

SUPPORTING INFORMATION

Thermal unfolding simulations of NBD1 domain variants reveal structural motifs associated with the impaired folding of F508del-CFTR

Sílvia G. Estácio*¹, Hugo F. M. C. Martiniano¹, Patrícia F. N. Faísca*^{1,2}

¹BioISI - Biosystems & Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, C8, 1749-016 Lisboa, Portugal

²Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, C8, 1749-016 Lisboa, Portugal

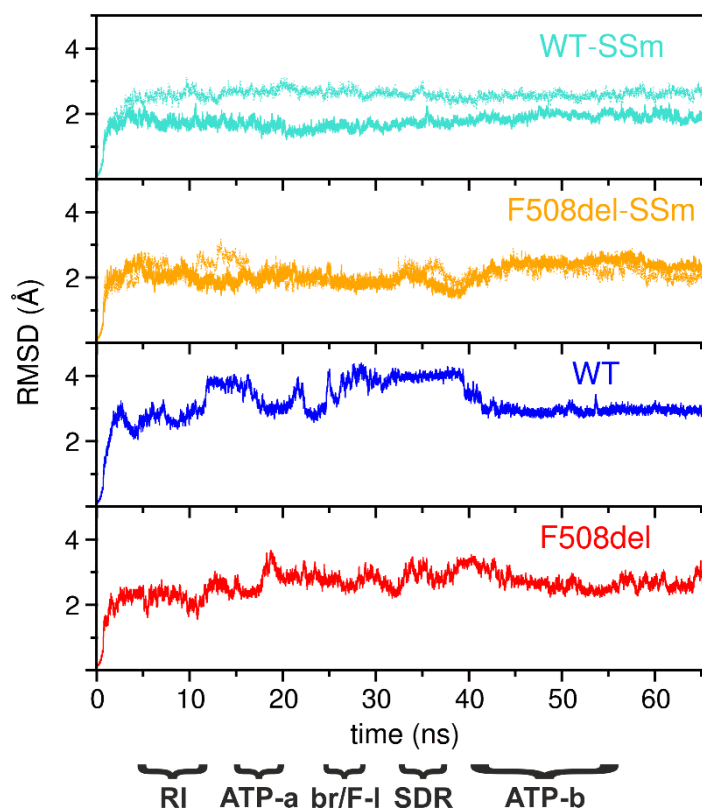


Figure S1. Root-mean-square deviation of the backbone C α atoms (C α -RMSD) with respect to the native structure along the 300-K trajectories of the four NBD1 variants. The brackets signal major regions along the primary sequence.

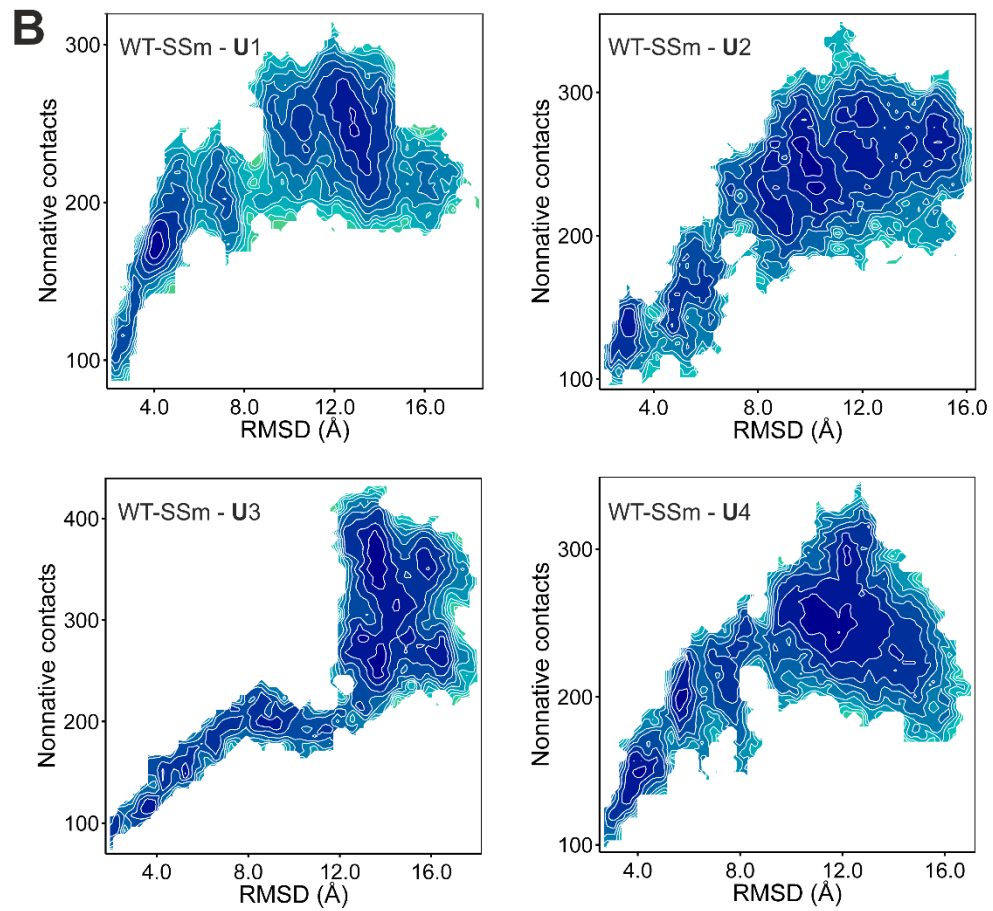
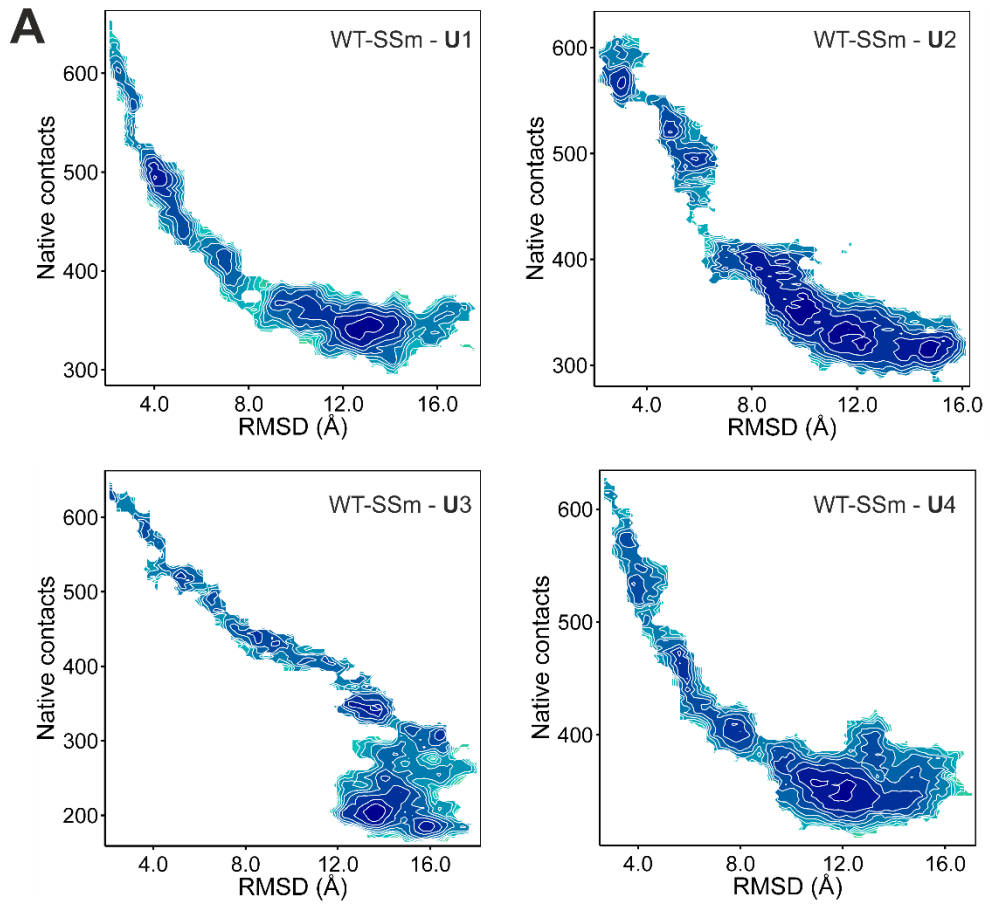


Figure S2. Two-dimensional property space histograms for the four WT-SSm unfolding simulations. Each property-based landscape was generated using the protein's core native (**A**)/nonnative (**B**) (residue) contacts and $\text{C}\alpha$ -RMSD as reaction coordinates. All the properties refer to the protein's core.

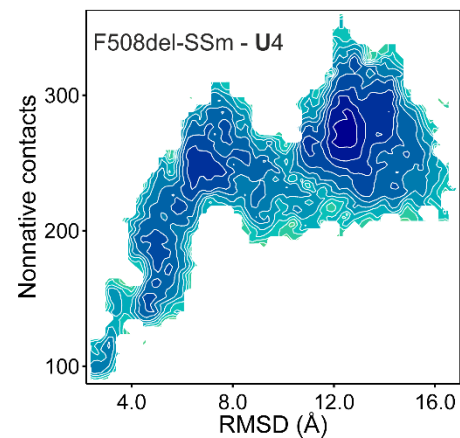
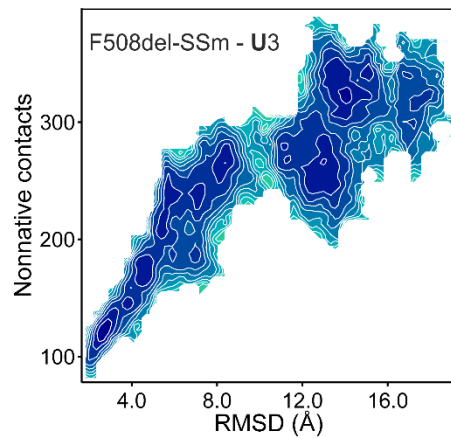
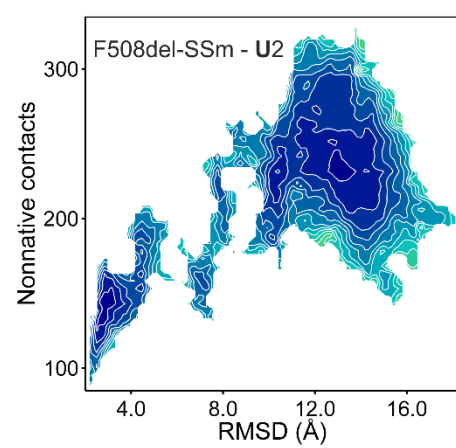
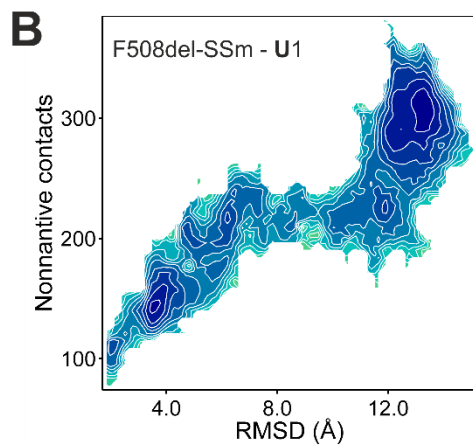
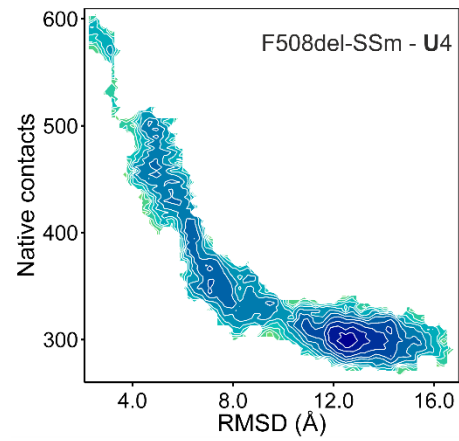
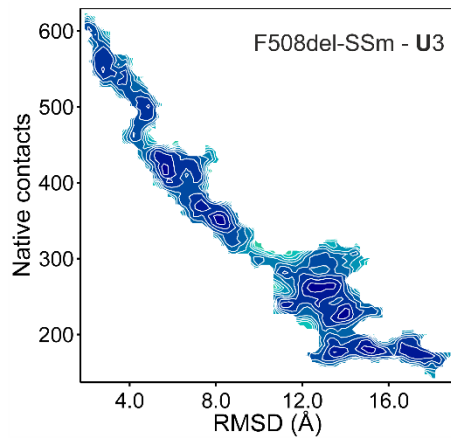
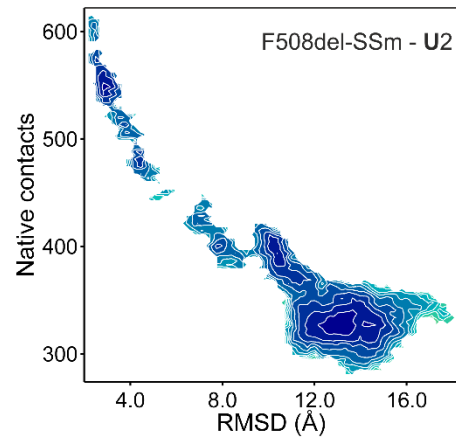
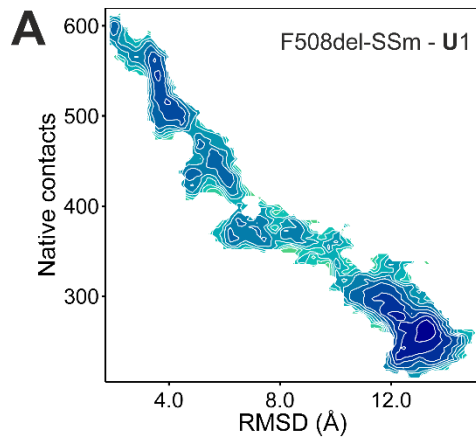


Figure S3. Two-dimensional property space histograms for the four F508del-SSm unfolding simulations. Each 2D property-based landscape was generated using the protein's core (inter-residue) native (A)/nonnative (B) contacts and C α -RMSD as reaction coordinates. All the properties refer to the protein's core.

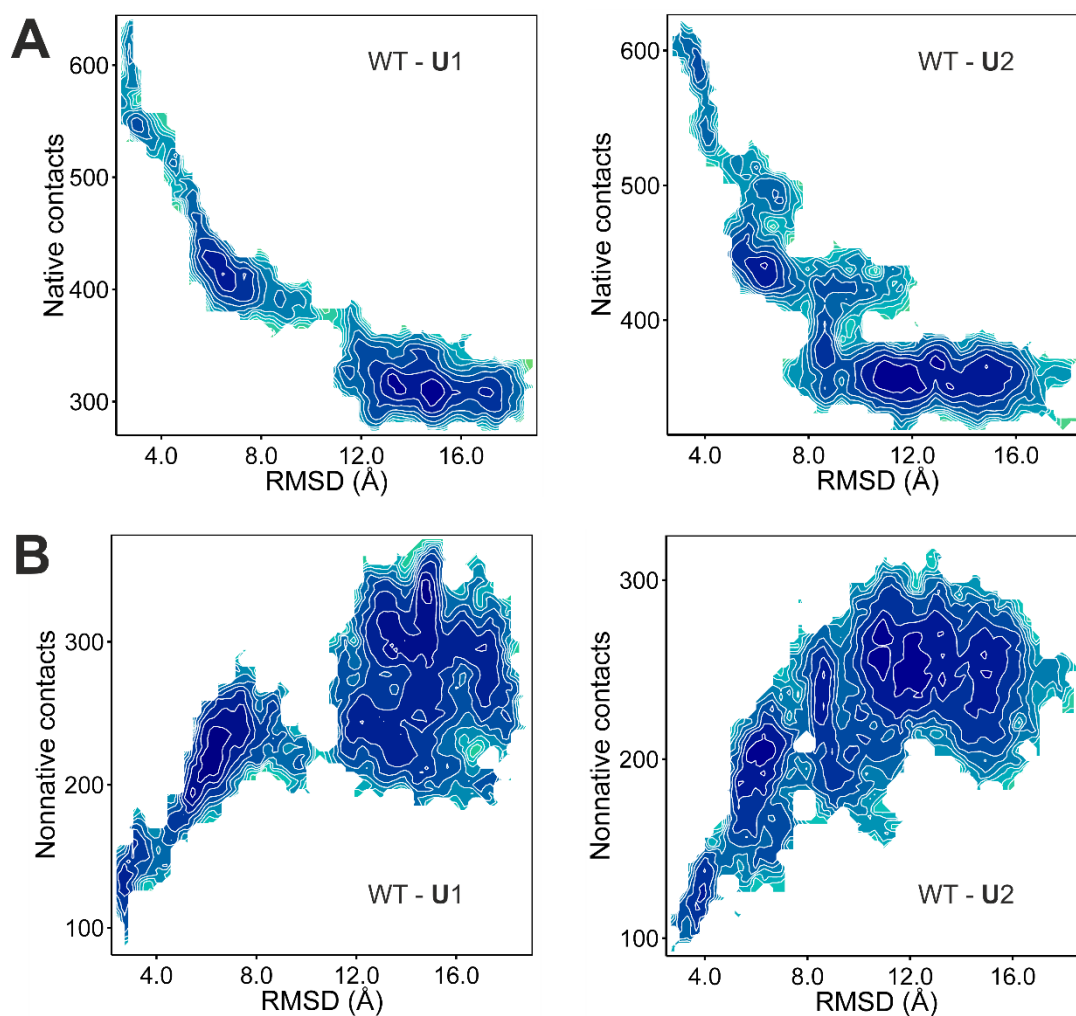


Figure S4. Two-dimensional property space histograms for the two WT unfolding simulations. Each 2D property-based landscape was generated using the protein's core (inter-residue) native (A)/nonnative (B) contacts and C α -RMSD as reaction coordinates. All the properties refer to the protein's core.

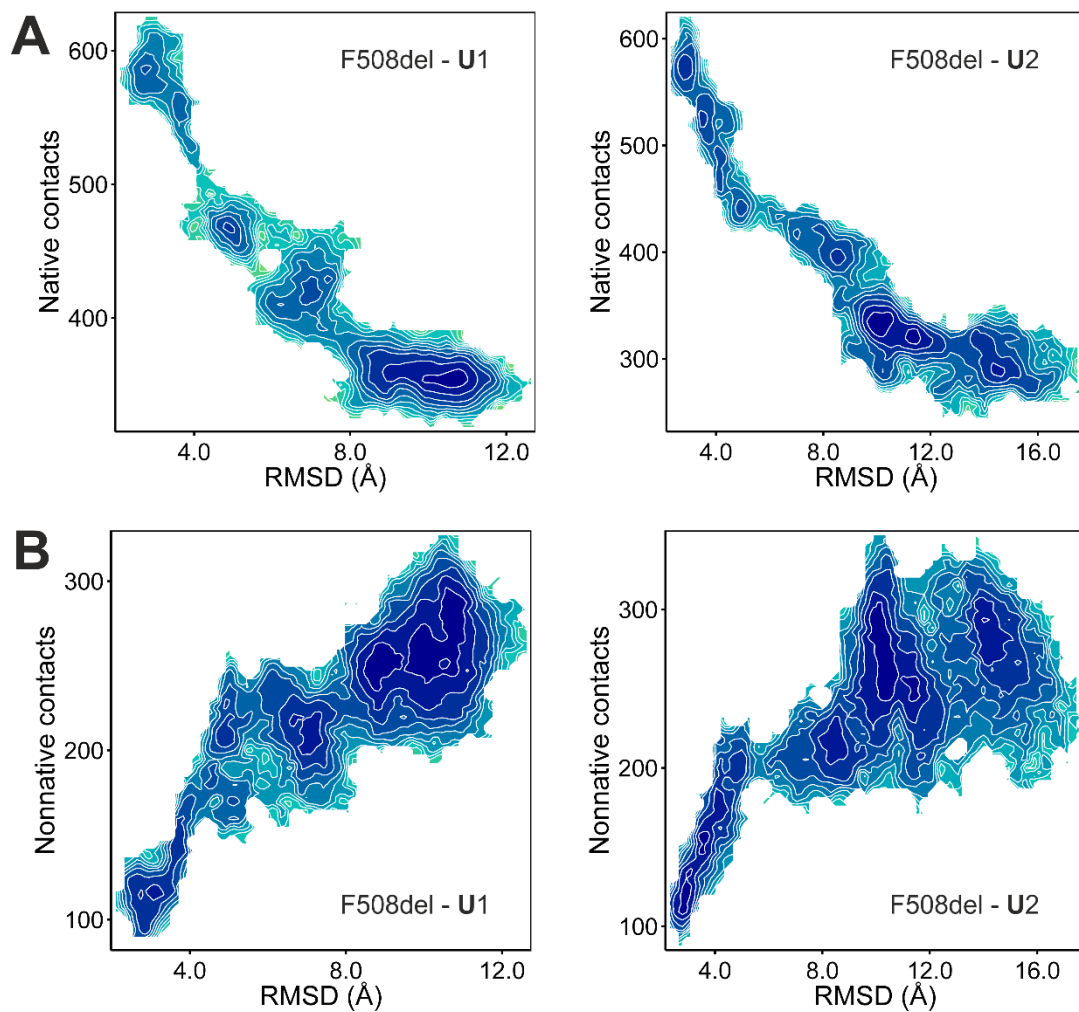


Figure S5. Two-dimensional property space histograms for the two F508del unfolding simulations. Each property-based landscape was generated using the protein's core native (A)/nonnative (B) (residue) contacts and $\text{C}\alpha$ -RMSD as reaction coordinates. All the properties refer to the protein's core.

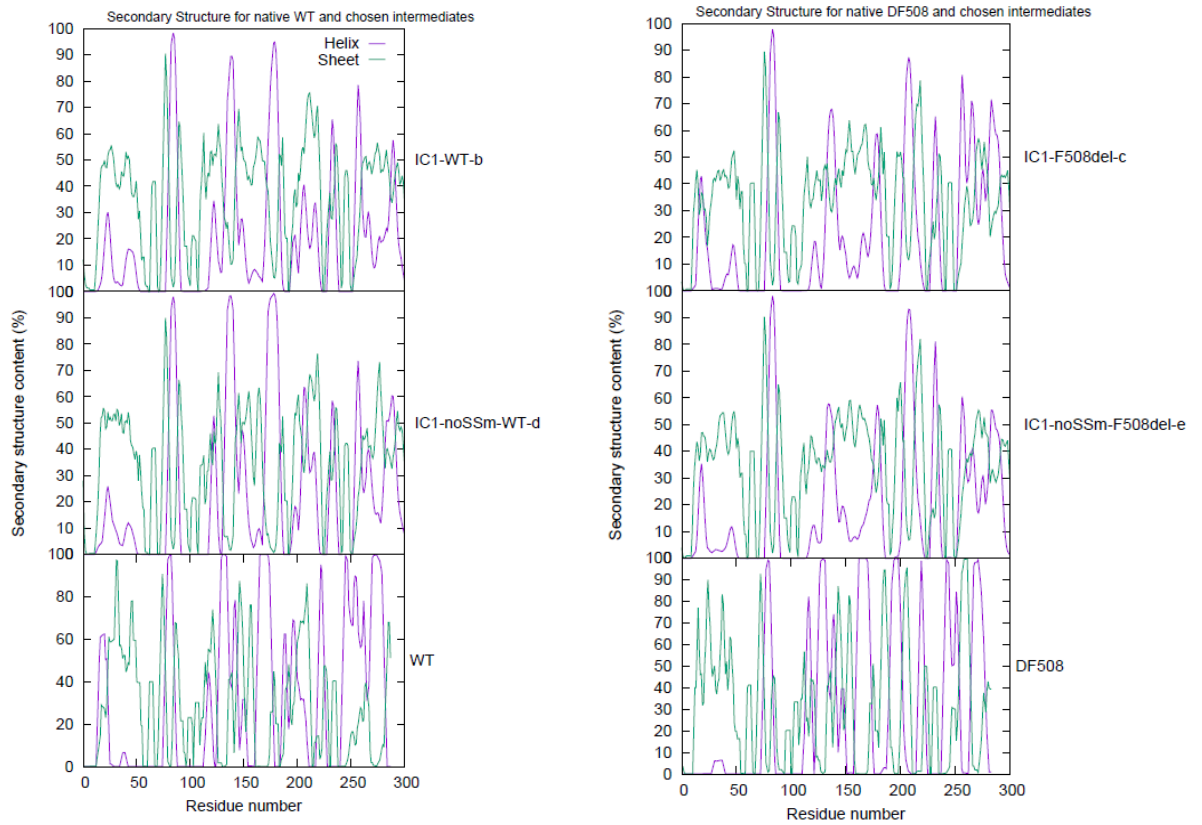


Figure S6. Structural characterization of selected intermediate states. Secondary structure content for selected intermediates of the WT NBD1 (left) and selected intermediates of the F508Del NBD1 (right). The secondary structure content was calculated with DSSP (<http://swift.cmbi.ru.nl/gv/dssp/>) and reported data represents a mean value calculated over the ensemble of conformations representative of the selected intermediates.

SI Table 1. The most statistically-relevant unfolding intermediates. This table lists the most relevant intermediates states in the unfolding trajectories of the four NBD1 variants. The labels I1, I2, etc or I1-SSm, I2-SSm, etc identify intermediate states exclusively populated by one variant, respectively WT/F508del or WT-SSm/F508del-SSm. IC1-SSm, IC2-SSm, etc are associated to intermediate states populated by both SSm variants while IC1-noSSm, IC2-noSSm, etc pinpoint intermediate states common to both WT and F508del variants. Intermediate states shared indistinctively by two or more variants (with and without additional single-point mutations) are labeled IC1, IC2, etc. The numbers added to the labels indicate the (decreasing) order of the intermediates' populations (1 identifies the most populated intermediate, 2 the second most populated, etc). The populations (in percentage) of the individual variants' intermediates were estimated as ratios to the total number of configurations in each variant's unfolding trajectories.

Intermediate state		$C\alpha$ RMSD - core (Å)	Native contacts - core	Non-native contacts - core	Hydro SASA (Å ²)	Q - core
IC2	WT-SSm-a (2%)	3.4 (0.6)	558 (31)	140 (20)	6114 (249)	0.89 (0.06)
	WT-a (5%)	3.7 (0.8)	555 (44)	142 (27)	6539 (288)	0.90 (0.08)
	F508del-a (10%)	3.3 (0.6)	554 (40)	135 (27)	6188 (306)	0.89 (0.07)
I3-SSm	F508del-SSm-a (4%)	3.9 (0.5)	522 (27)	151 (16)	6097 (354)	0.86 (0.05)
IC5	WT-SSm-b (4%)	4.8 (0.8)	486 (27)	177 (22)	6491 (257)	0.77 (0.06)
	F508del-b (6%)	4.7 (0.6)	473 (24)	192 (26)	6384 (385)	0.76 (0.05)
I2-SSm	F508del-SSm-b (8%)	5.3 (0.6)	449 (30)	195 (34)	6371 (256)	0.74 (0.07)
IC1	WT-SSm-c (1%)	6.6 (0.6)	419 (11)	209 (22)	6614 (245)	0.67 (0.03)
	WT-b (13%)	6.2 (0.6)	431 (17)	209 (22)	6858 (267)	0.69 (0.04)
	F508del-c (9%)	7.1 (0.8)	418 (18)	212 (17)	6333 (306)	0.67 (0.04)
IC2-SSm	WT-SSm-d (1%)	7.4 (0.6)	401 (12)	212 (25)	6818 (292)	0.64 (0.03)
	F508del-SSm-c (6%)	6.8 (0.6)	394 (25)	228 (32)	6626 (301)	0.65 (0.06)

IC7	F508del-SSm-e (2%)	10.4 (0.7)	388 (17)	228 (22)	6483 (248)	0.64 (0.04)
	WT-c (4%)	8.4 (0.7)	386 (17)	235 (20)	7202 (320)	0.62 (0.04)
IC3	WT-SSm-e (7%)	9.8 (1.0)	369 (18)	244 (26)	6788 (340)	0.59 (0.05)
	F508del-d (9%)	9.0 (0.6)	367 (20)	233 (21)	6714 (413)	0.59 (0.06)
IC1-noSSm	WT-d (18%)	11.5 (1.1)	357 (12)	253 (22)	7047 (423)	0.58 (0.04)
	F508del-e (44%)	10.3 (0.8)	347 (19)	265 (24)	6674 (381)	0.56 (0.06)
IC4	F508del-SSm-g (8%)	12.0 (0.8)	304 (17)	257 (26)	6856 (342)	0.50 (0.06)
	F508del-f (6%)	11.6 (0.8)	312 (13)	239 (25)	7553 (323)	0.50 (0.04)
I1-SSm	F508del-SSm-h (14%)	13.1 (0.7)	254 (18)	295 (29)	7015 (382)	0.42 (0.07)
IC1-SSm	WT-SSm-f (20%)	12.5 (1.2)	343 (19)	253 (27)	6889 (350)	0.54 (0.06)
	F508del-SSm-i (28%)	13.3 (1.2)	320 (18)	252 (34)	6810 (358)	0.53 (0.06)
IC2-noSSm	WT-e (12%)	13.7 (0.9)	307 (13)	290 (38)	7414 (489)	0.50 (0.04)
	F508del-g (8%)	14.0 (0.9)	302 (22)	282 (22)	7224 (305)	0.49 (0.07)
I1	WT-f (12%)	14.4 (1.2)	364 (11)	258 (19)	6923 (286)	0.59 (0.03)
IC8	WT-SSm-j (1%)	15.5 (0.9)	360 (13)	214 (21)	7186 (336)	0.57 (0.04)
	WT-h (5%)	14.7 (1.0)	345 (10)	229 (19)	7263 (358)	0.56 (0.03)
IC6	WT-SSm-k (2%)	14.6 (0.8)	316 (11)	266 (18)	6659 (268)	0.50 (0.04)
	WT-g (8%)	15.0 (0.6)	312 (13)	304 (32)	7090 (432)	0.50 (0.04)
IC3-SSm	F508del-SSm-j (3%)	14.2 (0.9)	183 (13)	334 (24)	7436 (328)	0.30 (0.07)

	WT-SSm-i (2%)	15.8 (0.7)	184 (8)	348 (20)	7472 (265)	0.29 (0.05)

The relative standard deviations of the fraction of native contacts (Q) were obtained using the rule for error propagation in a quotient which uses the (individual) standard deviation errors of both quantities used in the calculation of Q , i.e., the mean number of native contacts in the intermediate state ensemble and the mean number of native contacts in the 300-K ensemble.

SI Table 2. Structural characterization of the WT-SSm intermediates.

	Cα RMSD (Å)	Cα RMSD – core (Å)	native contacts	native contacts – core	atomic native contacts	atomic native contacts – core	nonnative contacts	nonnative contacts – core	atomic nonnative contacts	atomic nonnative contacts – core	Hydro SASA (Å²)
WT-SSm-a (2%)	5.5 (0.7)	3.4 (0.6)	567 (35)	558 (31)	4141 (312)	4095 (295)	207 (25)	140 (20)	3776 (258)	3108 (199)	6114 (249)
WT-SSm-b (4%)	6.9 (0.6)	4.8 (0.8)	493 (26)	486 (27)	3494 (250)	3458 (254)	249 (23)	177 (22)	4024 (242)	3328 (225)	6491 (257)
WT-SSm-c (1%)	8.4 (0.7)	6.6 (0.6)	425 (11)	419 (11)	2958 (113)	2923 (111)	279 (24)	209 (22)	4138 (275)	3442 (226)	6614 (245)
WT-SSm-d (1%)	9.3 (0.7)	7.4 (0.6)	406 (12)	401 (12)	2810 (113)	2787 (114)	276 (26)	212 (25)	3997 (280)	3379 (240)	6818 (292)
WT-SSm-e (7%)	11.5 (1.2)	9.8 (1.0)	374 (18)	369 (18)	2598 (142)	2569 (140)	314 (33)	244 (26)	4251 (324)	3555 (251)	6788 (340)
WT-SSm-f (20%)	13.5 (1.2)	12.5 (1.2)	349 (19)	343 (19)	2440 (160)	2405 (163)	327 (31)	253 (27)	4279 (315)	3563 (282)	6889 (350)
WT-SSm-g (4%)	13.6 (0.7)	13.5 (0.7)	206 (11)	201 (10)	1437 (89)	1412 (89)	452 (33)	365 (27)	4949 (347)	4125 (275)	7073 (452)

WT-SSm-h (4%)	15.8 (1.3)	14.5 (1.2)	246 (25)	241 (25)	1666 (180)	1640 (176)	392 (44)	304 (36)	4557 (389)	3750 (321)	7118 (389)
WT-SSm-i (2%)	15.8 (0.7)	15.8 (0.7)	190 (8)	184 (8)	1298 (78)	1273 (79)	425 (23)	348 (20)	4717 (247)	3883 (224)	7472 (265)
WT-SSm-j (1%)	16.0 (0.9)	15.5 (0.9)	366 (12)	360 (13)	2598 (123)	2566 (126)	281 (26)	214 (21)	3883 (307)	3251 (221)	7186 (336)
WT-SSm-k (2%)	16.5 (1.0)	14.6 (0.8)	325 (12)	316 (11)	2262 (95)	2206 (90)	364 (24)	266 (18)	4558 (274)	3610 (236)	6659 (268)

Mean values and respective standard deviations obtained for conformational ensembles of the unfolding intermediates identified via structural clustering. The second and third columns display the mean C α -RMSD of the full-length or core chains fitted to the corresponding regions of the native structure. The populations (in percentage) of the individual unfolding intermediates were estimated as ratios to the total number of configurations in the four unfolding trajectories.

SI Table 3. Structural characterization of the F508del-SSm intermediates.

	Cα RMSD (Å)	Cα RMSD – core (Å)	native contacts	native contacts – core	atomic native contacts	atomic native contacts – core	nonnative contacts	nonnative contacts – core	atomic nonnative contacts	atomic nonnative contacts – core	Hydro SASA (Å²)
F508del-SSm-a (4%)	7.4 (0.6)	3.9 (0.5)	529 (29)	522 (27)	3882 (264)	3850 (256)	214 (19)	151 (16)	3753 (217)	3126 (169)	6097 (354)
F508del-SSm-b (8%)	7.9 (1.0)	5.3 (0.6)	453 (30)	449 (30)	3179 (245)	3163 (242)	256 (37)	195 (34)	4029 (326)	3421 (292)	6371 (256)
F508del-SSm-c (6%)	8.5 (0.7)	6.8 (0.6)	397 (24)	394 (25)	2760 (198)	2748 (200)	288 (34)	228 (32)	4114 (319)	3528 (280)	6626 (301)
F508del-SSm-d (3%)	9.3 (0.6)	8.3 (0.4)	350 (15)	346 (15)	2390 (116)	2376 (119)	330 (21)	256 (20)	4352 (253)	3615 (238)	6744 (296)

F508del-SSm-e (2%)	10.7 (0.6)	10.4 (0.7)	392 (16)	388 (17)	2785 (161)	2777 (163)	315 (29)	228 (22)	4374 (284)	3481 (218)	6483 (248)
F508del-SSm-f (2%)	12.5 (0.9)	12.4 (1.1)	238 (12)	233 (12)	1583 (93)	1568 (91)	366 (32)	293 (31)	4387 (254)	3672 (255)	7411 (303)
F508del-SSm-g (8%)	12.6 (0.6)	12.0 (0.8)	308 (17)	304 (17)	2116 (148)	2100 (150)	335 (29)	257 (26)	4281 (288)	3514 (266)	6856 (342)
F508del-SSm-h (14%)	13.6 (0.7)	13.1 (0.7)	257 (17)	254 (18)	1759 (143)	1745 (148)	376 (35)	295 (29)	4472 (328)	3703 (282)	7015 (382)
F508del-SSm-i (28%)	13.9 (0.9)	13.3 (1.2)	324 (19)	320 (18)	2235 (151)	2221 (150)	329 (32)	252 (34)	4293 (330)	3554 (323)	6810 (358)
F508del-SSm-j (3%)	14.7 (0.7)	14.2 (0.9)	186 (13)	183 (13)	1167 (111)	1156 (112)	408 (26)	334 (24)	4629 (278)	3934 (262)	7436 (328)
F508del-SSm-k (2%)	17.2 (0.7)	17.3 (0.7)	182 (12)	177 (11)	1135 (92)	1119 (94)	401 (32)	318 (27)	4507 (319)	3753 (276)	7612 (330)

Mean values and respective standard deviations obtained for conformational ensembles of the unfolding intermediates identified via structural clustering. The second and third columns display the mean C α -RMSD of the full-length or core chains fitted to the corresponding regions of the native structure. The populations (in percentage) of the individual unfolding intermediates were estimated as ratios to the total number of configurations in the four unfolding trajectories.

SI Table 4. Structural characterization of the WT intermediates.

	Cα RMSD (Å)	Cα RMSD – core (Å)	native contacts	native contacts – core	atomic native contacts	atomic native contacts – core	nonnative contacts	nonnative contacts – core	atomic nonnative contacts	atomic nonnative contacts – core	Hydro SASA (Å²)
WT-a (5%)	6.1 (1.1)	3.7 (0.8)	565 (46)	555 (44)	4145 (384)	4081 (370)	189 (30)	142 (27)	3624 (278)	3085 (249)	6539 (288)
WT-b (13%)	9.5 (1.4)	6.2 (0.6)	437 (17)	431 (17)	3099 (144)	3061 (144)	276 (23)	209 (22)	4151 (232)	3468 (221)	6858 (267)
WT-c (4%)	10.2 (0.8)	8.4 (0.7)	390 (18)	386 (17)	2761 (140)	2737 (139)	291 (23)	235 (20)	4120 (257)	3561 (229)	7202 (320)
WT-d (18%)	12.5 (1.1)	11.5 (1.1)	362 (12)	357 (12)	2554 (111)	2528 (112)	325 (28)	253 (22)	4283 (273)	3556 (240)	7047 (423)
WT-e (12%)	14.2 (1.0)	13.7 (0.9)	312 (14)	307 (13)	2156 (116)	2132 (114)	345 (43)	290 (38)	4306 (419)	3719 (372)	7414 (489)
WT-f (12%)	15.3 (1.1)	14.4 (1.2)	368 (11)	364 (11)	2630 (93)	2612 (90)	333 (25)	258 (19)	4312 (248)	3599 (198)	6923 (286)
WT-g (8%)	15.3 (0.9)	15.0 (0.6)	318 (14)	312 (13)	2234 (129)	2203 (122)	372 (39)	304 (32)	4630 (339)	3923 (287)	7090 (432)
WT-h (5%)	16.3 (1.1)	14.7 (1.0)	350 (10)	345 (10)	2502 (94)	2477 (93)	320 (22)	229 (19)	4211 (260)	3315 (227)	7263 (358)

Mean values and respective standard deviations obtained for conformational ensembles of the unfolding intermediates identified via structural clustering. The second and third columns display the mean C α -RMSD of the full-length or core chains fitted to the corresponding regions of the native structure. The populations (in percentage) of the individual unfolding intermediates were estimated as ratios to the total number of configurations in the two unfolding trajectories.

SI Table 5. Structural characterization of the F508del intermediates.

	Cα RMSD (Å)	Cα RMSD – core (Å)	native contacts	native contacts – core	atomic native contacts	atomic native contacts – core	nonnative contacts	nonnative contacts – core	atomic nonnative contacts	atomic nonnative contacts – core	Hydro SASA (Å²)
F508del- a (10%)	5.5 (0.7)	3.3 (0.6)	561 (42)	554 (40)	4170 (378)	4141 (367)	188 (27)	135 (27)	3645 (218)	3060 (204)	6188 (306)
F508del- b (6%)	7.5 (0.8)	4.7 (0.6)	478 (25)	473 (24)	3466 (206)	3449 (203)	253 (37)	192 (26)	4048 (319)	3427 (225)	6384 (385)
F508del- c (9%)	9.6 (0.6)	7.1 (0.8)	424 (19)	418 (18)	3029 (192)	3002 (187)	297 (23)	212 (17)	4316 (250)	3468 (200)	6333 (306)
F508del- d (9%)	10.5 (0.8)	9.0 (0.6)	371 (20)	367 (20)	2592 (157)	2575 (154)	313 (27)	233 (21)	4323 (269)	3496 (204)	6714 (413)
F508del- e (44%)	11.8 (0.8)	10.3 (0.8)	350 (19)	347 (19)	2461 (181)	2446 (180)	345 (28)	265 (24)	4529 (306)	3721 (246)	6674 (381)
F508del- f (6%)	12.8 (0.9)	11.6 (0.8)	316 (14)	312 (13)	2130 (108)	2113 (108)	294 (26)	239 (25)	3923 (260)	3356 (238)	7553 (323)
F508del- g (8%)	14.5 (0.8)	14.0 (0.9)	305 (21)	302 (22)	2064 (176)	2048 (179)	350 (25)	282 (22)	4399 (270)	3709 (213)	7224 (305)
F508del- h (5%)	15.6 (0.8)	15.3 (0.9)	290 (13)	287 (13)	1901 (116)	1889 (116)	330 (35)	265 (33)	4178 (346)	3533 (317)	7469 (335)

Mean values and respective standard deviations obtained for conformational ensembles of the unfolding intermediates identified via structural clustering. The second and third columns display the mean C α -RMSD of the full-length or core chains fitted to the corresponding regions of the native structure. The populations (in percentage) of the individual unfolding intermediates were estimated as ratios to the total number of configurations in the two unfolding trajectories.

