Nanomechanics combined with HDX reveal allosteric drug binding sites of CFTR NBD1

Rita Padányi[1,#], Bianka Farkas[1,2,#], Hedvig Tordai[1], Bálint Kiss[1], Helmut Grubmüller[3], Naoto Soya[4], Gergely Lukács[4], Miklós Kellermayer[1], Tamás Hegedűs[1,5,*]

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

[2] *Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, HU,*

[3] *Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany*

[4] *Department of Physiology and Biochemistry, McGill University, Montréal, Quebec, Canada*

[5] *TKI, Eötvös Loránd Research Network, Budapest, HU*

[#] equally contributed

*Correspondence: Tamás Hegedűs
email: hegedus@hegelab.org

Abstract

Cystic fibrosis is most frequently caused by the deletion of F508 (ΔF508) in CFTR's nucleotide binding domain 1 (NBD1), thereby compromising CFTR folding, stability and domain assembly. Limitation to develop a successful therapy has been attributed to the lack of molecules that synergistically facilitate folding by targeting distinct structural defects of ΔF508-CFTR. To improve drug efficacy by targeting the ΔF508-NBD1 folding and stability, and to study potential ΔF508-NBD1 allosteric corrector binding sites at the atomic level, we combined molecular dynamics (MD) simulations, atomic force spectroscopy (AFM) and hydrogen-deuterium exchange (HDX) experiments to elucidate the mechanical and thermal stabilization mechanisms of ΔF508-NBD1 by 5-bromoindole-3-acetic acid (BIA). MD and AFM allowed us to describe unfolding intermediates and forces acting during NBD1 mechanical unfolding. Application of the low-potency corrector BIA increased the mechanical resistance of the ΔF508-NBD1 α-subdomain, which was confirmed as a binding site by computational modeling and HDX experiments. Our results underline the complementarity of MD and AFM despite their different pulling speeds and provide a possible strategy to improve folding correctors.

Introduction

Cystic fibrosis is a lethal disease caused by the functional defect of CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) chloride channel in the apical membrane of epithelial cells[1,2]. Biochemical, *in silico*, and structural studies of CFTR have contributed to understanding the effect of mutations ranging from misfolding to impaired regulation and channel gating[2–4]. CFTR is a member of the ABC (ATP Binding Cassette) protein superfamily that provides an ion conductance pathway through the cell membrane via two transmembrane domains (TMDs) each consisting of six TM helices[5–7]. CFTR possesses two nucleotide-binding domains (NBDs), which bind ATP, form a "dimer" via transient interactions, and regulate channel gating. An NBD is composed of a β-subdomain that binds ATP, and an α-subdomain that contains the ABC signature motif. The subdomains are not formed by sequential sequence regions but are intertwined (Figure 1a). The binding and hydrolysis events are communicated towards the TMDs by the so-called coupling helices, which are the regions of intracellular "loops" that interact with the NBDs[5]. Most of these structural features have recently been confirmed by cryo-electron microscopy[6–9]. Mutations cover every region of the protein, and many of them are located in the N-terminal nucleotide binding domain, NBD1[4].

Sequence alterations in CFTR likely affect protein function, folding or stability. The effect of a mutation, such as the most frequent deletion of F508 residue (ΔF508 or F508del) located in NBD1, is usually complex and alters many of the above processes[4]. This deletion impairs the global domain-domain assembly of CFTR and affects the interactions between transmembrane domains, NBD1/TMDs, and NBD1/NBD2[10–12]. A couple of specific amino acid interactions have been revealed to play an important role in CFTR domain assembly.

We have shown earlier that the hydrophobic side chain of F508 is crucial for the NBD1/CL4 interface, and the hydrophobic pocket of NBD1 around this amino acid could be a primary drug target for corrector development[5]. Furthermore, the NBD1/TMD assembly problem affecting global CFTR folding originates from local effects of ΔF508 in NBD1, and in the presence of this deletion NBD1 has shown decreased thermal stability[13–15].

Detailed studies of isolated wild type and ΔF508NBD1 demonstrated a substantially decreased melting temperature of the mutant domain[13,16] while suggesting a similar folding pathway compared to the wild type[17]. It was also proposed that a partially unfolded state is responsible for the aggregation propensity of NBD1. As primary objective in drug development the restoration of the ΔF508-CFTR folding was advocated by either exclusive correction of the NBD1 stability[18] or targeting both the NBD1 and NBD1-CL4 interface instability[19,20]. Both methods are likely effective only on an already folded NBD1 subpopulation. Notably, the only approved drug that can target the isolated NBD1 was suggested to stabilize an (un)folding intermediate and not the native fold[21].

Early in CFTR research, altered folding of ΔF508-NBD1 has been demonstrated, and it was shown that the deletion affects a folding step prior to the ATP binding site formation[22,23]. Studies on the co-translational NBD1 folding revealed that the synonymous codon for I507 upon F508 deletion results in lowering the speed of translation and has an effect on CFTR conformation[24,25]. The folding of the nascent NBD1 polypeptide chain emerging from the ribosome was also extensively investigated using FRET[17,26,27].

In contrast to the previous results, these studies indicated that NBD1 folding starts on the ribosome and suggested that the ΔF508 defect occurs late in the NBD1 folding pathway.

An effective correction of any CF mutation must correct all the impaired steps of CFTR folding and/or its function. The first success in CF drug development is represented by Ivacaftor (Kalydeco, VX-770) which targets the natively folded G551D CFTR gating mutant[28]. The cryo-EM structure of the drug-bound WT CFTR complex provided the first hint for Ivacaftor's mechanism of action at atomic resolution[29], launching the design of novel scaffolds for more efficient drugs. This would not only be important for G551D targeting, but also for other mutants with impaired function since the efficacy of Ivacaftor depends on the specific mutation type in the patient (https://pi.vrtx.com/files/uspi_ivacaftor.pdf). Lumacaftor (VX-809) has also emerged as a potential corrector of the ΔF508-CFTR folding defect[30,31]. Although VX-809 has a low efficacy in restoring the ΔF508-CFTR folding efficiency[19], serious efforts have been devoted to identify its binding site[32]. Experiments suggested that VX-809 exerts its action on the TMD1 and TMD1/NBD1 interaction and also binds to the cleft formed by the C-terminal helices and β-strands S3, S8, and S9[32–34].

The recently approved treatment for F508del, Trikafta, is a combination of VX-661 (Tezacaftor), VX-445 (Elexacaftor), and the gating potentiator VX-770 (Ivacaftor)[31]. VX-661 is a known Type I corrector that restores the NBD1/TMDs interface. VX-445 was shown to be a Type III corrector that acts on NBD1[21,35]. Long-term studies of Ivacaftor indicated that adult G551D patients experience bacterial infections and progressive loss of lung function in spite of an initial partial normalization of the lung function[36,37]. As the Trikafta treatment also exhibits only partial restoration of lung function, a long-term decline in effectivity of this drug combination can also be expected[31].

Since the lack of a type III corrector, which promotes NBD1 folding, was likely the bottleneck in developing an effective combination therapy such as Trikafta[31], improving existing and developing novel, more efficient folding correctors would be critical for long-term CF therapy. Therefore, the major objective of our study was to gain insights on NBD1 folding and its folding intermediates at a higher resolution compared to earlier studies. To identify unfolding pathways and intermediates, we performed both force-probe molecular dynamics simulations and force spectroscopy experiments on the wild type and ΔF508-NBD1. Since BIA (5-bromoindole-3-acetic acid) has been suggested as a type III corrector, we also investigated its effect on unfolding and determined its binding site using hydrogen-deuterium exchange (HDX).

Results

ΔF508 and WT S6-α-S8 core unfolding exhibit differences in MD simulation. Since pulling a large system such as the 250 a.a. long CFTR NBD1 in regular steered molecular dynamics simulations (MD) is highly limited, first we investigated the unfolding of WT and ΔF508 CFTR NBD1 using an all-atom Gō model (SMOG[38]). To identify the unfolding steps of NBD1, we analyzed the hierarchy in the detachment of secondary structural elements (SSE), which was calculated by the fraction of native contacts (Q) as a function of pulling time (Figure S1-S2). Although these pathways helped in the interpretation of AFM experiments

(see below), WT and mutant domains did not exhibit significant unfolding differences in the Gō simulations. Because of this observation and experiments, which have indicated the S6-α-S8 core (a.a. 487-603, Figure 1) to be crucial for NBD1 folding[17,38], we studied the unfolding of this core region in fully solvated atomistic force field pulling MD simulations, gaining higher resolution and accuracy of the late unfolding pathway.

In the MD simulations of the S6-α-S8 core we analyzed the timing of SSE detachments and the rupture forces at which these detachments occur. The order of SSE detachments was determined by monitoring the fraction of the native contacts as was done for the Gō simulations (Figure S1). Detachment sequences of the secondary structure units were determined in each simulation, and the frequencies of the resolved pathways were calculated (Figure 2a-c). In order to simplify the comparison between the WT and mutant pathways, we separated the S6-α-S8 core unfolding into two stages. In the first stage (Figure 2b), the β-strands S6, S7 and S8 and the helix H6 unfold. Most frequently S8 decoupled in the first step, followed by H6 unfolding and the concurrent detachment of S6 and S7. The frequency of this pathway is much lower in ΔF508 S6-α-S8 core than in WT (54% (27/50) versus 78% (39/50), $p < 0.05$, χ^2 test). The main divergence between the WT and ΔF508 protein was the different timing of the H6 detachment. An increased frequency of the pathways, in which the detachment of H6 occurs after the decoupling of S6 and S7, can be observed in ΔF508 S6-α-S8 core (ΔF508: 40%, WT: 22%, $p=0.0517$, χ^2 test). This could be caused either by weaker binding of S6 or by stronger binding of H6 to the folded part of the core when compared to the WT. Comparing the S6 rupture forces of WT and mutant shows that in the case of ΔF508 a greater proportion of S6 detached at lower forces, while the distribution of H6 rupture forces were unchanged (Figure 2d), indicating that the mutation weakened the S6 interactions. At the second stage of the S6-α-S8 core unfolding, the α-helices H5, H4, H4b and H3 detached (Figure 2c). H5 and H3 in the WT core showed a synchronized unfolding slightly more frequently (20%) than in the ΔF508 core (6%) ($p=0.074$, χ^2 test). Comparing the time points of the SSE detachments of the WT and the mutant core, we observed faster detachment of secondary structural units (S6, H4 and H5) in the mutant compared to WT (Figure S3).

To assess the effect of speed on the unfolding forces, we performed pulling simulations also at a lower speed, which resulted in smaller forces and loading rates (Figures S5 and S6). Nevertheless, the main pathways remained the same with somewhat shifted ratios; and some SSE detached earlier in the mutant than in the wild type (Figures S4 and S7).

An increased number of non-native contacts are characteristic for late WT NBD1 intermediates. To characterize the details of the unfolding steps, intermediate structures were determined and analyzed by using a special contact-based metric[39], which is efficient to compare highly different structures developing along pulling trajectories. Accordingly, the conformations from each pulling simulation were clustered separately based on contact RMSD. Then, the centroids of these clusters were pooled as intermediates, which were clustered again to yield all observable intermediates from every simulation. We identified four intermediate clusters describing the structural changes during unfolding in the case of both constructs (Figure 3). However, there is a well-defined wild type cluster (cluster #3), which was not observed for the ΔF508

S6-α-S8 core. Instead, the mutant conformations in the corresponding period (18-25 ns) changed their structure continuously into the totally unfolded state and were clustered into the last unfolding group.

To analyze the intramolecular interactions contributing to the formation of the well-defined intermediate structure observed in the WT cluster #3, the native and non-native contacts were calculated during unfolding. As expected, the number of native contacts decreased monotonically during unfolding (Figure S8) and appeared very similar for both constructs. In contrast, the number of non-native contacts increased at the beginning of pulling simulations (Figures 4a and S8), probably reflecting the equilibration of the structure under pulling conditions, then exhibited a decreasing trend. Interestingly, a secondary increase was observed in conformations between 18 and 25 ns and it was more pronounced in the WT than in the mutant, coinciding with an unfolding intermediate state detected only in the WT (cluster #3).

In the next step, we analyzed which residues were involved in the non-native contacts observed in these conformations between 18 and 25 ns (Figures S9 and S10). Per residue proximity values were exploited to determine those residues which participated in non-native contact formation (Figures 4b, S10 and S11). Residues exhibiting pronounced differences between the WT and mutant proteins were located around F508 and in the H3-H4 and H4-H5 loops. Most of the residues in non-native contacts were hydrophobic and located between helices H3, H4 and H5 (Figure 4c). The above studied non-native intermediate was the last one during unfolding, thus it is likely formed as the first intermediate in the reverse, folding process. This suggests that F508 may play an important role in the interaction network of a non-native intermediate during the early folding that was also observed in our folding simulations (Figure S12). Clustering and contact analysis were also performed on simulations using 0.1 m/s pulling speed (Figures S8, S9 and S10) and indicated some differences from simulations at higher pulling speed. We discussed this in the supplementary materials.

AFM experiments revealed distinct unfolding steps of NBD1 and decreased proportion of natively folded S6-α-S8 core in the case of ΔF508 mutant. Our MD simulations provided a set of unfolding pathways of natively folded NBD1 that supported the interpretation of the unfolding pattern observed with AFM for *in vitro* expressed NBD1. Cysteine-less NBD1 domain, carrying a deletion of the regulatory insertion (Cys-less NBD1 ΔRI) was used, since ΔRI improves the stability and solubility of the protein that was especially important for the ΔF508 mutant[15,16,40]. For simplicity, we will subsequently refer to this construct as NBD1. The NBD1 N-terminus was tagged with SUMO (small ubiquitin-like modifier) protein (Figure 5a), which has been reported to enhance expression and possesses a well-characterized unfolding pattern[41]. This SUMO fingerprint provided a specific force-spectroscopic signature for successful pulling events represented by typical sawtooth profiles with peaks indicating rupture events, and by a total length of ~110-130 nm and a terminal contour length increment of ~25 nm (Figure 5b).

In order to quantitatively compare the pulling experiments of WT and ΔF508-NBD1 constructs, contour length increments (ΔL), detected before the unfolding of SUMO, were collected from force-extension curves. For WT, the ΔL histogram exhibited well-defined peaks at 12 nm, 30 nm, and 42 nm corresponding to the last steps of unfolding (Figure 5d). Our simulations and the NBD1 structure indicated

that these steps derived from the breaking of the S6-α-S8 core (groups gS8 and gS6 in Figures S1 and S2) and corresponded to the detachment of gS8 and gS6 separately or together, involving 32 and 83 amino acid long regions (32 x 0.35 nm ~ 12 nm and 83 x 0.35nm ~ 30 nm, respectively) (Figure 5c). For natively folded NBD1, the unfolding of these segments gives rise to three different unfolding patterns including the sequential unfolding of gS8 and gS6, unfolding of gS6 followed by gS8, and simultaneous unfolding of gS8 and gS6. Two peaks at 55 and 70 nm were also observed. Albeit these data were noisy, since they include tip positions close to the mica surface, with adverse electrostatic interactions, the changes in contour length correspond quite well to specific breaking events of NBD1. The unfolding of a 152 a.a. segment (55 nm) was initiated by breaking of group gS3, and 74 or 72 nm long sections were potentially manifested upon breaking of gS10 or from gS9, respectively, indicated by our pulling simulations. In contrast, the histogram built from ΔF508 NBD1 data did not contain the well-defined peaks detected in the WT NBD1 and exhibited a wider distribution of length changes. Additionally, in a large number of pulling experiments with the ΔF508 mutant, unexpected contour length increments were observed as peaks around 20 and 35 nm (Figure 5d), suggesting that a significant number of ΔF508-NBD1 exhibit modified mechanical resistance or incorrect folding.

In order to correlate MD simulations with experiments at a more basic and fundamental level, we calculated contour length increments from force-extension curves of simulations using the WLC model as in experiments and found that the ΔL *in silico* peaks of β-strand unfolding overlap with experimental data. Unfolding of α-helices as force peaks can be observed in the simulations, albeit they did not emerge in experiments likely because of the buffered unfolding of α-helices in *in vitro* experiments[42] (see below, Supplementary Text, and Figures 5). To further compare the effect of ΔF508 on S6-α-S8 core folding, we counted every S6-α-S8 core unfolding pattern (Table S1). While 56% of the curves show the unfolding signatures of groups g8 and g6 (12 and 30 nm ΔL in any combination) in the case of WT NBD1, this value is reduced to 28% in the case of ΔF508 mutant. This observation suggested that ΔF508 decreased the proportion of the correctly folded S6-α-S8 core either by affecting the folding in bacteria or by destabilizing it during the experiment.

The corrector molecule BIA acts on the α-helical subdomain. A small compound, BIA (5-bromoindole-3-acetic acid) has been demonstrated to promote ΔF508-CFTR maturation and modestly stabilize the ΔF508-NBD1 against thermal unfolding[43]. Therefore, we investigated the effect of BIA on the unfolding of the mutant NBD1 in AFM experiments. We compared the contour length increments of ΔF508-NBD1 in the presence and absence of BIA and found the appearance of a pronounced peak around 24 nm in the presence of this compound (Figures 6a and S13). The typical WT peaks of 42 and 30 nm were not observed in the presence of BIA, suggesting that this molecule did not restore the native, WT-like conformation or mechanical properties of NBD1, but it bound to and stabilized an NBD1 region, resulting in a peak around 26 nm. Since this peak is close to the H5-H4-H3 peak observed in simulations we propose that BIA binds to this part of NBD1. This observation also suggests that BIA binding provides improved stabilization of this

region than α-helices would alone in the native structure, since with our AFM setup the helix unfolding cannot be detected as happened in the case of WT (Figure 5d).

Two peaks around 70 nm and 55 nm were also detected in the presence of BIA, which likely correspond to the β-subdomain and were observed in WT but not in ΔF508, pointing at some differences in simulations with higher pulling speed. Thus, we aimed to narrow the potential BIA binding sites using *in silico* methods. FTMap[44] identified several potential drug binding sites in NBD1 (Figure 6b), but sites at NBD1/CL4 and NBD1/NBD2 interface were excluded, since binding to those locations would interfere with CFTR assembly and maturation. Two out of three sites in the α-subdomain were strengthened by Autodock Vina[44] docking (Figures 6c and S14). One includes ends of H4, H5, H6 and loops, and another involves H6, the loop between S2 and S3, and H7. To determine the BIA binding site more accurately, the deuteration kinetics of the backbone NH in ΔF508-NBD1 was determined by the HDX-MS technique in the absence and presence of BIA at 37 °C. In the presence of BIA, the HDX was inhibited at four α-subdomain peptides, including parts of H4 (a.a. 512-525), H5 (a.a. 559-568), H6 (a.a. 591-594), and at a group with several secondary structure elements (end of H6, S8, H7, S9 and S10; a.a. 595-624) (Figure 6d). Since BIA decreased HDX of S2 and S3, which are in the vicinity of the second group, only to a smaller extent, the BIA binding site is likely confined at the ends of H4, H5, H6 and their loops. The strongly decreased dynamics of S8, H7, S9, and S10 could be caused by allosteric stabilization (Figure S16).

Discussion

In this study, we explored the mechanical unfolding of CFTR NBD1 domain using both atomistic molecular dynamics (MD) simulations and single molecule force spectroscopy (AFM) experiments. We found that deletion of F508 has a significant effect on NBD1 unfolding pathways and accelerated secondary structure element detachment. Our results suggest that these effects were likely exerted by hindering the formation of crucial non-native intermediate states in the late stages of unfolding, thus potentially in the early stage of folding (Figures 3, 4 and S8-S11).

Although the first step of CFTR biogenesis, which is affected by ΔF508, is the folding of NBD1, there are only a handful of studies investigating this process. Qu *et al.* demonstrated that ΔF508 increased the probability of off-pathway intermediates and affected an early folding step before the formation of the ATP binding site[22,23]. They used isolated NBD1 and measured its folding yield by light scattering and intrinsic Trp fluorescence, showing that F508 affects the rate of maturation and suggested that F508 makes crucial contacts during the folding process. Interestingly, it has also been demonstrated that deletion of the three nucleotides resulting in ΔF508 mutation causes alteration in mRNA structure, leading to a decreased rate of translation[24].

The authors also described that the lower translation speed resulted in altered CFTR conformations in metabolic pulse chase experiments[25]. A set of other experiments targeted NBD1 folding on ribosomes using truncation constructs and FRET[17,26,27]. These studies revealed that the folding of the N-terminal part involving β-strands S1-S6 took place while F508 was in the ribosome tunnel and the deletion affected a later

stage of domain folding[17]. The authors found that the ribosome delayed the α-subdomain folding which was essential in the correct insertion of S7/S8 strands into the β-sheet core[26]. These results seem to be contradictory to that of Qu *et al.*[22,23]. However, the N-terminal β-subdomain may fold rapidly and independently according to FRET experiments by Kim *et al.*[26], but the state of this subdomain is likely not sufficiently mature to provide an environment for forming a correct ATP binding site[22,23]. Most likely the different levels or timescales of experiments provided data on different aspects of the folding process.

While the above studies on NBD1 folding were fundamental and agreed in the involvement of altered intermediate states, their resolution was low. To overcome this, and because of successful application of simplified models in pulling simulations[45,46], first we used a native structure based Gō model, which overcomes computational limitations associated with pulling simulations using proteins with the size of NBD1. Although this simplified model was indispensable for force-extension curve analysis and, despite its simplicity, agreed well with FRET experiments performed by Kim *et al.*[26] (Supplementary Text), its accuracy is limited by the lack of explicit water molecules and non-native contacts. Therefore, we also performed regular, fully solvated atomistic force field pulling simulations with a smaller part of NBD1, the S6-α-S8 core. We detected altered pathway frequencies and faster detachment of certain secondary structure elements in the mutant core, suggesting differences in the interaction network around the β-strand S6 and in the final folded unit containing α-helices H4, H5 and H3. Importantly, our results showed that F508 remained a component of the folded part of the wild type NBD1 almost until the end of unfolding. The WT core exhibited a higher number of non-native interactions at a late stage of unfolding (around 18-25 ns) compared to ΔF508, suggesting that non-native interactions contribute to the stability of the late unfolding intermediate detected in the wild type core. These interactions also included positions with known CF-causing mutations (ΔI507, V520F, L558S and A559T), which were shown to cause α-core compaction by Shishido *et al.*[27].

Taken together, F508 supports the development and persistence of non-native interactions that may be an important factor for off-pathway avoidance and self-chaperoning. The non-native contacts, which have been described to influence the folding free-energy barrier[47] and can become the rate-limiting step of protein folding[48], likely serve as a deceleration mechanism to provide time for the NBD1 polypeptide to acquire the right intermediate state before engaging the next step of folding. This was also supported by the *in vitro* translation experiments of Kim *et al.*, showing that faster codons inhibited folding[26]. The same residues participating in the non-native contact formation during the unfolding (Figure 4 and S10) also appeared to be in contact in our folding simulations (Figure S12) verifying that the residue F508 and its surroundings may serve as a folding nucleus. Earlier, a decreased folding time was observed for the ΔF508-NBD1 in Gō folding simulations, suggesting that the NBD1 self-chaperoning was diminished[49].

Importantly, combining experiments and computer simulations, we identified drug binding sites that are located on the NBD1 surface distant from the protein axis and exposed to the solvent. Therefore, drug binding to these regions is unlikely to interfere with CFTR domain-domain assembly and maturation (Figure 6b and 6c). In addition, drugs acting on these regions may rescue not only the volatile folding and stability of the α-subdomain, but potentially stabilize the β-subdomain allosterically, experimentally confirmed by BIA binding (Figure 6a, d). We also demonstrated by computational methods that secondary site mutations

either in the β-subdomain or in the α-subdomain restored the WT-like allosteric network in the absence of F508 (Figure S16). Because of this allosteric subdomain coupling we propose that a drug, rationally designed for α-subdomain binding, will not only correct ΔF508 and other mutations in the α-subdomain, but it also has the potential to rescue CF mutations localized in the β-subdomain. This symmetricity of allosteric communication for promoting folding and stability can be observed not only between subdomains of CFTR NBD1, but also between domains of CFTR and of other ABC proteins (manuscript under preparation).

In summary, we found that during unfolding the detachments of certain SSE were accelerated in the ΔF508-NBD1 compared to the wild type, the main pathways were similar if unfolding started from a folded NBD1. However, in a significant portion of the wild type NBD1 the S6-α-S8 core is folded incorrectly, and its misfolding is enhanced by the F508 deletion. We conclude that the α-subdomain exhibits an inherent property for folding instability and propose that various NBD1 mutations may be corrected by small molecules binding to the α-subdomain, stabilizing the full domain in an allosteric manner. Furthermore, our results confirm that F508 maintains a network of non-native contacts and suggest a role in slowing down the translation and self-chaperoning.

Methods

Structural models. Wild type NBD1 structure based on an X-ray structure (PDBID: 2BBO) from an earlier study[50] was used as the starting point. In order to match the construct used in our experiments, regulatory insertion (a.a. 405-435) was removed and the gap was sealed by loop modeling of Modeller[51], setting residues 403, 404, 433, and 434 as a loop region. ΔF508 mutation was modeled similarly. The missense mutations were generated using the mutagenesis tool of VMD[52].

Molecular dynamics simulations. Conventional all-atom MD simulations were performed with the WT and ΔF508 S6-α-S8 core region of the NBD1 structure. The S6-α-S8 core (a.a. 487-604) consists of three β-sheets (S8, S7, S6) and five α-helices (H3, H4, H4b, H5, H6), including F508. MD simulations were run using GROMACS 2019[53] with the CHARMM36m[54] force field and the TIP3P water model. A 150 mM KCl concentration was used. Hydrogen atoms were replaced with virtual interaction sites to speed up the calculations, such that a 4 fs time step could be used[55]. Electrostatic interactions were calculated using the fast smooth PME algorithm[56]; the LINCS algorithm[57] was used to constrain bond lengths.

All structures were energy minimized using the steepest descent integrator in the first step, then equilibration (NVT, NPT) procedure was performed prior to each pulling simulation to generate inputs for independent simulations with different starting velocities (T=310 K, p=1 bar). The Nose–Hoover thermostat and the Parrinello-Rahman barostat with isotropic coupling were employed for the production runs. Time constants for the thermostat and the barostat were set to 2 picoseconds and 5 picoseconds, respectively. The C-terminus of the S6-α-S8 core structure was restrained, and the N-terminus was subjected to constant velocity pulling. Two pulling velocities, 1 m/s and 0.1 m/s were used, requiring 40 ns and 400 ns long

trajectories, respectively. 50 simulations were performed with both constructs and for each of the two pulling speeds. MD parameter files can be downloaded from http://resources.hegelab.org.

Protein expression and purification. We used a cysteine-less[58], His6-tagged SUMO-fusion NBD1 carrying a deletion of regulatory insertion (ΔRI), which improves the protein stability and solubility[59]. Cysteine-less construct was used to avoid interfering with immobilization via terminally introduced cysteines. SUMO-NBD1 constructs were purified from *E. coli* Rosetta 2 (DE3) pLysS strain. His6-tagged proteins were purified using an Ni-NTA affinity column (Profinity IMAC Ni-Charged, BioRad).

Single-molecule force spectroscopy experiments and analysis. Freshly cleaved mica surface was functionalized with APTES. The terminal cysteine of NBD1 was crosslinked to the APTES-coated mica using Sulfo-SMCC. Force spectroscopy was carried out on a Cypher atomic force microscope (AFM) instrument (Asylum Research) using PNP-TR cantilevers (spring constant: 100-200 pN/nm, NanoWorld). Experiments were performed at 25 ºC in PBS buffer (pH 7.2)[60]. Unfolding of NBD1-SUMO was carried out by first attaching the protein to the tip non-specifically by applying a constant force of 1 nN for 1 s to the tip on the mica surface, then followed by constant speed retraction with a pulling velocity of (1 μm/s).

Data was fitted using the worm-like chain (WLC) model of polymer elasticity using Igor Pro (Wavemetrics) extended with the Asylum Research AFM driving software. Since AFM tip and protein adhesion occurred at random locations, most of the retraction curves did not show the unfolding of the full protein. Force-extension curves exhibiting an overall length compatible with a completely unfolded NBD1-SUMO protein (total length of ~110-120 nm) and including the SUMO unfolding fingerprint as a terminal contour length (Lc) increase of ~25 nm were selected. All calculated Lc values were collected and summarized in histograms.

Analysis of unfolding pathways. The analysis was completed using GROMACS tools[53], the MDAnalysis package[61] and in-house Python scripts. For identification of secondary structural units' detachment over the pulling trajectories, the native contacts (contacts in the initial folded structure) of every secondary structural unit were determined and the fraction of native contacts (Q) was calculated as a function of time over each pulling trajectory using the *soft_cut* metric from the MDAnalysis package[61,62]. We defined a detachment event if Q decreased below a threshold value, which was set 0.2 in the case of Gō simulations with NBD1 and 0.1 in the case of conventional MD simulations with the S6-α-S8 core region. The structures of the unfolding simulations were clustered to identify intermediate states during unfolding. We used a pairwise contact-based RMSD with a cutoff value of 0.8 nm as a distance metric for clustering as described by Mercadante *et al.*[39,63]. The pairwise residue distance matrix of all unfolding conformations along every pulling trajectory was calculated and all values above the 0.8 nm cutoff value were set to 0.8 nm. The pairwise contact-based root-mean-square deviation (RMSD) values were calculated from these distance matrices, as

$$RMSD(f_i, f_j) = \frac{1}{N}\sum_{i,j}(f_i - f_j)^2$$

, where N is the number of residues and f_i is the pairwise a.a. distance matrix of a given frame with position index i. Using this type of RMSD values for clustering led to ignoring the changes between distant residues that would have masked changes in important contacts within the remaining folded part. We applied a density-based clustering algorithm, DBSCAN for clustering[64]. The structures of each pulling simulation were clustered separately using the contact-based RMSD as a distance measure and the centroids of these clusters were pooled. The centroid structures from all simulations were clustered again to yield all observable intermediates from every simulation. Clusters with fewer members than 5% of the clustered structures were omitted from the re-clustering step. The DBSCAN clustering parameter, Eps, was set based on the Elbow method[64] and considering the Silhouette Score[65]. The cluster centroids were selected based on a calculated similarity score[66] as

$$S_{ij} = e^{-D_{ij}/D_{std}}$$

$$argmax_i = \sum_j S_{ij}$$

, where D: pairwise distances of the structures of the cluster (contact based RMSD values), S_{ij}: pairwise similarity of two structures, D_{ij} ($RMSD_{ij}$): pairwise distance of structures i and j from the conformational ensemble (cluster), D_{std}: standard deviation of D, argmax: the structure most similar to all other structures within the cluster (centroid).

<u>*Proximity calculations.*</u> The number of native and non-native contacts during each unfolding trajectory were calculated and normalized. Amino acid residues were in contact if the distance between any atom of the two residues is smaller than the 0.45 nm cutoff value. Contacts between residues that were present in 75% of the initial structure of the 50 simulations were labeled as native. Residues that were not native contacts but got closer than the cutoff in the course of the unfolding simulations were assigned as non-native contacts. Proximity values[67] were used to identify important non-native contacts where their cumulative number was increased during the simulations (18-25 ns and 180-250 ns range at pulling velocities of 1 m/s and 0.1 m/s, respectively). Proximity is minimal (zero) for the cutoff value or larger distances and maximal (one) if the distance between two residues is zero:

$$proximity(a_i, a_j) = \begin{cases} (cut - d_{ij})/cut & d_{ij} < cut \\ 0 & d_{ij} \geq cut \end{cases}$$

, where cut: distance cutoff between amino acids (0.45 nm), a_i: amino acid residue in position i, d_{ij}: distance between amino acids in position i and j. Proximity values for amino acid pairs in each structure were summed for all the WT and the ΔF508 simulations (separately for both pulling velocities) in the time interval of interest (Figure S9). The summed 2D proximity values for each amino acid residue were also calculated by summing the previously calculated proximity values of the contacts of the given residue (Figure 4, S10).

Calculation of rupture forces and contour length increments in simulations. Rupture forces were derived from the pulling trajectories. Peaks of the force curves were collected for each SSE around the detachment time point of the given SSE. Normalized frequency of rupture forces from all and from individual SSE detachment events and their Gaussian density estimates were calculated and visualized. Contour lengths and their increments (ΔL) were calculated using the simple polynomial worm-like chain (WLC) interpolation formula[68].

Identification of drug binding sites in NBD1. FTMap webserver[69] (http://ftmap.bu.edu/) was used with default options to identify potential binding pockets on the NBD1 surface. Docking of BIA was performed with Autodock Vina[44]. Default options were applied except exhaustiveness, which was increased to 128. The search space was defined as described and shown in Figure S14.

Visualization. Structures are visualized using PyMOL (The PyMOL Molecular Graphics System, Version 1.8.4 Schrödinger, LLC). Figures were generated by Matplotlib[70].

Hydrogen-deuterium exchange (HDX) experiments. The ΔF508-NBD1 of human CFTR was purified as described[13]. Deuteriation time course of the ΔF508-NBD1 was measured by HDX coupled with mass spectrometry (HDX-MS) technique[71]. The sample concentration was 5 μM in buffer containing: 10 mM HEPES, 150 mM NaCl, 1 mM ATP, 2 mM $MgCl_2$ and 1 mM TCEP at pH 7.5. The deuterium uptake was performed in D_2O-based buffer in the presence and absence of 2 mM BIA. For each deuteration time, NBD1 was mixed with 1:14 dilution ratio into D_2O-based buffer, resulting more than 90% D_2O contents, and incubated for 10 s, 40 s, 2 min, 4 min and 10 min. HDX reaction was quenched by adding chilled quenching buffer (300 mM glycine and 8 M urea at pH 2.4) with 1:2 ratio. Quenched solution was flash frozen in MeOH containing dry ice and stored at -80 °C until use. 10 μL of quenched sample was injected into the sample loop, followed by 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 flow rate of 180 μL/min, the peptides were loaded onto a C8 analytical column (1 mm i.d. × 50 mm length, Thermo Fisher Scientific) and separated with Agilent 1290 Infinity II UHPLC system. Separated peptides were detected by LTQ Orbitrap XL (Thermo Fisher Scientific) in positive-ion mode for m/z 200 – 2000 using electrospray ionization. For peptide identification, tandem MS (MS/MS) analysis in data-dependent acquisition mode with collision-induced dissociation was performed in separate measurements. All MS/MS data were analyzed in Proteome Discoverer 1.4 (Thermo Fisher Scientific). The deuteration were determined from triplicate measurements and the collected data were analyzed using HDExaminer 2.3 (Sierra Analytics). The relative deuterium uptake (%D) for each peptide was calculated by comparing the centroids of the isotope envelopes of the deuterated samples against the undeuterated controls. Deuterium uptake plots were generated using Prism 6 (Graphpad).

Acknowledgments

We thank M. Habibi and S. Plotkin (University of British Columbia, Canada) for their help in setting up simplified pulling simulations. We acknowledge the computational resources made available on the GPU cluster of the Governmental Information-Technology Development Agency (https://kifu.gov.hu), the Grubmüller laboratory at Max Planck Institute (https://www.mpibpc.mpg.de/grubmueller), and Wigner GPU Laboratory (http://gpu.wigner.mta.hu). This work was supported by funds to T. Hegedus from the Cystic Fibrosis Foundation (CFF): HEGEDU18I0, HEGEDU20I0; and from NRDIO/NKFIH: K127961; to G. Lukacs from CCF LUKACS20G0, CIHR, CFI and Canada Research Chair Program to G. Lukacs.

References

1. Riordan, J. R. *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**, 1066–1073 (1989).
2. Csanády, L., Vergani, P. & Gadsby, D. C. Structure, Gating, and Regulation Of The CFTR Anion Channel. *Physiol. Rev.* **99**, 707–738 (2019).
3. Saint-Criq, V. & Gray, M. A. Role of CFTR in epithelial physiology. *Cell Mol Life Sci* **74**, 93–115 (2017).
4. Veit, G. *et al.* From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. *Mol Biol Cell* **27**, 424–433 (2016).
5. Serohijos, A. W. R. *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. U.S.A.* **105**, 3256–3261 (2008).
6. 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).
7. Zhang, Z., Liu, F. & Chen, J. Conformational Changes of CFTR upon Phosphorylation and ATP Binding. *Cell* **170**, 483-491.e8 (2017).
8. Zhang, Z. & Chen, J. Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator. *Cell* **167**, 1586-1597.e9 (2016).
9. Fay, J. F. *et al.* Cryo-EM Visualization of an Active High Open Probability CFTR Anion Channel. *Biochemistry* **57**, 6234–6246 (2018).
10. 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).
11. Du, K. & Lukacs, G. L. Cooperative assembly and misfolding of CFTR domains in vivo. *Mol Biol Cell* **20**, 1903–1915 (2009).

12. Thibodeau, P. H., Brautigam, C. A., Machius, M. & Thomas, P. J. Side chain and backbone contributions of Phe508 to CFTR folding. *Nat Struct Mol Biol* **12**, 10–16 (2005).
13. 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).
14. Wang, W., Okeyo, G. O., Tao, B., Hong, J. S. & Kirk, K. L. Thermally unstable gating of the most common cystic fibrosis mutant channel (ΔF508): 'rescue' by suppressor mutations in nucleotide binding domain 1 and by constitutive mutations in the cytosolic loops. *J Biol Chem* **286**, 41937–41948 (2011).
15. Aleksandrov, A. A. *et al.* Regulatory insertion removal restores maturation, stability and function of DeltaF508 CFTR. *J Mol Biol* **401**, 194–210 (2010).
16. 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).
17. Khushoo, A., Yang, Z., Johnson, A. E. & Skach, W. R. Ligand-driven vectorial folding of ribosome-bound human CFTR NBD1. *Mol Cell* **41**, 682–692 (2011).
18. He, L. *et al.* Restoration of NBD1 thermal stability is necessary and sufficient to correct ΔF508 CFTR folding and assembly. *J Mol Biol* **427**, 106–120 (2015).
19. Okiyoneda, T. *et al.* Mechanism-based corrector combination restores ΔF508-CFTR folding and function. *Nat Chem Biol* **9**, 444–454 (2013).
20. Veit, G. *et al.* Structure-guided combination therapy to potently improve the function of mutant CFTRs. *Nat Med* **24**, 1732–1742 (2018).
21. Veit, G. *et al.* Allosteric folding correction of F508del and rare CFTR mutants by elexacaftor-tezacaftor-ivacaftor (Trikafta) combination. *JCI Insight* **5**, 139983 (2020).
22. Qu, B. H. & Thomas, P. J. Alteration of the cystic fibrosis transmembrane conductance regulator folding pathway. *J Biol Chem* **271**, 7261–7264 (1996).
23. Qu, B. H., Strickland, E. H. & Thomas, P. J. Localization and suppression of a kinetic defect in cystic fibrosis transmembrane conductance regulator folding. *J Biol Chem* **272**, 15739–15744 (1997).
24. Bartoszewski, R. A. *et al.* A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J Biol Chem* **285**, 28741–28748 (2010).
25. Bali, V. *et al.* A synonymous codon change alters the drug sensitivity of ΔF508 cystic fibrosis transmembrane conductance regulator. *FASEB J* **30**, 201–213 (2016).
26. Kim, S. J. *et al.* Protein folding. Translational tuning optimizes nascent protein folding in cells. *Science* **348**, 444–448 (2015).

27. Shishido, H., Yoon, J. S., Yang, Z. & Skach, W. R. CFTR trafficking mutations disrupt cotranslational protein folding by targeting biosynthetic intermediates. *Nat Commun* **11**, 4258 (2020).
28. 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).
29. Liu, F. *et al.* Structural identification of a hotspot on CFTR for potentiation. *Science* **364**, 1184–1188 (2019).
30. 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).
31. 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).
32. Hudson, R. P. *et al.* Direct Binding of the Corrector VX-809 to Human CFTR NBD1: Evidence of an Allosteric Coupling between the Binding Site and the NBD1:CL4 Interface. *Mol Pharmacol* **92**, 124–135 (2017).
33. Ren, H. Y. *et al.* VX-809 corrects folding defects in cystic fibrosis transmembrane conductance regulator protein through action on membrane-spanning domain 1. *Mol Biol Cell* **24**, 3016–3024 (2013).
34. Loo, T. W. & Clarke, D. M. Corrector VX-809 promotes interactions between cytoplasmic loop one and the first nucleotide-binding domain of CFTR. *Biochem Pharmacol* **136**, 24–31 (2017).
35. Keating, D. *et al.* VX-445-Tezacaftor-Ivacaftor in Patients with Cystic Fibrosis and One or Two Phe508del Alleles. *N Engl J Med* **379**, 1612–1620 (2018).
36. Kirwan, L. *et al.* Longitudinal Trends in Real-World Outcomes after Initiation of Ivacaftor. A Cohort Study from the Cystic Fibrosis Registry of Ireland. *Ann Am Thorac Soc* **16**, 209–216 (2019).
37. Keown, K. *et al.* Airway Inflammation and Host Responses in the Era of CFTR Modulators. *Int J Mol Sci* **21**, E6379 (2020).
38. Kim, S. J. & Skach, W. R. Mechanisms of CFTR Folding at the Endoplasmic Reticulum. *Front Pharmacol* **3**, 201 (2012).
39. Mercadante, D., Gräter, F. & Daday, C. CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps. *Biophys J* **114**, 1267–1273 (2018).

40. Hudson, R. P. *et al.* Conformational changes relevant to channel activity and folding within the first nucleotide binding domain of the cystic fibrosis transmembrane conductance regulator. *J Biol Chem* **287**, 28480–28494 (2012).
41. Kotamarthi, H. C., Sharma, R. & Koti Ainavarapu, S. R. Single-molecule studies on PolySUMO proteins reveal their mechanical flexibility. *Biophys J* **104**, 2273–2281 (2013).
42. Takahashi, H., Rico, F., Chipot, C. & Scheuring, S. α-Helix Unwinding as Force Buffer in Spectrins. *ACS Nano* **12**, 2719–2727 (2018).
43. He, L. *et al.* Restoration of NBD1 Thermal Stability Is Necessary and Sufficient to Correct ΔF508 CFTR Folding and Assembly. *Journal of Molecular Biology* **427**, 106–120 (2015).
44. 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).
45. Habibi, M., Rottler, J. & Plotkin, S. S. As Simple As Possible, but Not Simpler: Exploring the Fidelity of Coarse-Grained Protein Models for Simulated Force Spectroscopy. *PLoS Comput Biol* **12**, e1005211 (2016).
46. Habibi, M., Rottler, J. & Plotkin, S. S. The unfolding mechanism of monomeric mutant SOD1 by simulated force spectroscopy. *Biochim Biophys Acta Proteins Proteom* **1865**, 1631–1642 (2017).
47. Clementi, C. & Plotkin, S. S. The effects of nonnative interactions on protein folding rates: Theory and simulation. *Protein Sci.* **13**, 1750–1766 (2004).
48. Shao, Q. & Zhu, W. Nonnative contact effects in protein folding. *Phys. Chem. Chem. Phys.* **21**, 11924–11936 (2019).
49. Serohijos, A. W. R., Hegedus, T., Riordan, J. R. & Dokholyan, N. V. Diminished self-chaperoning activity of the DeltaF508 mutant of CFTR results in protein misfolding. *PLoS Comput Biol* **4**, e1000008 (2008).
50. Serohijos, A. W. R. *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 U S A* **105**, 3256–3261 (2008).
51. Fiser, A. & Sali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* **374**, 461–491 (2003).
52. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33–38, 27–28 (1996).
53. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).

54. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. *J Chem Theory Comput* **8**, 3257–3273 (2012).
55. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry* **20**, 786–798 (1999).
56. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089–10092 (1993).
57. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463–1472 (1997).
58. Cui, L. *et al.* The role of cystic fibrosis transmembrane conductance regulator phenylalanine 508 side chain in ion channel gating. *J Physiol* **572**, 347–358 (2006).
59. Aleksandrov, A. A. *et al.* Regulatory insertion removal restores maturation, stability and function of DeltaF508 CFTR. *J. Mol. Biol.* **401**, 194–210 (2010).
60. Pires, R. H., Saraiva, M. J., Damas, A. M. & Kellermayer, M. S. Z. Force spectroscopy reveals the presence of structurally modified dimers in transthyretin amyloid annular oligomers: Force spectroscopy of misfolded dimers in amyloid TTR oligomers. *J Mol Recognit* **30**, e2587 (2017).
61. 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).
62. Best, R. B., Hummer, G. & Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci U S A* **110**, 17874–17879 (2013).
63. Gromiha, M. M. & Selvaraj, S. Inter-residue interactions in protein folding and stability. *Prog Biophys Mol Biol* **86**, 235–277 (2004).
64. Ester, M., Kriegel, H.-P. & Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. 6.
65. Rousseeuw, P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. (1987) doi:10.1016/0377-0427(87)90125-7.
66. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J* **109**, 1528–1532 (2015).
67. Stasik, P. M. Euclidean proximity function in image processing. in *2016 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)* 254–258 (IEEE, 2016). doi:10.1109/SPA.2016.7763623.
68. Bustamante, C., Marko, J., Siggia, E. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599–1600 (1994).

69. Kozakov, D. *et al.* The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* **10**, 733–755 (2015).
70. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90–95 (2007).
71. Soya, N., Roldan, A. & Lukacs, G. L. Differential Scanning Fluorimetry and Hydrogen Deuterium Exchange Mass Spectrometry to Monitor the Conformational Dynamics of NBD1 in Cystic Fibrosis. *Methods Mol Biol* **1873**, 53–67 (2019).

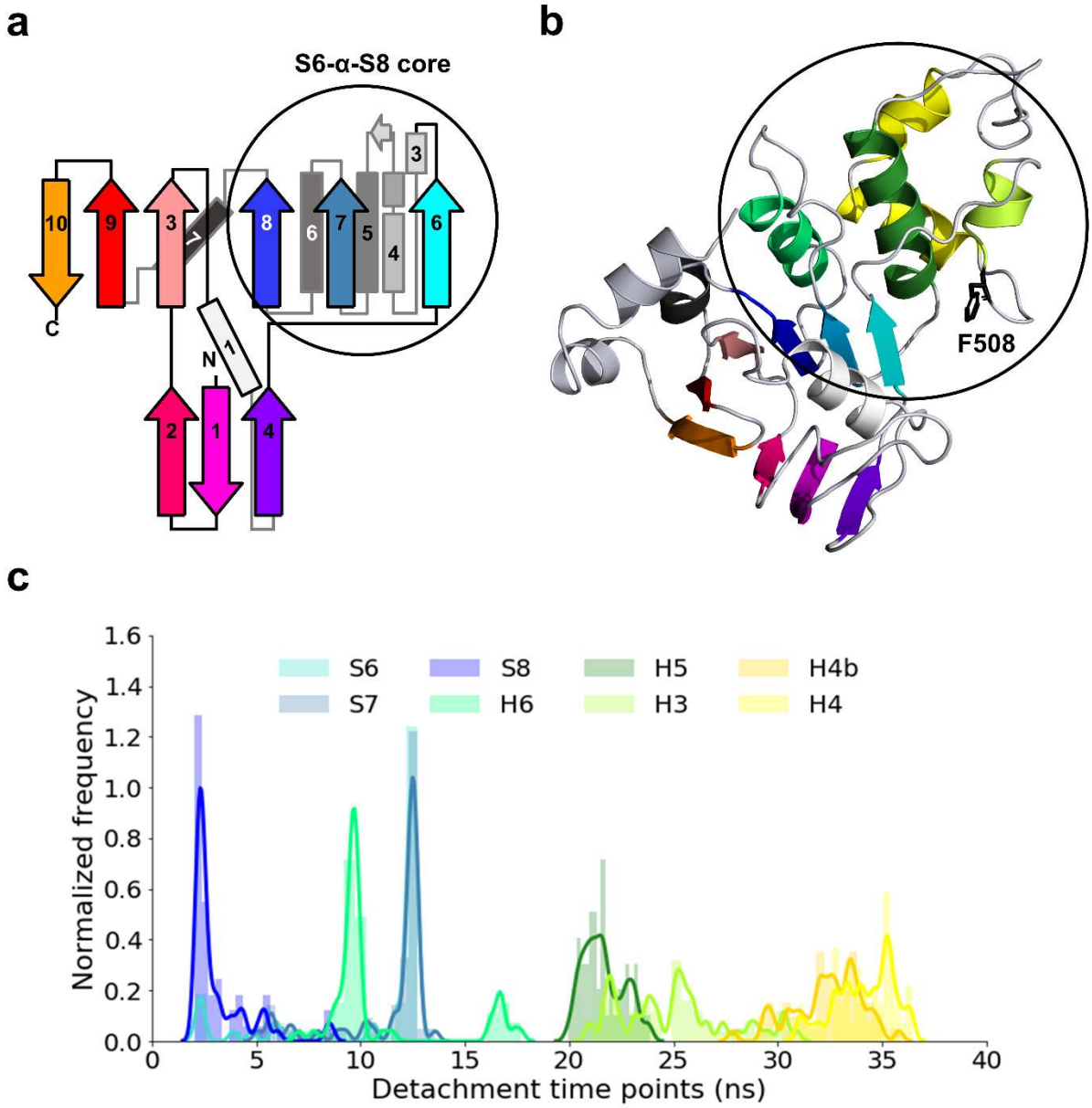

Fig. 1: Organization of CFTR NBD1 structure and secondary structural element detachment. Topology (**a**) and structure (**b**) of CFTR NBD1. β-strands are colored, and α-helices are marked in shades of grey. Dashed black circle labels the S6-α-S8 core region. Green and yellow helices: α-subdomain. (**c**) Distribution of the detachment time points of each secondary structural unit calculated from all MD trajectories with the wild type S6-α-S8 core (pulling velocity was 1 m/s).

a

#1				#2			WT	ΔF508
S8	H6	S6-S7		H5	H3	H4	54%	44%
S8	H6	S6-S7		H3-H5		H4	20%	6%
S8	H6	S6-S7		H3	H5	H4	4%	4%
S8	H6	S6	S7	H5	H3	H4		6%
S8	S6-S7		H6	H5	H3	H4	10%	14%
S8	S6	S7	H6	H5	H3	H4		8%
S6	S8	S7	H6	H5	H3	H4	4%	6%
outliers							8%	12%

b

#1				WT		ΔF508	
S8	H6	S6-S7		78%	78%	54%	60%
S8	H6	S6	S7			6%	
S8	S6-S7		H6	10%	14%	14%	28%
S8	S6	S7	H6			8%	
S6	S8	S7	H6	4%		6%	
outliers					8%		12%

c

#2		WT	ΔF508
H5	H3	68%	78%
H3	H5	4%	4%
H3-H5		20%	6%
outliers		8%	12%

d

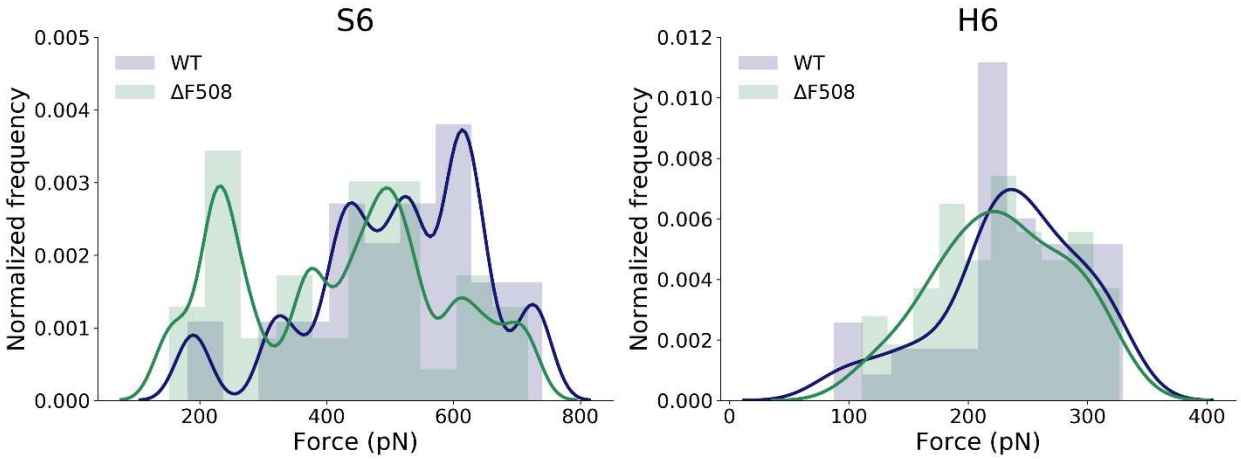

Fig. 2: Alternative unfolding pathways of S6-α-S8 core in pulling simulations. (**a**) Pathways were determined by the detachment sequence of secondary structure units. Synchronized unfolding of two elements is marked by hyphenation and enclosing them in one cell. H4 labels both the H4 and H4b helices since they always unfold at the same time. (**b**) Summary of the pathway frequencies of the first stage of the S6-α-S8 core unfolding. Helix H6 unfolds last in all outlier pathways. (**c**) Summary of the pathway frequencies of the second stage of the S6-α-S8 core. (**d**) Unfolding force distribution of secondary structural elements S6 and H6.

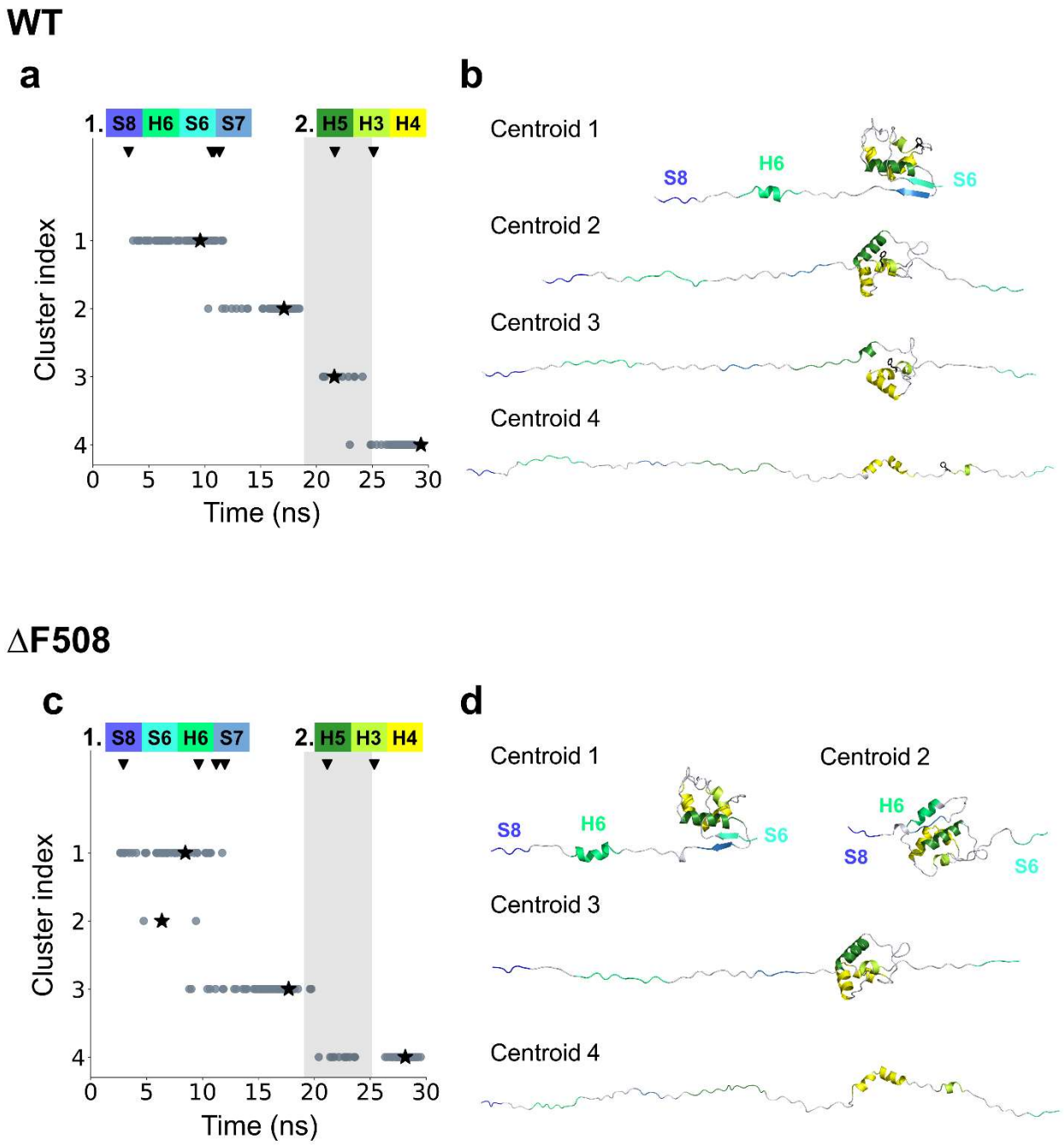

Fig. 3: Clusters of S6-α-S8 core intermediates during unfolding. Intermediate structures from WT (**a**) and F508del (**c**) pulling simulations were clustered using contact RMSD as a pairwise distance metric. Cluster centroids are indicated by stars and their structures are shown on the right (**b, d**). The grey area highlights the cluster with intermediate structures present in the wild type core but not in the mutant core. Arrowheads mark the averaged detachment time points of each secondary structural element calculated based on its fraction of native contacts.

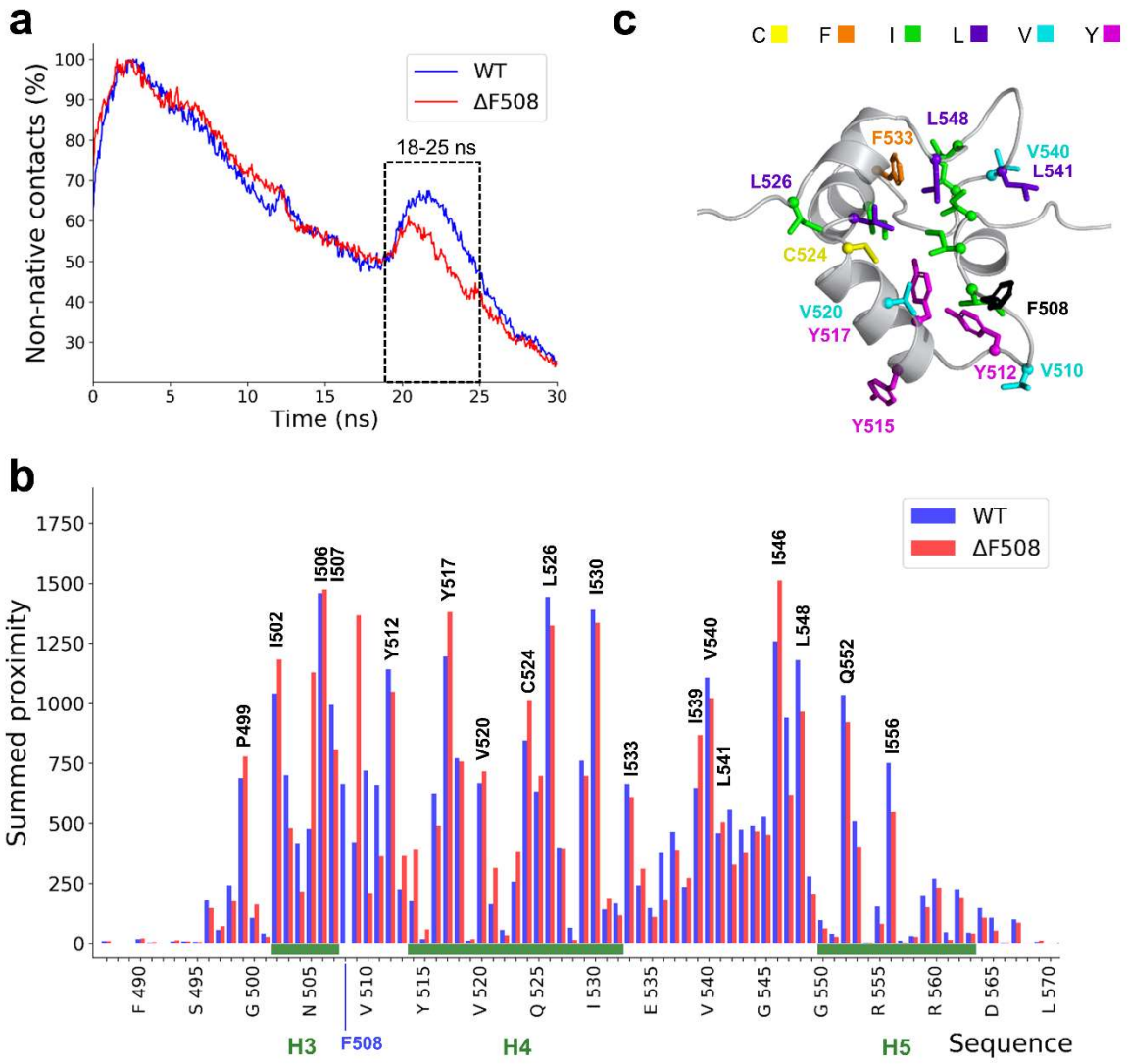

Fig. 4: An increased number of non-native contacts in WT may support self-chaperoning. (**a**) The number of non-native contacts normalized to the maximal value during pulling (n=50-50) is plotted. (**b**) Summed proximities of non-native contacts from all trajectories along the sequence of S6-α-S8 core. A higher value indicates that the residue has many close interactions during the investigated time period of unfolding (18-25 ns), which interactions are not present in the native structure. Blue and red columns represent the wild type and ΔF508 mutant, respectively. (**c**) Amino acid positions with the highest proximity values forming a hydrophobic core in the wild type protein are shown on the S6-α-S8 core structure (centroid of cluster #3) by stick representation.

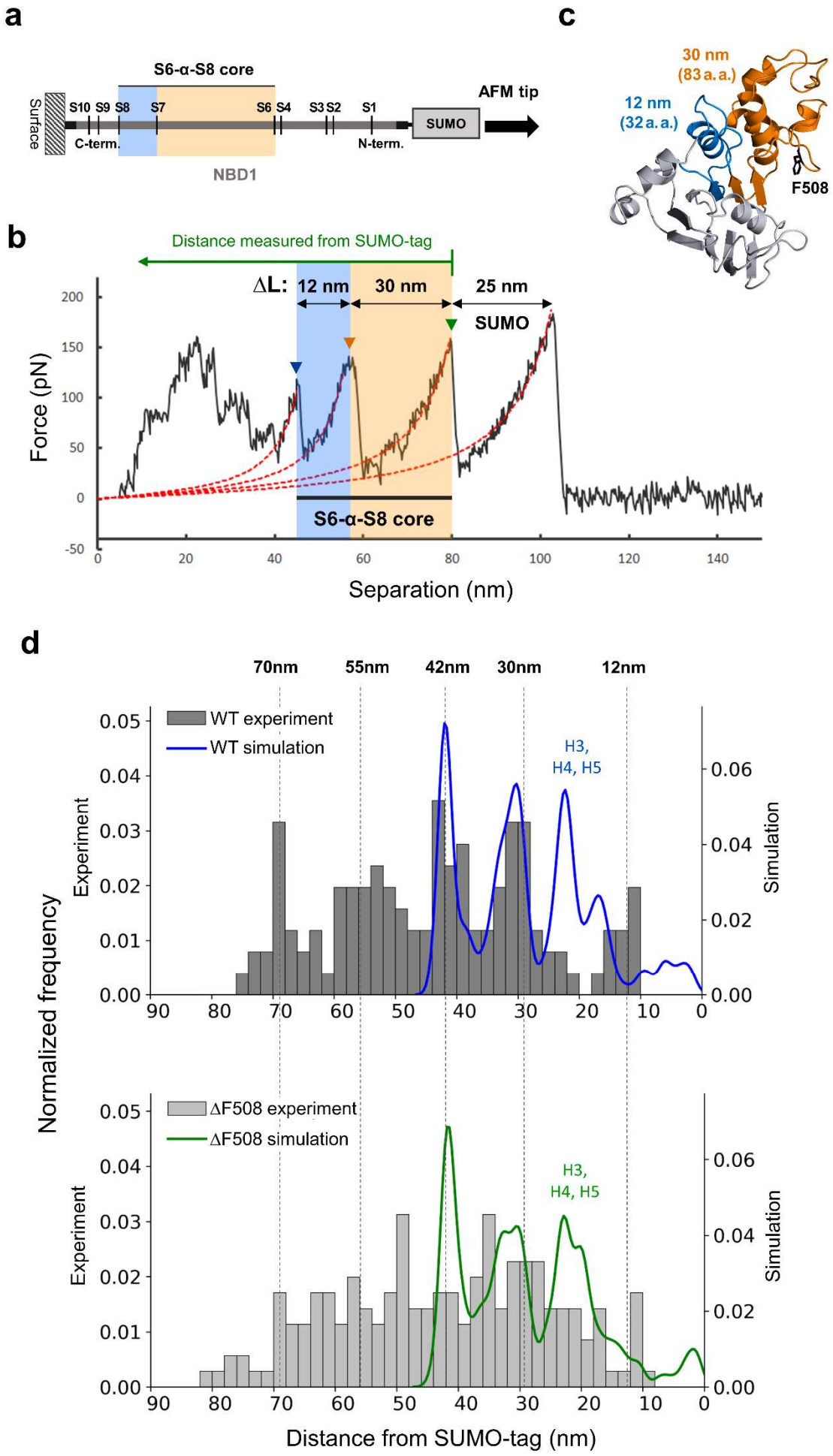
a
S6-α-S8 core
Surface
S10 S9 S8 S7 S6 S4 S3 S2 S1
C-term.
N-term.
SUMO
AFM tip
NBD1
c
30 nm
(83 a.a.)
12 nm
(32 a.a.)
F508
b
Distance measured from SUMO-tag
ΔL: 12 nm 30 nm 25 nm
SUMO
S6-α-S8 core
Force (pN)
Separation (nm)
d
70nm 55nm 42nm 30nm 12nm
WT experiment
WT simulation
H3,
H4, H5
Experiment
Simulation
Normalized frequency
ΔF508 experiment
ΔF508 simulation
H3,
H4, H5
Distance from SUMO-tag (nm)

Fig. 5: ΔF508 mutant exhibits a dispersed contour length distribution, acquired by AFM, when compared to wild type. (**a**) Schematic illustration of the cysteine-less ΔRI construct. SUMO-tagged NBD1 is immobilized on mica surface via a disulfide bond. The approximate position of β-strands is marked with S letters. (**b**) A force-extension curve demonstrating a characteristic sawtooth profile. Two unfolding events are marked by triangles. Peaks are fitted with the worm-like chain model (red dashed lines), providing contour length values (ΔL) from the unfolding of intermediates. (**c**) The structural regions corresponding to the unfolding regions with the given unfolding lengths are colored blue and orange on the NBD1 structure. (**d**) Histograms show the experimental contour length increments measured from the SUMO-tag. Increments corresponding to the expected length increases based on structure and simulations are indicated with dashed lines. Kernel density estimates (kde) show the distribution of contour length increases from S6-α-S8 core pulling MD simulations (blue -WT, green - ΔF508).

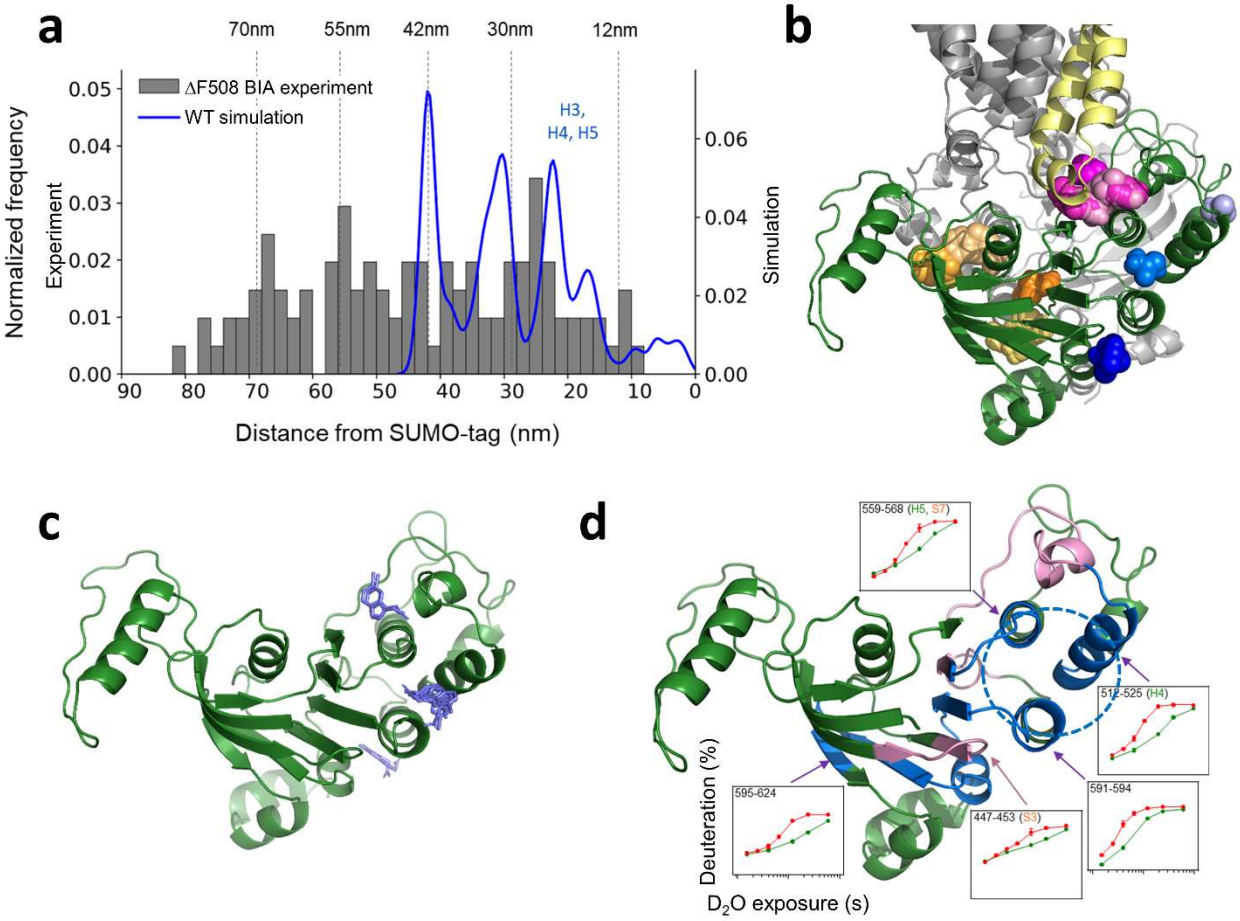

Fig. 6: BIA partially restores the WT unfolding pattern and binds to the α-subdomain. (**a**) The histogram shows the experimental contour length increments measured from the SUMO-tag, in the presence of BIA. Increments corresponding to the expected length increases based on WT pulling are indicated with dashed lines. Kernel density estimates show the distribution of contour lengths increases from WT S6-α-S8 core pulling MD simulations (blue). (**b**) Potential drug binding pockets in NBD1 were identified by FTMap. Some small molecule fragments bind to CL4 binding pocket (pink spheres) or NBD1/NBD2 interface (orange spheres). Green: NBD1, yellow: CL4, gray: other parts of CFTR. Binding of fragments (blue spheres) to rational pockets does not interfere with CFTR assembly. (**c**) BIA (blue sticks) was docked to NBD1 using Autodock Vina. (**d**) ΔF508 NBD1 regions, which exhibit lower dynamics upon BIA binding, are colored pink and blue in the structure, corresponding to slight and pronounced HDX rate differences, respectively. HDX rates in the absence (red) and presence (green) of BIA for selected peptide segments are shown in graphs. Blue dashed circle: BIA binding site.