å around protein COM). An upper wall at 38 Å around protein COM was also defined to prevent the distancing of uric acid from the protein. A grid between 0 Å and 40 Å was created and used to speed up calculations. Force at the walls was set to 2,000 kJ/mol/nm. For the biased CV the Gaussian height and sigma were set to 0.6 kJ/mol and 0.06 nm, respectively. The most appropriate width value was calculated from equilibration simulations as the half of the standard deviation of the distance CV. Gaussians were deposited every picosecond. A bias factor of 10 and a temperature of 310 K were set. The same GROMACS options were used as in equilibrium simulations. The time point when uric acid passed the Leu-valve was determined by visual inspection of the six trajectories. PLUMED tools were used for integrating energy (summing the hills).

Characterization of bilayer properties

100 frames were extracted from each trajectory for analysis. GridMAT-MD [41] was used to calculate bilayer thickness and area per lipid values. The grid was set to 200 and other parameters were the defaults. The following atoms located in the membrane interface at similar z coordinates were set as *atomnames*$_i$: N of POPC, O3 of cholesterol, and O3 of sitosterol. An example parameter file is included for download (https://www.hegelab.org/resources.html).

Pocket detection

Every 10th frame of simulations with the apo conformation (PDB ID: 6HCO) in POPC:cholesterol bilayer was submitted for mdpocket software [42]. The pockets, which were present at least in 75% of the frames, were extracted from the mdpout_freq_grid.dx grid file, using the extractISOPdb.py script included in mdpocket package.

Analysis and molecular visualization

Simulations were analyzed by the MDAnalysis Python package [43] and in-house Python scripts (e.g. extraction of cholesterol positions and TH ends). Plots were made by matplotlib Python package [44]. Visualization of structures was performed by PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). Specifically, the mobility of cholesterol molecules was characterized by the x and y coordinates of the O3 atom, extracted from the trajectories. Cholesterol molecules located in the extracellular and intracellular membrane leaflets were distinguished by clustering into two groups using the z coordinate of the O3 atom of cholesterol.

Results

ABCG2 bound cholesterol molecules promote the closure of TM2 and TM2′

Several recent ABCG2 structures exist in the RCSB database representing different conformations. Associated lipids (cholesterol and DOPE, 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine) were also resolved in the case of one cryo-EM map (EMD-4256, PDBID: 6HIJ, Fig. 1) [18]. There is no obvious explanation, why this particular map exhibits immobile lipids molecules since other ABCG2 structures (PDB IDs: 6HCO, 6ETI, 6FEQ, and 6FFC) have also been determined in the same or similar lipid environments, in the absence of nucleotides and the presence of various inhibitors [18, 23]. To gain insight into the interaction between the tightly bound lipid molecules and the ABCG2 protein, we performed equilibrium molecular dynamics simulations using the 6HIJ protein structure. ABCG2 and the associated lipid molecules were inserted into a POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) bilayer. The

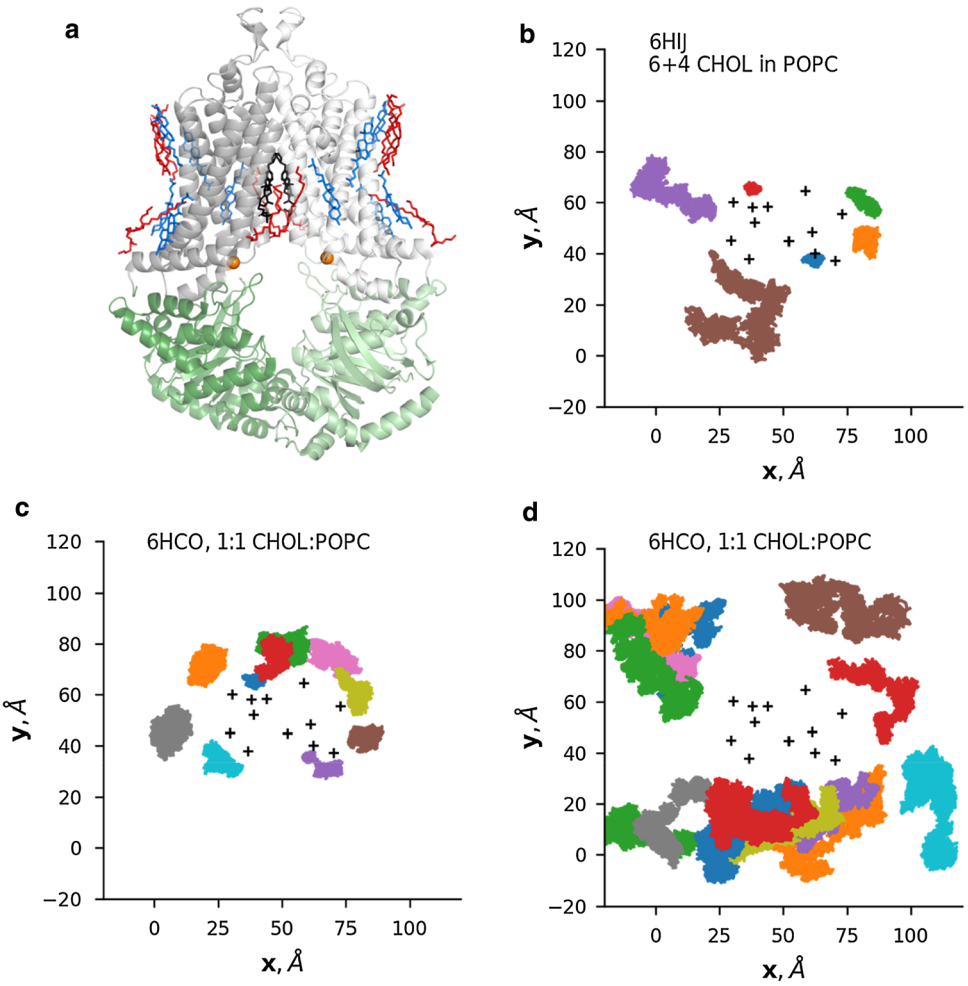


Fig. 1 The movement of ABCG2 bound cholesterol molecules is spatially confined. **a** The structure of ABCG2 and associated molecules (PDB ID: 6HIJ). Gray: TM domains, green: nucleotide-binding domains, blue: cholesterol, red: 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine, black: inhibitor MZ29, orange spheres: Cα atoms of V450 and V450′. **b** The movement of cholesterol molecules in the 6HIJ structure, located in the intracellular membrane leaflet was projected onto the x/y plane. The location of representative cholesterol molecules with more (**c**) and less (**d**) confined motion in a mixed bilayer (1:1 CHOL:POPC) is shown. Black cross (+) signs: position of TM helices in the starting structure

movement of the 6 and 4 cholesterol molecules located in the intracellular and extracellular membrane leaflets, respectively, was monitored throughout the 0.5 µs long trajectories ($n = 3$). Projection of their location onto the x/y plane (Fig. 1, Fig. S2) revealed a tight but dynamic interaction. A fraction of the cholesterol molecules remained protein-bound throughout the simulation, while other cholesterol molecules exhibited increased mobility.

Next, we investigated the effect of cholesterol on ABCG2 dynamics. In vivo cholesterol levels of membranes are highly dependent on cell type and organelle and vary between ~ 20 and 50% [45]. To saturate ABCG2 cholesterol sites in our simulations and to avoid limitations set by diffusion in all-atom simulations, we inserted the inward-facing ABCG2 structure (PDBID: 6HCO) into 50–50% POPC-cholesterol, 50–50% POPC-sitosterol, or 100% POPC bilayers using CHARMM-GUI [34, 46, 47]. Sitosterol was used as a control, since this sterol has a different effect on ABCG2 function compared to cholesterol [19, 20]. The initial position of sterol molecules was random, the lipids were evenly distributed, and some of the cholesterol molecules interacting with ABCG2 exhibited restricted motions as in the case of simulations with the 6HIJ system (Fig. 1, Fig. S2). Importantly, both sterols had similar effects on bilayer properties, such as thickness and lipid density (Fig. S3).

We characterized the mobility of TM helices, which were expected to exhibit altered movements in the presence of cholesterol, in bilayers of different lipid compositions. The position of Cα atoms of amino acids at the boundaries of TM helices (Table S2) was extracted and 2D histograms were calculated and plotted from the x, y components, separately for the distribution of intracellular and extracellular ends. Interestingly, a decreased size of the area visited by the intracellular ends of the inward-facing structure was observed in some of the simulations with cholesterol or sitosterol compared to simulations with a bilayer containing only POPC (Fig. 2, Fig. S4). While in some mixed-lipid simulations the movements of these ends were not so restricted, the bottom part of the helices moved towards the central axis in the presence of cholesterol. Sitosterol also exhibited a restricting effect in one of the simulations, but the ABCG2 TM helices did not move towards the center in a manner observed for

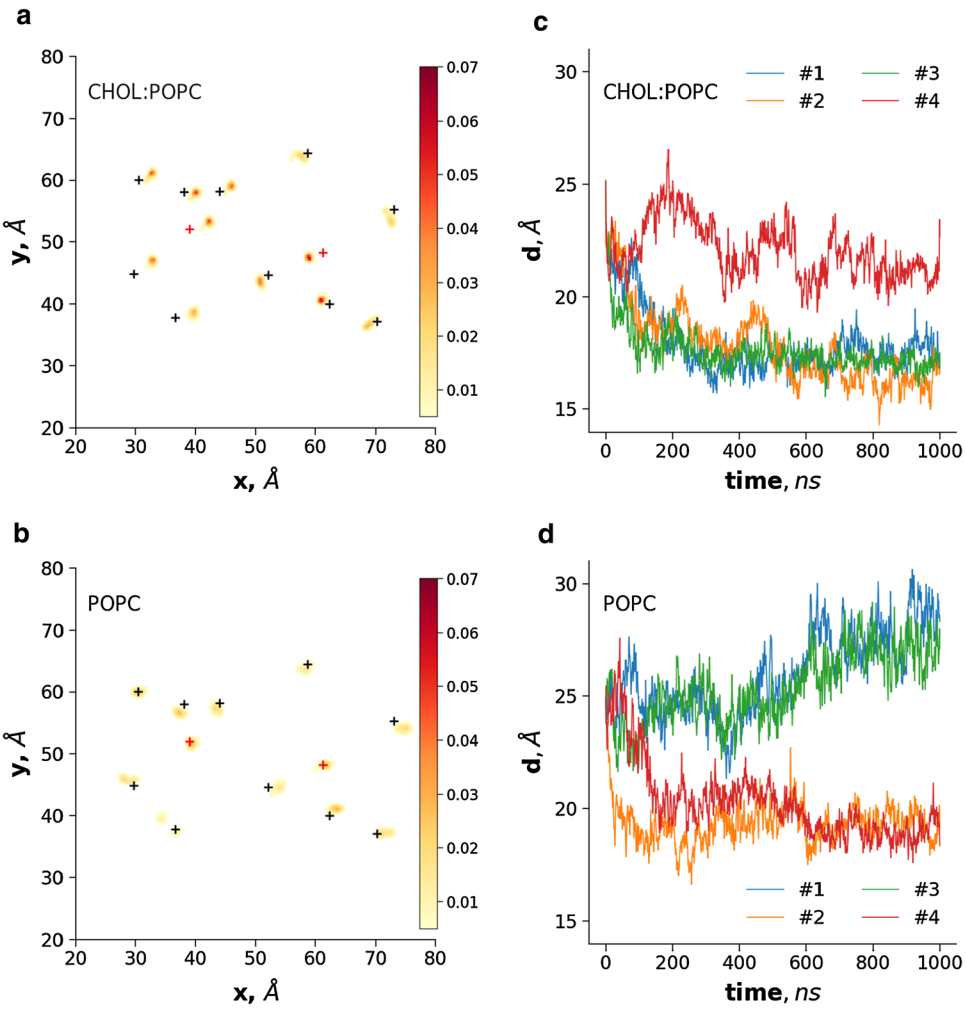


Fig. 2 Cholesterol promotes the closure of the intracellular ends of transmembrane helices. Simulations were performed with the inward-facing 6HCO structure in 1:1 cholesterol:POPC mixed (**a**) and POPC-only membrane bilayer (**b**). The x, y coordinates of the Cα atom of the first and last amino acids in each TM helices were recorded in frames after 100 ns. The x and y values were binned and their frequency was plotted in 2D histograms. One out of four simulations is shown for each bilayer. Black crosses: the position of the Cα atoms in the initial structure, red crosses: transmembrane helices 2 and 2′. **c**, **d** The distance between the center of mass of V450 and V450′ Cα atoms, located in the TM2 helices, were plotted

cholesterol. The most specific effect of cholesterol was the closure of TM2 and TM2′, characterized by the decreased distance between the intracellular ends of these helices (Cα atoms of V450 from both protomers, Fig. 2). The TM2 movement showed alterations when ABCG2 was placed in pure POPC bilayer. TM2 and TM2′ exhibited either no closure at all or their ends approached each other to a smaller extent than in simulations in the presence of cholesterol. This distance of the intracellular TM helix ends also resulted in fewer interactions between the two NBDs (Fig. S5). It is important to note that POPC tended to infiltrate the TM helices and to interfere with helix closure. The open NBDs of ABCG2 exhibited rigid body motions in these simulations with the apo ABCG2 structure (Fig. S6), and this phenomenon was also observed in simulations with MDR1-like structures in the absence of ATP [48–50].

The above-described cholesterol effects can be observed for the intracellular but not for the extracellular helix ends (Fig. S4), where transmembrane helices are tightly packed. The differences vanish in the case of the inward-closed structure with transmembrane helices tightly packed at both sides (Fig. S4 and Fig. S6).

Drug access and binding pocket volumes are affected by dynamic fluctuations

To test the effect of conformational fluctuations on the internal cavities sufficient for drug binding, pockets were identified through trajectories using the MDpocket software [42]. Identified cavities present with a frequency of 0.75 or more were analyzed and are shown in the context of the initial structure (Fig. 3a–c). In two simulations a large cavity involving Site 2, Site 2′, and Site 3 was observed. This complex cavity was exposed both to the membrane bilayer (Fig. 3b) and towards the cytosol (Fig. 3c). However, the cytosolic TM2 closure prevented the lateral opening (Fig. 3d) and also limited the access to the central cavity (Site 3) from the cytosol (Fig. 3e). In some of the parallel simulations exhibiting a similar degree of closure, the slightly different conformation of TM helices resulted in smaller drug binding cavities (Fig. 3f).

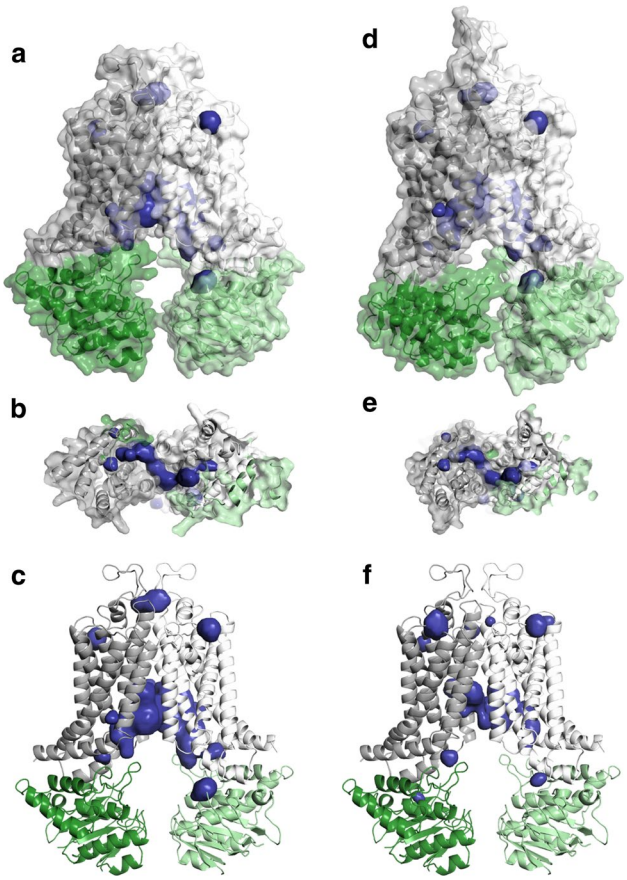


Fig. 3 ABCG2 drug binding cavities in MD simulations correspond to pockets identified by in silico docking. **a–c** Binding pockets determined in an MD trajectory are shown in the context of the starting structure. **d**, **e** The closure of the NBD and TM helices limits access to the translocation pathway. **f** In some of the simulations, spaces forming a translocation pathway are less frequent and less continuous. Gray and white: transmembrane domains, green and light green: nucleotide-binding domains, blue: pockets

Equilibrium simulations reveal substrate engagement along the translocation pathway

To discover the details of substrate translocation, we performed equilibrium simulations with the inward-facing ABCG2 structure (PDBID: 6HCO) embedded in a POPC/cholesterol bilayer in the presence of a substrate. We used uric acid, which is a physiological substance transported by ABCG2 [8, 9]. Since our simulation time scale was relatively short (0.5–1 μs), we increased the possibility of substrate binding and translocation by incorporating 30 uric acid molecules in the simulations (Fig. 4).

During one of the 0.5 μs long simulations, the distance between the cytoplasmic sides of the TM helices decreased to an extent that abolished the substrate entry point. During another 0.5 μs long simulation, a single uric acid molecule out of 30 entered Site 2′ (a pocket between TM1, TM2, and TM3 in chain B) from the large space between the partially separated NBDs via an entry point surrounded by amino acids N391′, Q393′, E446′, and S535 (Fig. 4). The substrate interacted with amino acids Q398′, S440′, S443′, R482′, and L539 with the highest frequency in Site 2′ and moved forward, towards the extracellular space to Site 3 between TM2, TM2′, TM5, and TM5′. The most

dc_2020_22

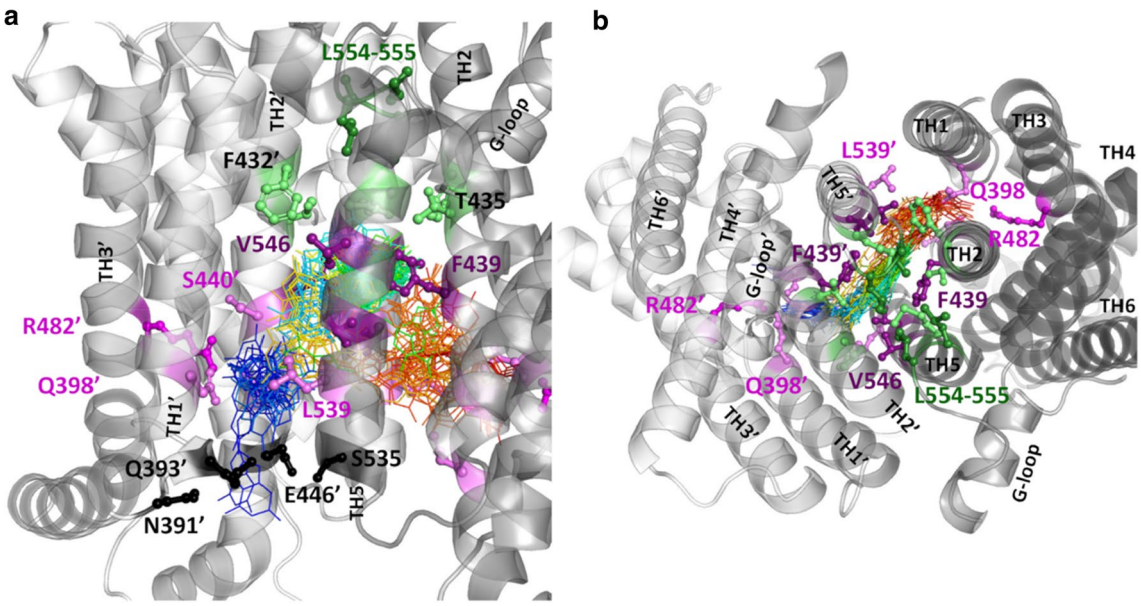


Fig. 4 The movements of the substrate, uric acid along the translocation pathway. In an equilibrium simulation with the inward-facing ABCG2 structure (PDBID: 6HCO) in the presence of uric acid molecules, a single uric acid molecule entered Site 2′ (a pocket between TM1, TM2, and TM3 in chain B, magenta and violet sticks) from the cytosolic region (black sticks). Uric acid moved forward from Site 2′ (R482′) to Site 3 between TM2, TM2′, TM5, and TM5′, finally to Site 2 at the opposite side in chain A. Line colors of blue-green-yellow–red correspond to uric acid and encode increasing simulation time points; deep purple: N436, F439, T542, and V546; light green: F432, T435, T538, and M549; green: Leu-valve, gray: transmembrane helices; **a** side view; **b** top view from the extracellular space

frequent substrate interaction partners included amino acids F439, T542, and V546 from both protomers and also N436 with a lower frequency. The uric acid molecule also moved further towards the extracellular space interacting with F432, T435, and M549 from both transporter halves, but did not reach the Leu-valve (residues 554 and 555). In the final phase, uric acid changed its moving direction towards the intracellular space and reached Site 2 in chain A at the opposite side of its entry point. The most frequent interactions (Q398, S440, S443, R482, and L539′) corresponded to the same residues observed at the opposite Site 2'.

The critical translocation step through the Leu-valve is characterized by metadynamics

While the backward movement of uric acid towards the intracellular space was partly inhibited by the closure of the intracellular TM helix ends, the substrate did not move into the extracellular space. This was expected since the inward-facing conformation after closing motions is still more open compared to known ATP-bound conformations. Therefore, we placed two Mg-ATP molecules to the nucleotide-binding domains and docked a single uric acid molecule into Site 3, and used a targeted MD protocol for "morphing" this inward-facing conformation with ligands into the inward-closed, 6HZM-like conformation (Fig. S7). The protocol was driven by minimizing the difference (RMSD) between the simulated 6HCO complex and the static, inward-closed conformation. This ATP- and substrate-bound model was used in 1 μs long simulations to test the exit of uric acid. During one of the simulations, the substrate moved towards the intracellular space and escaped in the direction of the central axis between TM3, TM3′, TM5, and TM5′ (not shown). During two additional simulations, the substrate moved further towards the extracellular space and on a single occasion, it also moved down to Site 2, as in the bottom open conformation discussed above (Fig. 4). Since we were unable to observe the final step of substrate transport to the extracellular space, we set up accelerated simulations using metadynamics.

Since convergence was not achieved in longer metadynamics simulations due to the inability of the substrate to find the path back to the central pocket, we were unable to calculate the free energy surface of the substrate export. Therefore, we aimed to perform several short simulations to the point when uric acid passed the Leu-valve and to calculate the energy barrier associated with each trajectory. We ran six metadynamics simulations with a reaction coordinate (collection variable, CV) defined by the distance between the center of mass of Cα of F439, F439′, T542, and T542′ and the center of mass of the uric acid molecule (Fig. 5 and

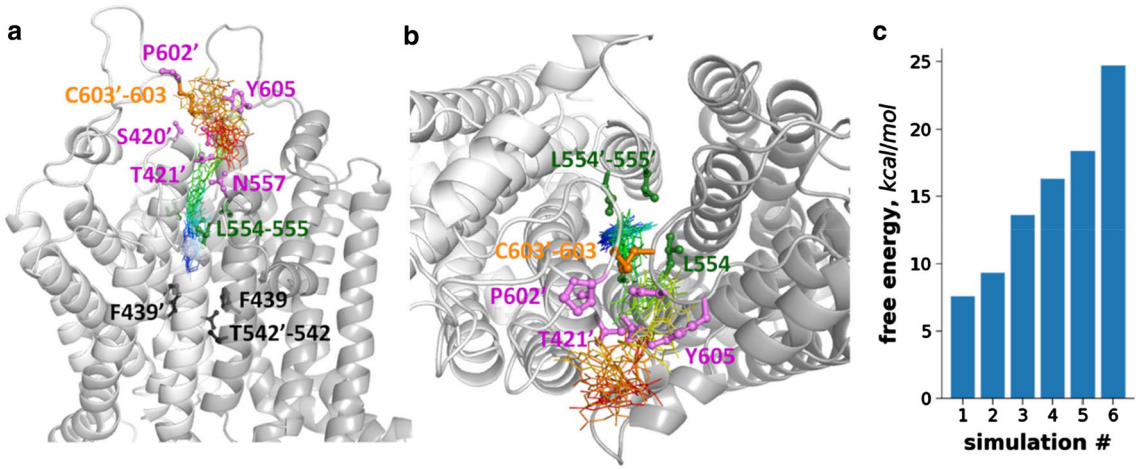


Fig. 5 The transit through the Leu-valve as characterized by metadynamics. **a**, **b** A 6HZM-like structure generated by a targeted MD simulation with two ATP molecules and a uric acid in Site 3 was subjected to metadynamics simulations. The resultant conformation was used in six metadynamics simulations with a reaction coordinate defined by the distance between the center of mass of Cα atoms of four amino acids (F439, T542 from both protomers, black sticks) and the center of mass of the uric acid molecule. Blue-green-yellow–red lines represent uric acid and correspond to increasing simulation time points. Uric acid passed through the Leu-valve (green sticks) in all simulations. Amino acids in the extracellular gate are labeled with violet. **c** Biasing energy accumulated until the time point, when the uric acid molecule passed the Leu-valve

Fig. S8). During metadynamics simulations, the CV was calculated at regular intervals (here, every picosecond) and a positive Gaussian potential is added to the energy at the current position in the energy landscape of the system. This increasing potential in time and along the CV discourage uric acid to stay at the same location, thus the small-molecule substrate was expected to move to other positions. We set a lower wall of 5 Å for this distance CV to inhibit exit towards the intracellular space. We also set an upper wall of 38 Å to restrict the substrate in a volume close to the protein that permits the sampling of the CV space between 5 Å and 38 Å. During all of the simulations, the uric acid molecule passed the Leu-valve and then left the intra-protein space (Site 4) after some lateral movements, since the space above the Leu-valve was limited by the lid formed by the extracellular loops. The substrate interacted with residues N601′, P602′ from one protomer and residues N604, Y605 from the other. The amino acids around the exit included 556–559, 616–618, and the extracellular end of the opposite TM2 (residues T421 and N425). To gain insights into the energetics of the substrate passage, we summed the hills of the bias up to the time point when the uric acid molecule passed the Leu-valve (Fig. 5c). In three out of six simulations the free energy input for passing the leucine amino acids was between 7–13 kcal/mole, which is comparable to the energy liberated in association to the hydrolysis of an ATP molecule. It is important to note that we could not consider an important degree of freedom, the conformation freedom of the protein, for biasing in metadynamics that results in large differences in the height of the energy barriers.

Discussion

Earlier, we described the effect of mutations (e.g. Q141K and R482G) on ABCG2 dynamics and identified potential hot spots along a drug translocation pathway using a homology model [29]. Since there were only small differences between our homology model and the first published experimental ABCG2 structure [24], most of our results and conclusions are still valid. However, several important questions remained unanswered and motivated our current work. The most challenging issue was how to reconcile experimental data on cholesterol binding motifs and a static structural model, as we did not perform simulations in the presence of cholesterol in our previous work. Here, we performed simulations in the presence of sitosterol and cholesterol, which sterols promoted an increased ABCG2 transport function in wet experiments to different levels. We observed an increasing probability and extent of closure of the cytoplasmic ends of TM helices as a consequence of lipid environment changes. In pure POPC, the distance between TH2 and TH2′ decreased to a value close to 20 Å in two out of four simulations (Fig. 2). In the presence of sitosterol, the closure could be observed to a similar level as in pure POPC bilayers (Fig. S4). However, in the absence of cholesterol, lipids tended to infiltrate the helices compared to simulations in the presence of cholesterol. During some of the simulations, both sterols decreased the lateral movement of TM helices as expected as a consequence of altered bilayer properties, such as the increased order of lipid chains (Fig. S3) [51–53]. Our

results strongly suggest that cholesterol limits the conformational space of ABCG2 and facilitate the closure of the TM helices. In addition, the effect of cholesterol is likely not to be exclusively attributed to the physical changes of the bilayer, since both sitosterol and cholesterol had similar general effects on membrane properties in our simulations (Fig. S3). Ordered cholesterol molecules at specific sites were observed in a cryo-EM density (PDB ID: 6HIJ), and these lipids remained bound also for a longer period at those sites during MD simulations. Thus further studies are required to elucidate the effects of lipids on ABCG2, similar to the in silico investigations performed by Sansom et al. [54–56] for other membrane proteins.

Regarding the substrate translocation pathway within the ABCG2 protein, the present MD simulations reinforce the presence of a pathway from Site 1 to Site 4, described in our former in silico docking study. The existence of the central Site 3 in this pathway has been confirmed by several cryo-EM structures with the bound substrate and inhibitor molecules. In addition to structural studies, amino acids indispensable for substrate binding and transport have been identified in this site, using various biochemical methods [23, 57]. Surprisingly, Gose *et al.* highlighted F439 as a single residue necessary for engaging transport [57]. The transport of small molecules used in their study was affected by mutations at position F439 but not at position N436. In another study by Manolaridis *et al*. [23], mutating the latter position was reported to abolish the export of a different molecule, estron-3-sulphate. Importantly, the small uric acid molecule used in our study contacted F439 with the highest frequency, and several other residues, including N436 with moderate frequency.

Another thorough experimental work supplemented with molecular dynamics simulations addressed the interactions of the longest extracellular loop (EL3) and the role of leucine amino acids at the constriction between Site 3 and Site 4 [27]. It was demonstrated that EL3 constrained by disulfide bonds interacts with the reentrant G-loop and all extracellular loops forming a lid to cover the upper cavity, thus creating Site 4 (off-site). Based on molecular dynamics simulations in the absence of any substrate molecule, they proposed that transport between the two sites is driven by the squeezing motion of leucine 555 residues, indicating a peristaltic transport mechanism [58, 59].

By metadynamics simulations (Fig. 5), here we demonstrate that the passage of a small substrate molecule, uric acid through the Leu-valve requires energy not greater than the hydrolysis of an ATP molecule. Importantly, this does not mean that moving a single substrate molecule requires one ATP, only indicates that a relatively small amount of energy is able to drive the substrate through the valve in some of our simulations. Interestingly, the inward-closed conformation is regarded as a structure developed after transport, since the substrate, which was included during the structure determination process (PDB ID:6HZM) [23], was not present in the constricted binding pocket of this structure. Therefore, the low energy required for substrate passage without biasing the protein conformation in our simulations strongly suggests the absence of large energy differences between the pre- and post-transport steps, further supporting a transport mechanism involving peristaltic motions with rather small conformational changes.

The entry of substrates from the intracellular direction to the central binding pocket has not been studied yet. The general view is that the inward-facing conformation provides unlimited access to the central binding pocket from the bottom. Even if this was the case, a molecule would not necessarily diffuse into the binding pocket without interactions with other residues along its path. To observe the entry of a substrate into the pocket we performed equilibrium simulations (no extra force applied) when uric acid molecules were also present. Surprisingly, a uric acid molecule did not immediately enter the central binding pocket, but first visited Site 2′ involving R482. Even after uric acid reached the central pocket and remained close to the crucial F439 residue, the substrate turned backward but this time became located to Site 2 at the opposite half transporter polypeptide chain. These additional interactions further support that the R482-site is likely to be an important location along the translocation pathway. Although we characterized the first steps of substrate transport, the initial step of substrate entry and possibly the substrate recognition process are likely to be more complex. First of all, the small uric acid molecule did not enter directly the central binding site, since it is possible that the gate was not sufficiently open at the critical time point when the uric acid started its penetration into the protein. In addition, the unresolved flexible intracellular loop regions not present during our simulations are likely to have an effect on substrate entry. The tip of the loop between the first and second NBD β-strands is not resolved in cryo-EM structures and may act as a gating helix analogous to structurally similar regions in bacterial transporters [60–62]. This role is supported by mutations, which are located in this loop and affect substrate specificity (Fig. S9). Importantly, the recent apo-closed ABCG2 structure strongly suggests that the substrates have a profound role in opening up the translocation pathway [28].

In summary, our results presented here emphasize that substrate specificity is unlikely to be defined by a single central binding pocket but several additional sites are involved along the translocation pathways. In our opinion, multiple pathways may exist, which should be imagined not as several channels through the protein, but as one or more potential channels with many translocation patterns, similar to a route with several alternative handles on a climbing wall. Along a single route, climbers with different heights or different

finger strengths may use different handles. Similarly, in the case of ABCG2, one substrate may enter into Site 2 and interact with R482, and as a consequence, its transport will be influenced by R482 mutations. Another substrate may not fully enter Site 2 or may enter but not interact with R482, thus the transport of this substrate will be unaffected in the presence of the R482G variation. For similar reasons, N439 is likely to be important for E1S transport but not for some other substrates as discussed above. While the C55S amino acid substitution does not alter the transport of SN-38, further experiments with other substrates may highlight the influence of this residue on substrate specificity. Even the proposed multiple ABCG2 drug binding pockets [63] may not represent different drug-binding pockets in a classical sense, but alternative interaction surfaces along the translocation pathway with different affinities for various substrates. In conclusion, we propose that interactions of small molecules with ABCG2 shall be characterized along their translocation pathway to predict the transport behavior of different substrates. Moving towards the paradigm of utilizing translocation pathways or patterns instead of single drug binding pockets may potentially lead to more efficient and higher validity predictions of drug-drug interactions.

Acknowledgements This work was supported by NKFIH K127961 and the Semmelweis Science and Innovation Fund. We acknowledge the computational resources made available on the GPU cluster of the Governmental Information-Technology Development Agency and Wigner GPU Laboratory (https://gpu.wigner.mta.hu).

Author contributions TN, AToth, and HT took part in acquisition and analysis. ATelbisz participated in writing. BS and ATordai wrote and revised the manuscript. TH conceived and designed the work, participated in analysis and writing.

Funding Open access funding provided by Semmelweis University.

Data availability Input structures and parameter files can be downloaded from https://www.hegelab.org/resources.html

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

References

1. Noguchi K, Katayama K, Sugimoto Y (2014) Human ABC transporter ABCG2/BCRP expression in chemoresistance: basic and clinical perspectives for molecular cancer therapeutics. Pharmgen Pers Med 7:53–64
2. Polgar O, Robey RW, Bates SE (2008) ABCG2: structure, function and role in drug response. Expert Opin Drug Metab Toxicol 4:1–15
3. Robey RW, To KK, Polgar O et al (2009) ABCG2: a perspective. Adv Drug Deliv Rev 61:3–13
4. Hillgren KM, Keppler D, Zur AA et al (2013) Emerging transporters of clinical importance: an update from the International Transporter Consortium. Clin Pharmacol Ther 94:52–63
5. Zamek-Gliszczynski MJ, Taub ME, Chothe PP et al (2018) Transporters in drug development: 2018 ITC recommendations for transporters of emerging clinical importance. Clin Pharmacol Ther 104:890–899
6. Kapoor P, Horsey AJ, Cox MH et al (2018) ABCG2: does resolving its structure elucidate the mechanism? Biochem Soc Trans 46:1485–1494
7. Safar Z, Kis E, Erdo F et al (2019) ABCG2/BCRP: variants, transporter interaction profile of substrates and inhibitors. Expert Opin Drug Metab Toxicol 15:313–328
8. Nakayama A, Matsuo H, Takada T et al (2011) ABCG2 is a high-capacity urate transporter and its genetic impairment increases serum uric acid levels in humans. Nucleosides Nucleotides Nucleic Acids 30:1091–1097
9. Woodward OM, Kottgen A, Coresh J et al (2009) Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. Proc Natl Acad Sci U S A 106:10338–10342
10. Kim M, Turnquist H, Jackson J et al (2002) The multidrug resistance transporter ABCG2 (breast cancer resistance protein 1) effluxes Hoechst 33342 and is overexpressed in hematopoietic stem cells. Clin Cancer Res 8:22–28
11. Volk EL, Farley KM, Wu Y et al (2002) Overexpression of wild-type breast cancer resistance protein mediates methotrexate resistance. Cancer Res 62:5035–5040
12. Chen ZS, Robey RW, Belinsky MG et al (2003) Transport of methotrexate, methotrexate polyglutamates, and 17beta-estradiol 17-(beta-D-glucuronide) by ABCG2: effects of acquired mutations at R482 on methotrexate transport. Cancer Res 63:4048–4054
13. Giri N, Agarwal S, Shaik N et al (2009) Substrate-dependent breast cancer resistance protein (Bcrp1/Abcg2)-mediated interactions: consideration of multiple binding sites in in vitro assay design. Drug Metab Dispos 37:560–570
14. Telbisz A, Hegedus C, Ozvegy-Laczka C et al (2012) Antibody binding shift assay for rapid screening of drug interactions with the human ABCG2 multidrug transporter. Eur J Pharm Sci 45:101–109
15. Ozvegy-Laczka C, Laczko R, Hegedus C et al (2008) Interaction with the 5D3 monoclonal antibody is regulated by intramolecular rearrangements but not by covalent dimer formation of the human ABCG2 multidrug transporter. J Biol Chem 283:26059–26070
16. Ozvegy-Laczka C, Varady G, Koblos G et al (2005) Function-dependent conformational changes of the ABCG2 multidrug transporter modify its interaction with a monoclonal antibody on the cell surface. J Biol Chem 280:4219–4227
17. Hegedus C, Telbisz A, Hegedus T et al (2015) Lipid regulation of the ABCB1 and ABCG2 multidrug transporters. Adv Cancer Res 125:97–137
18. Jackson SM, Manolaridis I, Kowal J et al (2018) Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. Nat Struct Mol Biol 25:333–340
19. Telbisz A, Ozvegy-Laczka C, Hegedus T et al (2013) Effects of the lipid environment, cholesterol and bile acids on the function

of the purified and reconstituted human ABCG2 protein. Biochem J 450:387–395
20. Telbisz A, Muller M, Ozvegy-Laczka C et al (2007) Membrane cholesterol selectively modulates the activity of the human ABCG2 multidrug transporter. Biochim Biophys Acta 1768:2698–2713
21. Bhatia A, Schafer HJ, Hrycyna CA (2005) Oligomerization of the human ABC transporter ABCG2: evaluation of the native protein and chimeric dimers. Biochemistry 44:10893–10904
22. Henriksen U, Fog JU, Litman T et al (2005) Identification of intra- and intermolecular disulfide bridges in the multidrug resistance transporter ABCG2. J Biol Chem 280:36926–36934
23. Manolaridis I, Jackson SM, Taylor NMI et al (2018) Cryo-EM structures of a human ABCG2 mutant trapped in ATP-bound and substrate-bound states. Nature 563:426–430
24. Taylor NMI, Manolaridis I, Jackson SM et al (2017) Structure of the human multidrug transporter ABCG2. Nature 546:504–509
25. Ferreira RJ, Bonito CA, Cordeiro M et al (2017) Structure-function relationships in ABCG2: insights from molecular dynamics simulations and molecular docking studies. Sci Rep 7:15534
26. Mozner O, Bartos Z, Zambo B et al (2019) Cellular processing of the ABCG2 Transporter-Potential Effects on Gout and Drug Metabolism. Cells 8:1215
27. Khunweeraphong N, Szollosi D, Stockner T et al (2019) The ABCG2 multidrug transporter is a pump gated by a valve and an extracellular lid. Nat Commun 10:5433
28. Orlando BJ, Liao M (2020) ABCG2 transports anticancer drugs via a closed-to-open switch. Nat Commun 11:2264
29. Laszlo L, Sarkadi B, Hegedus T (2016) Jump into a new fold-a homology based model for the ABCG2/BCRP Multidrug Transporter. PLoS ONE 11:e0164426
30. Khunweeraphong N, Stockner T, Kuchler K (2017) The structure of the human ABC transporter ABCG2 reveals a novel mechanism for drug extrusion. Sci Rep 7:13767
31. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(33–8):27–28
32. Huang J, Rauscher S, Nawrocki G et al (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14:71–73
33. Van Der Spoel D, Lindahl E, Hess B et al (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–1718
34. Jo S, Kim T, Iyer VG et al (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29:1859–1865
35. Wu EL, Cheng X, Jo S et al (2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J Comput Chem 35:1997–2004
36. Lomize MA, Lomize AL, Pogozheva ID et al (2006) OPM: orientations of proteins in membranes database. Bioinformatics 22:623–625
37. Morris GM, Huey R, Lindstrom W et al (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791
38. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455–461
39. Sutto L, Marsili S, Gervasio FL (2012) New advances in metadynamics. WIREs Comp Mol Sci 2:771–779
40. Tribello G, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) PLUMED 2: New feathers for an old bird. Computer Phys Commun 185:604–613
41. Allen WJ, Lemkul JA, Bevan DR (2009) GridMAT-MD: a grid-based membrane analysis tool for use with molecular dynamics. J Computat Chem 30(12):1952–1958
42. Schmidtke P, Bidon-Chanal A, Luque FJ et al (2011) MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. Bioinformatics 27:3276–3285
43. Michaud-Agrawal N, Denning EJ, Woolf TB et al (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem 32:2319–2327
44. Hunter J (2007) Matplotlib: A 2D graphics environment. Comp Sci Eng 9:90–95
45. van Meer G (1989) Lipid traffic in animal cells. Annu Rev Cell Biol 5:247–275
46. Jo S, Cheng X, Lee J et al (2017) CHARMM-GUI 10 years for biomolecular modeling and simulation. J Comput Chem 38:1114–1124
47. Lee J, Patel DS, Stahle J et al (2019) CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. J Chem Theory Comput 15:775–786
48. Corradi V, Gu RX, Vergani P et al (2018) Structure of transmembrane helix 8 and possible membrane defects in CFTR. Biophys J 114:1751–1754
49. Gyimesi G, Ramachandran S, Kota P et al (2011) ATP hydrolysis at one of the two sites in ABC transporters initiates transport related conformational transitions. Biochim Biophys Acta 1808:2954–2964
50. Tordai H, Leveles I, Hegedus T (2017) Molecular dynamics of the cryo-EM CFTR structure. Biochem Biophys Res Commun 491:986–993
51. Hofsass C, Lindahl E, Edholm O (2003) Molecular dynamics simulations of phospholipid bilayers with cholesterol. Biophys J 84:2192–2206
52. Petrache HI, Tu K, Nagle JF (1999) Analysis of simulated NMR order parameters for lipid bilayer structure determination. Biophys J 76:2479–2487
53. Galvan-Hernandez A, Kobayashi N, Hernandez-Cobos J et al (2020) Morphology and dynamics of domains in ergosterol or cholesterol containing membranes. Biochim Biophys Acta Biomembr 1862:183101
54. Corey RA, Stansfeld PJ, Sansom MSP (2020) The energetics of protein-lipid interactions as viewed by molecular simulations. Biochem Soc Trans 48:25–37
55. Hedger G, Sansom MSP (2016) Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations. Biochim Biophys Acta 1858:2390–2400
56. Song C, de Groot BL, Sansom MSP (2019) Lipid bilayer composition influences the activity of the antimicrobial peptide dermcidin channel. Biophys J 116:1658–1666
57. Gose T, Shafi T, Fukuda Y et al (2020) ABCG2 requires a single aromatic amino acid to "clamp" substrates and inhibitors into the binding pocket. FASEB J 34:4890–4903
58. Locher KP (2016) Mechanistic diversity in ATP-binding cassette (ABC) transporters. Nat Struct Mol Biol 23:487–493
59. Pan L, Aller SG (2015) Equilibrated atomic models of outward-facing P-glycoprotein and effect of ATP binding on structural dynamics. Sci Rep 5:7880
60. Bi Y, Mann E, Whitfield C et al (2018) Architecture of a channel-forming O-antigen polysaccharide ABC transporter. Nature 553:361–365
61. Caffalette CA, Corey RA, Sansom MSP et al (2019) A lipid gating mechanism for the channel-forming O antigen ABC transporter. Nat Commun 10:824
62. Chen L, Hou WT, Fan T et al (2020) Cryo-electron microscopy structure and transport mechanism of a wall teichoic acid ABC transporter. mBio 11:e02749–19
63. Clark R, Kerr ID, Callaghan R (2006) Multiple drugbinding sites on the R482G isoform of the ABCG2 transporter. Br J Pharmacol 149:506–515

DATABASES dc_2020_22

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

ABCMdb: A Database for the Comparative Analysis of Protein Mutations in ABC Transporters, and a Potential Framework for a General Application

Gergely Gyimesi,[1,3] Dávid Borsodi,[1,3] Hajnalka Saránkó,[1,3] Hedvig Tordai,[1,3] Balázs Sarkadi,[1–3] and Tamás Hegedűs[1–3]*

[1] *Membrane Research Group, Hungarian Academy of Sciences, Budapest, Hungary;* [2] *Institute of Molecular Pharmacology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary;* [3] *Department of Biophysics and Radiation Biology, Semmelweis University Budapest, Hungary*

Communicated by John McVey

Received 12 March 2012; accepted revised manuscript 29 May 2012.

Published online 12 June 2012 in Wiley Online Library (www.wiley.com/humanmutation).DOI: 10.1002/humu.22138

ABSTRACT: To overcome the pathological phenomena caused by altered function of ABC (ATP Binding Cassette) proteins, their mechanisms of action are extensively investigated, often involving the design of mutant constructs for experiments. Designing mutagenetic constructs, interpreting the result of mutagenetic experiments, and finding individual genetic variants require an extensive knowledge of previously published mutations. To aid the recapitulation of mutations described in the literature, we set up a database of ABC protein mutations (ABCMdb) extracted from full-text papers using an automatic mining approach. We have also developed a Web application interface to compare mutations in different ABC proteins using sequence alignments and to interactively map the mutations to 3D structural models. Currently our database contains protein mutations published for ABCB1, ABCB11, ABCC1, ABCC6, ABCC7, and the proteins of the ABCG subfamily. The database will be extended to include other members and subfamilies, and to provide information on whether or not a mutation is disease causing, represents a high-incidence polymorphism, or was generated only in vitro. The ABCMdb database should already help to compare the effects of mutations at homologous positions in different ABC proteins, and its interactive tools aim to advance the design of experiments for a wider range of proteins.

Hum Mutat 33:1547–1556, 2012.

KEY WORDS: protein mutations; ABC proteins; database; text mining

Introduction

ABC proteins form one of the largest protein families, represented in living organisms from bacteria to plants and humans. The human genome project identified 48 human ABC proteins, most of which serve as membrane-bound transporters of various substrates [Dean and Annilo, 2005]. ABC proteins are categorized into seven subfamilies (A-G) based on their sequence similarity [Dean and Annilo, 2005]. A common mechanism of their action is the binding and hydrolysis of ATP, which in ABC transporters facilitates the relocation of substrate molecules across the membrane. ABC transporters take part in critical cellular processes, and thus their physiological role and pathological phenomena arising from their hampered expression or malfunction are extensively studied [Gottesman and Ambudkar, 2001].

One of the phenomena mediated by ABC transporters is the extrusion of toxic molecules and xenobiotics from cells, which impacts the bioavailability of a wide range of pharmaceutical compounds interfering with their effect [Szakacs et al., 2008]. This phenomenon was first described as the multidrug resistance of tumor cells against chemotherapeutic treatment [Bell et al., 1985; Debenham et al., 1982], which was linked to the high expression levels of P-glycoprotein (MDR1, multidrug resistance protein 1, ABCB1). Later other members, such as MRP1 (multidrug resistance-associated protein 1, ABCC1) and ABCG2 have been cloned, which also play important roles in the clinically observed multidrug resistance of cancers [Cole et al., 1992; Doyle et al., 1998]. Apart from their role in cancer cell drug resistance, multidrug transporters have been proposed to form part of an innate chemoimmunity system [Sarkadi et al., 2006], protecting cells from the penetration of drugs and xenobiotics. The action of MDR transporters reduces the oral availability of drugs by decreasing their uptake from the digestion system [Szakacs et al., 2008]. The circumvention of this mechanism by the inhibition of MDR transporters is a major pharmacological issue, which requires a thorough understanding of multidrug recognition and transport. In addition, natural variants of multidrug transporters may cause altered substrate specificity or transport [Honjo et al., 2001; Morisaki et al., 2005], which can lead to altered pharmacokinetics or even sensitivity to certain drugs. The collection and description of MDR transporter mutations and their effects on substrate specificity can thus be a step toward personalized medical therapy.

Several pathological conditions have also been described and linked to mutations of certain ABC transporters. One of the most frequently studied inherited monogenic diseases is cystic fibrosis, which has been associated with the malfunction of the cystic fibrosis transmembrane conductance regulator (CFTR, ABCC7) protein [Riordan et al., 1989], acting as a chloride channel in the cell membrane. Although the most frequent and severe mutation is the

*Correspondence to: Tamás Hegedűs E-mail: tamas@hegelab.org

Contract grant sponsors: NKTH-OTKA-FP7 Mobility (MB08-C 80039); FP7-IRG Marie Curie (239270).

deletion of F508, over 1,600 disease-causing mutations of ABCC7 have been deposited in the Canadian Cystic Fibrosis Mutation Database (http://www.genet.sickkids.on.ca). A large number of mutagenetic studies have been devoted not simply to understand the mechanism of ABCC7 chloride transport, but also to seek for so-called rescue mutations that can overcome the impaired expression and function of the p.Phe508del (ΔF508) mutant.

Another member of the ABCC subfamily, MRP6 (ABCC6), was originally discovered as a transporter homologous to multidrug-resistance-associated proteins [Belinsky and Kruh, 1999], functioning as an organic anion transporter [Belinsky et al., 2002]. Later, it was discovered that mutations in the ABCC6 protein are responsible for the development of pseudoxanthoma elasticum (PXE) [Le Saux et al., 2000], and a large number of disease-causing mutations have been identified since then [Varadi et al., 2011].

The ABCG2 transporter, apart from contributing to multidrug resistance, also plays a role in physiological urate transport [Woodward et al., 2009], and certain dysfunctional mutants have been shown to induce the risk of developing gout [Dehghan et al., 2008; Matsuo et al., 2009]. Other members of the ABCG subfamily (ABCG1, ABCG4, ABCG5, and ABCG8) are involved in the transport of cholesterols and other sterols [Dean and Annilo, 2005]. The ABCG5 and ABCG8 transporters function as heterodimers that export sterols into the bile, preventing the accumulation of plant sterols and cholesterol [Woodward et al., 2011]. It has also been described that mutations affecting the *abcg5* and *abcg8* genes can cause dramatically increased plasma plant sterol concentrations that lead to sitosterolemia [Berge, 2003; Berge et al., 2000]. In contrast, the precise physiological role of ABCG1 and ABCG4 still needs to be elucidated, and no mutations of these proteins have yet been linked to any human diseases.

In addition to studying disease-associated mutations, a significant number of in vitro mutagenetic studies have been performed. Mutagenesis is a widely used method to probe the biochemical role of individual amino acids in substrate recognition and transport, interdomain communication or ATP binding, and hydrolysis. Mutating amino acids can also facilitate the expression, solubilization and subsequent crystallization, and structure determination of whole ABC transporters or their individual domains. When designing experimental constructs, it is important to know beforehand whether similar constructs have already been created, or mutants affecting the same regions or interactions have been described. In addition, if the functional role of a residue in a certain ABC protein has been well characterized, the role of equivalent amino acids in other ABC transporters can be inferred based on homology and amino acid conservation. Similar formulation can be used to interpret the mechanism of action of pathogenic natural variants, or to design mutant constructs of ABC domains that increase solubility based on homologous constructs for crystallization studies.

Several databases targeted at the collection of mutant variants of certain ABC proteins are available. The Cystic Fibrosis Mutation Database (CFTR1, http://www.genet.sickkids.on.ca) is devoted to the collection of individual mutations in the human *cftr* gene, originating mostly from clinical studies. Similarly, the PXE Mutation Database (PXE, http://www.pxe.org/mutation-database) aims to serve as a catalogue of all published mutations of the ABCC6 protein causing PXE. Other, more general databases also contain information on mutant variants, for example, the Swiss-Prot database [Bairoch and Apweiler, 1996; Boutet et al., 2007], or the Human Genetic Mutations Database [Cooper and Krawczak, 1996]. These databases employ manual extraction and curation of mutations from literature and clinical studies, providing high-quality data. However, due to their focus on disease-causing mutations and polymorphic variants, mutant constructs from in vitro studies are often overlooked and not collected. In addition, manual processing of literature tends to be slow and may not keep up with the exponentially growing number of available biomedical publications. Databases relying solely on human information processing thus might miss recently published or not often cited mutations.

To alleviate the burden of reading thousands of papers, automatic methods have been devised and applied to extract information from literature. Biomedical text mining is a large field with a wide range of applications such as the automatized extraction of protein–protein interactions [Krallinger et al., 2011], protein mutations [Caporaso et al., 2007; Doughty et al., 2011; Lee et al., 2007; Rebholz-Schuhmann et al., 2004], pharmacokinetic relationships between genes, drugs, and diseases [Frijters et al., 2010; Garten et al., 2010; Rubin et al., 2005] or protein and gene annotation [Camon et al., 2005]. Several methods have been implemented and applied successfully to mine human kinase mutations [Krallinger et al., 2009], mutations of G-protein coupled receptors (GPCRs), nuclear hormone receptors (NRs) [Horn et al., 2004], vitamin K-dependent coagulation serine proteases [Saunders and Perkins, 2008], and α-galactosidase A mutations related to Fabry disease [Kuipers et al., 2010]. A database that aims to gather protein level enzyme mutations having an effect on protein stability or function from PubMed abstracts is also available [Yeniterzi and Sezerman, 2009].

Our aim was to generate a database of ABC protein mutations using automated tools while still maintaining high-quality data. We employed the *MutMiner* automatic mutation extraction pipeline that we developed in our laboratory, which offers 95% precision and 84% recall in finding protein mutations mentioned in full-text papers. It should be noted that the *MutMiner* pipeline only searches for explicit mentions of protein point mutations, while any underlying genetic variations are disregarded. This also means that insertions, deletions, as well as any genetic variations that have no effect at the protein level (e.g., intronic or synonymous mutations) are not considered in the current study. Mutation mentions are harvested from literature and supplemented with a Web framework that aims to integrate extracted mutational data, sequence alignments, and available 3D structures of ABC transporters. This integrated framework (http://abcmutations.hegelab.org) will help in designing new mutations and interpreting the effects of existing mutations in the context of homologous proteins, sequence conservation, and spatial structure.

Methods

Data Mining from Full-Text Papers

A list of the PMIDs of papers corresponding to each protein under investigation was retrieved from PubMed using a keyword search. The full-text version of the papers was downloaded using a modified version of the *pdfetch* automated tool (http://code.google.com/p/pdfetch). The mentions of mutations were identified and extracted from the full-text papers using our framework, *MutMiner* (http://mutminer.hegelab.org). *MutMiner* is built on *MutationFinder* [Caporaso et al., 2007], a tool that uses regular expression patterns to recognize and extract protein point mutation mentions from English text. The pipeline supplements *MutationFinder* with text preprocessing tools, and algorithms to bind mutations in the text to protein names. For each target ABC protein, *MutMiner* was provided with the whole retrieved literature corresponding to the protein, and the UniProtKB/Swiss-Prot record downloaded in FASTA format. The UniProtKB accession

Table 1. Retrieving Full-Text Articles Corresponding to Each of the 10 ABC Transporters Investigated

Protein	PubMed keywords	Resulting papers	Retrieved papers
ABCB1	ABCB1 or MDR1 or P-GP or PGP	19,437	10,494
ABCB11	ABCB11 or BSEP	539	355
ABCC1	ABCC1 or MRP1	1,775	1,336
ABCC6	ABCC6 or MRP6	290	248
ABCC7	ABCC7 or CFTR	6,735	3,476
ABCG1	ABCG1	426	294
ABCG2	ABCG2 or BCRP or MXR	1,928	1,692
ABCG4	ABCG4	40	31
ABCG5	ABCG5	347	240
ABCG8	ABCG8	311	218

The keywords used to query for articles from the PubMed database are shown, with the number of papers resulting from the query, and the number of papers effectively retrieved by the full-text PDF download tool. The number of articles varies across three orders of magnitude, and in some cases, little more than half of the papers could be retrieved.

numbers used were P08183 (ABCB1), O95342 (ABCB11), P33527 (ABCC1), O95255 (ABCC6), P13569 (ABCC7), P45844 (ABCG1), Q9UNQ0 (ABCG2), Q9H172 (ABCG4), Q9H222 (ABCG5), and Q9H221 (ABCG8).

Software and Tools

The database uses the *MySQL* relational database backend (http://www.mysql.com) for data storage. For data access, the *SQLAlchemy* (http://www.sqlalchemy.org) and *Elixir* (http://elixir.ematia.de) object-relational mapper libraries were used. The Web interface was created based on the *TurboGears* Web framework (http://turbogears.org) and the *Genshi* templating library (http://genshi.edgewall.org).

Sequence Alignments

Alignments were created using *ClustalW* 2.0.10 [Larkin et al., 2007] with default options. Several multiple sequence alignments were created: (1) One alignment of full proteins for each subfamily where all subfamily members have a similar topology (ABCA, ABCC, and ABCG); (2) all half-transporter units of the ABCB subfamily, full transporters were considered as comprising two individual half-transporter units; (3) NBDs of transporters in each subfamily; (4) NBDs of all transporters; (5) NBDs of transporters of the ABCC and ABCG subfamilies; and (6) NBDs of transporters of the ABCB, ABCC, and ABCG subfamilies.

Structural Models

Homology models of the outward facing conformation for ABCB1, ABCC1, and ABCC6 were kindly provided by M. Wiese [Globisch et al., 2008], S.P. Cole [DeGorter et al., 2008], and A. Váradi [Fulop et al., 2009], respectively. For ABCC7, the extended homology model of Serohijos et al. was used [Serohijos et al., 2008]. Experimentally determined structures from the Protein Data Bank [Bernstein et al., 1977] with PDB IDs 2BBO (ABCC7 NBD1), 2BBS, 2BBT (ABCC7 NBD1 with F508 deletion), 2CBZ (ABCC1 NBD1), 2PZE (ABCC7 NBD dimer), 2PZF (ABCC7 NBD dimer with F508 deletion), 2PZG (ABCC7 NBD1) were included in the database. To visually map the mutations on the structure, the *Jmol* open source Java viewer for chemical structures (http://www.jmol.org) was employed.

ABC Protein Regions

The nucleotide binding domains were identified using the PS50893 ProSite profile, representing ABC transporter-type domains, and the *ps_scan* utility, downloadable from the ProSite Website (http://prosite.expasy.org). Highly conserved regions were assigned according to their consensus sequences, for the Walker A and the ABC signature motifs the ProSite patterns PS00017 and PS00211 were used, respectively, and for the Walker B motif the consensus sequence "hhhhD" was used, where "h" denotes hydrophobic amino acids. The membrane topology, the boundaries of the transmembrane helices were determined using TMDET [Tusnady et al., 2005] for the available 3D structures of human ABCB1 [Globisch et al., 2008], ABCC7 [Serohijos et al., 2008], and the dimeric biological unit of the ABCB10 X-ray structure [PDB:2YL4]. The transmembrane helix boundaries for other ABC proteins were predicted using HMMTOP [Tusnady and Simon, 2001], and corrected or supplemented with information from TMDET predictions of available 3D structures of other, homologous ABC proteins.

Reference Databases

The following databases were used as references the PXE Mutation Database (PXE, http://www.pxe.org/mutation-database, as of 12 January, 2012), the Cystic Fibrosis Mutation Database (CFTR1, http://www.genet.sickkids.on.ca/app, as of 13 January, 2012), the Swiss-Prot database at UniProt (UniProtKB/Swiss-Prot, http://www.uniprot.org, release 2012_01 [Bairoch and Apweiler, 1996; Boutet et al., 2007]), the public version of the Human Gene

Table 2. Results of Automatic Mutation Extraction

			Number of mutation mentions (hits)		
Protein	Number of papers	Number of unique mutations	Total	True	False
ABCB1	10,494	641	1,527	n/d	n/d
ABCB11	355	156	485	n/d	n/d
ABCC1	1,336	403	882	n/d	n/d
ABCC6	248	207	778	740 (95%)	38 (5%)
ABCC7	3,476	1,353	8,874	n/d	n/d
ABCG1	294	5	7	3 (43%)	4 (57%)
ABCG2	1,692	186	1,391	1,317 (95%)	74 (5%)
ABCG4	31	0	0	0	0
ABCG5	240	25	77	77 (100%)	0 (0%)
ABCG8	218	43	147	144 (98%)	3 (2%)

The number of unique mutations extracted from the set of downloaded full-text papers is shown. Hits are considered to be pairs consisting of a mutation and a paper that were found by the pipeline. In the case of the ABCC6 and ABCG proteins, the hits were manually verified whether or not they represent actual mutant variants of the corresponding protein, and categorized as being true or false positives, respectively. Based on manual verification, the precision of the automatic mining method can be estimated to be above 95%. Mutations of the ABCG1 protein constitute an exception, but the relatively high ratio of false positives in that case is likely due to the small sample size.

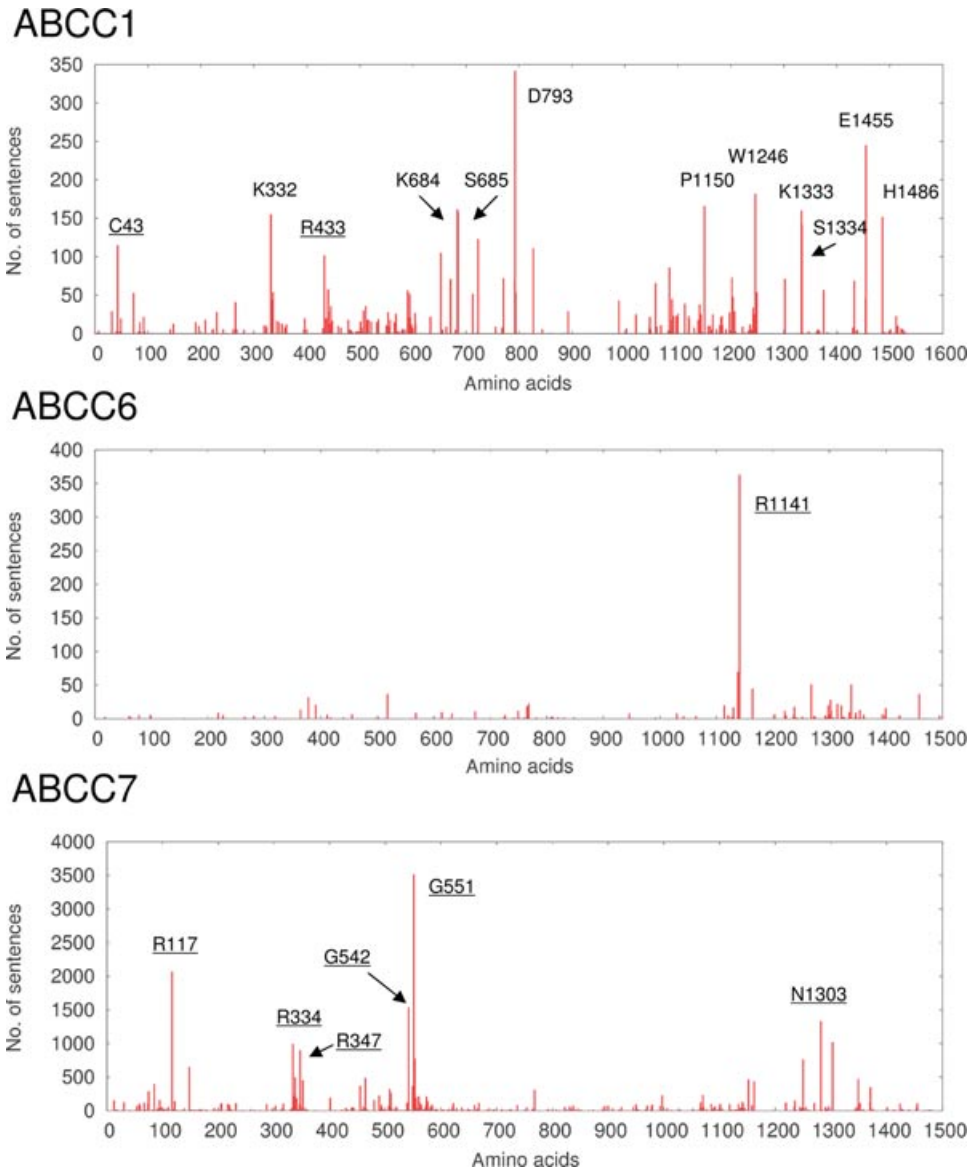


Figure 1. Distribution of mutation mentions along the protein sequence in ABCC proteins. The number of sentences mentioning mutant variants of a given amino acid position in the sequence is shown. The most frequently cited positions are indicated, with natural variants underlined. Most of the mutations investigated for ABCC6 and ABCC7 correspond to natural variations of the protein, whereas in the case of ABCC1, mostly induced mutations are studied. This is in accordance with the fact that ABCC1 has not been reported as a disease-causing gene, whereas the most often studied natural variants of ABCC6 and ABCC7 cause severe illnesses. Extracted mutations followed a similar distribution in the case of other analyzed proteins (data not shown).

Mutation Database (HGMD, http://www.hgmd.org, as of 27 January, 2012 [Cooper and Krawczak, 1996]).

Results

Establishment of the ABC Mutation Database

We selected 10 ABC transporter proteins as a basis for the database. Relevant papers were retrieved from the PubMed database by keyword queries for each protein (Table 1). To retrieve the full-text version of these papers, a modified version of the *pdfetch* utility was used. The number of papers for each protein varied on a wide range, and in many cases, the full-text PDF could not be downloaded (e.g., because of a lack of subscription). Statistics about the retrieved papers for each protein are summarized in Table 1. After running the PDF files through the *MutMiner* pipeline, extracted mutations for each protein were collected and uploaded into our mutation database.

The number of unique mutations extracted for each protein (shown in Table 2) also showed large variations, with 1,353 different mutations extracted for the ABCC7 protein, to ABCG4, for which no mutations could be extracted from the available 31 papers. For ABCB1, even though more than 10,000 papers were retrieved, only 641 mutations could be identified. For ABCC7 and ABCG2, the number of mutation mentions is 6–8 times the number of unique mutations, indicating that mutant variants of these proteins are much more intensively studied or more frequently found in the population than others in our dataset. A plot of the number of citing sentences found for mutant variants in each ABC protein

Table 3. Assessing the Overlap of Automatically Extracted Mutations with Mutations Described in Existing Databases

Database		Number of mutations retrieved		Number of mutations not retrieved		Number of new mutations
PXE (ABCC6)	159	(87%)	24	(13%)	48	
CFTR1 (ABCC7)	562	(62%)	338	(38%)	791	
Swiss-Prot	ABCB1	24	(96%)	1	(4%)	617
	ABCB11	33	(94%)	2	(6%)	123
	ABCC1	35	(97%)	1	(3%)	368
	ABCC6	37	(93%)	3	(7%)	170
	ABCC7	172	(96%)	7	(4%)	1,181
	ABCG1	0	(0%)	1	(100%)	5
	ABCG2	21	(95%)	1	(5%)	165
	ABCG4	0	(0%)	1	(100%)	0
	ABCG5	11	(92%)	1	(8%)	14
	ABCG8	16	(80%)	4	(20%)	27
	Total	**349**	**(94%)**	**22**	**(6%)**	**2,688**
HGMD	ABCB1	5	(71%)	2	(29%)	636
	ABCB11	62	(85%)	11	(15%)	94
	ABCC1	3	(100%)	0	(0%)	400
	ABCC6	144	(92%)	12	(8%)	63
	ABCC7	582	(71%)	238	(29%)	771
	ABCG1	0		0		5
	ABCG2	13	(100%)	0	(0%)	173
	ABCG4	0		0		0
	ABCG5	10	(83%)	2	(17%)	15
	ABCG8	22	(88%)	3	(12%)	21
	Total	**841**	**(76%)**	**268**	**(24%)**	**2,178**

Mutations stored in the PXE Mutation Database (PXE, ABCC6 mutations), the Cystic Fibrosis Mutation Database (CFTR1, ABCC7 mutations), the Swiss-Prot database, and the Human Gene Mutation Database (HGMD) were considered. The results show that for databases relying mostly on published data (PXE, Swiss-Prot, HGMD), a large percent of the mutations could be recovered using our automated method. In the case of databases containing a high amount of mutant variants from unpublished clinical studies (CFTR1, HGMD/ABCC7), the recovery ratio was lower (62–71%). However, in all cases, automatic extraction yielded a large number of unique mutations that were not present in either of the databases considered.

indicates certain positions whose variations are highly cited, most of which are disease-causing natural variants (shown for the ABCC subfamily in Fig. 1).

Overlap and Novelty of the ABC Database Compared to Other Databases

To assess the recovery of mutations stored in manually curated databases, the extracted mutations were compared to those found in the PXE Mutation Database (ABCC6 mutations), CFTR1 database (ABCC7 mutations), and mutations corresponding to other ABC proteins found in the Swiss-Prot database [Bairoch and Apweiler, 1996; Boutet et al., 2007] and Human Genome Mutation Database (HGMD, [Cooper and Krawczak, 1996]). The comparison was performed on the sets of unique mutations derived from our results and from each of the reference databases. The results are summarized in Table 3.

The PXE Mutation Database

Altogether the PXE database contains 183 missense and nonsense mutations of the ABCC6 transporter, out of which 159 (87%) was found by the *MutMiner* pipeline. In the case of the 24 mutations that the automatic method failed to extract, the papers cited as reference for these mutations in the PXE database were examined by hand to discover the reason the mutations were missed. The most frequent cause of miss (in the case of 7 mutations) was due to errors in the PDF to text conversion of a single paper. In five cases, no reference was provided for the mutations, and in one case only HapMap was mentioned as reference. Two mutations were not detected because *MutationFinder* did not recognize them in the text. In one case, the mutation was not found because the corresponding paper could not be downloaded automatically. Remarkably, the inspection of missed mutations also highlighted certain errors in the database. In four cases, the mutations were in fact synonymous, and incorrectly classified as missense. In three cases, the database contained a mistyped version of the mutation mentioned in the papers, and in one case, the database mutation was not found in the reference paper at all. The automatic method, however, found 48 unique mutations that were not present in the PXE database.

The Cystic Fibrosis Mutation Database

The CFTR1 database contains a collection of mutations that were found in the human ABCC7 protein. From over 1,600 mutations contained in the database, all 900 missense and nonsense mutations were examined, out of which 562 were recovered by data mining. An inspection of some of the remaining 338 cases revealed that most of these mutations are likely to originate from unpublished clinical studies, and therefore no reference paper could be obtained. The *MutMiner*-based method also found 791 mutations that were not present in the database.

Swiss-Prot

A comparison of extracted mutations with the Swiss-Prot database for all 10 ABC transporters shows that most of the mutations stored in Swiss-Prot could be retrieved from the literature using our automated method. In cases where the retrieval failed, the database mutations were manually examined by checking the reference publication if one was available. The examination revealed that in 9 of the 22 failed cases the mutation could not be discovered in the reference paper even by hand. In 4 cases, no reference publication was specified in either Swiss-Prot or in the dbSNP entry corresponding to the mutation. In four cases, the PDF downloader failed to fetch the reference publication, but the *MutMiner* pipeline was able to find the mutations after the paper was manually downloaded. In the other cases, the mutations were missing because they

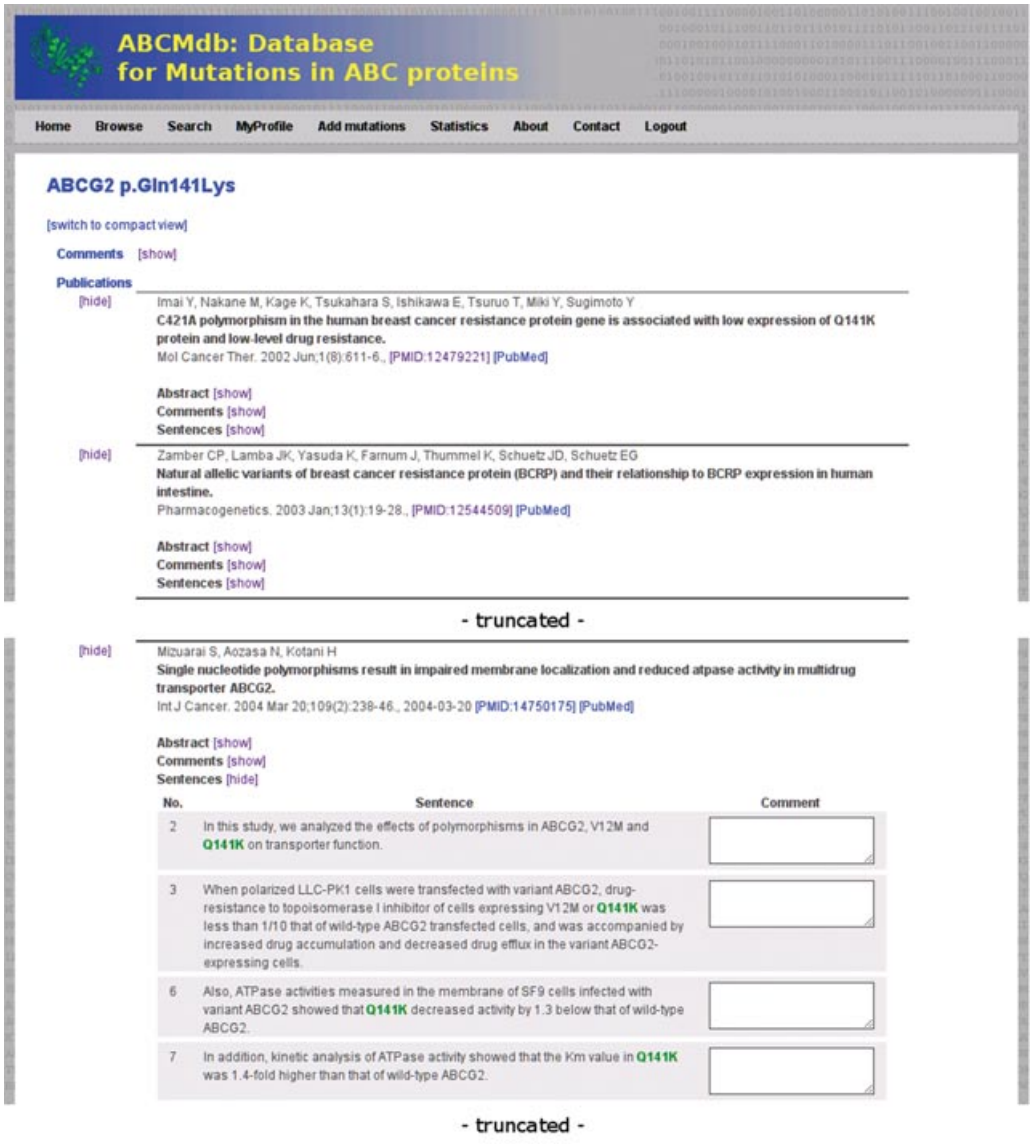


Figure 2. Detailed listing of publications and their sentences mentioning a mutation. The ABCG2 p.Gln141Lys mutation is shown. Bibliographic information and the abstract of the publications citing the mutation are displayed with a link to the PubMed record corresponding to the article. Reading the sentences mentioning the mutation makes it easy to get a brief overview of the functional effects of the mutation.

were either located in the supplementary material or we had no access to the reference publication. There were two cases where the reverse of the mutation was mentioned in the reference paper, such as "A632V" instead of "V632A" (p.Val632Ala), which might indicate a sequence ambiguity. In addition, literature mining yielded over 2,600 mutations that were not recorded in the Swiss-Prot protein annotations.

The Human Gene Mutation Database

The comparison to HGMD yielded similar results to Swiss-Prot, with a slightly higher rate of database mutations that were not retrieved automatically. One remarkable case is the ABCC7 protein, where the database contained 238 mutations that our method could not retrieve from the literature, out of which 173 were unpublished results, mostly pointing to the CFTR1 database for reference. For those mutations where a journal reference was cited, the PDF retrieval failed in most cases during the automatic download step.

Even though the complete recovery of mutations stored in databases was not possible, our mutation extraction pipeline was still able to find 94% of all mutant variants listed in Swiss-Prot and 76% of those listed in HGMD. In addition to recovering most of the database mutations, a considerable number of unlisted mutations were found compared to each database. It should be noted, however, that in this study we are only focusing on missense and nonsense protein mutations. Other variants, for example, insertions and deletions, are often stored in databases but are not recognized by our automatic extraction method. Also, it should be noted that only open and freely accessible databases were chosen as reference in this study, which might not be as up-to-date or complete as commercial alternatives (e.g., HGMD Professional).

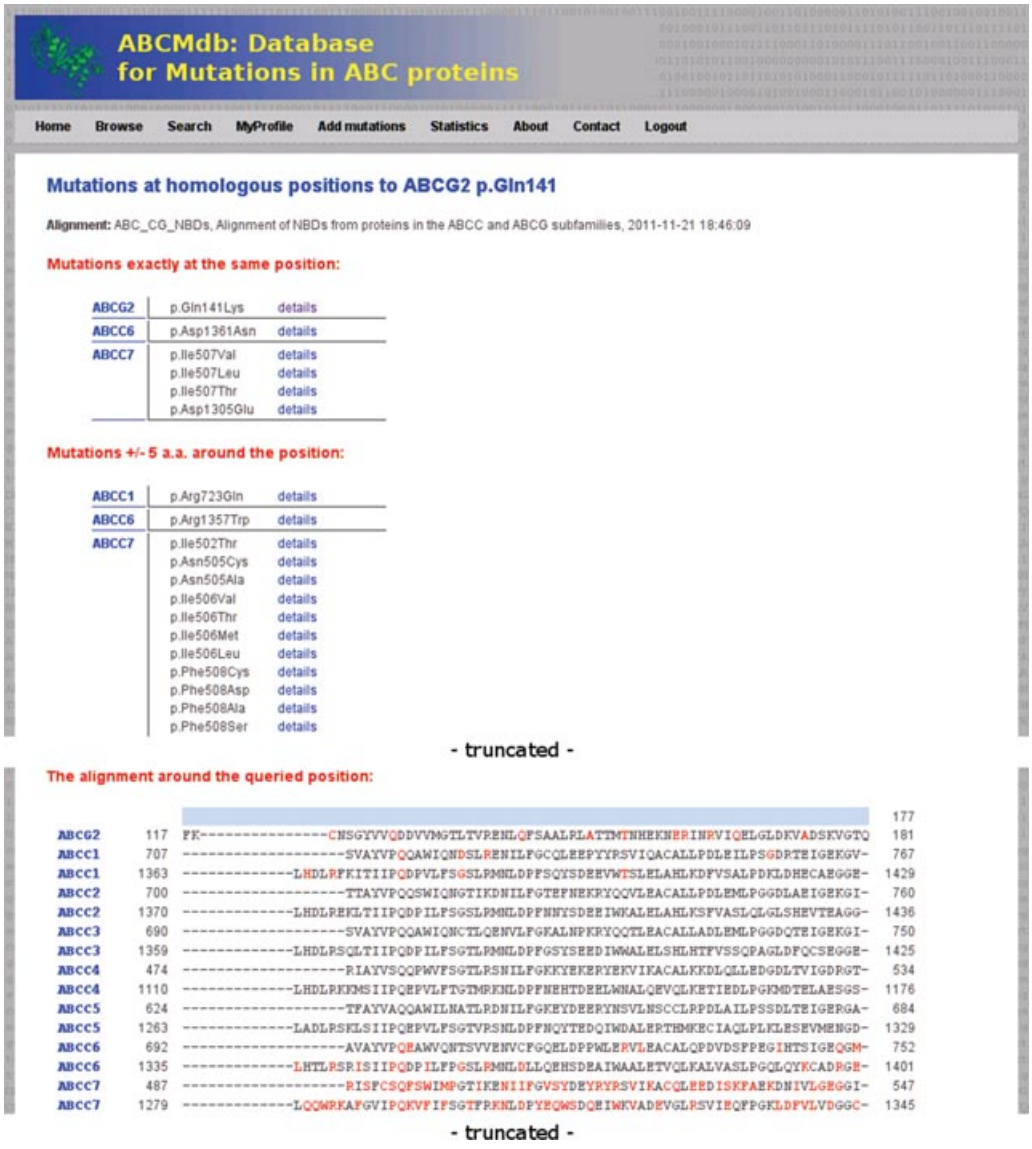


Figure 3. Mutations in and near homologous positions in multiple proteins. Specifying an amino acid position or an existing mutation can be used to search for mutations in homologous positions in other proteins. Exact homologues and mutations within five amino acids of the exact homologous position are listed. The multiple sequence alignment used for the search is also displayed showing the region around the selected position. In the alignment, positions for which mutational information is available are displayed in red online, and the list of corresponding mutations can be accessed by clicking on the amino acid letter.

The Web Interface for Comparative Studies

Browsing mutations and containing sentences

The Web interface provides a browsable listing of extracted mutations for specific ABC proteins, as well as functions to find homologous mutations in other ABC proteins, and to view mutations mapped onto a 3D structure of a certain ABC protein. For each mutation, a detailed listing of the publications where the mutation was found and all sentences mentioning the mutation are also displayed (Fig. 2). A concise list of all sentences mentioning the mutation is also available, allowing a quick overview of the context of each mutation mention. In many cases, the list of sentences can help to determine at a glance the functional or structural effect of the mutation or its disease-causing nature (Fig. 2). In addition to browsing mutations, several other views for the comparative analysis of homologous and nearby mutations and for 3D visualization are also available.

Finding homologous mutations using sequence alignments

For the comparative analysis of homologous proteins, a biologically correct sequence alignment is crucial. The database contains various multiple sequence alignments of full and half ABC transporters, and their nucleotide binding domains. In addition to these alignments, any custom alignment of ABC proteins present in the database can be uploaded after user registration on the Web site. The uploaded alignments can then be used in the same way as those provided with the system. Homologous mutations can be searched for either by position or by homology to an existing mutation appearing in the database. The first method is provided by the Search menu, where mutations homologous to a specified position in an ABC

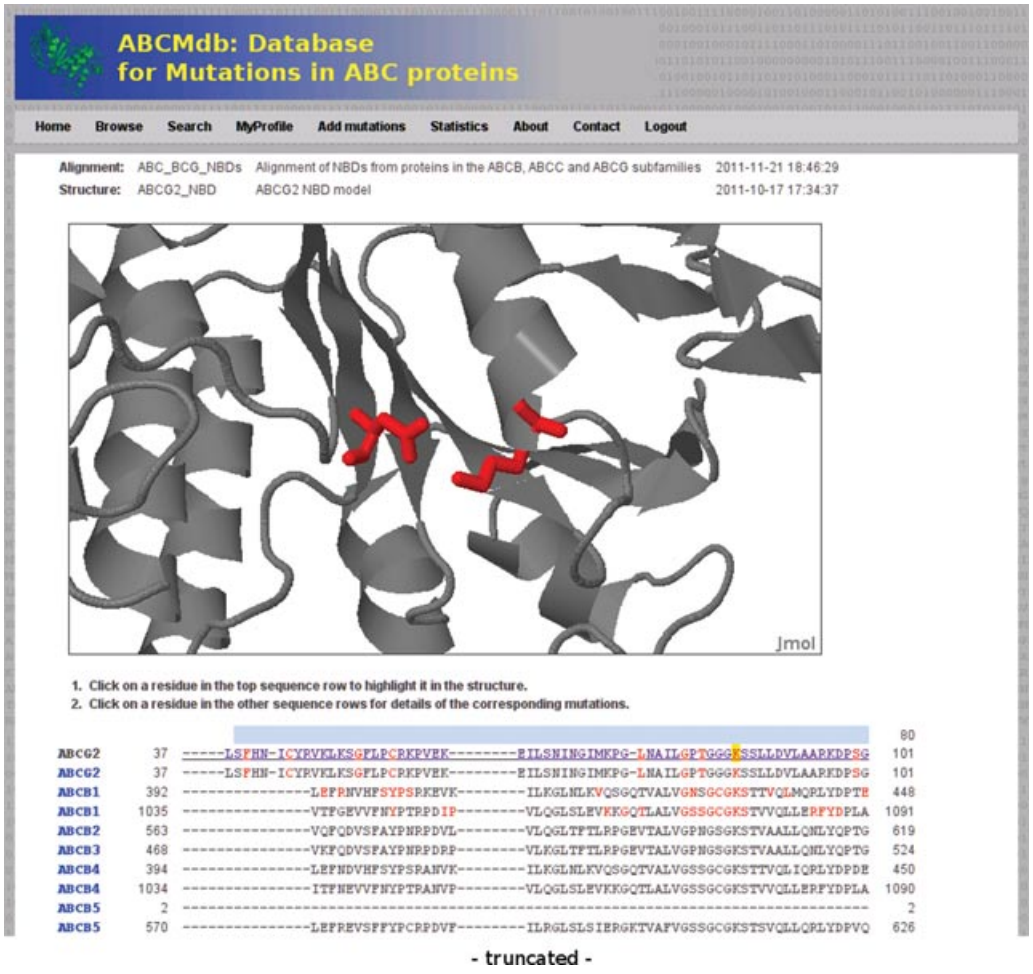


Figure 4. Displaying amino acid residues in the context of the 3D protein structure. An interactive 3D view can be displayed for proteins with an available structure. A multiple sequence alignment is also shown. Amino acids can be highlighted in the structure by clicking on the corresponding amino acid letter in the top row of the alignment. The ABCG2 catalytic p.Lys86 residue with its salt bridge partner p.Asp210 is shown in the figure.

transporter protein can be listed. The second method can be accessed by the "homologous mutations" link appearing next to each mutation in the list of mutations of a specific ABC protein. Both methods present a list of known mutations of other ABC proteins in exactly homologous and in nearby positions, along with the sequence alignment of residues in the neighborhood of the selected position (Fig. 3). Amino acids for which known mutations are available are highlighted in the alignment.

Visualizing mutations in context of the 3D structure

The Web interface provides an option to map mutations onto available structures of ABC proteins. Similarly to sequence alignments, custom structures in PDB format can be uploaded for ABC proteins in the database, and used besides those provided with the system. The mutation mapping feature can be accessed via the "map mutations" link appearing next to each protein when browsing the list of proteins. After selecting an alignment and a structure to be used for visualization, a page with an interactive 3D view of the selected protein structure is presented, and the selected alignment of ABC proteins (Fig. 4). A specific residue can be highlighted on the 3D structure by clicking on the corresponding amino acid in the top row of the alignment. Similarly to homologous mutation search, residues with an associated mutation in the database are highlighted in the sequence alignment.

Discussion

In summary, a comprehensive assessment of all described mutations of 10 ABC proteins was created using the *MutMiner* automated literature mining pipeline. Mentions of mutations were extracted from full-text papers and were placed in a database system for easy retrieval. A Web application was also created to provide a user-friendly interface to the database, along with tools to aid the comparative analysis and visualization of mutations in homologous positions.

Comparing the set of automatically extracted mutations with existing databases shows that most of the literature-cited mutations in databases can be recovered using our automated protocol. For some databases (e.g., Swiss-Prot), almost all mutations can be automatically extracted from the literature, whereas for others only 62% of the mutations can be recovered. The relatively low overlap of our extracted data with locus specific databases reflects the difference in approach taken during data collection. A significant number of mutant variants from genotyping studies are deposited in various locus-specific databases without being published. On the other

hand, these databases are often missing artificial mutant constructs of proteins that were generated for various biochemical studies. Our aim was to provide a repository of mutant variants where constructs originating from biochemical studies are also represented to facilitate the design of further biochemical experiments. In fact, our extraction pipeline yielded a significant number of mutations that are not present in the examined databases. It is thus clear that the role of ABCMdb is complementary to existing locus-specific databases that store mostly natural genetic variants. This complementary role can be exploited for the benefit of both approaches, for example, the automated method was also able to pinpoint possible errors and inconsistencies in existing manually curated databases. Similarly, manual curation of automatically extracted data can be used to eliminate false positives, which helps to improve the quality of extracted data. In this regard, the ABCMdb builds on the community effort of expert scientists of particular ABC transporters, providing functionality that enables registered users to comment and annotate database records.

One of the major issues hampering the recovery of database mutations is the incomplete retrieval of PDF files. While this could be overcome by improving the automatic text retrieval system, access to full-text articles will always be limited by the lack of institutional subscription to some journals. Similar studies have shown that valuable data can also be extracted from freely available abstracts, although they cannot replace mining full-text articles [Krallinger et al., 2009]. To overcome the limitation of accessing full-text papers, registered users can upload articles that will be processed and the extracted mutations will be added to the database.

Our results show that automatic data mining methods are useful tools to extract data from the literature. Because the sentences containing hits are available for each extracted mutation, automatic methods could be developed to supplement current data with various predicted properties of the mutations. Of particular importance would be the discrimination of DNA and protein mutations, as this would markedly reduce false positives generated by the extraction pipeline. Other valuable information that could be mined from the context of mutations include the host organism, the physiological, functional or structural effect of the mutation (e.g., disease causing, loss of function, change in substrate affinity, and incorrect folding), or whether it is a natural variant, a polymorphism, or an induced mutation. We plan to implement most of these features in future versions of our database.

Besides automatic data mining, our framework allows registered users to leave comments on specific mutations, papers or individual mentions of mutations, and to mark mutations as true or false hits, specifying the reason of the error in case the mutation was falsely predicted. We believe this system will help to improve the quality of data in the database. It is also notable that the extraction pipeline and Web interface is not specific to ABC proteins. A similar framework and database with minor changes could be set up for literature mining and display of mutations of other proteins and protein families. Based on the manual verification of the extracted ABCG family and ABCC6 mutations, the ABCMdb mutations are likely to contain over 95% of true positive hits, which we believe can make the database a reliable tool for the design of experiments and the comparative analysis of mutations.

Acknowledgments

Valuable suggestions of Paula Duek Roggli (CALIPHO group headed by Prof. Amos Bairoch, Swiss Institute of Bioinformatics) and Lajos Kalmár (Institute of Enzymology, Research Center of Natural Sciences, Hungarian Academy of Sciences) are greatly acknowledged. TH is a Bolyai Fellow of the Hungarian Academy of Sciences.

References

Bairoch A, Apweiler R. 1996. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. Nucleic Acids Res 24:21–25.

Belinsky MG, Chen ZS, Shchaveleva I, Zeng H, Kruh GD. 2002. Characterization of the drug resistance and transport properties of multidrug resistance protein 6 (MRP6, ABCC6). Cancer Res 62:6172–6177.

Belinsky MG, Kruh GD. 1999. MOAT-E (ARA) is a full-length MRP/cMOAT subfamily transporter expressed in kidney and liver. Br J Cancer 80:1342–1349.

Bell DR, Gerlach JH, Kartner N, Buick RN, Ling V. 1985. Detection of P-glycoprotein in ovarian cancer: a molecular marker associated with multidrug resistance. J Clin Oncol 3:311–315.

Berge KE. 2003. Sitosterolemia: a gateway to new knowledge about cholesterol metabolism. Ann Med 35:502–511.

Berge KE, Tian H, Graf GA, Yu L, Grishin NV, Schultz J, Kwiterovich P, Shan B, Barnes R, Hobbs HH. 2000. Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. Science 290:1771–1775.

Bernstein FC, Koetzle TF, Williams GJ, Meyer Jr EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank. A computer-based archival file for macromolecular structures. Eur J Biochem 80:319–324.

Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. 2007. UniProtKB/Swiss-Prot. Methods Mol Biol 406:89–112.

Camon EB, Barrell DG, Dimmer EC, Lee V, Magrane M, Maslen J, Binns D, Apweiler R. 2005. An evaluation of GO annotation retrieval for BioCreAtIvE and GOA. BMC Bioinformatics 6(Suppl 1):S17.

Caporaso JG, Baumgartner Jr WA, Randolph DA, Cohen KB, Hunter L. 2007. MutationFinder: a high-performance system for extracting point mutation mentions from text. Bioinformatics 23:1862–1865.

Cole SP, Bhardwaj G, Gerlach JH, Mackie JE, Grant CE, Almquist KC, Stewart AJ, Kurz EU, Duncan AM, Deeley RG. 1992. Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. Science 258:1650–1654.

Cooper DN, Krawczak M. 1996. Human Gene Mutation Database. Hum Genet 98:629.

Dean M, Annilo T. 2005. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genomics Hum Genet 6:123–142.

Debenham PG, Kartner N, Siminovitch L, Riordan JR, Ling V. 1982. DNA-mediated transfer of multiple drug resistance and plasma membrane glycoprotein expression. Mol Cell Biol 2:881–889.

DeGorter MK, Conseil G, Deeley RG, Campbell RL, Cole SP. 2008. Molecular modeling of the human multidrug resistance protein 1 (MRP1/ABCC1). Biochem Biophys Res Commun 365:29–34.

Dehghan A, Kottgen A, Yang Q, Hwang SJ, Kao WL, Rivadeneira F, Boerwinkle E, Levy D, Hofman A, Astor BC, Benjamin EJ, van Duijn CM, Witteman JC, Coresh J, Fox CS. 2008. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. Lancet 372:1953–1961.

Doughty E, Kertesz-Farkas A, Bodenreider O, Thompson G, Adadey A, Peterson T, Kann MG. 2011. Toward an automatic method for extracting cancer- and other disease-related point mutations from the biomedical literature. Bioinformatics 27:408–415.

Doyle LA, Yang W, Abruzzo LV, Krogmann T, Gao Y, Rishi AK, Ross DD. 1998. A multidrug resistance transporter from human MCF-7 breast cancer cells. Proc Natl Acad Sci USA 95:15665–15670.

Frijters R, van Vugt M, Smeets R, van Schaik R, de Vlieg J, Alkema W. 2010. Literature mining for the discovery of hidden connections between drugs, genes and diseases. PLoS Comput Biol 6.

Fulop K, Barna L, Symmons O, Zavodszky P, Varadi A. 2009. Clustering of disease-causing mutations on the domain–domain interfaces of ABCC6. Biochem Biophys Res Commun 379:706–709.

Garten Y, Coulet A, Altman RB. 2010. Recent progress in automatically extracting information from the pharmacogenomic literature. Pharmacogenomics 11:1467–1489.

Globisch C, Pajeva IK, Wiese M. 2008. Identification of putative binding sites of P-glycoprotein based on its homology model. ChemMedChem 3:280–295.

Gottesman MM, Ambudkar SV. 2001. Overview: ABC transporters and human disease. J Bioenerg Biomembr 33:453–458.

Honjo Y, Hrycyna CA, Yan QW, Medina-Perez WY, Robey RW, van de Laar A, Litman T, Dean M, Bates SE. 2001. Acquired mutations in the MXR/BCRP/ABCP gene alter substrate specificity in MXR/BCRP/ABCP-overexpressing cells. Cancer Res 61:6635–6639.

Horn F, Lau AL, Cohen FE. 2004. Automated extraction of mutation data from the literature: application of MuteXt to G protein-coupled receptors and nuclear hormone receptors. Bioinformatics 20:557–568.

Krallinger M, Izarzugaza JM, Rodriguez-Penagos C, Valencia A. 2009. Extraction of human kinase mutations from literature, databases and genotyping studies. BMC Bioinformatics 10(Suppl 8):S1.

Krallinger M, Vazquez M, Leitner F, Salgado D, Chatr-Aryamontri A, Winter A, Perfetto L, Briganti L, Licata L, Iannuccelli M, Castagnoli L, Cesareni G. 2011. The Protein-Protein Interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text. BMC Bioinformatics 12(Suppl 8):S3.

Kuipers R, van den Bergh T, Joosten HJ, Lekanne dit Deprez RH, Mannens MM, Schaap PJ. 2010. Novel tools for extraction and validation of disease-related mutations applied to Fabry disease. Hum Mutat 31:1026–1032.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

Le Saux O, Urban Z, Tschuch C, Csiszar K, Bacchelli B, Quaglino D, Pasquali-Ronchetti I, Pope FM, Richards A, Terry S et al. 2000. Mutations in a gene encoding an ABC transporter cause pseudoxanthoma elasticum. Nat Genet 25:223–227.

Lee LC, Horn F, Cohen FE. 2007. Automatic extraction of protein point mutations using a graph bigram association. PLoS Comput Biol 3:e16.

Matsuo H, Takada T, Ichida K, Nakamura T, Nakayama A, Ikebuchi Y, Ito K, Kusanagi Y, Chiba T, Tadokoro S, Takada Y, Oikawa Y. 2009. Common defects of ABCG2, a high-capacity urate exporter, cause gout: a function-based genetic analysis in a Japanese population. Sci Transl Med 1:5ra11.

Morisaki K, Robey RW, Ozvegy-Laczka C, Honjo Y, Polgar O, Steadman K, Sarkadi B, Bates SE. 2005. Single nucleotide polymorphisms modify the transporter activity of ABCG2. Cancer Chemother Pharmacol 56:161–172.

Rebholz-Schuhmann D, Marcel S, Albert S, Tolle R, Casari G, Kirsch H. 2004. Automatic extraction of mutations from Medline and cross-validation with OMIM. Nucleic Acids Res 32:135–142.

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.

Rubin DL, Thorn CF, Klein TE, Altman RB. 2005. A statistical approach to scanning the biomedical literature for pharmacogenetics knowledge. J Am Med Inform Assoc 12:121–129.

Sarkadi B, Homolya L, Szakacs G, Varadi A. 2006. Human multidrug resistance ABCB and ABCG transporters: participation in a chemoimmunity defense system. Physiol Rev 86:1179–1236.

Saunders RE, Perkins SJ. 2008. CoagMDB: a database analysis of missense mutations within four conserved domains in five vitamin K-dependent coagulation serine proteases using a text-mining tool. Hum Mutat 29:333–344.

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.

Szakacs G, Varadi A, Ozvegy-Laczka C, Sarkadi B. 2008. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). Drug Discov Today 13:379–393.

Tusnady GE, Dosztanyi Z, Simon I. 2005. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. Bioinformatics 21:1276–1277.

Tusnady GE, Simon I. 2001. The HMMTOP transmembrane topology prediction server. Bioinformatics 17:849–850.

Varadi A, Szabo Z, Pomozi V, de Boussac H, Fulop K, Aranyi T. 2011. ABCC6 as a target in pseudoxanthoma elasticum. Curr Drug Targets 12:671–682.

Woodward OM, Kottgen A, Coresh J, Boerwinkle E, Guggino WB, Kottgen M. 2009. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. Proc Natl Acad Sci USA 106:10338–10342.

Woodward OM, Kottgen A, Kottgen M. 2011. ABCG transporters and disease. Febs J 278:3215–3225.

Yeniterzi S, Sezerman U. 2009. EnzyMiner: automatic identification of protein level mutations and their impact on target enzymes from PubMed abstracts. BMC Bioinformatics 10 (Suppl 8):S2.

dc_2020_22

Computational and Structural Biotechnology Journal 16 (2018) 396–403

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Quantitative comparison of ABC membrane protein type I exporter structures in a standardized way

Georgina Csizmadia [a,b], Bianka Farkas [b,c], Zoltán Spagina [b,c], Hedvig Tordai [b], Tamás Hegedűs [a,b,*]

[a] MTA-SE Molecular Biophysics Research Group, Hungarian Academy of Sciences, Budapest, Hungary
[b] Department of Biophysics and Radiation Biology, Semmelweis University, Budapest, Hungary
[c] Faculty of Information Technology, Pázmány Péter Catholic University, Budapest, Hungary

ARTICLE INFO

Article history:
Received 3 September 2018
Received in revised form 12 October 2018
Accepted 13 October 2018
Available online 18 October 2018

Keywords:
ABC proteins
Membrane proteins
Structure comparison
Structure validation
Quantitative structural properties

SUMMARY

An increasing number of ABC membrane protein structures are determined by cryo-electron microscopy and X-ray crystallography, consequently identifying differences between their conformations has become an arising issue. Therefore, we propose to define standardized measures for ABC Type I exporter structure characterization. We set conformational vectors, conftors, which describe the relative orientation of domains and can highlight structural differences. In addition, continuum electrostatics calculations were performed to characterize the energetics of membrane insertion illuminating functionally crucial regions. In summary, the proposed metrics contribute to deeper understanding of ABC membrane proteins' structural features, structure validation, and analysis of movements observed in a molecular dynamics trajectory. Moreover, the concept of standardized metrics can be applied not only to ABC membrane protein structures (http://conftors.hegelab.org).

1. Introduction

ABC membrane proteins play important roles in many physiological processes from bacteria to man. They translocate substrates through the membrane bilayer or regulate channel function involving ATP binding and hydrolysis [1–3]. The functional expression of ABC membrane proteins can be altered either by mutations or regulatory processes [2], leading to various pathological states. The most known disorder caused by an ABC membrane protein is cystic fibrosis. Several mutations in the CFTR (cystic fibrosis transmembrane conductance regulator; ABCC7) chloride channel cause cystic fibrosis, a monogenic disease with high morbidity and mortality [4–6]. Most of the mutations affect protein folding and stability [7]. A decrease in the functional expression of the CFTR channel leads to a reduced chloride conductance in the epithelia resulting in imbalanced salt and water homeostasis. Although high throughput screening efforts resulted in molecules rescuing some CFTR variants (e.g. the G551D mutant), none of the identified drugs sufficiently restores the functional expression of ΔF508, the most frequent CFTR mutant [8–10]. The lack of high-resolution structural information has hindered drug development [11,12]. Because of cystic fibrosis and other ABC protein related disorders, understanding the relations between the conformations and function of ABC membrane proteins is of great importance. An increasing number of ABC exporter structures have been solved recently, which can be divided into two groups [13,14]. There are a higher number of Type I (P-glycoprotein-like) structures determined, which exhibit two large transmembrane domains, usually each built from six helices and two cytosolic nucleotide binding domains (Fig. S1). Their specific feature compared to ABCG-like transporters is the so called intracellular loops or domains, which are the continuation of TM helices in the cytosol. These "loops" contain coupling helices interacting with the nucleotide binding domains [15]. Structures in the absence of ATP exhibit highly separated NBDs or NBDs with contacts only at the opposite site of coupling helix interactions. In this apo conformation the TM domains expose a cavity towards the cytoplasm [14]. Therefore, these conformations are called "bottom-open, inward-facing" and "bottom-closed, inward-facing", respectively. Upon binding of two ATP molecules, NBDs form a tight interaction and rearrange the transmembrane helices to close the cavity at the cytosolic side and open it to the extracellular space. Structures without opening at the extracellular region are also observed and suggested to form an intermediate of the transport cycle [16,17]. These conformations are named "bottom-closed, outward-facing" and "occluded" (bottom- and top-closed), accordingly.

Although all of the ABC membrane protein structures contain valuable information, there are several debates in the field. The "bottom-open, inward-facing" conformations are questioned based on the constant-contact model of the mechanism of function [18]. In addition,

Abbreviations: ABC, ATP binding cassette; CFTR, cystic fibrosis transmembrane conductance regulator; CG, coarse grained; CH, coupling helix; COG, center of geometry; ICD, intracellular domain; NBD, nucleotide binding domain; TH, transmembrane helix; TM, transmembrane; TMD, transmembrane domain.

* Corresponding author.
E-mail address: hegedus@hegelab.org (T. Hegedűs).

https://doi.org/10.1016/j.csbj.2018.10.008

the wide separation of NBDs was not stable in molecular dynamics simulations and the large distance may be contributed to crystal contacts [14,19]. Similarly, the largely top-opened "bottom-closed, outward-facing" structures have also been criticized and suggested to be formed by the lack of the lateral pressure originating from the membrane bilayer during crystallization [20]. It is also challenging to interpret the unusual break of a transmembrane helix in the ABCC7/CFTR chloride channel structures determined by Chen et al. [21–23], since it is not present in other CFTR structures (R. Ford et al. unpublished and [24]) or other transporters of the same subfamily.

Despite the large efforts of transmembrane structure determination in the last decades, ABC membrane proteins, as well as membrane proteins in general are significantly underrepresented in the PDB database compared to globular proteins (http://pdbtm.enzim.hu, http://blanco.biomol.uci.edu/mpstruc). Even the revolution in cryo-electron microscopy (EM) methods [25,26] did not significantly decrease this difference in the last years. Moreover, diffracting crystals and cryo-EM images of membrane proteins exhibit low resolution confining their structure determination [27]. Therefore structure validation and assessing the quality of membrane protein structures are crucial not only in the field of ABC proteins but for all types of membrane proteins. For structure curation and validation the wwPDB has launched a tool, OneDep, based on recommendations of experts in crystallography, NMR, and cryo-EM [28]. These metrics (e.g. clashes between atoms, Wilson B-value [29,30], and the fit between R and R_{free} [31]) are extremely important and are the basis for a high quality database of curated 3D structures. However, they cannot provide higher-level information on the validity or possible distortion of domain-domain orientation caused by experimental conditions. Higher level comparison of protein structures have been initiated and organized in the CoDNaS database, but this tool exposes the limitation of being able to compare the structures of the same protein [32]. A comparative molecular dynamics (MD) study has been recently performed to assess various conformations of ABCB1/MDR1 [33], but the application of MD for a higher number of structures are highly resource intensive.

In order to address issues associated with the increasing number of ABC membrane protein structures and conformations, we aim to define metrics that can characterize structural properties at a higher level. We demonstrate that various vectors defined based on specific structural features of a protein family can highlight specific differences in conformations and alterations in structures. In addition, membrane solvation energy calculations can draw attention to functionally important regions.

2. Methods

2.1. ABC Membrane Protein Structures

Coordinates of ABC Type I exporters and information on membrane orientations were downloaded from the OPM database or from the PDB followed by the calculation of membrane region using the PPM server (Table S1) [34,35]. The structure determination method was collected from the PDB files, while the conformational state was determined by visual inspection. Transmembrane helix boundaries were extracted from OPM and PPM. Other regions, such as coupling helices, were identified semi-manually. PyMOL was used for structure and surface electrostatics visualization (The PyMOL Molecular Graphics System, Version 1.8.4 Schrödinger, LLC). The orientation of the structures was standardized by the rotation and translation of structures to a selected reference structure. As a reference we selected a structure of TM287/288 (PDBID: 3QF4), which exhibits intermediate 3D properties, such as inward-facing TMDs and bottom-closed NBDs [36].

2.2. Comparison of Transmembrane Region Localizations

Selected resources include OPM [34,35], PDBTM [37,38], and MEMPROTMD [39] databases. The positioning of the structures in the membrane bilayer was characterized by the tilting angle and the localization along the Z axis. We compared the values from each set to the values from the OPM as a reference. The protein tilting was calculated by the angle between the membrane normal and the principal axis of the protein, which axis is set by the bisector between the THX1 and THX2 conftors (CONFormational vecTORs; TH4–5 and TH10–11). In the case of the CG structures the membrane normal was calculated by subtracting the center of geometry of PO4 and NC3 beads in the upper leaflet from that in the lower leaflet. The vector normal of the plane defined by three DUMMY atoms were used for OPM entries. The difference in the membrane centers was determined after structural alignment. Regarding the CG structures, the back-bone beads were aligned to the Cα atoms of the all-atom structures. The reference membrane center was calculated as the center of geometry (COG) of the DUMMY membrane atoms from OPM. The membrane center of systems from CG simulations was determined by the COG of the PO4 and NC3 beads.

2.3. Calculation of Helix Properties and Vectors

The bending, rotation, and twist of helices were calculated using MDAnalysis and HELANAL [40,41]. Conftors were calculated using Python scripts combining the numpy and MDAnalysis packages (Table S2). To provide a simplified and visual comparison of structures, conftors were plotted using Python Matplotlib.

2.4. Molecular Dynamics Simulations

Coarse-grained simulations were performed with structures of PDB entries 2HYD, 3Qf4, 4KSB, 5UAK, 5TSI, 5UJ9, and 4PL0 using GROMACS with the MARTINI force field (elnedyn22) [42,43]. Trajectories of all-atom simulations have been obtained using GROMACS as described earlier [44]. More details on simulation parameters are in the supplementary material.

2.5. Electrostatics Calculations

PDB2PQR [45] was run with PARSE force field [46], pH 7.0 and the option to create input template. Structural preparation needed for low resolution structures was made by VMD's Automatic PSF Builder. The input template for APBS [47] was modified to add 150 mM Na^+ and Cl^- ions with charge 1 and −1, and radius 0.95 and 1.81 Å, respectively. For membrane solvation calculations APBSmem [48] was run with the following parameters (Table S4): Grid dimensions and fine grid size for x and y coordinates was collected from the input file of APBS. The z coordinate of the fine grid size was $-2^*z_{min} + 40$ Å, where z_{min} is the smallest z coordinate from the PQR file. This way the whole protein was included in the fine grid, even when membrane was moved with ±20 Å. Medium grid size was 2 times and coarse grid size was 5 times of the fine grid size. Grid dimension was 161 for each axis. The membrane thicknesses from OPM were used and the flooding algorithm of APBSmem was used as the membrane filling method.

3. Results

3.1. Positioning of ABC Type I Exporters in Membrane Bilayers

The tilting of the protein relative to the bilayer normal and the location of the hydrophobic bilayer core are important features of protein conformation. The experimental information on tilting and insertion at the atomic resolution is highly limited, thus we assessed the membrane orientation of ABC membrane proteins using various computational methods. We compared data on membrane bilayer boundaries from OPM [34,35], PDBTM [37,38], and MEMPROTMD [39] databases. OPM calculates and minimizes the transfer free energy of transmembrane proteins at different values of distance from the bilayer center, bilayer thickness, and tilting. PDBTM's TMDET algorithm is a

geometrical approach utilizing an objective function dependent on amino acid hydrophobicity. In the MEMPROTMD database membrane protein structures in a bilayer are generated using coarse grain simulations. Since we found ABC membrane proteins with large conformational changes in MEMPROTMD (e.g. transition from the bottom-open to the bottom-closed conformation), which changes may influence the interactions with lipid molecules, we performed CG simulations using MARTINI [43,49] on a selected set of ABC membrane protein structures as well.

We extracted the tilt angle and the location of the bilayer around each ABC membrane protein structure (Fig. 1, Fig. S2, and Table S1). The tilting was calculated as the angle between the membrane normal and the principal axis of the protein. This axis is set by the bisector between the vectors defined by two pairs of TM helices (TH4–5 and TH10–11), which cross-over to the opposite NBD, discussed below as THX1 and THX2 conftors (CONFormational vecTORS).

In most of the cases the difference in tilting obtained from different databases is negligible (below 6°). The largest differences were 10–12°, which seem significant by visual inspection, but correspond to a difference in immersion by only a few amino acids (1–2 helical turns) for the helices most distant from the central z-axis.

The bilayer location around the protein was characterized by the distance along the z-axis between the COG of transmembrane helices based on OPM annotation and the bilayer center. Most of the differences were negligible in z-location, except for a few structures, including 2HYD, 3QF4, 4KSB, and 5UAK. Therefore, we assessed the membrane embedment of these structures by membrane solvation calculation using APBS (Fig. 2).

Membrane solvation energy of 3QF4 has a minimum at the 0 distance, which means an agreement with OPM. Membrane insertion predictions of 4KSB by APBS agree with that of MEMPROTMD. However, the membrane insertion for 2HYD derived from APBS calculations is similar to that from PDBTM.

We conclude that the OPM, PDBTM, MEMPROTMD, and our CG simulations exhibit only slight differences in the investigated measures and it cannot be ascertained which method provides a prediction correlating the best with in vivo conditions. In this study we used the OPM database, since it exhibited the less deviation from the other methods.

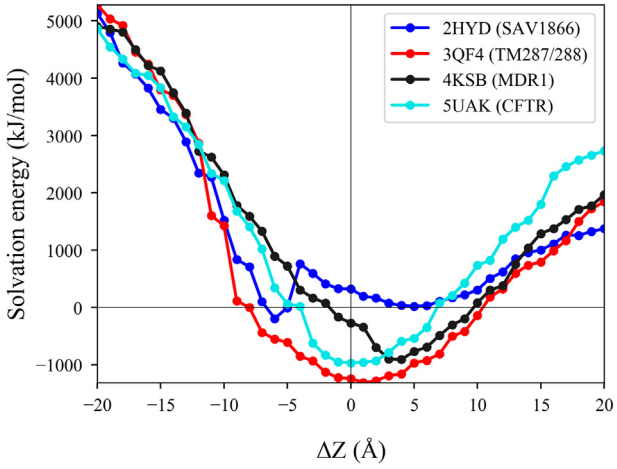


Fig. 2. Orientation in membranes assessed by APBS calculations. Membrane solvation energy was calculated using APBSmem and shown for selected ABC membrane proteins. Calculations were performed for each protein at different positions of the bilayer (1 Å steps). Zero point is the location of the bilayer defined by OPM.

3.2. Helix Bending and End-Locations Facilitate Structure Comparison and Highlight Critical Differences Between Conformations

In order to understand distinctive features of various ABC Type I exporter conformations and select critical 3D properties for the description of structures, we characterized various geometric properties of TM helices. We calculated the bending, the rotation, and the twist of helices [40] (TH2–5 and TH8–11), which have longer intracellular parts interacting with NBDs for each conformation class. In the "occluded" class there are positions with higher values of bending angles (Fig. S3). The inspection of these conformations revealed that there was a particular structure (T1SS, Type-1 secretion system, PDBID: 5L22) with high bending angle values originating from breaks in several TM helices.

In order to visualize the relative orientation of TM helices, their intracellular and extracellular end positions were projected into 2D (Fig. S4). The application of these projections is demonstrated on the

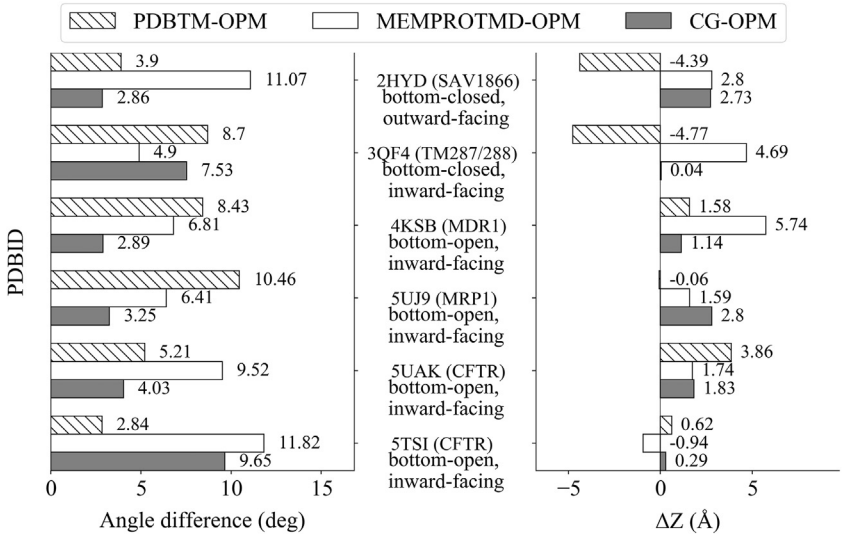


Fig. 1. Comparison of bilayer location around ABC membrane proteins by in silico methods. Differences in the tilting angle of proteins in the membrane and the z-positioning of the membrane bilayer around proteins are depicted for selected ABC Type I exporter structures. Values extracted from PDBTM, MEMPROTMD, and our CG simulations are compared to values from OPM. OWF: outward-facing, IWF: inward-facing. See also Fig. S2.

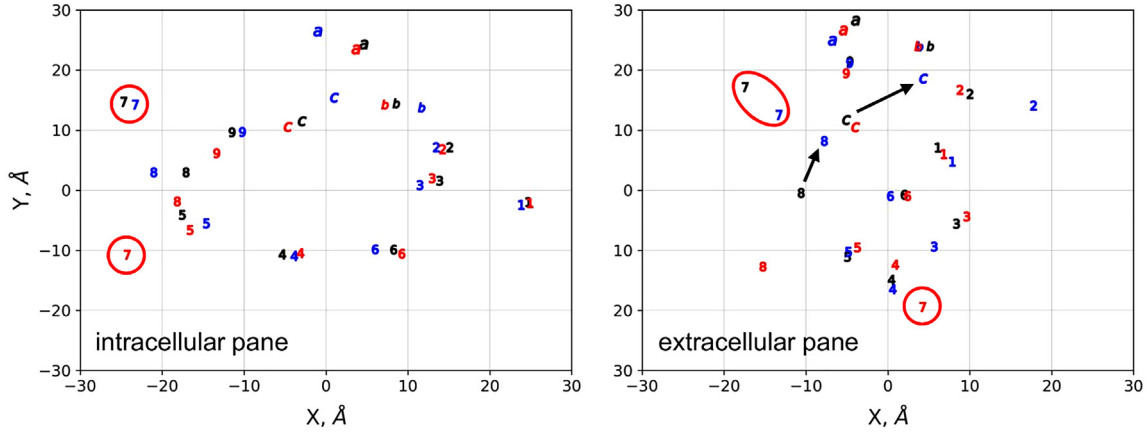


Fig. 3. The distance between the ends of TM helices highlight differences in conformations. The intracellular (left) and extracellular (right) ends of TM helices in apo (PDBID: 5UAK, black; Fay et al., red) and ATP-bound (PDBID: 5W81, blue) CFTR structures were projected into 2D. The unusual localization of TH7 is highlighted by red circles. Arrows indicate the altered extracellular position of TH8 and TH12 in the ATP-bound structure. See also Fig. S4.

CFTR (ABCC7) cryo-EM structures [21–24], which have been received an intense attention. The projections in Fig. 3 reveal that TH7 in the structure of Fay et al. [24] is located at a completely different position compared to other structures [21–23] (and also to an unpublished electron density map by Ford et al., University of Manchester, UK). The relocated TH7 and TH8 are claimed to be a result of the ~200 a.a. disordered regulatory domain connected to the N-terminus of TH7. Fig. 3 also reveals that there are only small differences in the relative localization of the intracellular ends of TM helices between the apo and ATP-bound conformations, while extracellular ends of TH8 and TH12 are repositioned in the ATP-bound structure (Fig. 3, arrows). The comparison of these two structures reveals the closure of the NBDs and the associated dislocation of some intracellular helices, while the TM helices are practically unchanged, except the above mentioned two extracellular ends.

3.3. Defining Conftors Sensitive to Overall and relative Domain Conformations

We clustered ABC membrane protein structures on pairwise RMSD values. However, because of the composite nature of RMSD, some members of a cluster also exhibit crucial structural variations compared to other members (Fig. S5). Therefore, we aimed to characterize conformations using vectors defined in a manner to pin differences in intra- and inter-domain arrangements (Fig. 4).

We propose that simplified representation of protein conformations using conftors (CONFormational vecTORS), carefully selected standardized vectors based on high resolution structures help to interpret differences between protein structures. As a proof of principle, we demonstrate the definition and application of conftors in the case of ABC transmembrane proteins.

Since the function of ABC membrane proteins is coupled to conformational changes in the TM domains, we selected vectors that are capable to describe differences in the orientation of transmembrane helices. For example, to compare the level of opening towards the extracellular space (e.g. the conformation of the outward-open 2HYD and the outward-closed, occluded conformation of 4PL0, Fig. S6) we defined the THV1–2, THX1–2 and THC3–9 conftors (Table S2). As the orientation and bending of an individual α-helix may be deviated in a specific structure, defining the ends of TMD conftors as the COG of Cα at the ends of two α-helices may be justified in some cases (THX, Fig. 4a). These THX conftors characterize the conformation of the transmembrane helices, which cross over from one TMD to the opposite NBD. While the angle between THV1 and THV2 conftors can separate

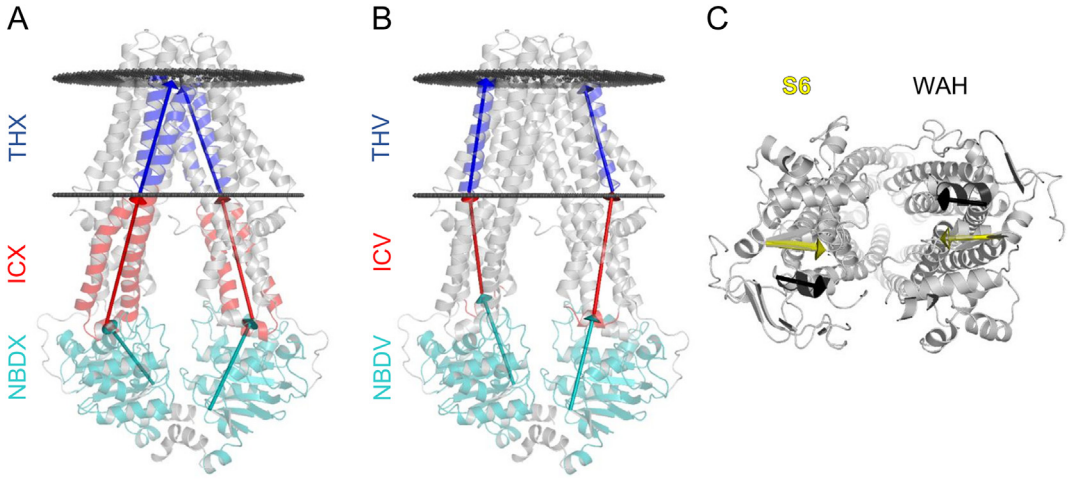


Fig. 4. ABC conftors: dedicated vectors to describe ABCType I exporter conformation. (a, b) Vectors are defined by either a single Cα or the center of mass of more Cα. Conftors in the membrane region point from the intracellular to the extracellular ends of helices (blue; a: THX1 and THX2, b: THV1 and THV2). Conftors for the intracellular domains (a: ICX1 and ICX2, b: ICV1 and ICV2) and nucleotide binding domains (a: NBDX1 and NBDX2, b: NBDV1 and NBDV2) are red and teal, respectively. **(c)** Conftors are also defined between the Walker A helices (black) and strand S6 (yellow with black line) of the opposite NBDs.

only the bottom-closed, top-closed conformations with an average value of 23° compared to all other conformations with values between 36 and 40°, angle enclosed by the THX conftors is able to make distinction between "bottom-open, inward-facing" (46°), "bottom-closed, top-closed" (26°), and the two other conformations (39° and 35°) (Table S3).

Since the so-called intracellular loops or domains (ICDs), which are the continuation of TM helices, can enclose an angle with the membrane-embedded parts of the TM helices, we set separate conftors, ICV1–2 and ICX1–2 and for the transmembrane and intracellular parts of the TM helices (Fig. 4, Table S2). The angle between ICV1 and ICV2 differentiate the inward-facing (43° and 38°), the outward-facing (60°), and occluded (53°) conformations (Table S3).

The closed and open conformation of NBDs is usually determined easily by visual inspection. However, the extent of their separation and especially their orientation and rotation relative to each other and to the TMDs remain hidden. Therefore, we defined conftors NBDV1–2 and NBDX1–2 pointing from the COG of coupling helices to the first residue of S9, the last strand in NBD with small deviation among structures (Fig. 4a–b). NBDX conftors have slightly lower values for the inward-facing conformations compared to outward-facing or top-closed structures. The NBDX1/NBDX2$_{ext}$ conftor's length (the distance of the NBD/TMD interface regions) reveals differences not only between inward-facing and outward-facing conformations, but also between "bottom-open, inward-facing" and "bottom-closed, inward-facing" conformations (Table S3). The S6 conftor defined by the opposite strands S6 and the WAH conftor based on the opposite α-helices incorporating the Walker A motif are anticipated to depict the orientation of NBDs relative to each other and also to the TMDs (Fig. 4c).

3.4. Application of Conftors

Conftors can also be used for visualization purposes. For example, the degree of opening is shown by the THC conftors (Fig. S6). A more exquisite example includes structural models of CFTR. The inward-facing apo cryo-EM structure exhibits properties similar to other inward-facing structures (PDBID: 5U71, Fig. 5). Since it does not exhibit an open pathway for chloride, a complex modeling complemented with experiments has been performed by Das et al. to generate a conformation with open channel [50]. This study is important, since even a phosphorylated and ATP bound CFTR structure (PDBID: 5W81) is not opened [23].

However, the conftors of these structures reveals that the models exhibit large deviations from known structures, which differences are hidden or attenuated in the 3D structure (Fig. 5). Their models show differences in the transmembrane regions, which can be anticipated because channel opening may require different conformational changes in the TM domains compared to active ABC transporters. In contrast, the differences in the intracellular domain orientations indicate inaccuracies.

Conftors can also be effectively used for analysis of trajectories from molecular dynamics simulations. We have recently investigated the dynamics of the inward-facing CFTR cryo-EM structure and noticed the closure of the nucleotide binding domains (Fig. 3a in [44]). As the measures we have calculated were not sufficient to fully understand the movements of the protein in detail, we analyzed the trajectories employing conftors. Several angles between various conftors were calculated over the trajectories and plotted in Fig. 6. The conformation of the membrane embedded parts of the TM helices did not change largely, albeit a small decrease in angles of THX1 and THX2 can be observed (Fig. 6a, black), that may arise from the lateral pressure of the lipid bilayer [20]. The increase in the NBDX1/2 angle indicates that the bottom of the NBDs (the opposite site of the TMD/NBD interface) gets closer to each other. This event is clearly shown by the length of the NBDX1/2$_{int}$ "distance" conftor (Fig. 6b, cyan), as NBDX1/2$_{int}$ highly fluctuates till ~18 ns, when it became stable around 50 Å. The conftors describing the distance between Walker A and Signature motifs (WAH1-SIG2 and WAH2-SIG1) indicate the higher separation of WAH1 and SIG2 in the nonfunctional degenerate ATP-binding site-1, when compared to that of the canonical site-2 (Fig. 6c). This observation indicates an asymmetry in the association of the two NBDs.

3.5. Electrostatics Calculations Highlight Structural Hot Spots

Measures representing physicochemical properties are also crucial for structure characterization. Protein surface electrostatics can be

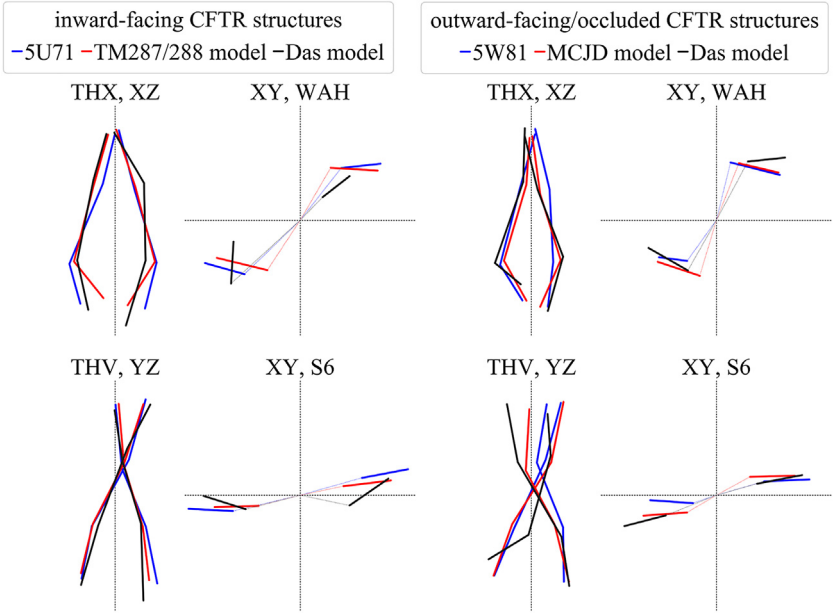


Fig. 5. Conftors highlight important similarities and differences among CFTR structural models. The open Das model exhibits large deviations in NBD rotations indicated by WAH and S9 conftors. The closed Das model shows differences in both TM helix and NBD conformations when compared to other structural models, and the orientation of an NBD relative to the TMDs is not observed in any other ABC Type I exporter structure. CFTR cryo-EM structures are PDBIDs 5U71 and 5W81. TM287/288 and McjD based homology models have been generated by Corradi et al. CFTR models by Das et al. are from http://troll.med.unc.edu/cftr/. See also Fig. S6.

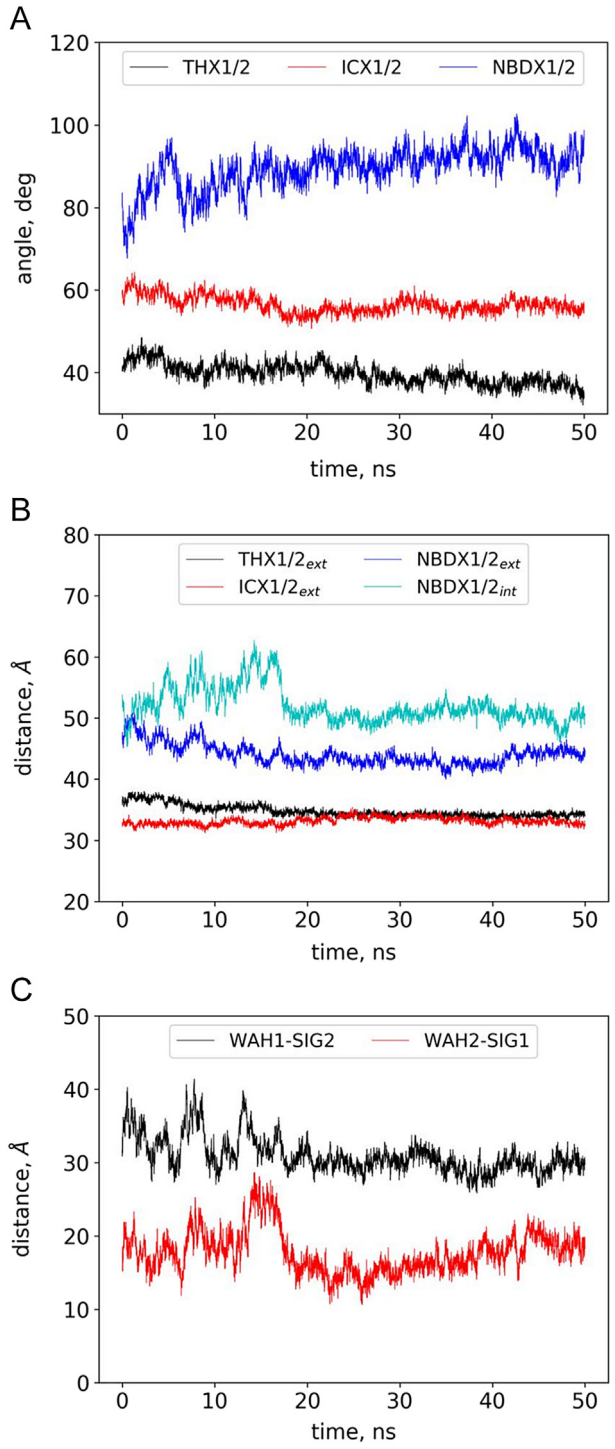


Fig. 6. Application of conftors for MD analysis. Angles between conftors (**a**) and lengths of conftors (**b, c**) were calculated over the trajectory of an MD simulation with the CFTR bottom-open, inward-facing structure (PDBID: 5UAK). NBD1 and NBD2 got into contact at around 18 ns of the simulation.

used to characterize the overall conformation of transmembrane domains. The positive-inside rule should manifest in most of the transmembrane proteins [51,52], thus in structural context the amino acids in the region of the inner membrane boundary are expected to build up a positively charged ring around the protein. This positive ring ("blue collar") is present in many experimental structures, such as calcium ATPases [53]. A quantitative description of this ring is quite challenging, but calculating and analyzing surface electrostatics in individual cases can be informative to localize protein-protein interactions. At interaction sites the positive ring is expected to be ceased. Indeed, at the intramolecular interaction site of the L0/Lasso motive preceding the first transmembrane domain in ABCC proteins, such as ABCC1/MRP1 (Fig. S7) and ABCC7/CFTR [44] the positive ring breaks with a hydrophobic spot. The functionally important amphipathic α-helix of the L0/Lasso motif binds to this patch as observed in several cryo-EM structures.

APBS can also be applied for membrane solvation calculations. We computed the membrane solvation energy for each ABC Type I exporters and for each transmembrane helix from per amino acid contribution to the solvation, using APBSmem (Fig. 7). The total solvation energy spans from low negative to high positive values and we could not detect any correlation between the energy and some other property, such as determination method or resolution. Interestingly, the structures with the lowest and highest solvation energies were solved by X-ray. The two structures with the highest energy (SAV1866, PDBID: 2HYD and MsbA, PDBID: 3B60) are the two outward-open conformations suggesting that these widely open conformations may be caused by the lack of a bilayer under crystallization conditions. In contrast, the solvation energy of the outward-open MRP1 (PDBID: 6BHU) [54] and MDR1 (PDBID: 6C0V) [55] conformations, which are less open towards the extracellular space, is small.

4. Discussion

An increasing number of membrane protein structures are being determined. In the accompanying papers the new structures are compared to previously known ones and this comparison is usually semi-quantitative and not complete. In addition, while basic important metrics, which are general for all types of proteins (e.g. phi/psi angles, fit of the model to experimental data) are required to be calculated for validation, no quantitative and standardized measures have been defined to characterize geometric and physicochemical properties of structures. Importantly, earlier it was relatively straightforward which known conformations should be used for comparison to the new conformation because of the low number of available high-resolution structures. With the increasing number of solved structures, a set of standardized measures help to avoid a bias in reference structure selection and also in selecting structural regions for demonstrating novel and intriguing properties of the newly determined conformation. At this moment the low number of structures in certain membrane protein families limits the definition of conftors. For example, only one conformation has been determined for the ABCG2 and ABCG5/ABCG8 Type II exporters [13,56,57], therefore we could not test the usefulness of any conftor for the ABCG subfamily. Since in the case of ABC Type I exporters there are a larger number of "bottom-open" and "bottom-closed" conformations, numerous conftors could be defined and validated. Importantly, the existence of various conformations enabled us to evaluate vectors as conftors and discard which do not deliver information (e.g. THV conftors are not discriminative for the outward-facing and the inward-facing conformations, while THX conftors can differentiate these conformations well; Table S3).

Using various quantitative measures, we show how to demonstrate crucial differences between CFTR conformations for researchers other than structural biologists, since these differences, even in an ambiguously-defined form, have generated uncertainty in the field regarding the validity of the experimental structures. Actually, the slight differences between the apo and ATP-bound conformations, the membrane solvation of TH8 [44], and the dislocated TH7 and TH8 in the CFTR structure of Fay et al. [24] (Fig. 3) suggest that most likely the lipid environment (micelle) can have a profound effect on CFTR structure [58]. In the case of structures determined in a micelle, it is hard to imagine other factor than the lipid/detergent environment, playing a role in maintaining the conformation of the TM helices in the ATP-bound conformation highly similar to the apo conformation, while the intracellular parts of these helices (ICDs) and the NBDs exhibit a

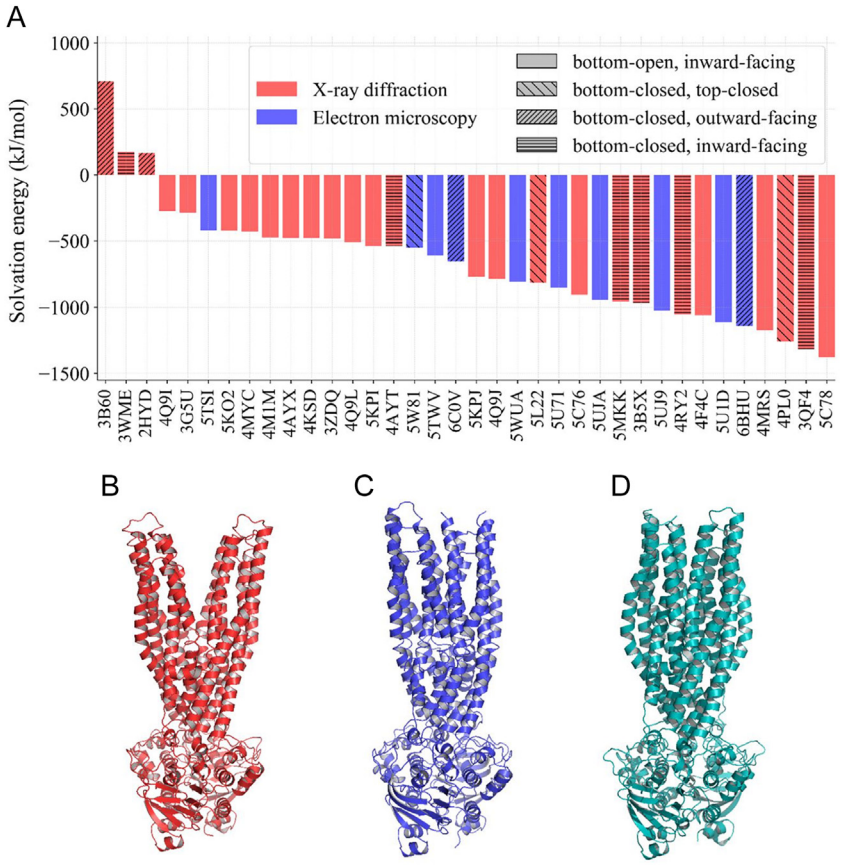


Fig. 7. Membrane solvation energy values are high for conformations with a large outward-facing cavity. (a) APBSmem was used to calculate membrane solvation energy. Conformations with smaller (MRP1, PDBID: 6BHU and MDR1, PDBID: 6C0V) and larger (SAV1866, PDBID: 2HYD and MRP1, PDBID: 3B60) outward-facing cavities exhibit negative and positive solvation energies, respectively. Differences in the level of the opening are shown for Sav1866 (**b**) and MDR1/ABCB1(**c**). The wide opening of Sav1866 has been questioned and an alternative ATP-bound conformation has been proposed (Protein Model Database: PM0075213) (**d**).

significant closure. On the other hand, the highly deviated TH7 and TH8 conformations of CFTR from different laboratories (Fig. 3) underscore the mobile nature of this region, which phenomenon has already been indicated by experiments. Most importantly, the above mentioned metrics can be useful not only for structure validation and comparison, but understanding the conformational changes associated to function (Fig. 3 and Fig. S4) and dynamics (Fig. 6). We emphasize that outliers of angles or membrane insertion energies may not indicate problems with a structure but may sign structurally or functionally important regions, as outlying phi/psi angles in the case of annexin [59].

The utilization of standardized metrics for structure validation and structure comparison aid the rigorous description of structural features and advance our knowledge on function-related conformations, thus help to understand the effect of mutations on protein structure and promote structure-based drug design. The proposed and similar metrics can be applied not only to the ABC membrane proteins. However, for other classes of proteins several vectors should be tested by an expert on the given protein family as long as no automatic algorithms are available. To overcome the difficulties of manual definitions of conftors, we are developing a web application and algorithms for generalized application of conftors (http://conftors.hegelab.org).

Acknowledgements

This work was supported by National Research, Development and Innovation Ofice, Hungary (NKFIH K 111678) and Cystic Fibrosis Foundation, USA(HEGEDU18I0). We acknowledge NIIF National Information Infrastructure Development Institute (http://www.niif.hu/en), and MTA Wigner GPU Laboratory (http://gpu.wigner.mta.hu) for awarding us access to resources based in Hungary, and the support of their stuff is gratefully acknowledged. Donation of a Quadro P6000 GPU by NVIDIA Corporation is greatly appreciated.

Author Contributions

G.C., B.F., Z.S., and T.H. conducted the experiments; H.T. and T.H. designed the experiments and wrote the paper.

Declaration of Interests

The authors declare no competing interests.

Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2018.10.008.

References

[1] Szakács G, Váradi A, Özvegy-Laczka C, Sarkadi B. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME–Tox). Drug Discov Today 2008;13:379–93. https://doi.org/10.1016/J.DRUDIS.2007.12.010.
[2] George AM, editor. ABC Transporters - 40 Years on. Cham: Springer International Publishing; 2016. https://doi.org/10.1007/978-3-319-23476-2.
[3] Locher KP. Review. Structure and mechanism of ATP-binding cassette transporters. Philos Trans R Soc Lond B Biol Sci 2009;364:239–45. https://doi.org/10.1098/rstb.2008.0125.
[4] Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989;245:1066–73. https://doi.org/10.1126/SCIENCE.2475911.
[5] Boucher RC. Airway Surface Dehydration in Cystic Fibrosis: Pathogenesis and Therapy. Annu Rev Med 2007;58:157–70. https://doi.org/10.1146/annurev.med.58.071905.105316.

[6] Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. Nat Rev Genet 2015;16:45–56. https://doi.org/10.1038/nrg3849.
[7] Veit G, Avramescu RG, Chiang AN, Houck SA, Cai Z, Peters KW, et al. From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. Mol Biol Cell 2016;27:424–33. https://doi.org/10.1091/mbc.e14-04-0935.
[8] Boyle MP, Bell SC, Konstan MW, McColley SA, Rowe SM, Rietschel E, et al. A CFTR corrector (lumacaftor) and a CFTR potentiator (ivacaftor) for treatment of patients with cystic fibrosis who have a phe508del CFTR mutation: a phase 2 randomised controlled trial. Lancet Respir Med 2014;2:527–38. https://doi.org/10.1016/S2213-2600(14)70132-8.
[9] Cholon DM, Quinney NL, Fulcher ML, Esther CR, Das J, Dokholyan NV, et al. Potentiator ivacaftor abrogates pharmacological correction of ΔF508 CFTR in cystic fibrosis. Sci Transl Med 2014;6:246ra96. https://doi.org/10.1126/scitranslmed.3008680.
[10] Veit G, Avramescu RG, Perdomo D, Phuan P-W, Bagdany M, Apaja PM, et al. Some gating potentiators, including VX-770, diminish ΔF508-CFTR functional expression. Sci Transl Med 2014;6:246ra97. https://doi.org/10.1126/scitranslmed.3008889.
[11] Meng X, Clews J, Martin ER, Ciuta AD, Ford RC. The structural basis of cystic fibrosis. Biochem Soc Trans 2018;46(5):1093–8. https://doi.org/10.1042/BST20180296.
[12] Veit G, Xu H, Dreano E, Avramescu RG, Bagdany M, Beitel LK, et al. Structure-guided combination therapy to potently improve the function of mutant CFTRs. Nat Med 2018:1. https://doi.org/10.1038/s41591-018-0200-x Oct 8;. [Epub ahead of print].
[13] Lee J-Y, Kinch LN, Borek DM, Wang J, Wang J, Urbatsch IL, et al. Crystal structure of the human sterol transporter ABCG5/ABCG8. Nature 2016;533:561–4. https://doi.org/10.1038/nature17666.
[14] Locher KP. Mechanistic diversity in ATP-binding cassette (ABC) transporters. Nat Struct Mol Biol 2016;23:487–93. https://doi.org/10.1038/nsmb.3216.
[15] Dawson RJP, Locher KP. Structure of a bacterial multidrug ABC transporter. Nature 2006;443:180–5. https://doi.org/10.1038/nature05155.
[16] Choudhury HG, Tong Z, Mathavan I, Li Y, Iwata S, Zirah S, et al. Structure of an antibacterial peptide ATP-binding cassette transporter in a novel outward occluded state. Proc Natl Acad Sci U S A 2014;111:9145–50. https://doi.org/10.1073/pnas.1320506111.
[17] Perez C, Gerber S, Boilevin J, Bucher M, Darbre T, Aebi M, et al. Structure and mechanism of an active lipid-linked oligosaccharide flippase. Nature 2015;524:433–8. https://doi.org/10.1038/nature14953.
[18] George AM, Jones PM. Perspectives on the structure-function of ABC transporters: the Switch and Constant Contact models. Prog Biophys Mol Biol 2012;109:95–107. https://doi.org/10.1016/j.pbiomolbio.2012.06.003.
[19] Gyimesi G, Ramachandran S, Kota P, Dokholyan NV, Sarkadi B, Hegedus T. ATP hydrolysis at one of the two sites in ABC transporters initiates transport related conformational transitions. Biochim Biophys Acta 2011;1808:2954–64. https://doi.org/10.1016/j.bbamem.2011.07.038.
[20] Stockner T, de Vries SJ, Bonvin AMJJ, Ecker GF, Chiba P. Data-driven homology modelling of P-glycoprotein in the ATP-bound state indicates flexibility of the transmembrane domains. FEBS J 2009;276:964–72. https://doi.org/10.1111/j.1742-4658.2008.06832.x.
[21] Liu F, Zhang Z, Csanády L, Gadsby DC, Chen J. Molecular Structure of the Human CFTR Ion Channel. Cell 2017;169:85–95. https://doi.org/10.1016/j.cell.2017.02.024 [e8].
[22] Zhang Z, Chen J. Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator. Cell 2016;167:1586–97. https://doi.org/10.1016/j.cell.2016.11.014 [e9].
[23] Zhang Z, Liu F, Chen J. Conformational Changes of CFTR upon Phosphorylation and ATP Binding. Cell 2017;170:483–91. https://doi.org/10.1016/j.cell.2017.06.041 [e8].
[24] Fay JF, Aleksandrov LA, Jensen TJ, Cui LL, Kousouros JN, He L, et al. Cryo-EM Visualization of an Active High Open Probability CFTR Anion Channel. Biochemistry 2018. https://doi.org/10.1021/acs.biochem.8b00763 Oct 3. [Epub ahead of print].
[25] Callaway E. The revolution will not be crystallized: a new method sweeps through structural biology. Nature 2015;525:172–4. https://doi.org/10.1038/525172a.
[26] Cheng Y. Single-Particle Cryo-EM at Crystallographic Resolution. Cell 2015;161:450–7. https://doi.org/10.1016/j.cell.2015.03.049.
[27] Robert X, Kassis-Sahyoun J, Ceres N, Martin J, Sawaya MR, Read RJ, et al. X-ray diffraction reveals the intrinsic difference in the physical properties of membrane and soluble proteins. Sci Rep 2017;7:17013. https://doi.org/10.1038/s41598-017-17216-1.
[28] Gore S, Sanz García E, Hendrickx PMS, Gutmanas A, Westbrook JD, Yang H, et al. Validation of Structures in the Protein Data Bank. Structure 2017;25:1916–27. https://doi.org/10.1016/j.str.2017.10.009.
[29] Read RJ, Adams PD, Arendall WB, Brunger AT, Emsley P, Joosten RP, et al. A new generation of crystallographic validation tools for the protein data bank. Structure 2011;19:1395–412. https://doi.org/10.1016/j.str.2011.08.006.
[30] Wilson AJC. Determination of Absolute from Relative X-Ray Intensity Data. Nature 1942;150:152. https://doi.org/10.1038/150152a0.
[31] Brünger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 1992;355:472–5.
[32] Monzon AM, Rohr CO, Fornasari MS, Parisi G, et al. CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state. Database (Oxford) 2016:2016.
[33] Condic-Jurkic K, Subramanian N, Mark AE, O'Mara ML. The reliability of molecular dynamics simulations of the multidrug transporter P-glycoprotein in a membrane environment. PLoS One 2018;13:e0191882. https://doi.org/10.1371/journal.pone.0191882.
[34] Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI. Positioning of proteins in membranes: a computational approach. Protein Sci 2006;15:1318–33. https://doi.org/10.1110/ps.062126106.
[35] Lomize AL, Pogozheva ID, Mosberg HI. Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. J Chem Inf Model 2011;51:930–46. https://doi.org/10.1021/ci200020k.
[36] Hohl M, Briand C, Grütter MG, Seeger MA. Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nat Struct Mol Biol 2012;19:395–402. https://doi.org/10.1038/nsmb.2267.
[37] Tusnády GE, Dosztányi Z, Simon I. PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res 2005;33:D275–8. https://doi.org/10.1093/nar/gki002.
[38] Kozma D, Simon I, Tusnády GE. PDBTM: Protein Data Bank of transmembrane proteins after 8 years. Nucleic Acids Res 2013;41:D524–9. https://doi.org/10.1093/nar/gks1169.
[39] Stansfeld PJ, Goose JE, Caffrey M, Carpenter EP, Parker JL, Newstead S, et al. MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. Structure 2015;23:1350–1361. doi:10.1016/j.str.2015.05.006.
[40] Bansal M, Kumar S, Velavan R. HELANAL: a program to characterize helix geometry in proteins. J Biomol Struct Dyn 2000;17:811–9. https://doi.org/10.1080/07391102.2000.10506570.
[41] Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem 2011;32:2319–27. https://doi.org/10.1002/jcc.21787.
[42] Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 2013;29:845–54. https://doi.org/10.1093/bioinformatics/btt055.
[43] Periole X, Cavalli M, Marrink S-J, Ceruso MA. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. J Chem Theory Comput 2009;5:2531–43. https://doi.org/10.1021/ct9002114.
[44] Tordai H, Leveles I, Hegedűs T. Molecular dynamics of the cryo-EM CFTR structure. Biochem Biophys Res Commun 2017;491:986–93. https://doi.org/10.1016/j.bbrc.2017.07.165.
[45] Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 2007;35:W522–5. https://doi.org/10.1093/nar/gkm276.
[46] Tang CL, Alexov E, Pyle AM, Honig B. Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. J Mol Biol 2007;366:1475–96. https://doi.org/10.1016/j.jmb.2006.12.001.
[47] Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci 2018;27:112–28. https://doi.org/10.1002/pro.3280.
[48] Callenberg KM, Choudhary OP, de Forest GL, Gohara DW, Baker NA, Grabe M. APBSmem: a graphical interface for electrostatic calculations at the membrane. PLoS One 2010;5:e12722. https://doi.org/10.1371/journal.pone.0012722.
[49] Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. J Phys Chem B 2007;111:7812–24. https://doi.org/10.1021/jp071097f.
[50] Das J, Aleksandrov AA, Cui L, He L, Riordan JR, Dokholyan NV. Transmembrane helical interactions in the CFTR channel pore. PLoS Comput Biol 2017;13:e1005594. https://doi.org/10.1371/journal.pcbi.1005594.
[51] Elazar A, Weinstein JJ, Prilusky J, Fleishman SJ. Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology. Proc Natl Acad Sci U S A 2016;113:10340–5. https://doi.org/10.1073/pnas.1605888113.
[52] von Heijne G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. Nature 1989;341:456–8. https://doi.org/10.1038/341456a0.
[53] Penniston JT, Padányi R, Pászty K, Varga K, Hegedus L, Enyedi A. Apart from its known function, the plasma membrane Ca^{2+}ATPase can regulate Ca^{2+} signaling by controlling phosphatidylinositol 4,5-bisphosphate levels. J Cell Sci 2014;127:72–84. https://doi.org/10.1242/jcs.132548.
[54] Johnson ZL, Chen J. ATP Binding Enables Substrate Release from Multidrug Resistance Protein 1. Cell 2018;172:81–9. https://doi.org/10.1016/j.cell.2017.12.005 [e10].
[55] Kim Y, Chen J. Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation. Science 2018;359:915–9. https://doi.org/10.1126/science.aar7389.
[56] Taylor NMI, Manolaridis I, Jackson SM, Kowal J, Stahlberg H, Locher KP. Structure of the human multidrug transporter ABCG2. Nature 2017;546:504–9. https://doi.org/10.1038/nature22345.
[57] Jackson SM, Manolaridis I, Kowal J, Zechner M, Taylor NMI, Bause M, et al. Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. Nat Struct Mol Biol 2018;25:333–40. https://doi.org/10.1038/s41594-018-0049-1.
[58] Warschawski DE, Arnold AA, Beaugrand M, Gravel A, Chartrand É, Marcotte I. Choosing membrane mimetics for NMR structural studies of transmembrane proteins. Biochim Biophys Acta 2011;1808:1957–74. https://doi.org/10.1016/j.bbamem.2011.03.016.
[59] Shao C, Zhang F, Kemp MM, Linhardt RJ, Waisman DM, Head JF, et al. Crystallographic analysis of calcium-dependent heparin binding to annexin A2. J Biol Chem 2006;281:31689–95. https://doi.org/10.1074/jbc.M604502200.

Bioinformatics, 36(8), 2020, 2595–2598
doi: 10.1093/bioinformatics/btz539
Advance Access Publication Date: 10 July 2019
Applications Note

OXFORD

Structural bioinformatics

MemBlob database and server for identifying transmembrane regions using cryo-EM maps

Bianka Farkas [1,2,3,†], Georgina Csizmadia[1,2,†], Eszter Katona[1,4], Gábor E. Tusnády[5] and Tamás Hegedűs[1,2,*]

[1]Department of Biophysics and Radiation Biology, Semmelweis University, Budapest 1094, Hungary, [2]MTA-SE Molecular Biophysics Research Group, Hungarian Academy of Sciences, Budapest 1094, Hungary, [3]Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest 1083, Hungary, [4]Faculty of Brain Sciences, University College London, London W1T 7NF, UK and [5]'Momentum' Membrane Protein Bioinformatics Research Group, Institute of Enzymology, RCNS, Hungarian Academy of Sciences, 1117 Budapest, Hungary

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Associate Editor: Lenore Cowen

Received on December 20, 2018; revised on May 31, 2019; editorial decision on June 28, 2019; accepted on July 9, 2019

Abstract

Summary: The identification of transmembrane helices in transmembrane proteins is crucial, not only to understand their mechanism of action but also to develop new therapies. While experimental data on the boundaries of membrane-embedded regions are sparse, this information is present in cryo-electron microscopy (cryo-EM) density maps and it has not been utilized yet for determining membrane regions. We developed a computational pipeline, where the inputs of a cryo-EM map, the corresponding atomistic structure, and the potential bilayer orientation determined by TMDET algorithm of a given protein result in an output defining the residues assigned to the bulk water phase, lipid interface and the lipid hydrophobic core. Based on this method, we built a database involving published cryo-EM protein structures and a server to be able to compute this data for newly obtained structures.

Availability and implementation: http://memblob.hegelab.org.

Contact: tamas.hegedus@hegelab.org

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Membrane proteins play an important role in many cellular processes and are highly significant drug targets (Santos *et al.*, 2017; Yin and Flynn, 2016). To understand their folding, maturation and function, and to develop new therapies targeting membrane proteins, determination of both high-resolution structures and their transmembrane (TM) region is crucial. NMR has been applied mostly for small regions of TM proteins, e.g. one TM helix with short flanking regions either in the absence or presence of a membrane mimetics (Berman *et al.*, 2000). Nevertheless, membrane interaction sites were usually not directly tested. In the case of crystallography, lipids in a crystal can be identified infrequently, and in most cases may have attached to non-physiological sites. Experiments, where tags were inserted at various positions around putative TM helices and their accessibility was tested, usually have provided low resolution data (Chang *et al.*, 1994; Zagotta *et al.*, 2016).

Due to the difficulties associated with experimental approaches, various *in silico* methods have been developed to determine the TM region. The most popular methods are the TMDET (Tusnády *et al.*, 2004) and the PPM (Lomize *et al.*, 2011) algorithms that were utilized to generate PDBTM (Kozma *et al.*, 2012) and OPM (Lomize *et al.*, 2012) databases, respectively. These methods deliver the membrane definition as a slab with two parallel planes. Another frequently used database, MEMPROTMD (Stansfeld *et al.*, 2015) provides predictions by building a membrane bilayer around the protein using molecular dynamics simulations.

The revolution in cryo-electron microscopy (cryo-EM) not only led to an increasing number of solved TM protein structures but also allowed the investigation of protein structures in the presence of a lipid environment (e.g. micelle, bicelle and nanodisc). Furthermore, the resulting electron microscopy density maps contain information about the membrane embedment of the protein. We developed a pipeline that extracts the edges of the blob that corresponds to the membrane boundary of the targeted TM protein. We built a database for this hidden information of cryo-EM maps with a resolution better than 4 Å and also a web application to allow the analysis of unpublished densities.

2 Materials and methods

Structures determined by cryo-EM at a resolution higher than 4 Å and their corresponding electron microscopy density maps were downloaded from RCSB and EMDB, respectively (July, 2018). TMDET translation matrices were collected either from PDBTM (Kozma *et al.*, 2012) or by submitting the PDB file to TMDET (Tusnady *et al.*, 2005). As it will be seen below, the center of the TMDET predicted bilayer plays an important role in searching the edge of the density corresponding to the lipid environment (membrane blob).

The pipeline was built on and managed by Python scripts (Fig. 1). In the first step, a non-protein density map was created by subtracting the modeled protein density map of the all-atom structure from the EMD density. The theoretical density map was generated using the VMD MDFF package (Trabuco *et al.*, 2009) at a resolution of 6 Å. We have simulated protein density at resolutions 4, 5, 6, and 8 Å with the 6 Å value providing a slightly cleaner membrane blob compared to the 4 and 5 Å values. This was most likely because the density maps contained regions with lower and higher resolution than 4 Å. Before subtraction, the theoretical map was scaled to the EMD map using the ratio of the largest density values in the two maps. After subtraction, the values below the 10% of the maximal density value were set to zero and the resulted MRC map was converted to 3 D points and corresponding density values.

We found that the start of the search for the blob boundaries was simpler from the inside of the membrane than from the opposite direction. To set the origin in the membrane region, we translated the coordinates by the TMDET matrix, which set the (0, 0, 0) into the middle of a predicted bilayer. Then x-y sections of this translated, experimental, non-protein density map were generated at a frequency of $\Delta z = 2$Å and these sections were slivered from 0 to 350° in angle slices of 10°. The density values in each slice were summed resulting in an array of density values for every z/angle pairs. This array was smoothed by a Savitzky–Golay filter in both dimensions with a window size of five and a polynomial order of three. The first minimum values were identified in both positive and negative z directions from $z = 0$ (set by TMDET) and proposed as the boundaries of the membrane blob. The established boundaries in each slice were projected to the all-atom structure to pair atoms with their localization. Surface atoms that were more distant from the bilayer center than the z-coordinate of the boundary of a given slice were considered as water accessible. By defaults, atoms in an interval of 8 Å from the boundaries toward $z = 0$ were defined to be located within the interface region. 8 Å was used at is it widely applied (Callenberg *et al.*, 2012; Marcoline *et al.*, 2015; Pabst *et al.*, 2000), but this value can be set by the user. Atoms closer to the center were considered to belong to the hydrophobic core. To classify atoms as buried or surface-exposed, DSSP (Kabsch and Sander, 1983) was run using the all-atom structure as an input. The output of DSSP, the boundaries in each slice and the atomic coordinates were combined to set the localization of a residue in the B-factor field of the corresponding PDB file. The values of -10, 0, 5, 10 and 15 sign any undefined localization (unknown residues or non-protein molecules), buried residues, surface residues in the hydrophobic core region, surface residues in the water phase and surface residues in the lipid interface, respectively. We also provided a text file for the easy assessment of the TM regions. PDB files were manipulated using the MDAnalysis Python package (Michaud-Agrawal *et al.*, 2011), while structural images were created using PyMol (The PyMOL Molecular Graphics System, Version 1.8.4 Schrödinger, LLC) and Chimera (Pettersen *et al.*, 2004). The plots with cross-sections and summed densities were generated by Python Matplotlib (Hunter, 2007).

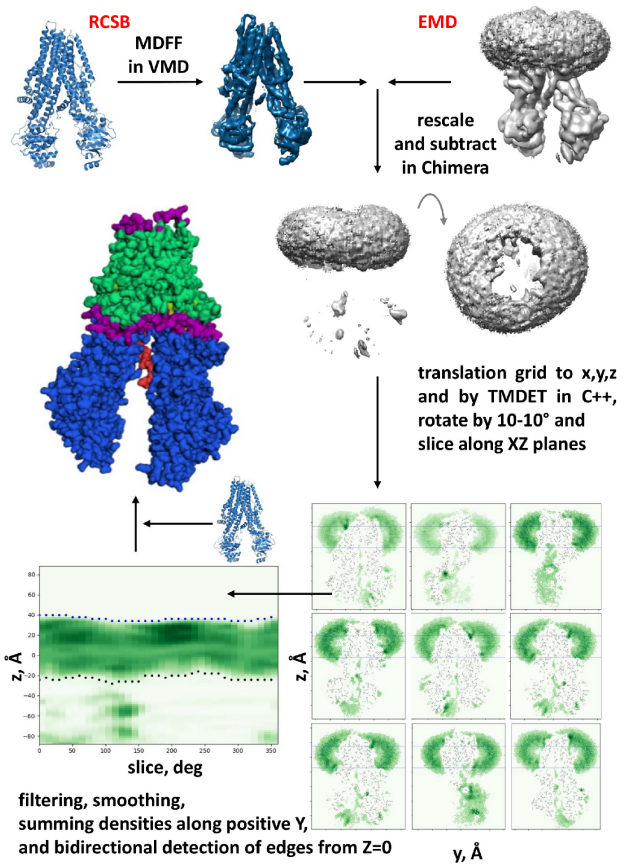


Fig. 1. Main steps of membrane region determination. The density of the protein calculated from the atomistic structure is subtracted from the whole density [CFTR, PDBID: 5UAK (Liu *et al.*, 2017)]. The remaining density is smoothed and projected to 2D. The boundaries of slices are determined from this matrix and mapped back to the all-atom structure

A database and a web application were created to make the pipeline and the results of our runs available. MariaDB (http://mariadb.org) is used as a database backend to store the parameters of submitted jobs and the calculated data of proteins with a resolution better than 4 Å. SQLAlchemy (http://www.sqlalchemy.org) is employed for the object/relational mapping and TurboGears web framework (http://www.turbogears.org) to tie the data, logic and presentation layers into a web application. The main calculation runs independently from the web application in a linear queue system and completed in 1–2 min, which is reasonable, considering the publication frequency of new experimental membrane protein structures.

3Dmol.js is used to visualize the 3D structure in the web page (Rego and Koes, 2015).

3 Results

The web application provides a graphical interface for submitting files and browsing the results, requiring the electron density file in MRC formant and the corresponding all-atom structure in PDB format as inputs. To translate the system and set the (0, 0, 0) coordinate inside the hydrophobic membrane region, the TMDET XML file generated based on the all-atom PDB file is also required. However, if this XML file is not provided by the user, the first four characters of the PDB file name will be treated as a PDBID and used to retrieve the XML file from PDBTM (Kozma *et al.*, 2012). If this process is unsuccessful, the PDB file is submitted to TMDET to obtain the required XML file (Tusnady *et al.*, 2005). On the submit page, the recalculation of the results from our dataset can be initiated by typing a PDB or EMD ID in the appropriate box.

The result page can be accessed upon submitting a new calculation, initiating the recalculation of existing results, and from the browse page of our web application. The result page includes images of *y*-*z* cross-sections at 0, 90, 180 and 270°, while the images of cross-sections at every 10° are packed for download. The plot of the summed and smoothed densities (Fig. 1) is placed into this page as well. The determined boundaries are indicated by blue and black circles. Boundary values outside of the ±1.5 interquartile distance calculated from all *end0* or *end1* boundary values are labeled by triangles. We also put an interactive structural model for visualization on this page to help to decide whether the automatically determined boundaries need manual adjustment. To aid the manual correction of edge detection, we implemented a set of simple commands combined with selection expressions. The three main commands are: (i) *slice_def* end_i ***around*** *z dz*, where *slice_def* is an integer corresponding to a given slice from 0 to 350°, end_i is end0 or end1, *z* is the z-coordinate to search around in the range of (*z-dz, z+dz)*. Slice definition can include comma separated and hyphen separated list of slices (e.g. 10, 40, 100–160) and also an asterisk for all slices. (ii) *slice_def* end_i ***average*** *slice_A slice_B*, which sets the edge of selected slices to the average value of slice A and slice B. (iii) *slice_def* end_i ***equal*** *slice_A*, which set the boundary of the selected slices to the value of slice A's boundary. All the commands are listed in Supplementary Table S1.

We ran the MemBlob pipeline on 92 TM protein structures determined by cryo-EM with a resolution of 4 Å or better. The calculations revealed that approximately 30% of the maps did not exhibit well-defined densities corresponding to the membrane environment (Supplementary Fig. S1). These structures have either been solved in the absence of a well-formed lipid environment or their electron microscopy density maps exhibited a very low signal to noise ratio preventing the detection of the membrane blob boundaries. A good signal to noise ratio of an experimental map is crucial to detect the membrane environment, since densities arising from lipids are significantly lower than those from proteins. In the case of a cation channel [PDBID: 5H3O, (Li *et al.*, 2017)] our pipeline detected rational TM regions in spite of the lack of a membrane blob. A closer look at the density map suggests that the amphipol environment in this case does not contribute to the cryo-EM density, but densities can be observed between the TM helices of the protein. These inter-helical densities indicate the presence of intercalated lipid molecules. We did not find any correlation between the visibility or other properties of the lipid environment and the type of the membrane mimetics (e.g. micelle, nanodisc and amphipol). For example, while the amphipol blob did not contribute to the cryo-EM density in the case of PDBID: 5H3O (Li *et al.*, 2017), it was visible in other instances, such as PDBID: 3J5P (Liao *et al.*, 2013). A summary of the runs is collected in Supplementary Table S2.

We compared the TM region definitions of our pipeline to TMDET predictions, since this *in silico* method has been indicated to provide more feasible boundaries compared to OPM (Koehler Leman *et al.*, 2017). We used it to guide the boundary search in our pipeline. In addition, we have not detected large differences in predictors when previously used for ABC proteins (Csizmadia *et al.*, 2018). For the comparison, first, to get the thickness of the hydrophobic core, we calculated the distance between the boundaries decreased by the thickness of the two interface regions (2 × 8 Å). Then, we averaged the *z*-coordinates of the boundaries for each of the sides resulting in a slab, similar to the output of *in silico* predictors (Supplementary Fig. S2). The membrane center of the MemBlob slab differs from the TMDET center by more than 5 Å only in four cases. In contrast, the MemBlob pipeline determines a thicker membrane environment compared to TMDET. This is often caused by the deep embedment of the protein into the lipids. As a consequence, the location of short regions, which have been considered extra- or intracellular by *in silico* predictors, is indicated intramembranous by our pipeline (Supplementary Fig. S3). This type of membrane-embedment, when the extracellular parts are located in a pit of the membrane, cannot be predicted by *in silico* methods. The physiological role of this embedment may be to provide better protection of the protein from extracellular effects, such as proteases.

4 Conclusions

Most of the cryo-EM studies focus on the determination and characterization of protein structures. However, density maps may contain valuable information other than the well-defined protein density, which has not been fully utilized yet. For example, electron densities derived from disordered protein segments are difficult to extract and interpret. Recently, a machine learning algorithm has been developed for automatic identification of density blobs of ligands from experimental electron microscopy density maps (Kowiel *et al.*, 2019). However, our pipeline is the first that allows the assessment of membrane localization of TM proteins from experimental data at a large scale, using cryo-EM densities. While learning algorithms may supersede the semi-automatic refinement of the boundaries in our pipeline, as of now, we cannot exploit an automatic detection method at this moment due to the low number of cryo-EM maps with sufficient membrane environment densities. Our pipeline possesses two major differences when compared to other existing methods providing TM region prediction. First, MemBlob is fully based on experimental data. While CCTOP supplements its prediction with a large amount of information from experiments (Dobson *et al.*, 2015), this data is coarse-grained (e.g. accessibility experiments), which helps the identification of extramembranous regions rather than the exact location of the bilayer boundaries. Second, MemBlob presents the membrane region as a volume with boundaries that follows the shape of the lipid environment, and not as a slab with parallel edges. MEMPROTMD provides a more realistic configuration of the membrane around the protein using molecular dynamics simulations compared to slab models, but it does not incorporate experimental data other than protein structures (Stansfeld *et al.*, 2015). Therefore, the MemBlob pipeline will be useful for researchers working on structure determination of membrane

proteins using cryo-EM and also for developers of membrane region predictors, who can apply MemBlob results as a true positive experimental set. Since the number of membrane protein structures are expected to rise, the output of our methods will most likely be the starting point to develop automatic methods for the identification of the membrane environment in density maps.

Acknowledgements

Thanks for the suggestions and critical notes to H. Tordai (Department of Biophysics and Radiation Biology, Semmelweis University). We acknowledge NIIF National Information Infrastructure Development Institute (http://www.niif.hu/en) and MTA Wigner GPU Laboratory (http://gpu.wigner.mta.hu) for awarding us access to resources based in Hungary, and the support of their staff is gratefully acknowledged.

Funding

This work was supported by the National Research, Development and Innovation Office [K111678, K119287, K125607, K127961], the Cystic Fibrosis Foundation [CFF HEGEDU18I0] and the Semmelweis Science and Innovation Fund.

Conflict of Interest: none declared.

References

Berman,H.M. *et al.* (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

Callenberg,K.M. *et al.* (2012) Membrane bending is critical for the stability of voltage sensor segments in the membrane. *J. Gen. Physiol.*, **140**, 55–68.

Chang,X.B. *et al.* (1994) Mapping of cystic fibrosis transmembrane conductance regulator membrane topology by glycosylation site insertion. *J. Biol. Chem.*, **269**, 18572–18575.

Csizmadia,G. *et al.* (2018) Quantitative comparison of ABC membrane protein type I exporter structures in a standardized way. *Comput. Struct. Biotechnol. J.*, **16**, 396–403.

Dobson,L. *et al.* (2015) CCTOP: a consensus constrained TOPology prediction web server. *Nucleic Acids Res.*, **43**, W408–W412.

Hunter,J.D. (2007) Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.*, **9**, 90–95.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–637.

Koehler Leman,J. *et al.* (2017) Computing structure-based lipid accessibility of membrane proteins with mp_lipid_acc in RosettaMP. *BMC Bioinformatics*, **18**, 115.

Kowiel,M. *et al.* (2019) Automatic recognition of ligands in electron density by machine learning. *Bioinformatics*, **35**, 452–461.

Kozma,D. *et al.* (2012) PDBTM: protein data bank of transmembrane proteins after 8 years. *Nucleic Acids Res.*, **41**, D524–D529.

Li,M. *et al.* (2017) Structure of a eukaryotic cyclic-nucleotide-gated channel. *Nature*, **542**, 60–65.

Liao,M. *et al.* (2013) Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*, **504**, 107–112.

Liu,F. *et al.* (2017) Molecular structure of the human CFTR ion channel. *Cell*, **169**, 85–95.

Lomize,A.L. *et al.* (2011) Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. *J. Chem. Inf. Model.*, **51**, 930–946.

Lomize,M.A. *et al.* (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.*, **40**, D370–D376.

Marcoline,F.V. *et al.* (2015) Membrane protein properties revealed through data-rich electrostatics calculations. *Structure*, **23**, 1526–1537.

Michaud-Agrawal,N. *et al.* (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.*, **32**, 2319–2327.

Pabst,G. *et al.* (2000) Structural information from multilamellar liposomes at full hydration: full q-range fitting with high quality x-ray data. *Phys. Rev. E*, **62**, 4000–4009.

Pettersen,E.F. *et al.* (2004) UCSF chimera? A visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

Rego,N. and Koes,D. (2015) 3Dmol.js: molecular visualization with WebGL. *Bioinformatics*, **31**, 1322–1324.

Santos,R. *et al.* (2017) A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.*, **16**, 19–34.

Stansfeld,P.J. *et al.* (2015) MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure*, **23**, 1350–1361.

Trabuco,L.G. *et al.* (2009) Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods*, **49**, 174–180.

Tusnady,G.E. *et al.* (2005) TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. *Bioinformatics*, **21**, 1276–1277.

Tusnády,G.E. *et al.* (2004) Transmembrane proteins in the protein data bank: identification and classification. *Bioinformatics*, **20**, 2964–2972.

Yin,H. and Flynn,A.D. (2016) Drugging membrane protein interactions. *Annu. Rev. Biomed. Eng.*, **18**, 51–76.

Zagotta,W.N. *et al.* (2016) Measuring distances between TRPV1 and the plasma membrane using a noncanonical amino acid and transition metal ion FRET. *J. Gen. Physiol.*, **147**, 201–216.

Cellular and Molecular Life Sciences (2022) 79:73
https://doi.org/10.1007/s00018-021-04112-1

Cellular and Molecular Life Sciences

ORIGINAL ARTICLE

Ins and outs of AlphaFold2 transmembrane protein structure predictions

Tamás Hegedűs[1,2] · Markus Geisler[3] · Gergely László Lukács[4] · Bianka Farkas[1,5]

Received: 13 September 2021 / Revised: 25 November 2021 / Accepted: 20 December 2021

Abstract

Transmembrane (TM) proteins are major drug targets, but their structure determination, a prerequisite for rational drug design, remains challenging. Recently, the DeepMind's AlphaFold2 machine learning method greatly expanded the structural coverage of sequences with high accuracy. Since the employed algorithm did not take specific properties of TM proteins into account, the reliability of the generated TM structures should be assessed. Therefore, we quantitatively investigated the quality of structures at genome scales, at the level of ABC protein superfamily folds and for specific membrane proteins (e.g. dimer modeling and stability in molecular dynamics simulations). We tested template-free structure prediction with a challenging TM CASP14 target and several TM protein structures published after AlphaFold2 training. Our results suggest that AlphaFold2 performs well in the case of TM proteins and its neural network is not overfitted. We conclude that cautious applications of AlphaFold2 structural models will advance TM protein-associated studies at an unexpected level.

Keywords Structure prediction · Transmembrane proteins · AlphaFold2 · Bioinformatics

Introduction

Although enormous resources were devoted to predict protein structures for many decades, building a protein structure from its sequence remained a challenging task [1]. There was a change at the 13th Critical Assessment of Protein Structure Prediction (CASP13) competition [2] when the neural network-based approach, AlphaFold excelled. The improved version, AlphaFold2 (AF2) achieved an accuracy level much higher than other predictors at CASP14 [3, 4]. Importantly, DeepMind released their code with deep learning models and deposited AF2-predicted structures for the human [5] and 20 other proteomes in collaboration with EBI (https://alphafold.ebi.ac.uk). Moreover, to ease the running of predictions for researchers, DeepMind [6] and community Google Collaboration notebooks [7] have been generated, albeit applying some simplifications.

AlphaFold2 was trained using multiple sequence alignments (MSA) and experimental protein structures deposited before 2018-04-30. Five different models were trained (e.g. with different random seeds, with or without structural templates) to promote an increased diversity in structure predictions [6]. The input for prediction is the sequence of a single protein chain, used for MSA generation and structural template search. The quality of the resulted structural models is characterized by the mean of per residue pLDDT (predicted Local Distance Difference Test) score (which takes values between 0 and 100, the higher value is better) and the structures are ranked accordingly [3]. The pLDDT confidence measure predicts the accuracy of the Cα Local Distance Difference Test (lDDT-Cα) for the corresponding prediction. Although this means that the high accuracy and reliability of AF2 observed in CASP14 can be transferred to predicting the structure of any protein sequences (or whole proteomes) [3, 5], this has not been validated yet and scientists do not have a clear indication how well AF2-predicted structures can be trusted. Moreover, AlphaFold2 structural prediction

✉ Tamás Hegedűs
hegedus@hegelab.org

1 Department of Biophysics and Radiation Biology, Semmelweis University, Budapest, Hungary

2 TKI, Eötvös Loránd Research Network, Budapest, Hungary

3 Department of Biology, University of Fribourg, Fribourg, Switzerland

4 Department of Physiology and Biochemistry, McGill University, Montreal, QC, Canada

5 Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary

Published online: 15 January 2022

of transmembrane proteins is treated with skepticism, as it remain challenging by both experimental and computational methods, especially because AlphaFold2 was not tuned for TM proteins. It is also not known, whether the structural model with the highest pLDDT score always corresponds to the native structure. To tackle these issues, we investigated if AF2-predicted human α-helical TM protein structures exhibit correctly located TM regions. To demonstrate at a higher resolution that the predicted TM folds are native, we compared predicted structures of the ATP Binding Cassette (ABC) superfamily from the AF2-predicted 21 proteomes to existing experimental ABC folds.

ABC proteins play a role in important cellular processes in all types of organisms and most of them transport substrates through the cell membrane in an ATP-dependent manner [8–10]. ABCC7/CFTR is a special member, which is an ATP-gated chloride channel and includes a long intrinsically disordered regulatory R domain [11, 12]. The functional form of ABC proteins is built from two highly conservative nucleotide-binding domains (NBDs) and two transmembrane domains (TMDs) which can be encoded in one or separate peptide chains. The low conservation of their TMDs are related to diverse functions and their currently known TM folds are also structurally divergent and can be classified into eight groups (Pgp-, ABCG2-, MalFG-, BtuC-, EcfT-, LptFG-, MacB-, and MlaE-like folds) [13, 14]. Our results suggest that AlphaFold2 provides protein structures for transmembrane proteins as reliable as for soluble proteins and can help to solve many issues associated with transmembrane protein structures.

Results

Transmembrane topology assignments in AlphaFold2 structures

First, pLDDT score distribution for soluble and transmembrane proteins were compared. We split the human AF2 structures to these two groups using the HTP (Human Transmembrane Proteome) database [15], calculated the mean pLDDT score for each protein, and plotted their distribution (Fig. 1a and Fig. S1). Mean pLDDT values were also calculated separately for the TM and non-TM regions of transmembrane proteins. Intriguingly, soluble proteins exhibited a broader distribution and a significant area at lower pLDDT values compared to TM proteins. This was unexpected, since the majority of the AlphaFold2 learning set inherently included more soluble protein templates and the algorithm was not tuned for transmembrane proteins. However, correlation between low pLDDT values and disordered segments was observed [5], thus our observation suggested that more soluble proteins possess disordered regions than TM proteins. Interestingly, a very large portion of TM regions (53%) were predicted with high pLDDT scores (> 90) (Fig. 1a) indicating that AF2 captured the rules governing protein structures within the hydrophobic region.

Next, we compared the spatial localization of TM helices in AF2 structures if helix orientation corresponds with rational and physiological orientation in a lipid bilayer slab using the Constrained Consensus Topology prediction (CCTOP) software [16], which includes information from both experimental and computational sources. We separated the start and end positions of predicted TM helices to two residue sets according to their localization relative to the opposite sides of the bilayer. The distance between the center of geometry of the two sets were calculated and its distribution is plotted (Fig. 1b). The majority of the membrane thickness values were in the range between 20 and 35 Å, which is in the range of the hydrophobic region thickness. To support this finding with experimental data, the hydrophobic thickness of experimentally determined human transmembrane protein structures was retrieved from the PDBTM database [17]. The AF2 and experimental distribution largely overlapped (Fig. 1b). These observations suggested that hydrophobic thickness values below 15 Å or above 35 Å may indicate an erroneous AF2 structure (725 out of 5,952, 12%, Table S1). An inaccurate TM topology prediction of CCTOP may provide an outlier hydrophobic thickness in the case of a correct AF2-predicted structure. The CCTOP reliability versus thickness plot (Fig. 1c) indicated that the topology of most proteins, whose AF2-predicted structure exhibited hydrophobic thickness within the 15–35 Å regime, was predicted with high reliability. Structures with lower hydrophobic thickness values and high CCTOP reliability were likely inaccurately predicted by AlphaFold2, while structure predictions with lower thickness and lower CCTOP scores were located in the twilight zone. Intriguingly, we observed that some of these entries may have low topology reliability because of their existence in protein–protein complexes, but AF2 predicted the monomeric form correctly (Fig. S2). This suggests that AF2 may also be used to identify and aid the correction of improper membrane topology predictions.

We also investigated the distribution of pLDDT scores versus hydrophobic thickness (Fig. 1d). This plot indicated that AF2 structures with non-physiological thickness values can process very high pLDDT scores, consequently, these scores alone may be insufficient to select correct TM structures in blind predictions.

Helix packing in AF2-predicted ABC models overlaps with experimental folds

To assess AF2 TM protein predictions at a higher resolution, we aimed to compare AF2-built ABC TM folds with experimentally determined folds. Structures of ABC

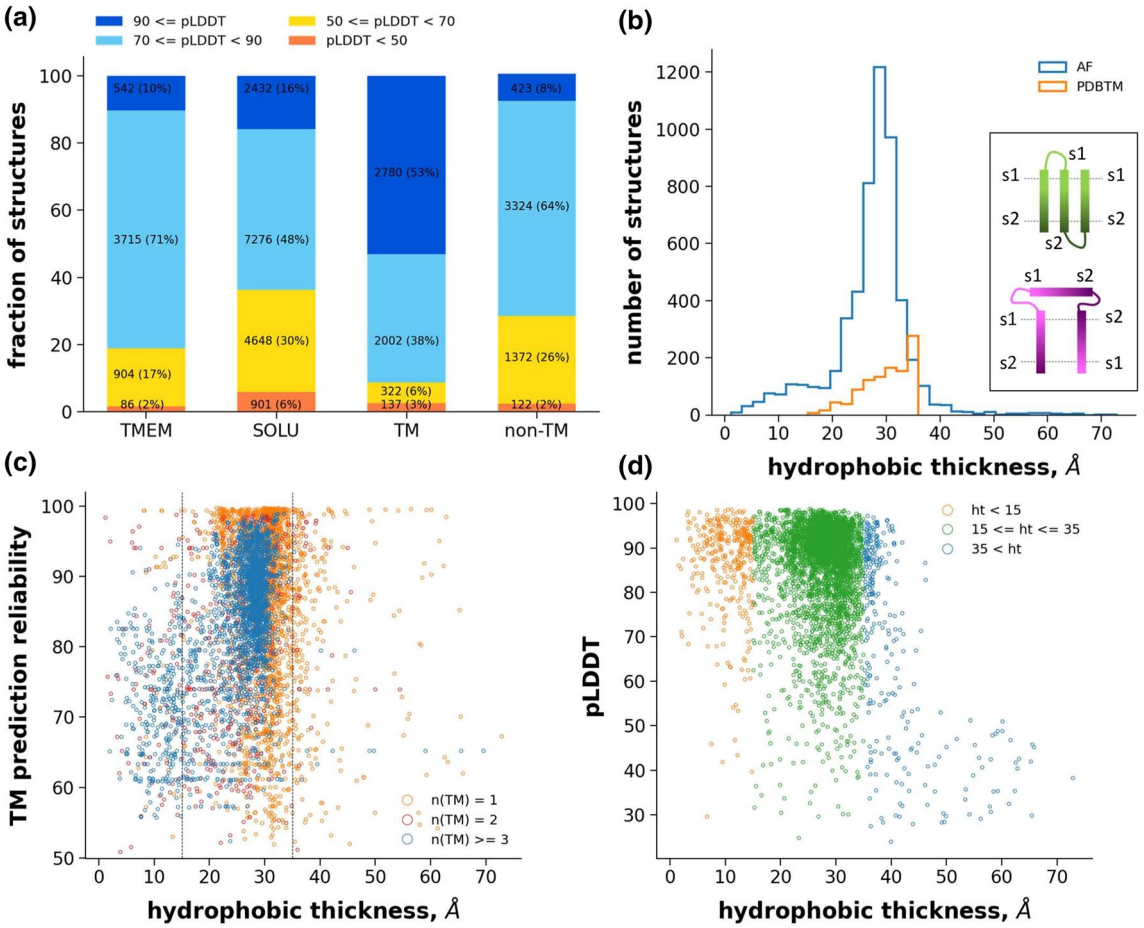


Fig. 1 Quantitative analysis of human AF TM structures. (**a**) Mean pLDDT scores were calculated for human transmembrane (TMEM), soluble (SOLU), TM regions of TM proteins, and non-TM regions of the same proteins. The fraction of structures in reliability ranges, used in the human proteome AlphaFold2 paper [5], are shown. (**b**) The hydrophobic thickness was calculated for human TM proteins as the distance between the center of geometry of Cα atoms of side1 and side2 of transmembrane helices. TM helices of AF2-structures were selected based on CCTOP predictions. The hydrophobic thickness of experimental structures was collected from PDBTM. The inset demonstrates how the distance calculation can be effected by a topology in the case of incorrectly built AF2 structures (purple; correct structure: green; s1 and s2: side1 and side2). (**c**) The hydrophobic thickness of each protein and the corresponding CCTOP reliability scores are shown. (**d**) The hydrophobic thickness of each protein and the corresponding pLDDT scores were plotted

superfamily members are a reasonable choice to investigate AlphaFold2 performance on TM proteins, since the currently available PDB entries, which include 675 chains with ABC transmembrane domains, are diverse and can be classified into 8 different structural folds (Fig. S3) [13, 14]. We characterized the similarity of each ABC transmembrane domain to every ABC reference fold using the Template Modeling score (TM-score) [18, 19] (Fig. 2a). If comparison of two structures results in a TM-score below 0.3 then they are structurally unrelated, while a TM-score above 0.5 indicates identical folds [19]. The range between 0.3 and 0.5 is the twilight zone. Each target transmembrane domain was classified according to the best match to an ABC reference fold and the TM-scores were above 0.5 in all cases. The observed variation of scores among these experimental ABC structures originated from differences in conformations (e.g. apo and ATP-bound structures).

In the next step, we selected ABC structures from the 21 proteomes with AF2 predictions by a stringent PFAM search, which was performed with 28 PFAM Hidden Markov Models (Table S2) that resulted in 1137 hits. For assessing the similarity of structures to the eight selected reference folds, we calculated TM-scores between the AF2-predicted transmembrane ABC structures and the reference structures. The best out of eight scores were saved for each structure. We found that all TM-score values were above 0.5 (Fig. 2b). One outlier protein (Q2G2E2), which matched the YitT_transmembrane PFAM entry, was somewhat similar to the aquaporin/GlpF fold (e.g. PDBID: 1fx8) suggesting that the YitT_transporter PFAM entry is wrongly classified. Indeed,

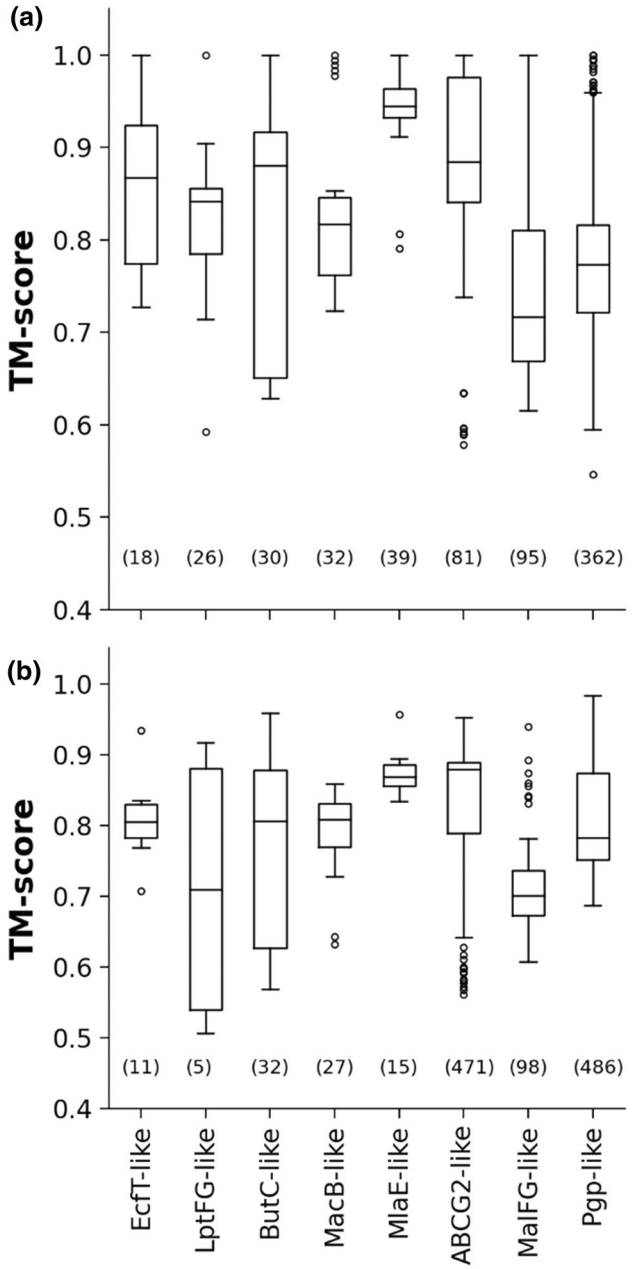


Fig. 2 All AF2-predicted ABC structures exhibit valid ABC TM folds. (**a**) The best TM-score for every experimental ABC TM structure and ABC reference fold comparisons are shown in the boxplot, grouped into ABC fold families, sorted by the total number of included transmembrane chains. Numbers indicate the sum of chains in experimental structures matching an ABC reference fold. Only MlaE fold family does not contain experimental structures released before 2018-04-30. (**b**) The same plot was generated for ABC protein structures from the 21 proteomes predicted by AF2. The number of matched structures within a fold family is indicated in parenthesis. A TM-score above 0.5 indicates that the compared structure and the reference fold exhibits the same architecture

this protein belongs to the non-ABC, Novobiocin Exporter (NbcE) Family based on the Transport Classification Database [20].

Some of the predicted ABC structures included additional N-terminal TM-like helices, which were somewhat distant from the core TM domain and likely are membrane-associated regions, such as the L0/Lasso motif of ABCC proteins [21–23]. In many cases, membrane-associated regions, loops, and mobile segments not resolved in experimental structures have been rationally modeled by AF2, based on visual inspection (see below and Fig. S2), thus the AF2 machine learning method may have grasped some knowledge on a lipid bilayer around TM proteins. However, in other cases, long loops with low pLDDT scores, which are likely disordered regions, were unrealistically crossing the bilayer region. Those in our eyes are not negatively affecting AF2 predictions and were thus not considered as an issue, since the localization of disordered regions also cannot be trusted in the case of AF2-predicted soluble protein structures.

Prediction of challenging and novel transmembrane folds

Importantly, the above and any retrospective analysis of AF2 predictions are limited by the fact that a significant portion of the AF2-predicted (transmembrane) protein structures deposited at EBI have corresponding experimental structures with either the same sequence or a homologous sequence, either included in the AF2 training set (up to 2018–04-30) or used as templates during prediction runs (up to mid of 2020). Therefore, we selected the challenging TM target of CASP14 (T1024, LmrP, PDBID: 6t1z released on 2019–10-07), which possessed homologous structures, and novel TM folds that were also released after 2018–04-30 for characterizing AF2 performance.

The prediction of the T1024 target, ranked #43 with GDT_TS score and RMSD of 60.29 and 5.61 Å, respectively (#1 by Arne Elofsson: 63.3 and 3.74 Å). However, LmrP has a hinge region that effects predictions and AF2 likely produced a functional conformation different from that observed in the 6t1z structure, supported by distance restraints from double electron–electron resonance spectroscopy [24]. Since the AF2 LmrP model submitted to CASP14 was created with an earlier version of AlphaFold [25], we rerun the LmrP prediction with disabled template usage. The top model exhibited 82.82, 1.74 Å, and 0.92 GDT_TS, RMSD and TM-score, respectively, when compared to 6t1z (Fig. 3a). These observations suggest that AF2 prediction of flexible targets should be interpreted carefully and AF2 may be utilized to discover novel conformations related to different functional states.

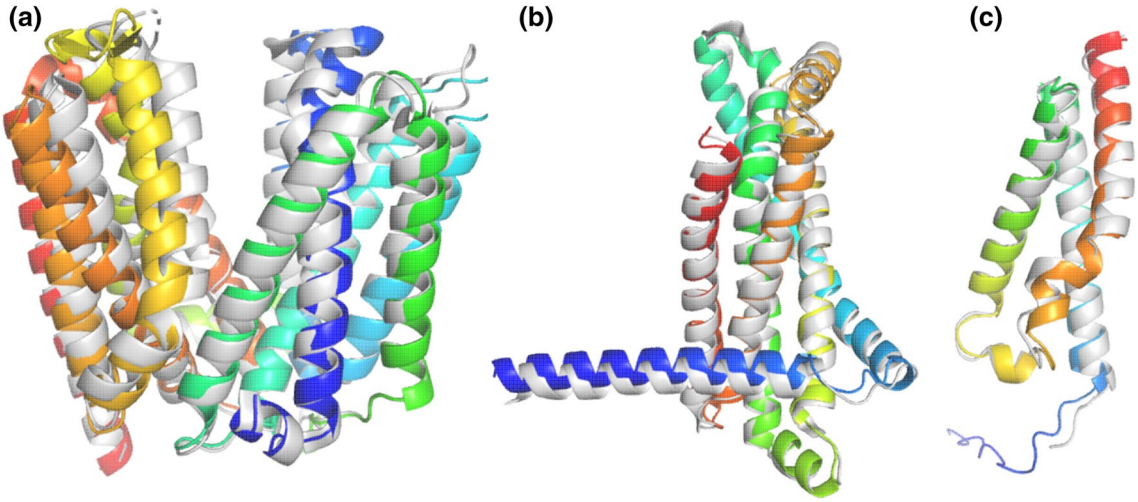


Fig. 3 Blind transmembrane fold predictions. The AF2-predicted structures (blue to red: N- to C-termini) of the challenging CASP14 TM target, LmrP (**a**), the MlaE (**b**), and the EMC6 (c) exhibit perfect alignments with their experimental structures (gray) 6t1z, 7ch0, and 6ww7, respectively. None of these experimental structures were published before 2018-04-30. Only structures with a sequence homologous to LmrP were in the AF2 training set

In the next step, we performed extensive literature, SCOP, and PFAM searches to identify transmembrane protein structures or their homologous structures, which were not inserted into the AF2 training set. We found the ABC transmembrane MlaE-like fold (7cge, 7ch0, and 7cgn were released on 2020-09-09; 7ch7 was released on 2021–05-19) [26], the ER membrane protein complex subunit sixfold (EMC6, PDBIDs: 6wb9, 6ww7, 6z3w, 7ado, 7adp, 7kra, and 7ktx, with the earliest release date of 2020-05-27) [27], and the MprF structure (PDBIDs: 6lvf and 7duw, released on 2021-02-03 and 2021-04-21, respectively) [28] as valid targets for blind AF2 TM protein predictions. AF2 runs without templates resulted in top models highly similar to the experimental structures of MlaE (PDBID: 7ch0, RMSD: 1.28 Å, TM-score: 0.95, Fig. 3b) or EMC6 (PDBID: 6ww7, RMSD: 0.96 Å, TM-score: 0.93, Fig. 3c). In contrast, the top prediction of the multiple peptide resistance factor (MprF) transmembrane domain sequence did not match the experimental structure (Fig. 4a). Therefore, we performed this prediction several times ($n = 6$) with different random seeds and compared the output to the transmembrane domain of 7duw using TM-score. Plotting the pLDDT scores versus TM-scores (Fig. 4b) indicated that among the 30 predicted structures the one with the best pLDDT score exhibited the highest TM-score, thus was the most similar to the target structure (Fig. 4c). Importantly, the difference in MprF conformations involves the separation of two subdomains (flippase and synthase) [29] and AF2 may have captured a functionally relevant state as in the case of LmrP.

AF2 can provide hints for investigating ABC structure-associated questions

To demonstrate possible contributions of AF2-predicted structural models to studies targeting membrane proteins, we assessed AF2 ABC models in various test cases. At first place, we tested half transporter ABCG proteins, which consist of an NBD and a TMD in a polypeptide chain and function in homodimeric or heterodimeric complexes [14]. The first experimentally determined ABCG2-like fold was the X-ray structure of the ABCG5/ABCG8 heterodimer (PDBID: 5do7) published in 2016 [30]. Our first observation with the AF2-generated ABCG8 structure was regarding its soluble NBD. After the publication of the first ABCG2 structure [31], structural alignment and sequence analysis indicated a registry shift in the first β-strand of ABCG8 NBD (Fig. 5a) that happened because of the low resolution of this region. Although the 5do7 structure was in the AF2 training set and was present in the pdb70 template database, the AF2-predicted ABCG8 structure deposited at EBI did not have this error (Fig. 5a). An ABCG5/ABCG8 structure with a correct registry was also released on 2021-04-07 (PDBID: 7jr7 [32]), but AF2 template search for building models deposited at EBI used pdb70 downloaded on 2021-02-10 [5].

To assess ABCG5/ABCG8 transmembrane domain (TMD) predictions, we ran AF2 without application of templates. First, the ABCG5 TMD predictions were of exceptionally good quality regarding the RMSD (root mean square deviation) and TM-score values of 0.61 Å and 0.94,

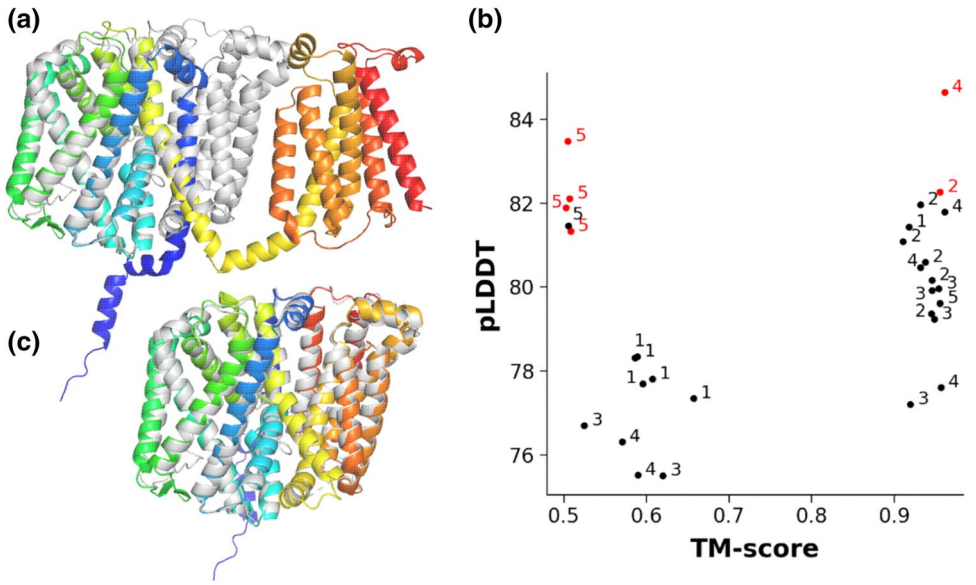


Fig. 4 AF2 predicts two conformations of a new transmembrane fold. (**a**) The top AF2-prediction of the novel MprF TM fold (blue to red: N- to C-termini) aligned to the experimental structure 7duw (gray). (**b**) pLDDT and TM-score values, calculated for every structural model from six runs, were plotted. Numbers (1–5) indicate the corresponding AF2 models. Red points were the top ranked hits from a given run. (**c**) Structural alignment of the prediction with the best pLDDT score (blue to red) and experimental structure (gray). 7duw and any other structure homologous to MprF were not included in the AF2 training set

respectively, when compared to the ABCG5 chain in the 7jr7 structure. Second, we investigated ABCG5/ABCG8 heterodimer predictions. Since only single chains can be submitted to AlphaFold2, we concatenated the two sequences with a part of the CFTR R domain sequence (a.a. 675-800). This disordered sequence was sufficiently long not to constrain the conformational space of the dimer and did not exhibit strong intramolecular interactions even in its native, AF2-predicted structural environment (Fig. S4). The predicted TMD dimer exhibited 2.18 Å RMSD and its individual chains showed 0.98 and 0.96 TM-score values when compared with the 7jr7 structure (Fig. 5b).

To investigate if AlphaFold2 can distinguish between intra- and intermolecular interactions in the case of homomeric complexes, we performed a prediction with ABCG2, which forms homodimers [33]. The complex of the two identical TMDs was also predicted exceptionally well (2.42 Å RMSD and 0.9 TM-score when compared to PDBID: 6vxf). Interestingly, cysteine residues forming intra- and intermolecular disulfide bonds were close to each other (Fig. S5).

We also examined how AF2 structural models can supplement or replace homology models in molecular dynamics (MD) simulations. The TM regions of distant ABC proteins exhibit low sequence conservation with good accordance of their dissimilar functions and substrates. However, their folds in a family are highly conserved, thus homology modeling can provide high-quality models [34–37]. We chose AtABCG36/PEN3/PDR8 [38] from the model plant *Arabidopsis thaliana*, which is a well-investigated full transporter of the ABCG subclass for that no structures yet exist. When the homology model exhibiting two ABCG2-like TMDs (Fig. 5c) was inserted into a membrane bilayer and subjected to a 50 ns long MD simulation, one portion of an α-helix, which is part of the central drug binding pocket, exhibited fast unfolding (~ 10 ns) in an equilibrium MD simulation. Then, the AF2-predicted AtABCG36 structure under the same conditions remained stable in a 500 ns long MD simulation (Fig. S6). However, one should be careful with simulations using AI-based structural models, since their conformation may be kinetically trapped into a specific state, inhibiting the study of conformational changes [39].

The CFTR/ABCC7 chloride channel is also a member of the ABC superfamily with a Pgp-like fold. The functional mechanism of this protein is of interest, since some mutations effect channel gating and cause cystic fibrosis [40]. One of its structures was determined using cryo-EM under activating condition, in the presence of ATP and phosphorylation, but the extracellular pore of the channel remained in a closed state, most likely due to a kink in TM8, corresponding to an unwound segment in the transmembrane region [41] (Fig. 5f). This kink is present in most CFTR structures (PDBIDs: 5uak, 5uar, 5o2p, 5w81, 6msm, and 6o1v) [41–44]. However, the kink is absent from the chicken CFTR structure (PDBIDs: 6d3s and 6d3r) [45] and such a conformation has not been detected in other ABC structures. We performed equilibrium simulations with the 5w81 structure [12] to detect channel opening, but appearance of tunnels with sufficient diameter to pass chloride ions were

rare events and was observed only once out of 22 simulations (6 × 100 ns + 16 × 35 ns, 427/116,000 frames, 0.36%). Intriguingly, many of the conformations provided a tunnel opened towards lipid molecules of the extracellular membrane leaflet (Fig. 5g). After correcting the kink by homology modelling based on the MRP1 structure (PDBID: 5uj9) (Fig. 5f), opening of the extracellular pore could be observed in five out of six simulations at a higher probability (6 × 100 ns, 2245/60,000 frames, 3.74%). Remarkably, modeling CFTR TMDs using AlphaFold2 without CFTR or any templates resulted in a conformation similar to that of MRP1 with a straight TM8 helix (Fig. 5f, h). Since TM8 has been suggested to be flexible regarding to its membrane embedment [46], it is likely sensitive to its environment and based on the functional assays and the structure determination protocol [42], the detergent added in the last step (3 mM fluorinated Fos-Choline-8) likely biased the experimental structure.

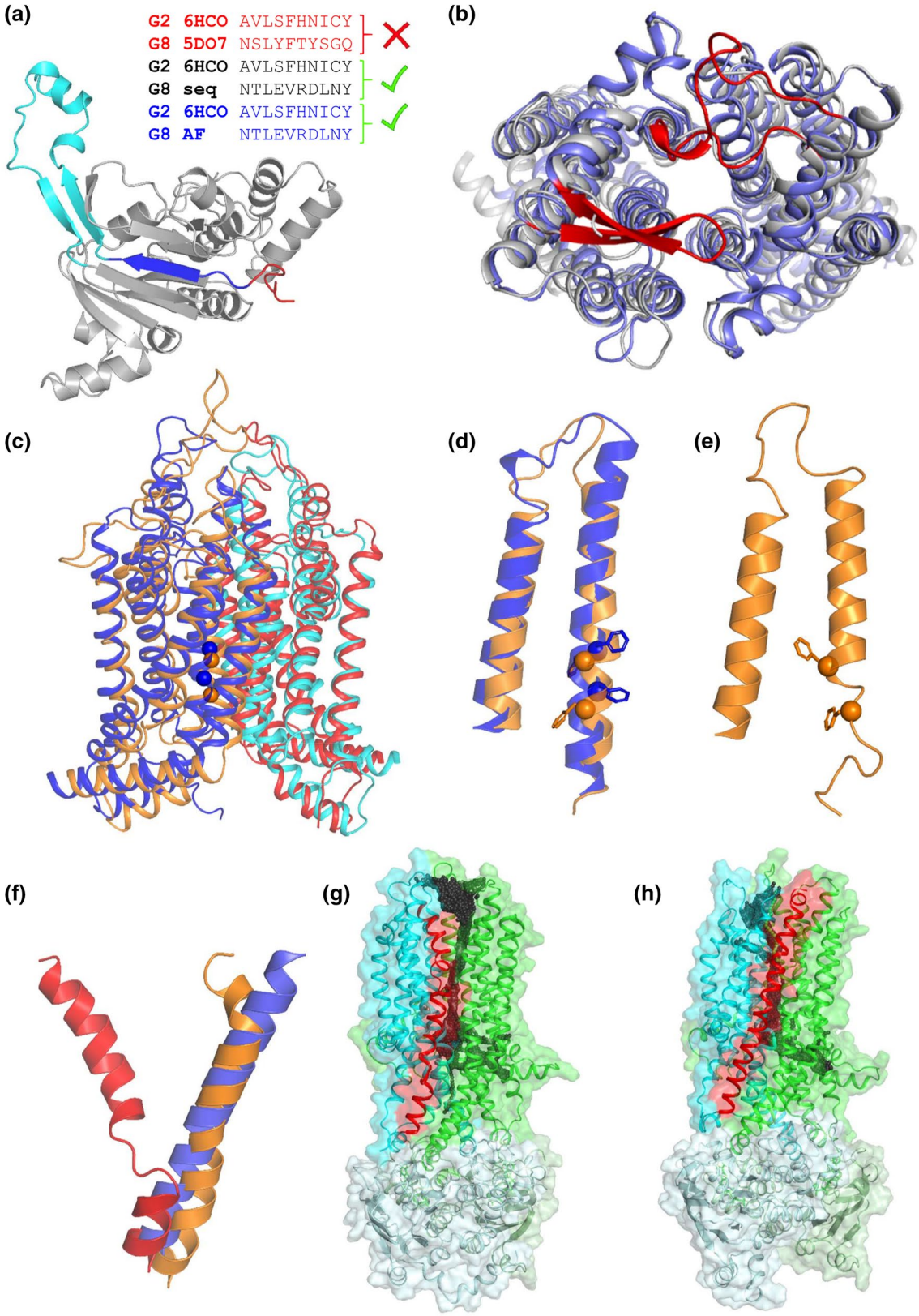
Discussion and conclusions

We demonstrated that at least ~ 90% of the AF2-predicted TM structures of the human proteome represented membrane-protein like structures, using the most available and reliable measure, the location of TM helices from consensus predictions and experimental structures, for assessing TM protein structure quality at a large scale. Since the pLDDT score distribution did not shift much to lower values compared to soluble proteins (Fig. S1), it is likely valid to state that AF2 predicts TM proteins as well as soluble proteins. However, predicted TM structures with low hydrophobic thickness and high pLDDT score (Fig. 1d) suggest that evaluation depending solely on pLDDT score may not be sufficient to select the best AF2-predicted model, at least in the case of TM proteins. A similar conclusion was drawn comparing the AF2-predicted and cryo-EM structures of the pump-like channelrhodopsin with structural features never seen before [47]. In specific cases, resource intensive molecular dynamics simulations may be used to asses AF2 models, since MD simulations were demonstrated to reveal erroneous structural models built using either homology modelling (Fig. S6) or experimental methods [48].

A very important issue is associated with retrospective studies, including ours, which assess AlphaFold2 performance based on AF2 structures deposited at EBI. Most likely a significant portion of the predicted models can be related to experimental structures with homologous sequences, included in the AF2 training set or used as templates during model building or both. In these cases, AF2 may be considered as a highly advanced homology modelling tool, which performs an automatic but high-quality sequence alignment and provides high-quality results even in the case of target sequences with low sequence similarity to any known structures. This is a very important property of AF2 and will advance structural biology studies of TM proteins, since the hydrophobic regions are usually not highly conserved (e.g. sequence identity between ABC transmembrane domains is usually below 20–30%; ABCG2 exhibits 27% and 26% identities when compared to the closely related ABCG5 and ABCG8, respectively). For the correct interpretation of retrospective studies and evaluation of AF2 performance, it is important to implement a versioning system for AF2 models. This objective seems to be more complicated than for experimental structures, since the structure prediction depends on the version of the deep learning models, various sequence databases, and the pdb70 structure database.

Taken together, investigating AF2 performance in blind predictions requires an experimental structure, which or structures with homologous sequences were not included in the training set. In addition, the AF2 prediction of such targets should be performed without using templates. In this way, predictions for a high number of homologous sequences and their systematic comparison to corresponding structures generated with templates could be informative regarding to blind predictions and to the effect of template usage. However, this type of large-scale studies using AlphaFold2 requires high resources, likely unavailable for most academic institutes. Here, we identified three transmembrane structures qualified for fully blind AF2 predictions (Fig. 3 and Fig. 4). The outputs suggested that AlphaFold2 can be reliably used for building TM structures in a blind setup. Intriguingly, both LmrP and MrpF predictions indicated that running AF2 with different random seeds may be a valid approach to predict structures corresponding to different conformational states.

Furthermore, our results demonstrate that AlphaFold2 is a highly valuable tool in many areas of TM protein research. The correction of the register shift by AF2 in ABCG8 NBD (Fig. 5a), supports the application of AlphaFold2 in molecular replacement protocols aiding experimental structure determination [49]. In addition, screening experimental structures with their corresponding AF2 structures may detect structural errors and contribute to improving PDB database quality. Similarly, the absence of the kink in CFTR TM8 in an AF2 model predicted with disabled template usage (Fig. 5f) raises novel questions that will lead us to a deeper understanding of CFTR channel function. Importantly, our runs resulting in the corrected registry shift in ABCG8 are indications against an overfitting in the neural network behind AlphaFold2 and for overcoming memory footprints originating from training. We also demonstrated that AF2 was capable of predicting transmembrane dimer structures independently of their homo- or heteromeric nature (Fig. 5b and Fig. S5), while AF2 was not trained for multimer predictions. Though, this success may be at least


(a)
G2 6HCO AVLSFHNICY
G8 5DO7 NSLYFTYSGQ
G2 6HCO AVLSFHNICY
G8 seq NTLEVRDLNY
G2 6HCO AVLSFHNICY
G8 AF NTLEVRDLNY
(b)
(c)
(d)
(e)
(f)
(g)
(h)

◂**Fig. 5** AF2 predictions and ABC structure-associated issues. (**a**) ABCG2 and ABCG8 NBD β1 strand sequence alignments generated by structural alignment of 6hco (ABCG2) and 5do7 (ABCG5/ABCG8), by ClustalW with manual adjustment of ABCG2 and ABCG8 sequences based on ABCG2 structures, and by structural alignment of ABCG2 and AF2-predicted ABCG8 NBDs. Structure: AF2 ABCG8 NBD, blue: β1 strand, red: the segment corresponding to the β1 strand in the registry shifted 5do7 NBD, cyan: gating loop or regulatory insertion. (**b**) Structural alignment of 7jr7 (gray) and AF2-predicted (blue) ABCG5/ABCG8 TM domains (top view). Non-conserved loops with low-quality predictions are red. (**c**) Aligned homology (orange: TMD1, red: TMD2) and AF2 (blue: TMD1, cyan: TMD2) models of AtABCG36. Blue and orange spheres label F589 and F592 in TM2 facing the substrate binding pocket. (**d**) The magnified view of AtABCG36 TM1 and TM2 indicates that the alignments are not shifted but that spatial localization and side chain packing differ. (**e**) TM2 in the homology model unwinds in MD simulations. (**f**) zfCFTR TM8 is kinked in PDBID:5w81 (red) along with other structures and it is straight in both MRP1-based model (orange) and AF2-predicted structure (blue). The helices are extracted from a full TM domain alignment for visualization. (**g**) Surface representation of zfCFTR (PDBID:5w81). Red: TM8, green: TMD1, cyan: TMD2, pale green: NBD1, pale cyan: NBD2, black spheres: CAVER spheres indicating channel opening towards the extracellular space and the extracellular boundary of the lipid bilayer. (**h**) Surface representation of zfCFTR with MRP1-modelled, straight TM8. No lateral opening to the extracellular membrane leaflet can be observed

partially caused by the footprint of these proteins themselves in the AF2 neural network, successful protein-peptide docking [50], when peptides were not involved in alignments, is an argument against this reasoning. Interestingly, the novel deep learning model, AlphaFold2-Multimer [51], trained for predicting protein complexes is reported to excel AlphaFold2 in heteromeric but not in homomeric predictions.

In summary, our study underscores that AlphaFold2 can provide reliable protein structures also for transmembrane proteins and perform well in many areas associated with structural analysis of TM proteins. While the artificial intelligence inside AlphaFold2 can predict valuable structural information and correct structure-related flaws (e.g. registry shift, alignments, TM topology prediction, etc.), the limited predictive power of structural models from blind predictions involving flexible regions retain experimental validation desirable.

Methods

Databases and associated software

AlphaFold2 structures predicted for 21 proteomes were downloaded from https://alphafold.ebi.ac.uk in July, 2021. Proteins and their structures are identified in the manuscript with their UniProt accession number. Human Transmembrane Protein database [15] (2021-06-02) was received as an XML file from http://htp.enzim.hu. The data also contained CCTOP [16] (http://cctop.enzim.ttk.mta.hu) predictions and their reliability values. The hydrophobic thickness of experimentally determined human TM protein structures was retrieved from the PDBTM database (http://pdbtm.enzim.hu, 2021-07-23) [17]. Python was used to parse their XML files.

ABC PFAM entries were identified at https://pfam.xfam.org ($n=28$) and extracted from the Pfam-A.hmm file. The selected entries and their accession numbers are listed in Table S2. The sequence of every AF2 structure was searched using HMMER hmmsearch (http://hmmer.org) [52]. The E parameter was set to 0.001 and the match length was restricted to a minimum of 90% of the HMM profile length. The hmmsearch output was parsed using BioPython [53].

Novel structural folds for multi-pass α-helical transmembrane proteins were collected by extensive literature search (match: MprF) and by manual screening of the membrane protein selection of the SCOP database [54] (80 fold families and their subfamilies; http://scop.mrc-lmb.cam.ac.uk/term/2) and corresponding entries in the PFAM database [55] (matches: MlaE and EMC6).

Data analysis and visualization

MDAnalysis [56] and NumPy [57] Python packages were used for calculation of mean pLDDT values and hydrophobic membrane thickness. The pLDDT value of each residue were extracted from the B-factor column of AF2 structure files. For TM thickness calculation end positions of TM helices were retrieved from HTP/CCTOP and divided into two groups representing the two sides of the membrane. Plotting was done with Matplotlib (https://matplotlib.org) [58].

TM-score was calculated with TMalign [59]. Reference ABC structures are listed and shown in Fig. S3. Their TM domains were selected manually.

Molecular visualization and RMSD calculation were performed using PyMOL (The PyMOL Molecular Graphics System, Version 2.4.0 Schrödinger, LLC). RMSD of MD trajectories was calculated with the GROMACS rms tool.

Running AlphaFold2

AlphaFold2 was downloaded from github and installed as described (https://github.com/deepmind/alphafold) on a Debian 10 box with an AMD Ryzen Threadripper 2950X 16-Core Processor. 96 GB RAM was installed and ~ 75 GB peak usage was observed during jackhmmer run. The calculation was accelerated by an NVidia Quadro P6000 GPU with 24 GB RAM, which was almost fully utilized when the predicted sequence length was 1571. The required databases were located on two 2 TB HDD in a RAID0 setup. Typical run timings were: "features": 25–60 min, "predict_and_compile_model_*": 3–50 min, "relax_model_*": 1 min—6 h based on input sequences between 290 and 1571 a.a. length.

To exclude CFTR structures as templates from predictions, we modified run_alphafold.py, docker/run_docker.py, and alphafold/data/templates.py scripts to implement a -skip function. The modified scripts can be downloaded from http://alphafold.hegelab.org. Template usage was disabled by setting –max_template_date option to 1900-01-01. Dimer predictions were run by concatenating sequences with a part of the intrinsically disordered CFTR R domain, a.a. 675–800. pLDDT scores and ranking of predicted structures were extracted from the ranking_debug.json file.

Homology modelling

AtABCG36 (UniProt ACC: Q9XIE2) was homology modeled based on an ABCG2 homodimer structure (PDBID: 6hzm) using Modeller [60]. Sequence alignment was generated using ClustalW [61] and adjusted manually. One hundred structures were generated and the one with the best DOPE score was selected for MD simulations.

zfCFTR TM7 and TM8 was homology modeled similarly. The two helices were set for modelling based on the corresponding regions of MRP1 (PDBID: 5uj9 [23]) and the rest was kept static and based on the 5w81 zfCFTR structure.

Molecular dynamics simulations

MD simulations with AtABCG36 were performed using GROMACS 2019 with the CHARMM36m force field [62, 63]. Simulation systems were prepared using CHARMM-GUI [64, 65]. Structural models were oriented according to the OPM (Orientations of Proteins in Membranes) database [66] and all N- and C-termini were patched with ACE (acetyl) and CT3 (N-Methylamide) groups, respectively. The proteins were inserted in a bilayer with 1:1 POPC:PLPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine: 1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine) in the extracellular leaflet and 45:40:10:5 POPC:PLPC:POPS:PIP2 (POPS: 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine, PIP2: phosphatidylinositol 4,5-bisphosphate) in the intracellular leaflet. Both systems with the homology model or the AF2 structure were energy minimized using the steepest descent integrator (values for max. steps 50,000 and max. force 500 kJ/mol/nm were set). Six equilibration steps, according to the standard CHARMM-GUI protocol, were applied with decreasing position restraints. In the 50 ns (homology model) and 500 ns (AF2 model) long production runs, Nosé-Hoover thermostat and Parrinello-Rahman barostat with semiisotropic coupling were employed. Time constants for the thermostat and the barostat were set to 1 picosecond and 5 picosecond, respectively. The fast smooth PME algorithm [67] and LINCS algorithm [68] were used to calculate electrostatic interactions and to constrain bonds, respectively. GROMACS rmsf tools were used to calculate RMSF (root mean square fluctuation).

Simulations with the zfCFTR structure containing the kinked TM8 have been published and the protocol and parameters were described there [12]. The structure with the straightened, MRP1-based TM8 was subjected to MD simulations using the same protocol, including the same version of GROMACS, force field, and lipid composition. Channel pathways were determined using CAVER [69] as described in [12].

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00018-021-04112-1.

Acknowledgements We thank J. Jumper (DeepMind, UK), H. Tordai, R. Padányi (Semmelweis University, Hungary) and G. Gyimesi (University of Bern, Switzerland) for their helpful suggestions. We acknowledge the computational resources made available B. Babics (Boblem IT Co.), Governmental Information-Technology Development Agency (https://hpc.kifu.hu), the Grubmüller laboratory at Max Planck Institute (https://www.mpibpc.mpg.de/grubmueller), and Wigner GPU Laboratory (http://gpu.wigner.mta.hu) and we thank the help of their members.

Author contributions TH, MG, and GL conceived ideas and wrote the manuscript. TH and BF performed calculations and their analysis.

Funding Open access funding provided by Semmelweis University. This work was supported by funds to T. Hegedűs from the Cystic Fibrosis Foundation (CFF): HEGEDU20I0 and from NRDIO/NKFIH: K127961; to G. Lukacs from CCF LUKACS20G0, CIHR, CFI and Canada Research Chair Program to G. Lukacs; to M. Geisler from the Swiss National Funds (310030_197563).

Data availability All input data are available from public resources and their accession numbers are listed.

Code availability Modified AlphaFold2 scripts can be downloaded from http://alphafold.hegelab.org.

Declarations

Conflict of interest None.

References

1. Kuhlman B, Bradley P (2019) Advances in protein structure prediction and design. Nat Rev Mol Cell Biol 20:681–697. https://doi.org/10.1038/s41580-019-0163-x
2. Won J, Baek M, Monastyrskyy B et al (2019) Assessment of protein model structure accuracy estimation in CASP13: challenges in the era of deep learning. Proteins 87:1351–1360. https://doi.org/10.1002/prot.25804
3. Jumper J, Evans R, Pritzel A et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature. https://doi.org/10.1038/s41586-021-03819-2
4. Pereira J, Simpkin AJ, Hartmann MD et al (2021) High-accuracy protein structure prediction in CASP14. Proteins. https://doi.org/10.1002/prot.26171
5. Tunyasuvunakool K, Adler J, Wu Z et al (2021) Highly accurate protein structure prediction for the human proteome. Nature. https://doi.org/10.1038/s41586-021-03828-1
6. (2021) https://github.com/deepmind/alphafold. DeepMind. Accessed 30 July 2021
7. Mirdita M, Ovchinnikov S, Steinegger M (2021) ColabFold—making protein folding accessible to all. bioRxiv. https://doi.org/10.1101/2021.08.15.456425
8. Hamdoun A, Hellmich UA, Szakacs G, Kuchler K (2021) The incredible diversity of structures and functions of ABC transporters. FEBS Lett 595:671–674. https://doi.org/10.1002/1873-3468.14061
9. Sarkadi B, Homolya L, Hegedűs T (2020) The ABCG2/BCRP transporter and its variants—from structure to pathology. FEBS Lett. https://doi.org/10.1002/1873-3468.13947
10. Nagy T, Tóth Á, Telbisz Á et al (2020) The transport pathway in the ABCG2 protein and its regulation revealed by molecular dynamics simulations. Cell Mol Life Sci. https://doi.org/10.1007/s00018-020-03651-3
11. Csanády L, Vergani P, Gadsby DC (2019) Structure, gating, and regulation of the CFTR anion channel. Physiol Rev 99:707–738. https://doi.org/10.1152/physrev.00007.2018
12. Farkas B, Tordai H, Padányi R et al (2019) Discovering the chloride pathway in the CFTR channel. Cell Mol Life Sci. https://doi.org/10.1007/s00018-019-03211-4
13. Srikant S (2020) Evolutionary history of ATP-binding cassette proteins. FEBS Lett 594:3882–3897. https://doi.org/10.1002/1873-3468.13985
14. Thomas C, Aller SG, Beis K et al (2020) Structural and functional diversity calls for a new classification of ABC transporters. FEBS Lett 594:3767–3775. https://doi.org/10.1002/1873-3468.13935
15. Dobson L, Reményi I, Tusnády GE (2015) The human transmembrane proteome. Biol Direct. https://doi.org/10.1186/s13062-015-0061-x
16. Dobson L, Reményi I, Tusnády GE (2015) CCTOP: a consensus constrained TOPology prediction web server. Nucleic Acids Res 43:W408-412. https://doi.org/10.1093/nar/gkv451
17. Kozma D, Simon I, Tusnády GE (2013) PDBTM: protein data bank of transmembrane proteins after 8 years. Nucleic Acids Res 41:D524-529. https://doi.org/10.1093/nar/gks1169
18. Zhang Y, Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. Proteins 57:702–710. https://doi.org/10.1002/prot.20264
19. Xu J, Zhang Y (2010) How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics 26:889–895. https://doi.org/10.1093/bioinformatics/btq066
20. Saier MH, Reddy VS, Moreno-Hagelsieb G et al (2021) The transporter classification database (TCDB): 2021 update. Nucleic Acids Res 49:D461–D467. https://doi.org/10.1093/nar/gkaa1004
21. Bakos E, Evers R, Calenda G et al (2000) Characterization of the amino-terminal regions in the human multidrug resistance protein (MRP1). J Cell Sci 113(Pt 24):4451–4461
22. Zhang Z, Chen J (2016) Atomic structure of the cystic fibrosis transmembrane conductance regulator. Cell 167:1586-1597.e9. https://doi.org/10.1016/j.cell.2016.11.014
23. Johnson ZL, Chen J (2017) Structural basis of substrate recognition by the multidrug resistance protein MRP1. Cell 168:1075-1085.e9. https://doi.org/10.1016/j.cell.2017.01.041
24. Del Alamo D, Govaerts C, Mchaourab HS (2021) AlphaFold2 predicts the inward-facing conformation of the multidrug transporter LmrP. Proteins 89:1226–1228. https://doi.org/10.1002/prot.26138
25. Jumper J, Evans R, Pritzel A et al (2021) Applying and improving AlphaFold at CASP14. Proteins. https://doi.org/10.1002/prot.26257
26. Chi X, Fan Q, Zhang Y et al (2020) Structural mechanism of phospholipids translocation by MlaFEDB complex. Cell Res 30:1127–1135. https://doi.org/10.1038/s41422-020-00404-6
27. Structural and mechanistic basis of the EMC-dependent biogenesis of distinct transmembrane clients - PubMed. https://pubmed.ncbi.nlm.nih.gov/33236988/. Accessed 22 Nov 2021
28. Song D, Jiao H, Liu Z (2021) Phospholipid translocation captured in a bifunctional membrane protein MprF. Nat Commun 12:2927. https://doi.org/10.1038/s41467-021-23248-z
29. Ernst CM, Peschel A (2019) MprF-mediated daptomycin resistance. Int J Med Microbiol 309:359–363. https://doi.org/10.1016/j.ijmm.2019.05.010
30. Lee J-Y, Kinch LN, Borek DM et al (2016) Crystal structure of the human sterol transporter ABCG5/ABCG8. Nature 533:561–564. https://doi.org/10.1038/nature17666
31. Taylor NMI, Manolaridis I, Jackson SM et al (2017) Structure of the human multidrug transporter ABCG2. Nature 546:504–509. https://doi.org/10.1038/nature22345
32. Zhang H, Huang C-S, Yu X et al (2021) Cryo-EM structure of ABCG5/G8 in complex with modulating antibodies. Commun Biol 4:526. https://doi.org/10.1038/s42003-021-02039-8
33. Henriksen U, Fog JU, Litman T, Gether U (2005) Identification of intra- and intermolecular disulfide bridges in the multidrug resistance transporter ABCG2. J Biol Chem 280:36926–36934. https://doi.org/10.1074/jbc.M502937200
34. Serohijos AWR, Hegedus T, Riordan JR, Dokholyan NV (2008) Diminished self-chaperoning activity of the ΔF508 mutant of CFTR results in protein misfolding. PLoS Comput Biol 4:e1000008. https://doi.org/10.1371/journal.pcbi.1000008
35. László L, Sarkadi B, Hegedűs T (2016) Jump into a new fold—a homology based model for the ABCG2/BCRP multidrug transporter. PLoS ONE 11:e0164426. https://doi.org/10.1371/journal.pone.0164426
36. Khunweeraphong N, Stockner T, Kuchler K (2017) The structure of the human ABC transporter ABCG2 reveals a novel mechanism for drug extrusion. Sci Rep 7:13767. https://doi.org/10.1038/s41598-017-11794-w
37. Ferreira RJ, Bonito CA, Cordeiro MNDS et al (2017) Structure-function relationships in ABCG2: insights from molecular dynamics simulations and molecular docking studies. Sci Rep 7:15534. https://doi.org/10.1038/s41598-017-15452-z
38. Aryal B, Huynh J, Schneuwly J et al (2019) ABCG36/PEN3/PDR8 Is an exporter of the auxin precursor, indole-3-butyric acid, and involved in auxin-controlled development. Front Plant Sci 10:899. https://doi.org/10.3389/fpls.2019.00899
39. Heo L, Janson G, Feig M (2021) Physics-based protein structure refinement in the era of artificial intelligence. Proteins. https://doi.org/10.1002/prot.26161
40. Veit G, Avramescu RG, Chiang AN et al (2016) From CFTR biology toward combinatorial pharmacotherapy: expanded

classification of cystic fibrosis mutations. Mol Biol Cell 27:424–433. https://doi.org/10.1091/mbc.E14-04-0935
41. Zhang Z, Liu F, Chen J (2017) Conformational changes of CFTR upon phosphorylation and ATP binding. Cell 170:483-491.e8. https://doi.org/10.1016/j.cell.2017.06.041
42. Liu F, Zhang Z, Csanády L et al (2017) Molecular structure of the human CFTR Ion channel. Cell 169:85-95.e8. https://doi.org/10.1016/j.cell.2017.02.024
43. Zhang Z, Liu F, Chen J (2018) Molecular structure of the ATP-bound, phosphorylated human CFTR. Proc Natl Acad Sci USA. https://doi.org/10.1073/pnas.1815287115
44. Liu F, Zhang Z, Levit A et al (2019) Structural identification of a hotspot on CFTR for potentiation. Science 364:1184–1188. https://doi.org/10.1126/science.aaw7611
45. Fay JF, Aleksandrov LA, Jensen TJ et al (2018) Cryo-EM visualization of an active high open probability CFTR anion channel. Biochemistry 57:6234–6246. https://doi.org/10.1021/acs.biochem.8b00763
46. Carveth K, Buck T, Anthony V, Skach WR (2002) Cooperativity and flexibility of cystic fibrosis transmembrane conductance regulator transmembrane segments participate in membrane localization of a charged residue. J Biol Chem 277:39507–39514. https://doi.org/10.1074/jbc.M205759200
47. Kishi KE, Kim YS, Fukuda M et al (2021) Structural basis for channel conduction in the pump-like channelrhodopsin ChRmine. bioRxiv. https://doi.org/10.1101/2021.08.15.456392
48. Ivetac A, Sansom MSP (2008) Molecular dynamics simulations and membrane protein structure quality. Eur Biophys J 37:403–409. https://doi.org/10.1007/s00249-007-0225-4
49. McCoy AJ, Sammito MD, Read RJ (2021) Possible implications of AlphaFold2 for crystallographic phasing by molecular replacement. BioRxiv. https://doi.org/10.1101/2021.05.18.444614
50. Tsaban T, Varga J, Avraham O et al (2021) Harnessing protein folding neural networks for peptide-protein docking. bioRxiv. https://doi.org/10.1101/2021.08.01.454656
51. Evans R, O'Neill M, Pritzel A et al (2021) Protein complex prediction with AlphaFold-multimer. bioRxiv. https://doi.org/10.1101/2021.10.04.463034
52. Eddy SR (2011) Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195
53. Cock PJA, Antao T, Chang JT et al (2009) Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423. https://doi.org/10.1093/bioinformatics/btp163
54. Andreeva A, Kulesha E, Gough J, Murzin AG (2020) The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. Nucleic Acids Res 48:D376–D382. https://doi.org/10.1093/nar/gkz1064
55. Mistry J, Chuguransky S, Williams L et al (2021) Pfam: the protein families database in 2021. Nucleic Acids Res 49:D412–D419. https://doi.org/10.1093/nar/gkaa913
56. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem 32:2319–2327. https://doi.org/10.1002/jcc.21787
57. Harris CR, Millman KJ, van der Walt SJ et al (2020) Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2
58. Hunter JD (2007) Matplotlib: A 2D graphics environment. Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55
59. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33:2302–2309. https://doi.org/10.1093/nar/gki524
60. Fiser A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol 374:461–491. https://doi.org/10.1016/S0076-6879(03)74020-8
61. Larkin MA, Blackshields G, Brown NP et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. https://doi.org/10.1093/bioinformatics/btm404
62. Huang J, Rauscher S, Nawrocki G et al (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14:71–73. https://doi.org/10.1038/nmeth.4067
63. Abraham MJ, Murtola T, Schulz R et al (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001
64. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29:1859–1865. https://doi.org/10.1002/jcc.20945
65. Wu EL, Cheng X, Jo S et al (2014) CHARMM-GUI membrane builder toward realistic biological membrane simulations. J Comput Chem 35:1997–2004. https://doi.org/10.1002/jcc.23702
66. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. Bioinformatics 22:623–625. https://doi.org/10.1093/bioinformatics/btk023
67. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092. https://doi.org/10.1063/1.464397
68. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3c1463::AID-JCC4%3e3.0.CO;2-H
69. Petrek M, Otyepka M, Banás P et al (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities. BMC Bioinform 7:316. https://doi.org/10.1186/1471-2105-7-316

Published online 29 October 2020 dc_2020_22 Nucleic Acids Research, 2021, Vol. 49, Database issue D355–D360
doi: 10.1093/nar/gkaa954

The MemMoRF database for recognizing disordered protein regions interacting with cellular membranes

Georgina Csizmadia [1,†], Gábor Erdős [2,†], Hedvig Tordai [1], Rita Padányi [1], Silvio Tosatto [3], Zsuzsanna Dosztányi [2] and Tamás Hegedűs [1,*]

[1]Department of Biophysics and Radiation Biology, Semmelweis University, Budapest 1094, Hungary, [2]MTA-ELTE Lendület Bioinformatics Research Group, Department of Biochemistry, Eötvös Loránd University, Budapest 1117, Hungary and [3]Department of Biomedical Sciences, University of Padua, Padua 35131, Italy

Received August 08, 2020; Revised September 25, 2020; Editorial Decision October 06, 2020; Accepted October 28, 2020

ABSTRACT

Protein and lipid membrane interactions play fundamental roles in a large number of cellular processes (e.g. signalling, vesicle trafficking, or viral invasion). A growing number of examples indicate that such interactions can also rely on intrinsically disordered protein regions (IDRs), which can form specific reversible interactions not only with proteins but also with lipids. We named IDRs involved in such membrane lipid-induced disorder-to-order transition as MemMoRFs, in an analogy to IDRs exhibiting disorder-to-order transition upon interaction with protein partners termed Molecular Recognition Features (MoRFs). Currently, both the experimental detection and computational characterization of MemMoRFs are challenging, and information about these regions are scattered in the literature. To facilitate the related investigations we generated a comprehensive database of experimentally validated MemMoRFs based on manual curation of literature and structural data. To characterize the dynamics of MemMoRFs, secondary structure propensity and flexibility calculated from nuclear magnetic resonance chemical shifts were incorporated into the database. These data were supplemented by inclusion of sentences from papers, functional data and disease-related information. The MemMoRF database can be accessed via a user-friendly interface at https://memmorf.hegelab.org, potentially providing a central resource for the characterization of disordered regions in transmembrane and membrane-associated proteins.

INTRODUCTION

Many proteins contain intrinsically disordered regions (IDRs) that do not fold into well-defined structures in isolation and are best represented by conformational ensembles (1,2). The flexibility enables IDRs to participate in highly specific and reversible interactions, which form a common molecular basis for a wide range of physiological processes, including signalling, gene regulation, cell cycle regulation, scaffolding, or chaperoning and pathological conditions (3–5). Such segments, which show disorder-to-order transition upon forming protein–protein interactions, are called molecular recognition features (MoRFs) (6–8). Recent evidence suggests that IDRs are also important components of membrane-associated proteins (MAPs) and transmembrane proteins (TMPs). In these proteins disordered regions of variable sizes can be located in loops between transmembrane segments or in N- and C-terminal regions (9). Many of these disordered regions possess a common feature, namely, they exhibit lipid induced alteration of ordered/disordered status. Here we suggest to term the segments involved in protein–lipid interactions as MemMoRFs, since they represent a distinct category of context-dependent behaviour when compared to MoRFs.

The fundamental role of MemMoRFs in various cellular functions is demonstrated by their ubiquitous presence in proteins associated with a wide range of cellular processes. A set of proteins with MemMoRFs participates in the regulation of cell cycle, apoptosis and phagocytosis (e.g. NOTCH1 (10) and interferon alpha-inducible protein 27-like protein 1 (11)). Other MemMoRF containing proteins play important roles in cellular trafficking and shaping the cell membrane (e.g. Myc box-dependent-interacting protein 1 (12), α-synuclein (13) and pro-neuregulin-1 (14)). MemMoRFs are expected to be abundant in signalling proteins and were detected in both transmembrane receptors (e.g. integrin β3 (15), PTEN tumor suppressor (16) and receptor tyrosine kinases, such as EGFR (17) and erbB-2 (18)) or membrane-associated proteins (e.g. tyrosine-protein ki-

*To whom correspondence should be addressed. Tel: +36 1 4591500/60233; Fax: +36 1 266 6656; Email: hegedus@hegelab.org
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

nase Src (19)). MemMoRFs also play important roles in pathological processes. An important MemMoRF group is associated with neurodegenerative diseases and includes α-synuclein, amyloid-β precursor protein, and prion protein (20). In addition, MemMoRFs have been identified in several viral proteins, such as capsid proteins of HIV1 and HCV (21,22), and TMPs, including a potassium channel of the Influenza A virus (23). Most likely, these viral proteins and various toxins, such as melittin (24) and the GPCR impairing toxin of *Gila monster* (25), require MemMoRF for engaging with or penetrating through membrane bilayers.

The specific conformational properties of these IDRs are shaped by the chemical composition and physical properties of the membrane, and the relevant IDRs can also shape the properties of the membrane (26–28). An important mode to regulate these protein–lipid interactions is phosphorylation. For example, the intracellular domain of the T-cell receptor ζ forms transient helices in the intracellular membrane leaflet (Supplementary Figure S1) (29). One of the transient helices (LID region) possesses an ITAM (immunoreceptor tyrosine-based activation motif), which includes a tyrosine immersed in the membrane bilayer and is a target for phosphorylation (29). When phosphorylated, the ITAM MemMoRF undocks from the membrane and loses its α-helical character. A similar feature can be observed in transporter regulation, involving single pass TMPs possessing MemMoRFs, such as phospholamban (PLN) and FXYD proteins (e.g. phospholemman, FXYD domain-containing ion transport regulator 4 and Na^+/K^+ ATPase subunit γ) (Supplementary Figure S1) (30–32). The small size of these proteins makes them attractive for drug targeting (e.g. PLN in cardiac diseases) (33).

Although disorder-to-order transition was detected in numerous studies targeting protein–lipid interactions, this phenomenon has not been examined in a generalized manner. As a consequence, information about these protein segments are scattered in the literature, and the detailed structural properties of MemMoRFs are difficult to interpret in the context of additional structural and functional modules. Here, we report a novel, comprehensive database of experimentally validated MemMoRFs, currently containing 131 examples in 96 proteins as a gold standard set. The MemMoRF database complements existing databases of disordered proteins in general (34,35) and also databases of interactions of disordered proteins with ordered or other disordered proteins (36–38). A separate database was built for MoRFs of protein-protein interactions of MAPs and TMPs (38). In contrast, we collected those flexible protein regions into our MemMoRF database, which participate in reversible membrane binding, associated with conformational changes.

IDENTIFICATION AND SYSTEMATIC COLLECTION OF MemMoRFs

The main aim of this database is to facilitate the investigations and targeting of proteins involving MemMoRFs. As a first step, we collected nuclear magnetic resonance (NMR) structures of TMPs and MAPs, since IDRs of these proteins can be located in the close vicinity of biological membranes. We strongly relied on NMR data, since this spectroscopy method is the most widely used approach to determine lipid interaction of proteins at residue level. NMR structures of TMPs were gathered using the Membrane Protein Browser of Protein Data Bank (PDB) (39). MAPs were looked up in the UniProt database (40) using the 'peripheral membrane protein' subcellular location keyword (UniProt 2019.01.14). To characterize the transient secondary structures of IDRs, we did not rely solely on the deposited NMR structures, but also incorporated secondary structure propensity (SSPop; using δ2D) (41) and flexibility (1-S^2; using Random Coil Index, RCI) (42). We considered a region intrinsically disordered if the residues within the region exhibited 'coil' secondary structure population (>0.5) and high flexibility (>0.15). A region was also considered disordered in the case of corresponding disorder annotation in DisProt, DIBS, MFIB or PFAM databases (34,36–37,43). As a next stage, we also explored invisible regions of X-ray and cryo-EM structures, which belong to TMPs or MAPs possessing annotated disordered regions. We subjected the whole protein set to extensive literature analysis to identify IDRs and MemMoRFs or further validate MemMoRFs detected using NMR or other structural data. Thus, proofs of disorder could be derived from calculations, databases, structural information and literature evidence.

A total number of 538 proteins including 206 TMPs and 332 MAPs were screened for segments located in an IDR and exhibiting altered dynamics in a membrane mimetic. A total of 149 membrane interacting regions were identified in 107 proteins. Only 11 regions were derived exclusively from unresolved residues in X-ray or cryo-EM structures. 131 out of the membrane interacting regions are disordered in aqueous solution and become ordered ($n = 107$) or retain flexibility ($n = 19$) upon binding to membrane mimetics. Eighteen regions with stable secondary structure both in solution and in membrane bound state were found. Eighty-four out of 149 membrane interacting regions are in TMPs and the remaining 65 are in MAPs. Intracellular, extracellular and periplasmic localizations were observed in 121, 23 and 2 cases, respectively (Figure 1). SSPop and flexibility values were calculated from published chemical shift data with sufficient quality, available for 41% of our protein dataset. Using data from NMR experiments, X-ray and cyro-EM structures and literature, we found 92 regions among identified IDRs, which were not annotated in the DisProt database (Supplementary Tables S1 and S2).

We categorized the entries depending how their dynamical properties changed upon interaction with the lipid bilayer. In many cases, the corresponding NMR data indicates that the observed transitions resulted in increased α-helical propensity values. In addition, some disordered segments retained their conformational freedom when interacted with membrane mimetics or specific membrane lipids. Although these segments still likely sample a reduced conformational space in the presence of a membrane environment, we labeled their transition as 'disorder-to-disorder' to emphasize the dynamics of their membrane bound state. In the case of some of the MemMoRFs derived from X-ray or cryo-EM structures, there is no information on the lipid-bound structure, thus we labeled these entries as 'disorder-to-unkown'. We also included cases, where the binding helices were stable both in aqueous solution and interact-

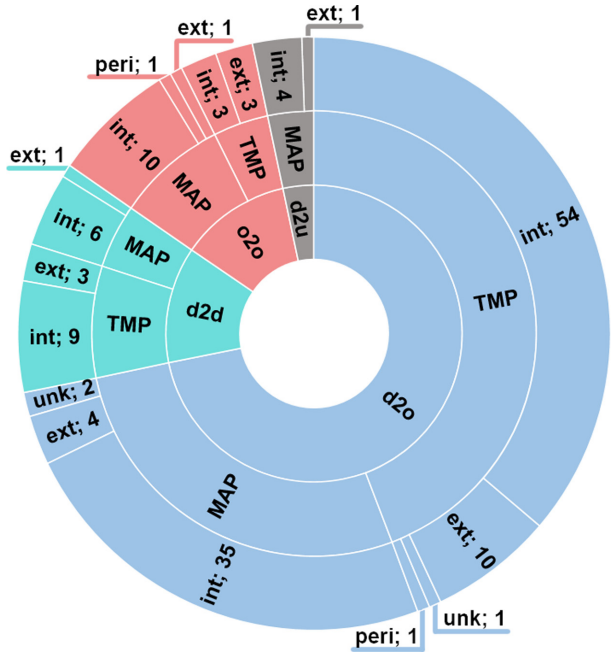


Figure 1. Distribution of MemMoRF types in the database. A total of 149 membrane interacting regions were identified in 107 proteins. One hundred and thirty-one out of these regions are disordered in aqueous solution and become ordered ($n = 107$) or retain flexibility ($n = 19$) upon binding to membrane mimetics. Eighteen regions with stable secondary structure both in solution and in membrane bound state were found. d2o: disorder-to-order, d2d: disorder-to-disorder, o2o: order-to-order, d2u: disorder-to-unknown, TMP: transmembrane protein, MAP: membrane associated protein, int: intracellular side, ext: extracellular side, peri: periplasmic side, unk: unknown location.

ing with the membrane (e.g. helices from kinase suppressor of Ras 1 and BH3-interacting domain death agonist). We labeled these entries as 'bistable helix'. While they are not classical MemMoRFs, these segments still exhibit a dynamic equilibrium between lipid and water phases and participate in regulatory interactions.

To demonstrate the analysis of various data for IDR and MemMoRF validation and phosphorylation-dependent regulatory disorder-to-order transition, we selected integrin β3, which plays a role in angiogenesis and tumor growth (44,45). Targeting integrin β3 associated signalling was shown to induce apoptosis of endothelial tumor cells (44) and cell permeable peptides derived from integrin β cytoplasmic tails (CT) were developed for angiogenesis inhibition (46). Since these peptides overlap with a MemMoRF, a detailed description of phosphorylation dependent protein and membrane interactions of this region will help to improve the potential therapeutic use of integrin β CT peptides. The complexity of various conformations of the C-terminal integrin β3 MemMoRF under different conditions is shown in Figure 2 and accessible at https://memmorf.hegelab.org/entry/P05106. The C-terminal MemMoRF region (a.a. 770–784) exhibits low helix propensity (<0.5) and high flexibility values (>0.15) in the absence of a membrane mimetic (PDB ID: 2KNC and BMRB ID: 16496) that indicates a disordered state for these residues (47). Although the helical propensity for a.a. 771–776 is lower than the commonly used SSPop threshold (0.5), a small stable helix is present in all of the 20 structures deposited in the 2KNC PDB structure. However, this part of the protein is marked disordered in the DisProt and PFAM databases. Importantly, in the presence of a lipid environment (PDB ID: 2KV9 and BMRB ID: 16771) (48) the decreased flexibility, which approaches the disorder threshold (0.15), confirms the presence of a MemMoRF. At the same time, the helical propensity increases and indicates a stable helix in a smaller part of the sequence (a.a. 776–781). This is in contrast with the long α-helix present in the structural ensemble deposited in the PDB. Moreover, in the paper accompanying these NMR structures (48) the authors noted the presence of transient helical contents and increasing disorder towards the C-terminus based on hydrogen-deuterium exchange experiments. Interestingly, phosphorylation of this MemMoRF promotes disorder (helical propensity <0.5 and flexibility >0.15) in the presence of a membrane mimetic as well (PDB ID: 2LJD and BMRB ID: 17930), thus serves as a conformational switch (49). Again, this NMR ensemble also exhibits a stable small helix between residues 775–780. The potential over-representation of helical content urged us to supplement the entries in the MemMoRF database with supporting statements from papers, and it also cautions database curators that making a decision about disorder should not rely solely on atomistic NMR structures. Phosphorylation-dependent regulation via MemMoRFs is further demonstrated by the T cell receptor ζ and phospholamban (Supplementary Figure S1).

WEB SERVER AND INTERFACE IMPLEMENTATION

In order to make MemMoRF data accessible and to facilitate research on proteins possessing MemMoRF regions, we developed a web application available at https://memmorf.hegelab.org.

Browsing and searching data

The database provides both browsing and searching functionalities in the 'Browse' page, where all entries are listed in a table format. Table columns contain the protein and the gene name, the source organism, the UniProt accession number type, the corresponding PDB and BMRB identifiers and MIM entry, if available, for each protein in our database. The table is sortable and can be filtered by simple queries or using dropdown lists. Searches independent from columns can be performed by typing a query in a text box located at the top of the page.

Entry pages

Each entry starts with the most important, basic information about the protein from UniProt (protein name, accession code, gene name, organism and keywords). Then, MemMoRFs are listed along with their boundaries, type of transition upon membrane binding (e.g. disorder-to-order), localization (e.g. intracellular) and statements supporting disordered and membrane interacting nature of the region.

The SSPop and flexibility calculations, if available, are shown in a graph, which can be saved using the controls next to it. Visibility of corresponding lines can

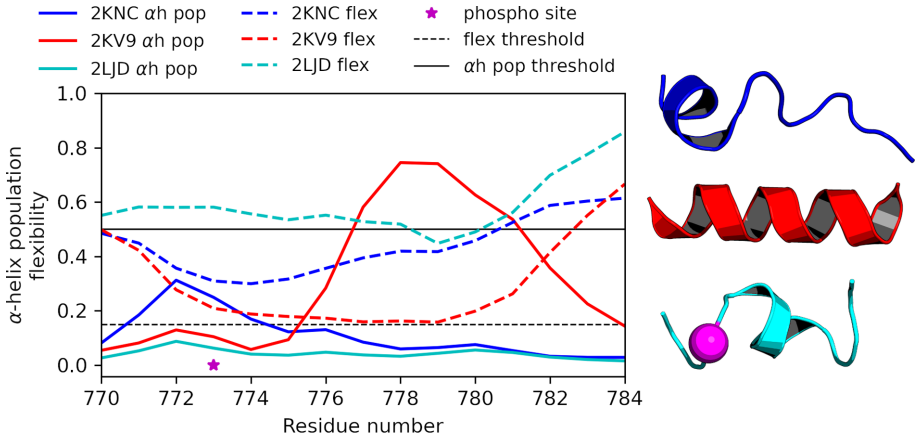


Figure 2. Experimental NMR data provide information on disorder level thus input for MemMoRF identification. Secondary structure population (e.g. helix and coil) and flexibility (1-S^2) values were utilized to characterize the per residue disorder-order propensity in NMR ensembles, as exemplified by a MemMoRF from integrin β3. Blue: in organic solvent; red: in DPC; cyan: phosphorylated in DPC; magenta: phosphorylation site; αh pop: α-helix population calculated by δ2D; flex: 1-S^2 calculated by RCI, αh pop threshold: 0.5, flex threshold: 0.15.

be toggled by clicking on their legend entry. NMR experiments associated with the entry can be selected in a box, right from the graph. This box lists the membrane mimetic used in the NMR experiment, the PDB ID of the calculated NMR ensemble, the BMRB ID and the '_Assigned_chem_shift_list.ID' of chemical shift data used for the calculation. The '_Assigned_chem_shift_list.ID' identifies a specific set of chemical shift values from the BMRB NMR-STAR files, which may contain multiple sets acquired under different experimental conditions.

Sequence specific data, including MemMoRF regions, transmembrane (TM) helices, PFAM domains (43), short linear motifs from ELM (50), post-translational modifications from PhosphoSitePlus (51), IDRs from DisProt, DIBS and MFIB (34,36–37), and segments covered in the PDB are also shown. To further characterize the sequences, Wimley-White hydrophobicity plots and IUPred short predictions were also included into the main graph (52,53). Structures strongly associated with the MemMoRF can be selected for display and manipulation using LiteMol (54). MemMoRFs and TM helices are labelled by default on the structure. In the bottom region of the entry page, additional information is listed to assess the role of MemMoRFs in pathophysiological states. These data include disease phenotypes from MIM (55), disease causing and polymorphic variations from dbSNP (56), and DrugBank records (57). Protein-protein interactions with a link to the IntAct (58) and STRING (59) entries of the protein are followed by an interactive, embedded STRING network graph.

Help and feedback pages

Although the web application is simple and self-explanatory, we provide a comprehensive help page with sections about the following items: i) MemMoRF definition, ii) information resources linked to the memMoRF database, iii) details of our data collection pipeline, iv) search possibilities of the browse page, v) the content of the entry page and vi) links to the statistics and download pages. Users are encouraged to submit comments or questions via a contact form or email located in the feedback page.

Server implementation and database structure

The web application is served via a DJANGO (version 2.1.1) based web interface, fuelled by an SQL database providing fast access to data even in the case of parallel queries. The SQL database contains all the information collected from the literature and various databases (e.g. UniProt, dbSNP, MIM and Drug Bank), organized into multiple tables. Each record in the UniProt table represents a single protein and has the UniProt accession number as a unique key. Other data, including records in the MemMoRF table, are linked to this table by UniProt accession number. In order to provide the best possible user experience on various devices and browsing options for users, the front-end compatibility of MemMoRF is supported by a combination of bootstrap (version 4.3.1) and JQuery (version 2.1.4).

In addition to data access through the web application interface, data can also be downloaded in JSON, XML or TSV formats, or by a RESTful API serving standard JSON format (e.g. https://memmorf.hegelab.org/rest/Q9NR61.json).

CONCLUSION

MemMoRF database delivers data on membrane interacting disordered regions of TMPs and MAPs. These regions exhibit dynamic structural alterations during transition from solvent to a membrane bound state. The collected information has a high potential to contribute to understanding crucial cellular functions and pathological states associated with membrane proteins. Our database provides significant benefits for the broad scientific community because of the following: i) it is a freely accessible, easy-to-use, and organized resource; ii) it includes a gold standard set for determining the sequence requirements for lipid interaction of disordered regions; iii) it represents a high-quality set for developing novel *in silico* pipelines for MemMoRF

identification; and iv) it promotes further experimental investigation and development of drug targeting approaches for MemMoRF containing proteins.

We also provide a new set of disordered protein regions, which are not currently present in other manually maintained databases, based on careful literature curation and NMR-based calculation of secondary structure propensities and flexibility. Our aim is to establish and maintain the MemMoRF database in the long term as a central resource for membrane proteins with lipid bilayer interacting disordered segments.

DATA AVAILABILITY

MemMoRF database and web application are available at http://memmorf.hegelab.org. Since the server is located in the EU, it fully adheres to the General Data Protection Regulation (GDPR).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Pietro Sormanni and Michele Vendruscolo (Department of Chemistry, University of Cambridge, Cambridge, UK) for providing their scripts for preparing and running δ2D calculations.

FUNDING

National Research, Development and Innovation Office [NKFIH 127961 to T.H.]; Higher Education Institutional Excellence Programme of the Ministry for Innovation and Technology in Hungary. Funding for open access charge: Semmelweis University, EFOP-3.6.3-VEKOP-16-2017-00009.
Conflict of interest statement. None declared.

REFERENCES

1. Sormanni,P., Piovesan,D., Heller,G.T., Bonomi,M., Kukic,P., Camilloni,C., Fuxreiter,M., Dosztanyi,Z., Pappu,R.V., Babu,M.M. *et al.* (2017) Simultaneous quantification of protein order and disorder. *Nat. Chem. Biol.*, **13**, 339–342.
2. Uversky,V.N. (2016) Dancing protein clouds: the strange biology and chaotic physics of intrinsically disordered proteins. *J. Biol. Chem.*, **291**, 6681–6688.
3. Dunker,A.K., Brown,C.J., Lawson,J.D., Iakoucheva,L.M. and Obradović,Z. (2002) Intrinsic disorder and protein function. *Biochemistry*, **41**, 6573–6582.
4. Dyson,H.J. and Wright,P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **6**, 197–208.
5. van der Lee,R., Buljan,M., Lang,B., Weatheritt,R.J., Daughdrill,G.W., Dunker,A.K., Fuxreiter,M., Gough,J., Gsponer,J., Jones,D.T. *et al.* (2014) Classification of intrinsically disordered regions and proteins. *Chem. Rev.*, **114**, 6589–6631.
6. Oldfield,C.J., Cheng,Y., Cortese,M.S., Romero,P., Uversky,V.N. and Dunker,A.K. (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry*, **44**, 12454–12470.
7. Mohan,A., Oldfield,C.J., Radivojac,P., Vacic,V., Cortese,M.S., Dunker,A.K. and Uversky,V.N. (2006) Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.*, **362**, 1043–1059.
8. Yang,J., Gao,M., Xiong,J., Su,Z. and Huang,Y. (2019) Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. *Protein Sci.*, **28**, 1952–1965.
9. Tusnády,G.E., Dobson,L. and Tompa,P. (2015) Disordered regions in transmembrane proteins. *Biochim. Biophys. Acta*, **1848**, 2839–2848.
10. Deatherage,C.L., Lu,Z., Kroncke,B.M., Ma,S., Smith,J.A., Voehler,M.W., McFeeters,R.L. and Sanders,C.R. (2017) Structural and biochemical differences between the Notch and the amyloid precursor protein transmembrane domains. *Sci. Adv*, **3**, e1602794.
11. Klammt,C., Maslennikov,I., Bayrhuber,M., Eichmann,C., Vajpai,N., Chiu,E.J.C., Blain,K.Y., Esquivies,L., Kwon,J.H.J., Balana,B. *et al.* (2012) Facile backbone structure determination of human membrane proteins by NMR spectroscopy. *Nat. Methods*, **9**, 834–839.
12. Löw,C., Weininger,U., Lee,H., Schweimer,K., Neundorf,I., Beck-Sickinger,A.G., Pastor,R.W. and Balbach,J. (2008) Structure and dynamics of helix-0 of the N-BAR domain in lipid micelles and bilayers. *Biophys. J.*, **95**, 4315–4323.
13. Ulmer,T.S., Bax,A., Cole,N.B. and Nussbaum,R.L. (2005) Structure and dynamics of micelle-bound human alpha-synuclein. *J. Biol. Chem.*, **280**, 9595–9603.
14. Chukhlieb,M., Raasakka,A., Ruskamo,S. and Kursula,P. (2015) The N-terminal cytoplasmic domain of neuregulin 1 type III is intrinsically disordered. *Amino Acids*, **47**, 1567–1577.
15. Vinogradova,O., Vaynberg,J., Kong,X., Haas,T.A., Plow,E.F. and Qin,J. (2004) Membrane-mediated structural transitions at the cytoplasmic face during integrin activation. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 4094–4099.
16. Nanda,H., Heinrich,F. and Lösche,M. (2015) Membrane association of the PTEN tumor suppressor: neutron scattering and MD simulations reveal the structure of protein-membrane complexes. *Methods*, **77-78**, 136–146.
17. Choowongkomon,K., Carlin,C.R. and Sönnichsen,F.D. (2005) A structural model for the membrane-bound form of the juxtamembrane domain of the epidermal growth factor receptor. *J. Biol. Chem.*, **280**, 24043–24052.
18. Bragin,P.E., Mineev,K.S., Bocharova,O.V., Volynsky,P.E., Bocharov,E.V. and Arseniev,A.S. (2016) HER2 transmembrane domain dimerization coupled with self-association of membrane-embedded cytoplasmic juxtamembrane regions. *J. Mol. Biol.*, **428**, 52–61.
19. Pérez,Y., Maffei,M., Igea,A., Amata,I., Gairí,M., Nebreda,A.R., Bernadó,P. and Pons,M. (2013) Lipid binding by the Unique and SH3 domains of c-Src suggests a new regulatory mechanism. *Sci. Rep.*, **3**, 1295.
20. Riek,R. and Eisenberg,D.S. (2016) The activities of amyloids from a structural perspective. *Nature*, **539**, 227–235.
21. Serrano,S., Araujo,A., Apellániz,B., Bryson,S., Carravilla,P., de la Arada,I., Huarte,N., Rujas,E., Pai,E.F., Arrondo,J.L.R. *et al.* (2014) Structure and immunogenicity of a peptide vaccine, including the complete HIV-1 gp41 2F5 epitope: implications for antibody recognition mechanism and immunogen design. *J. Biol. Chem.*, **289**, 6565–6580.
22. Penin,F., Brass,V., Appel,N., Ramboarina,S., Montserret,R., Ficheux,D., Blum,H.E., Bartenschlager,R. and Moradpour,D. (2004) Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J. Biol. Chem.*, **279**, 40835–40843.
23. Schnell,J.R. and Chou,J.J. (2008) Structure and mechanism of the M2 proton channel of influenza A virus. *Nature*, **451**, 591–595.
24. Appadu,A., Jelokhani-Niaraki,M. and DeBruin,L. (2015) Conformational changes and association of Membrane-Interacting peptides in myelin membrane models: a case of the C-Terminal peptide of proteolipid protein and the antimicrobial peptide melittin. *J. Phys. Chem. B*, **119**, 14821–14830.
25. Neidigh,J.W., Fesinmeyer,R.M., Prickett,K.S. and Andersen,N.H. (2001) Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. *Biochemistry*, **40**, 13188–13200.
26. Gustavsson,M., Traaseth,N.J. and Veglia,G. (2012) Probing ground and excited states of phospholamban in model and native lipid membranes by magic angle spinning NMR spectroscopy. *Biochim. Biophys. Acta*, **1818**, 146–153.
27. Fakhree,M.A.A., Blum,C. and Claessens,M.M.A.E. (2019) Shaping membranes with disordered proteins. *Arch. Biochem. Biophys.*, **677**, 108163.

28. Zeno,W.F., Baul,U., Snead,W.T., DeGroot,A.C.M., Wang,L., Lafer,E.M., Thirumalai,D. and Stachowiak,J.C. (2018) Synergy between intrinsically disordered domains and structured proteins amplifies membrane curvature sensing. *Nat. Commun.*, **9**, 4152.
29. Aivazian,D. and Stern,L.J. (2000) Phosphorylation of T cell receptor zeta is regulated by a lipid dependent folding transition. *Nat. Struct. Biol.*, **7**, 1023–1026.
30. Geering,K. (2005) Function of FXYD proteins, regulators of Na, K-ATPase. *J. Bioenerg. Biomembr.*, **37**, 387–392.
31. MacLennan,D.H. and Kranias,E.G. (2003) Phospholamban: a crucial regulator of cardiac contractility. *Nat. Rev. Mol. Cell Biol.*, **4**, 566–577.
32. Akin,B.L., Hurley,T.D., Chen,Z. and Jones,L.R. (2013) The structural basis for phospholamban inhibition of the calcium pump in sarcoplasmic reticulum. *J. Biol. Chem.*, **288**, 30181–30191.
33. Ablorh,N.-A.D. and Thomas,D.D. (2015) Phospholamban phosphorylation, mutation, and structural dynamics: a biophysical approach to understanding and treating cardiomyopathy. *Biophys. Rev.*, **7**, 63–76.
34. Hatos,A., Hajdu-Soltész,B., Monzon,A.M., Palopoli,N., Álvarez,L., Aykac-Fas,B., Bassot,C., Benítez,G.I., Bevilacqua,M., Chasapi,A. *et al.* (2020) DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Res.*, **48**, D269–D276.
35. Piovesan,D., Tabaro,F., Paladin,L., Necci,M., Micetic,I., Camilloni,C., Davey,N., Dosztányi,Z., Mészáros,B., Monzon,A.M. *et al.* (2018) MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res.*, **46**, D471–D476.
36. Schad,E., Fichó,E., Pancsa,R., Simon,I., Dosztányi,Z. and Mészáros,B. (2018) DIBS: a repository of disordered binding sites mediating interactions with ordered proteins. *Bioinformatics*, **34**, 535–537.
37. Fichó,E., Reményi,I., Simon,I. and Mészáros,B. (2017) MFIB: a repository of protein complexes with mutual folding induced by binding. *Bioinformatics*, **33**, 3682–3684.
38. Miskei,M., Antal,C. and Fuxreiter,M. (2017) FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Res.*, **45**, D228–D235.
39. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
40. UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
41. Camilloni,C., De Simone,A., Vranken,W.F. and Vendruscolo,M. (2012) Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry*, **51**, 2224–2231.
42. Berjanskii,M.V. and Wishart,D.S. (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.*, **127**, 14970–14971.
43. El-Gebali,S., Mistry,J., Bateman,A., Eddy,S.R., Luciani,A., Potter,S.C., Qureshi,M., Richardson,L.J., Salazar,G.A., Smart,A. *et al.* (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.*, **47**, D427–D432.
44. Brooks,P.C., Montgomery,A.M., Rosenfeld,M., Reisfeld,R.A., Hu,T., Klier,G. and Cheresh,D.A. (1994) Integrin alpha v beta 3 antagonists promote tumor regression by inducing apoptosis of angiogenic blood vessels. *Cell*, **79**, 1157–1164.
45. Brooks,P.C., Strömblad,S., Klemke,R., Visscher,D., Sarkar,F.H. and Cheresh,D.A. (1995) Antiintegrin alpha v beta 3 blocks human breast cancer growth and angiogenesis in human skin. *J. Clin. Invest.*, **96**, 1815–1822.
46. Cao,Z., Suo,X., Chu,Y., Xu,Z., Bao,Y., Miao,C., Deng,W., Mao,K., Gao,J., Xu,Z. *et al.* (2018) Peptides derived from the integrin β cytoplasmic tails inhibit angiogenesis. *Cell Commun. Signal.*, **16**, 38.
47. Yang,J., Ma,Y.-Q., Page,R.C., Misra,S., Plow,E.F. and Qin,J. (2009) Structure of an integrin alphaIIb beta3 transmembrane-cytoplasmic heterocomplex provides insight into integrin activation. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 17729–17734.
48. Metcalf,D.G., Moore,D.T., Wu,Y., Kielec,J.M., Molnar,K., Valentine,K.G., Wand,A.J., Bennett,J.S. and DeGrado,W.F. (2010) NMR analysis of the alphaIIb beta3 cytoplasmic interaction suggests a mechanism for integrin regulation. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 22481–22486.
49. Deshmukh,L., Meller,N., Alder,N., Byzova,T. and Vinogradova,O. (2011) Tyrosine phosphorylation as a conformational switch: a case study of integrin β3 cytoplasmic tail. *J. Biol. Chem.*, **286**, 40943–40953.
50. Kumar,M., Gouw,M., Michael,S., Sámano-Sánchez,H., Pancsa,R., Glavina,J., Diakogianni,A., Valverde,J.A., Bukirova,D., Čalyševa,J. *et al.* (2020) ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res.*, **48**, D296–D306.
51. Hornbeck,P.V., Zhang,B., Murray,B., Kornhauser,J.M., Latham,V. and Skrzypek,E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*, **43**, D512–D520.
52. Wimley,W.C. and White,S.H. (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.*, **3**, 842–848.
53. Mészáros,B., Erdos,G. and Dosztányi,Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, **46**, W329–W337.
54. Sehnal,D., Deshpande,M., Vařeková,R.S., Mir,S., Berka,K., Midlik,A., Pravda,L., Velankar,S. and Koča,J. (2017) LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nat. Methods*, **14**, 1121–1122.
55. Hamosh,A., Scott,A.F., Amberger,J.S., Bocchini,C.A. and McKusick,V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
56. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
57. Wishart,D.S., Feunang,Y.D., Guo,A.C., Lo,E.J., Marcu,A., Grant,J.R., Sajed,T., Johnson,D., Li,C., Sayeeda,Z. *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
58. Orchard,S., Ammari,M., Aranda,B., Breuza,L., Briganti,L., Broackes-Carter,F., Campbell,N.H., Chavali,G., Chen,C., del-Toro,N. *et al.* (2014) The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
59. Szklarczyk,D., Gable,A.L., Lyon,D., Junge,A., Wyder,S., Huerta-Cepas,J., Simonovic,M., Doncheva,N.T., Morris,J.H., Bork,P. *et al.* (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, **47**, D607–D613.