Journal of Bioenergetics and Biomembranes, Vol. 29, No. 5, 1997

Modeling of Nucleotide Binding Domains of ABC Transporter Proteins Based on a F_1-ATPase/recA Topology: Structural Model of the Nucleotide Binding Domains of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)*

Mario A. Bianchet,[1] Young Hee Ko,[2] L. Mario Amzel[1] and Peter L. Pedersen[2]

Received July 1, 1997; accepted November 1, 1997

Members of the ABC transporter superfamily contain two nucleotide binding domains. To date, the three dimensional structure of no member of this super-family has been elucidated. To gain structural insight, the known structures of several other nucleotides binding proteins can be used as a framework for modeling these domains. We have modeled both nucleotide binding domains of the protein CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) using the two similar domains of mitochondrial F_1-ATPase. The models obtained, provide useful insights into the putative functions of these domains and their possible interaction as well as a rationale for the basis of Cystic Fibrosis causing mutations. First, the two nucleotide binding domains (folds) of CFTR are each predicted to span a 240–250 amino acid sequence rather than the 150–160 amino acid sequence originally proposed. Second, the first nucleotide binding fold, is predicted to catalyze significant rates of ATP hydrolysis as a catalytic base (E504) resides near the γ phosphate of ATP. This prediction has been verified experimentally [Ko, Y.H., and Pedersen, P.L. (1995) **J. Biol. Chem. 268,** 24330-24338], providing support for the model. In contrast, the second nucleotide binding fold is predicted at best to be a weak ATPase as the glutamic acid residue is replaced with a glutamine. Third, F508, which when deleted causes ~70% of all cases of cystic fibrosis, is predicted to lie in a cleft near the nucleotide binding pocket. All other disease causing mutations within the two nucleotide binding domains of CFTR either reside near the Walker *A* and Walker *B* consensus motifs in the heart of the nucleotide binding pocket, or in the *C* motif which lies outside but near the nucleotide binding pocket. Finally, the two nucleotide binding domains of CFTR are predicted to interact, and in one of the two predicted orientations, F508 resides near the interface.
This is the first report where both nucleotide binding domains of an ABC transporter and their putative domain-domain interactions have been modeled in three dimensions. **The methods and the template used in this work can be used to analyze the structures and function of the nucleotide binding domains of all other members of the ABC transporter super-family.**

KEY WORDS: Nucleotide Binding Domain; Cystic Fibrosis; ABC transporters; Traffic ATPases; CFTR.

[1] Department of Biophysics and Biophysical Chemistry.
[2] Department of Biological Chemistry, The Johns Hopkins University, School of Medicine, 725 N Wolfe Street, Baltimore, MD, 21205-2185.

* Supported by Grants from the NIH (NIDDK) and the Cystic Fibrosis Foundation to PLP. Supported by Grant GM25432 from the NIH to LMA.

0145-479X/97/1000-0503$12.50/0

INTRODUCTION

Members of the ABC transporter super-family [Hyde et al. (1990); Higgins (1992)] share extensive sequence similarities. The family, which includes, the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), P-glycoprotein (MDR1, a permease involved in the transport of a wide variety of drugs), the Ste6 gene product, pfMDR, Histidine permease (HisP), and many others. All these proteins contain membrane domains as well as large hydrophilic domains associated with binding of ATP. Because several use hydrolysis of ATP to support substrate accumulation, they are also called "Traffic ATPases" [Doige & Ames (1993)]. Members of the ABC transporter protein family have three characteristic sequence signatures: (A) $GX_4GK[T/S]$ (also called "P-loop"), (B) $RX_{6-8}h_4D$ and (C) LSXGX[R/K] (X: any residue; and h: hydrophobic residue). The first two of these signatures, also called "Walker motifs" A and B, are involved in ATP binding [Walker et al. 1982].

One of the better known members of the ABC transporter super-family, the CFTR protein, is a polypeptide related to the disease Cystic Fibrosis (CF). CF, an autosomal recessive genetic disease predominant among Caucasians, affects numerous organs, including lung and airways, pancreas and sweat glands. The disease symptoms are presumably caused by a decrease in chloride ion conductance across the apical membrane in epithelial cells that results from mutations in the gene that encodes CFTR, a 1480 amino acid chloride channel [Riordan et al. (1989)]. Of the mutations found in the CF gene, 42% are missense, 21% are nonsense, and 10% alter nucleotides essential for codon splicing [Tsui (1992)]. Approximately 70% of all cases of CF are caused by a single deletion mutation (ΔF508) [Kerem et al. (1990)]. This mutation causes severe disease with symptomatology that includes severe lung disease, by preventing trafficking of the protein to its final position in the plasma membrane [Cheng et al. (1990), Denning et al. (1992), Dalemans et al. (1991), Qu et al. (1996)]. Another deletion close to codon 508, ΔI507 produces a very similar phenotype. Importantly, missense mutations such as F508C, I506V and I507V are benign and do not cause the disease [Kobayashi et al. (1990)]. Other missense mutations are found in or near the motif $GX_4GK[T/S]$, i.e., A455E and G458V. Significantly, mutation of G551, (G551D) in motif C (LSX*G*X[R/K]) is associated with high chloride levels in sweat, pancreatic insufficiency and variable lung disease, producing a phenotype very similar to ΔF508 [Hamosh et al. (1992)]. Interestingly, mutation of R553 can partially suppress the effect of ΔF508, [Teem et al. (1993)].

Optimal activation of the CFTR channel requires phosphorylation by *c*AMP dependent kinases [Cheng et al. (1991)], and recently, Li et al. (1996) have presented direct evidence that ATP hydrolysis is used to gate, or to modulate the CFTR chloride channel. The A and B motifs associated with the nucleotide binding site of CFTR and ABC transporters are found also in proteins of the nucleotide binding family [Walker et al. (1982)] with known three-dimensional structures, such as adenylate kinase (ADK) [Schulz et al. (1974)], elongation factor-tu [Morikawa et al. (1978)], ras p21 [Pai et al. (1990)], transducin (a G-protein) [Noel et al. (1993)], recA [Story et al. (1992)] and the F_1 sector of F_0F_1-ATPase (ATP synthase) [Abrahams et al. (1994), Bianchet et al. (1998)]. Most of these proteins use the energy of hydrolysis of (G/A)TP to bring about conformational changes that transmit signals or perform work. Of these proteins, only F_0F_1-ATPase is involved in ion transport across a biological membrane. In animal cells, this enzyme synthesizes ATP using a proton gradient across the inner mitochondrial membrane as the driving force (for reviews see [Weber and Senior (1997) and Pedersen and Amzel (1993)]. Operating in the opposite direction, F_0F_1-ATPase can pump protons across the membrane against their electrochemical gradient utilizing the free energy of ATP hydrolysis. F_1-ATPase, the soluble part of the ATP synthase complex is a multiple subunit protein made up of five different polypeptide chains. The two larger subunits, α and β, contain the sites for synthesis or hydrolysis of ATP. Recently, atomic resolution structures of substantial parts of F_1-ATPase were elucidated, for the bovine heart F_1 [Abrahams et al. (1994)], for the $\alpha_3\beta_3$ complex of the F_1 of *Thermophilium Bacterium* [Shirakihara et al. (1997)], and for the F_1 of rat liver [Bianchet et al. (1998)]. The features described above make F_1-ATPase a suitable template for modeling the regions of ABC transporters associated with binding of nucleotides. This paper presents the three-dimensional models of two such regions of CFTR based on F_1-ATPase. In developing this model we utilized features common to all ABC transporters. Therefore, modeling other members of this super-family (i.e., MDR, Permeases, etc.) using the alignment and assignment of struc-

tural elements presented in this paper should be straightforward.[3]

METHODS

Although NBD1 and NBD2 have been defined in the literature as residues F433 through S589, and Y1218 through R1386 respectively [Riordan et al. (1989)], we define them here as residues L441 through K684 and L1227 through L1480, respectively. The justification for this is that this is the minimum number of residues that can make a complete nucleotide binding domain using the modeling approach described here. The sequences of NBD1 and NBD2 were aligned manually with the sequences of the nucleotide domains of the $F_1\alpha$- and $F_1\beta$-subunits taking into consideration the criteria discussed in the following sections. Coordinates for the equivalent atoms in the aligned residues were obtained from the template molecule, rat liver F_1-ATPase. The non common atoms were generated automatically and side chain orientations were optimized by visual inspection. Insertions were built using the program O [Jones et al. (1990)] and the molecular modeling program Quanta [MSI Inc.] based on the predicted secondary structure when it was compatible with the topology assigned. The coordinates of the initial model are optimized using several picoseconds of molecular dynamics at 300 K, with the version 2.3 of CHARMM in the Quanta 41 software. For drawing and visualization the programs Setor [Evans (1993)] and O were used. The geometry and quality of the models was assessed using the protein package of Quanta 41. Polar and apolar accessible area were calculated using the programs AREAIMOL, SURFACE and DIFAREA of the CCP4 suite of program [CCP4 (1994)]. Accessibility surfaces and electrostatics superficial potential were obtained with the program GRASP [Nicholls et al. (1991)]. The sequences of ABC transporters were aligned using the program MAXHOM, a neural network multiple sequence alignment included in the PHD package.(http://www.embl.heidelberg.de/predictprotein) [Rost (1996)].

[3] This work was first presented in abstract form at both the Ninth Annual North American Cystic Fibrosis Conference (Dallas, Texas) [Bianchet et al. (1995a)], and the Biophysical Society Meeting (Baltimore Md.)[Bianchet et al. (1995b)].

RESULTS AND DISCUSSION

Domain Organization of ABC Transporters

Several groups have proposed models of complete ABC transporters based on analysis of hydropathy profiles [Riordan et al. (1989), Petsko (1990)]. Significantly, CFTR is a single-chain multiple domain protein, with a hydropathy profile (Figure 1) characteristic of the ABC super-family of proteins with the expected regions of defined or dominant hydropathy. The two halves of this profile seem to be similar to each other, suggesting gene duplication. However in the case of CFTR , the lack of conservation in the relative positions of exon-intron segments argues against this [Riordan et al. (1989)]. The CFTR regions between residues 50 to 404 and 850 to 1150 are predominantly hydrophobic and associated with transmembrane spanning domains (TMD1 & TMD2 respectively). A straightforward prediction of topological transmembrane helices suggests 6 helices per transmembrane domain [for a review of the methods see White (1994)]. Two amphipathic regions between residues 441 and 687 and between residues 1227 and 1480 (the C-terminus) contain the characteristic ABC sequences. These two regions have similar size, hydropathy characteristics, and sequences homologies beyond the ABC sequences. Probably, the regions fold in a similar way forming two domains, that we refer to here as nucleotide binding domains NBD1 and NBD2. Traditionally, on the basis of sequence homology with nucleotide binding domains of other proteins, the NBDs of CFTR are considered to span shorter regions, from F433 through S589 in the case of the first NBD, and from Y1218 though R1386 in the case of the

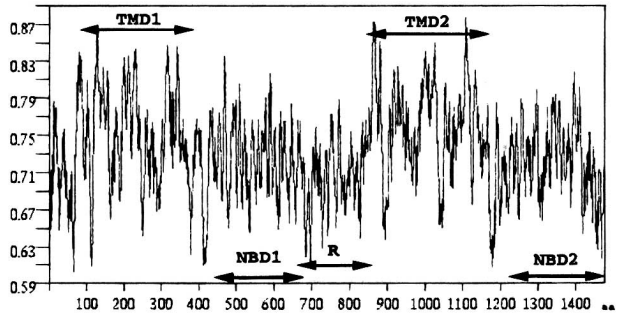


Fig. 1. *Hydropathy plot of the entire CFTR.* Hydropathy plot calculated using the Kite-Doolite algorithm. The double arrows point to the approximate positions of the transmembrane regions (TMD1 & TMD2), Nucleotide Binding Domains (NBD1 & NBD2) and Regulatory Domain (R domain) in the sequence.

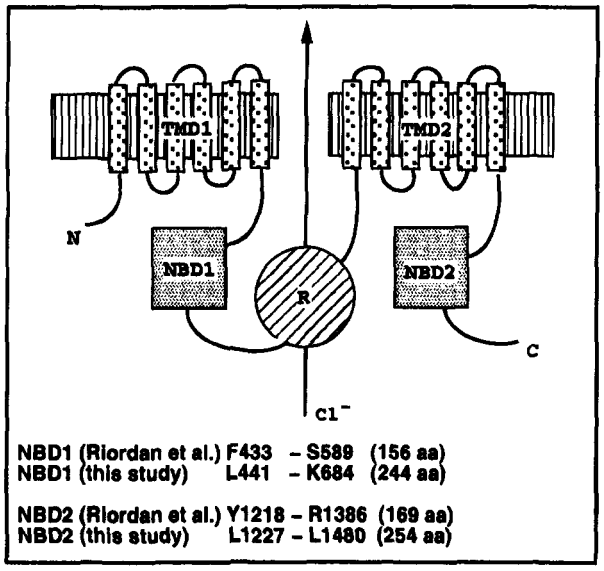


Fig. 2. *Schematic figure of a model for the entire CFTR.* The five distinct domains are depicted. The critical ΔF508 mutation causing 70% of all cases of cystic fibrosis lies in the NBD1 domain.

second NBD [Riordan et al. (1989)]. These two nucleotide binding domains, which contain many of the sites that upon mutation give rise to CF, bind ATP and, in the case of NBD1, also hydrolyze ATP [Ko et al. (1995)]. The long exon 13 of the CF gene encodes a region of accentuated hydrophilicity bridging the N-terminal and C-terminal portions of CFTR. This region has a large number of alternating charged groups and most of the putative phosphorylation sites. Because phosphorylation by protein kinase A and protein kinase C activates the CFTR chloride channel, this region is called the "regulatory (R) domain" [Riordan et al. (1989)]. This domain makes CFTR unique within the ABC transporter family. All these observations are summarized in the simplified model of CFTR shown in the Figure 2.

Alignment of the Nucleotide Binding Domains of CFTR with Those of other ABC Transporters

As described in Methods we modeled as NBDs 244/254 amino acid regions of CFTR. The boundaries of these regions were determined using the hydropathy profile and sequence alignment of NBD1 and NBD2. The beginning of NBD1 was set at residue L441 (18 amino acids before motif A); the end of NBD2 was set at the CFTR carboxy terminal, L1480. Alignment of NBD1 and NBD2 was used to define the end of NBD1 at residue K684 and the beginning of NBD2 at L1227. Tables 1 & 2 show an automatic multiple sequence alignment of the two NBDs in CFTR, with non CFTR proteins found in the SwissProt database. Only proteins with more than 30% sequence identity with the target sequence were selected: this percentage is the threshold for structural similarity proposed by Sander et al. (1994). The two NBDs of CFTR have a high percentage of sequence identity with each other (33%), although, each NBD aligns with different regions of other ABC proteins (Tables 1 & 2). The alignment divides the NBDs of the ABC transporters into two classes (NBD1-like and NBD2-like), suggesting for each class a conserved distinctive function. Also, each alignment shows three common conserved regions: 1) a high identity initial region of 36 amino acids, which includes the motif A, positions 18–25 Tables 1 and 2; 2) the motif C (positions 108–113/120–125, Table 1 and 2 respectively); and 3) the motif B (sequence positions range 120–132/132–144).

Alignments in Tables 1 and 2 also show a slightly different conserved stretch of at least eight residues (region D) after position 60, ending in an aromatic residue (F[508/1296] in CFTR). The sequences in each NBD class are ([T/S]h[K/R/Q][E/D]Nh_2[F/Y]) in the NBD1-like class and Xh_2XQ[D/E/Q/N/R/K]h_3F in the NBD2-like class. Another conserved feature consist of a conserved glycine at position 40 in NBD1 (39 in NBD2) that is preceded or followed (NBD2 or NBD1) by a zone of high frequency of insertion. Also, a highly conserved motif is observed at positions 103–105 (underline in Table 1) or 115–117 (underline in Table 2): an acidic residue followed by a glycine (E/DXG) just before the motif C (underline in CFTR). No sequence identities are found after position 200, deep inside the exon 13 encoded segment. The alignment of a representative set of ABC transporters based on these considerations, shown in Tables 1 and 2, suggests that a model of both NBDs of CFTR can be generalized to other members of the ABC transporter family.

Topology of Nucleotide Binding Domains

Analysis of known three-dimensional structures of proteins involved in nucleotide binding can be used to extract a set of conserved topological features. Certainly, a model for the NBDs must contain these features or have an explanation for their absence. Alignment of the secondary structural elements that form the nucleotide binding cassette (NBC) of a repre-

Table I. Automatic alignment [Rost (1996)] of 244 residues of NBD1 sequence of CFTR with 32 other sequences of non-CFTR proteins (listed below) found in the SwissProt sequence database. The amino acids at the beginning and the end of an insertion are indicated with lower case letters, the intervening amino acids are omitted

Code in Table 1	Identity %	No aa	No. of Ins. plus del.	Size of Ins.	Prot. Size	Swissprot Database Code	Description
cftr_human	100	244	0	0	1480	P13569	Dependent Chloride Channel (CFTR)
yhd5_yeast	40	223	4	19	1592	P38735	Probable ATP-Dependent Permease
mrp1_human	39	221	1	2	1531	P33527	Multidrug Resistance-Associate Protein
ycfi_yeast	38	226	2	8	1515	P39109	Metal Resistance Protein
mdr1_enthi	36	97	2	17	114	P16875	Multidrug Resistance Protein
yawb_schpo	34	244	2	10	1478	Q10185	Probable ATP-Dependent Permease
yfib_bacsu	34	212	4	22	573	P54718	Hypothetical ABC Transporter
mdr_leita	34	242	1	2	1548	P21441	Multidrug Resistance Protein
cydd_hacin	34	204	6	31	586	P45082	Transport ATP-Binding Protein
sur_cricr	33	244	3	35	1581	Q09427	Sulfonylurea Receptor.
Yfic_bacsu	33	202	4	16	604	P54719	Hypothetical ABC Transporter
yor1_yeast	33	242	2	2	1477	P53049	Oligomycin Resistance ATP-Binding Protein
heta_anasp	33	212	4	22	607	P22638	Heterocyst Differentiation
sur_rat	32	244	3	39	1580	Q09429	Sulfonylurea Receptor
Sfuc_serma	32	204	6	40	345	P21410	Iron(III)-Transport ATP-Binding Protein
y015_mycge	32	199	4	19	589	P47261	Hypothetical ABC Transporter
yk83_yeast	32	240	4	19	1218	P36028	Probable ATP-Dependent Permease
cydc_ecoli	32	204	3	15	573	P23886	Transport ATP-Binding Protein
coma_strpn	32	201	4	19	717	Q03727	Transport ATP-Binding Protein
prtd_erwch	32	212	2	17	575	P23596	Proteases Secretion ATP-Binding Protein
y099_haein	32	219	6	33	356	P44513	Hypothetical ABC Transporter
nata_bacsu	31	200	7	27	246	P46903	ATP-Binding Transport Protein
mesd_leume	31	200	5	20	722	Q10418	Mesentericin Y105 Transporter
ybba_haein	31	178	5	23	227	P45247	Hypothetical ABC Transporter
mbpx_marpo	31	201	5	18	370	P10091	Probable Transport Protein
ndva_rhime	31	219	4	27	616	P18767	Beta-(1—>2)Glucan Export
uvra_serma	36	55	2	5	74	P25735	Excinuclease ABC Subunit
ywja_bacsu	30	201	3	15	575	P45861	Hypothetical ABC transporter
potg_ecoli	30	234	9	28	404	P31134	Putrescine Transport ATP-Binding Protein
nist_lacla	30	211	3	20	600	Q03203	Nisin Transport Atp-Binding Protein
mdl2_yeast	30	218	4	19	820	P33311	ATP-Dependent Permease MDR
hlyb_actac	30	201	4	19	707	P23702	Leukotoxin Secretion ATP-Binding Protein
spab_bacsu	30	218	3	20	614	P33116	Subtilin Transport ATP-Binding Protein

	1		|-----A-----|		50
cftr_human	LKDINFKIER	GQLLAVAGST	GAGKTSLLMM	IMGELEPSEG	KIKHSGRISF
yhd5_yeast	LCGLNIKFQI	GKLNLILGST	GSGKSALLLG	LLGELNLISG	SIIvtNSFAY
mrp1_human	LNGITFSIPE	GALVAVVGQV	GCGKLSLLSA	LLAEMDKVEG	HVAIKGSVAY
ycfi_yeast	LKNINFQAKK	GNLTCIVGKV	GSGKTALLSC	MLGDRFRVKG	FATVHGSVAY
mdr1_enthi	.....	.....S	GCGKSTTIQL	IQRNYEPNGG	RVTLDgqIGL
yawb_schpo	LRDIDFVARR	GELCCIVGKV	GMGKSSLLEA	CLGNMQKHSG	SVFRCGSIAY
yfib_bacsu	LRNVSFSAKP	RETIAILGAT	GSGKSTLFQL	IPRLYQPDSG	RIYirRQIGY
mdr_leita	LRNVSLTIPK	GKLTMVIGST	GSGKSTLLGA	LMGEYSVESG	ELWAERSIAY
cydd_haein	.KPLNFQIPA	NHNVALVGQS	GAGKTSLMNV	ILGFL.PYEG	SLKINghIAW
sur_cricr	LSNITIRIPR	GQLTMIVGQV	GCGKSSLLLA	TLGEMQKVSG	AVFwrGPVAY
yfic_bacsu	LKHLQFTVPA	GQSIAFVGPT	GAGKTTVTNL	LARFYEPNDG	KILIDgnMGF
yor1_yeast	FKDLNFDIKK	GEFIMITGPI	GTGKSSLLNA	MAGSMRKTDG	KVEVNGDLLM
heta_anasp	LNNITLTIER	GKTTALVGAS	GAGKTTLADL	IPRFYDPTEG	QILVDgkMAV
sur_rat	LSNITIRIPR	GQLTMIVGQV	GCGKSSLLLA	TLGEMQKVSG	AVFwrGPVAY
sfuc_serma	LEHIDLQVAA	GSRTAIVGPS	GSGKTTLLRI	IAGFEIPDGG	QILLQggIGF
y015_mycge	LTGINFSVKH	GDVVAIVGPT	GAGKSTIINL	LMKFYKPFEG	KIYmrEKISI
yk83_yeast	LKNISIDFKL	NSLNAIIGPT	GSGKSSLLLG	LLGELNLLSG	KIYvtNSMAY
cydc_ecoli	LKGISLQVNA	GEHIAILGRT	GCGKSTLLQQ	LTRAWDPQQG	EILlrQTISV

Table I. Continued.

coma_strpn	LSDINLTVPQ	GSKVAFVGIS	GSGKTTLAKM	MVNFYDPSQG	EISLGgyINY
prtd_erwch	LQNIHFSLQA	GETLVILGAS	GSGKSSLARL	LVGAQSPTQG	KVRLDgtIGY
y099_haein	LHDISFSLQR	GEILFLLGSS	GCGKTTLLRA	IAGFEQPSNG	EIWLKERLIF
nata_bacsu	VRDVSLTIEK	GEVVGILGEN	GAGKTTMLRM	IASLLEPSQG	VITVDggVLF
mesd_leume	IDDVSLTITA	GEKIALVGIS	GSGKSTLVKL	LVNFFQPESG	TISlrGHINY
ybba_haein	LKGVSFSMEP	AELVAIVGSS	GSGKSTLLHT	LGGLDQPSSG	EVFINgyLGF
mbpx_marpo	LDRVSLYVPK	FSLIALLGPS	GSGKSSLLRI	IAGLDNCDYG	NIWLHgrMSF
ndva_rhime	VRNVSFKAKA	GQTIAIVGPT	GAGKTTLVNL	LQRVHEPKHG	QILIDgsIAT
uvra_serma	LKNINLIIPR	DKLIVVTGLS	GSGKSSLAFD	TLY....AEG	QRRYVESLsy
ywja_bacsu	LNDINLSIQA	GETVAFVGPS	GAGKSTLCSL	LPRFYEASEG	DITirGQIGV
potg_ecoli	VDDVSLTIYK	GEIFALLGAS	GCGKSTLLRM	LAGFEQPSAG	QIMLDgpINM
nist_lacla	LKNINLSFEK	GELTAIVGKN	GSGKSTLVKI	ISGLYQPTMG	IIQyqKNISV
mdl2_yeast	FKNLNFKIAP	GSSVCIVGPS	GRGKSTIALL	LLRYYNPTTG	TITirRHIGI
hlyb_actac	LNNINLDISQ	GEVIGIVGRS	GSGKSTLTKL	IQRFYIPEQG	QVLIDgqVGV
spab_bacsu	LKHINVSLHK	GERVAIVGPN	GSGKKTFIKL	LTGLYEVHEG	DILINgqIAA
	51	|---D---|			100
cftr_human	CSQFSWIMPG	TIKENIIFGV	SYDEYRYRSV	IKACQLEEDI	SKFAEKDNIV
yhd5_yeast	CSQSAWLLND	TVKNNIIFDN	FYNEDRYNKV	IDACGLKRDL	EILPAGDLTE
mrp1_human	VPQQAWIQND	SLRENILFGC	QLEEPYYRSV	IQACALLPDL	EILPSGDRTE
ycfi_yeast	VSQVPWIMNG	TVKENILFGH	RYDAEFYEKT	IKACALTIDL	AILMDGDKTL
mdr1_enthi	VGQEPVLFAG	TIRENIMLGA	KEGETLSKDE	MIEcnAHEFV	SKLAEGYDTL
yawb_schpo	AAQQPWILNA	TIQENILFGL	ELDPEFYEKT	IRACCLLRDF	EILADGDQTE
yfib_bacsu	VPQEVLLFSG	TIKENIAWGK	ENASleIMDA	AKLAQIHETI	LKLPNGYDTV
mdr_leita	VPQQAWIMNA	TLRGNILFFD	EERAEDLQDV	IRCCQLEADL	AQFCGGLDTE
cydd_haein	VGQNPLLLQG	TIKENLLLG.	..DVQANDEE	INQALMRSQA	KEFTDKLGLh
sur_cricr	ASQKPWLLNA	TVEENITFES	PFNKQRYKMV	IEACSLQPDI	DILPHGDQTQ
yfic_bacsu	VLQDSFLFQG	TIRENIRYGR	LDASDqvEAA	AKTANAHSFI	ERLPKGYDTV
yor1_yeast	CG.YPWIQNA	SVRDNIIFGS	PFNKEKYDEV	VRVCSLKADL	DILPAGDMTE
heta_anasp	VSQDTFIFNT	SIRDNIAYGT	saSEAEIREV	ARLANALQFI	EEMPEGFDTK
sur_rat	ASQKPWLLNA	TVEENITFES	PFNKQRYKMV	IEACSLQPDI	DILPHGDQTQ
sfuc_serma	VPQDGALFpf	TVAGNIGFGL	KGGKREKQRR	IEA......L	MEMVALDRRL
y015_mycge	VLQDSFLFSG	TIKENIRLGR	..QDATDDEI	IAACKTadFI	MRLPKGYDTY
yk83_yeast	CSQTPWLISG	TIKDNVVFGE	IFNKQKFDDV	MKSCCLDKDI	KAMTAGIRTD
cydc_ecoli	VPQRVHLFSA	TLRDNLLLAS	PGSSDEALSE	ILRRVGLEKL	LEDAGL.NSW
coma_strpn	LPQQPYVFNG	TILENLLLGA	KEGTTqiLRA	VELAEIREDI	ERMPLNYQTE
prtd_erwch	LPQDVQLFKG	SLAENIARFG	DADPEKVVAA	AKLAGVHELI	LSLPNGYDTE
y099_haein	GENFnlFPHL	NVYRNIAYGL	GNGKGkeKTR	IEQIMQLTGI	FELADR....
nata_bacsu	GGETGLYDRM	TAKENLQYFG	RLYGLN.RHE	IKA.RIEDLS	KRFGMRDYMN
mesd_leume	LPQEPFIFSG	SIMENLLLGA	KPGTTQ.EDI	IRAVEIAedI	EKMSQGFGTE
ybba_haein	VYQFHHLMAd	tALENVMMPM	L...IGHQNK	TEAKDRAEKM	.LSAVGLSHR
mbpx_marpo	VFQHYALFkm	TVYENISFGL	RLRGFSAQKI	TNKVNDLLNC	LRIAD....I
ndva_rhime	VFQDAGLMNR	SIGENIRLGR	EDASleVMAA	AEAAAASDFI	EDRLNGYDTV
uvra_serma	ARQFLSLME.	.....	.....	.....	.....
ywja_bacsu	VQQDVFLFSG	TLRENIAYGR	laSEEDIWQA	VKQAHLEELV	HNMPDGLDTM
potg_ecoli	MFQSYALFpm	TVEQNIAFGL	KQDKLpiASR	VNEMLGLVHM	QEFAKR....
nist_lacla	LFQDFVKYEL	TIRENIGLSD	LSSQWEDEKI	IKVLdlKTNN	QYVLDTQLGN
mdl2_yeast	VQQEPVLMSG	TIRDNITYGL	TYTPTkiRSV	AKQCFCHNFI	TKFPNTYDTV
hlyb_actac	VLQDNVLLNR	SIRENIALTN	..PGMPMEKV	IAAAKLadFI	SELREGYNTV
spab_bacsu	LFQDFMKYEM	TLKENIGFGQ	IDKLHqvLDI	VRADFLKSHS	SYQFDTQLGL
	101 |-----C-----|		|-------------B-----------|		150
cftr_human	LGEGGITLSG	GQRARISLAR	AVYKDADLYL	LDSPFGYLDV	LTEKEIFESC
yhd5_yeast	IGEKGITLSG	GQKQRISLAR	AVYSSAKHVL	LDDCLSAVDS	HTAVWIYENC
mrp1_human	IGEKGVNLSG	GQKQRVSLAR	AVYSNADIYL	FDDPLSAVDA	HVGKHIFENV
ycfi_yeast	VGEKGISLSG	GQKARLSLAR	AVYARADTYL	LDDPLAAVDE	HVARHLIEHV
mdr1_enthi	IGEKGALLSG	GQRQRI....	.....	.....	.....
yawb_schpo	VGEKGISLSG	GQKARISLAR	AVYSRSDIYL	LDDILSAVDQ	HVNRDLVRNL
yfib_bacsu	LGQRGVNLSG	GQKQRISIAR	ALIRKPAILL	LDDSTSALDL	QTEAKLLEAI
mdr_leita	IGEMGVNLSG	GQKARVSLAR	AVYANRDVYL	LDDPLSALDA	HVGQRIVQDV

Table I. Continued.

cydd_haein	iKDGGLGISV	GQAQRLAIAR	ALLRKGDLLL	LDEPTASLDA	QSENLVLQA.
sur_cricr	IGERGINLSG	GQRQRISVAR	ALYQQTNVVF	LDDPFSALDV	HLSDHLMQAG
yfic_bacsu	LTQNGSGISQ	GQKQLISIAR	AVLADPVLLI	LDEATSNIDT	VTEVNIQEA.
yor1_yeast	IGERGITLSG	GQKARINLAR	SVYKKKDIYL	FDDVLSAVDS	RVGKHIMDEC
heta_anasp	LGDRGVRLSG	GQRQRIAIAR	ALLRDPEILI	LDEATSALDS	VSERLIQES.
sur_rat	IGERGINLSG	GQRPGISVAR	ALYQHTNVVF	LDDPFSALDV	HLSDHLMQAG
sfuc_serma	AALWPHELSG	GQQQRVALAR	ALSQQPRLML	LDEPFSALDT	GLRAATRKAV
y015_mycge	ISNKADYLSV	GERQLLTIAR	AVIRNAPVLL	LDEATSSVDV	HSEKLIQES.
yk83_yeast	VGDGGFSLSG	GQQQRIALAR	AIYSSSRYLI	LDDCLSAVDP	ETALYIYEEC
cydc_ecoli	LGEGGRQLSG	GELRRLAIAR	ALLHDAPLVL	LDEPTEGLDA	TTESQILEL.
coma_strpn	LTSDGAGISG	GQRQRIALAR	ALLTDAPVLI	LDEATSSLDI	LTEKRIVDN.
prtd_erwch	LGDGGGGLSG	GQRQRIGLAR	AMYGDPCLLI	LDEPNASLDS	EGDQALMQAI
y099_haein	...FPHQLSG	GQQQRVALAR	ALAPNPELIL	LDEPFSALDE	HLRQQIRQEM
nata_bacsu	RRVGG..FSK	GMRQKVAIAR	ALIHDPDIIL	FDEPTTGLDI	.TSSNIFREF
mesd_leume	LAESG.NISG	GQKQRIALAR	AILVDSPVLI	LDESTSNLDV	LTEKKIIDNL
ybba_haein	ITHRPSALSG	GERQRVAIAR	ALVNNPSLVL	ADEPTGNLDH	KTTESIFELI
mbpx_marpo	SFEYPAQLSG	GQKQRVALAR	SLAIQPDFLL	LDEPFGALDG	ELRRHLSKWL
ndva_rhime	VGERGNRLSG	GERQRVAIAR	AILKNAPILV	LDEATSALDV	ETEARVKDA.
uvra_serma	....	....	....	....	....
ywja_bacsu	IGERGVKLSG	GQKQRLSIAR	MFLKNPSILI	LDEATSALDT	ETEAAIQKA.
potg_ecoli	...KPHQLSG	GQRQRVALAR	SLAKRPKLLL	LDEPMGALDK	KLRDRMQLEV
nist_lacla	WFQEGHQLSG	GQWQKIALAR	TFFKKASIYI	LDEPSAALDP	VAEKEIFDYF
md12_yeast	IGPHGTLLSG	GQKQRIAIAR	ALIKKPTILI	LDEATSALDV	ESEGAINYT.
hlyb_actac	VGEQGAGLSG	GQRQRIAIAR	ALVNNPRILI	FDEATSALDY	ESENIIMHN.
spab_bacsu	WFDEGRQLSG	GQWQKIALAR	AYFREASLYI	LDEPSSALDP	IAEKETFDTF
	151				200
cftr_human	VCKLMANKTR	ILVTSKMEHL	KKADKILILH	EGSSYFYGTF	SELQNLQPDF
yhd5_yeast	ItpLMKNRTC	ILVTHNVstL	RNAHFAIVLE	NGKVKNQGTI	TELQS.KGLF
mrp1_human	IggMLKNKTR	ILVTHSMSYL	PQVDVIIVMS	GGKISEMGSY	QELLARDGAF
ycfi_yeast	LggLLHTKTK	VLATNKVSAL	SIADSIALLD	NGEITQQGTY	DEITKDADSP
mdr1_enthi	....	....	....	....	....
yawb_schpo	LggLLRSRCV	ILSTNSLTVL	KEASMIYMLR	NGKIIESGSF	TQLSslLSEF
yfib_bacsu	S...TYHCTT	LIITQKITTA	MKADQILLLE	DGELIEKGTH	SELLS.....
mdr_leita	ILGRLRGKTR	VLATHQIHLL	PLADYIVVLQ	HGSIVFAGDF	AAFSalEETL
cydd_haein	LNEASQHQTT	LMITHRIEDL	KQCDQIFVMQ	RGEIVQQGKF	TELQHE....
sur_cricr	ILELldKRTV	VLVTHKLQYL	PHADWIIAMK	DGTIQREGTL	KDFQRSECQL
yfic_bacsu	LARLMEGRTS	VIIAHRLNTI	QRADQIVVLK	NGEMIEKGSH	DELIR.QKGF
yor1_yeast	LTGMLANKTR	ILATHQLSLI	ERASRVIVLG	TDGQVDIGTV	DELKARNQTL
heta_anasp	IEKLSVGRTV	IAIAHRLSTI	AKADKVVVME	QGRIVEQGNY	QELLEQRGKL
sur_rat	ILELldKRTV	VLVTHKLQYL	PHADWIIAMK	DGTIQREGTL	KDFQRSECQL
sfuc_serma	AELLtaKVAS	ILVThqSEAL	SFADQVAVMR	SGRLAQVGAP	QDL.....
y015_mycge	IGRLMKNKTS	FIISHRLSII	RDATLIMVIN	DGKVLEMGNH	DQLMKQNGFY
yk83_yeast	LcpMMKGRTC	IITSHNISLv	kRADWLVILD	RGEVKSQGKP	SDLIKSN.EF
cydc_ecoli	LAEMMREKTV	LMVTHRLRGL	SRFQQIIVMD	NGQIIEQGTH	AELLARQGRY
coma_strpn	..LIALDKTL	IFIAHRLTIA	ERTEKVVVLD	QGKIVEEGKH	ADLLA.QGGF
prtd_erwch	VALQKRGATV	VLITHRPALT	TLAQKILILH	EGQQQRMGll	TELQQRSAAN
y099_haein	LQALrsGASA	IFVThrDESL	RYADKIAIIQ	QGKILQIDTP	RTLY.....W
nata_bacsu	IQQLKREQKT	ILFSSHieVQ	ALCDSVIMIH	SGEVIYRGAL	ESLYESEriF
mesd_leume	M..KLTEKTI	IFVAHRLTIS	QRVDRILTMQ	SGKIIEDGTH	DTLLKAGGFY
ybba_haein	QqnQEQNIAF	LLVTHDMGLA	EKLSRRLVMQ	DG.....	....
mbpx_marpo	KRYLQDNktT	IMVTHDqeAI	SMADEIVILK	EGRLLQQGKP	KNLYDQPINF
ndva_rhime	IDALRKDRTT	FIIAHRLSTV	READLVIFMD	QGRVVEMGGF	HELSQSNGRF
uvra_serma	....	....	....	....	....
ywja_bacsu	LQELSEGRTT	LVIAHRLATI	KDADRIVVVT	NNGIEEQGRH	QDLIEAGGLY
potg_ecoli	VDILevGVTC	VMVTHDQeaM	TMAGRIAIMN	RGKFVQIGEP	EEIYEHPtrY
nist_lacla	V.ALSENNIS	IFISHSLNAA	RKANKIVVMK	DGQVEDVGSH	DVLLRRCQYY
md12_yeast	FGQLMKSKSM	TIVSIAHRlr	RSENVIVLGH	DGSVVEMGKF	KELYANPTSA
hlyb_actac	MHKICQNRTV	LIIAHRLSTV	KNADRIIVMD	KGEIIEQGKH	QELLKDEKGL
spab_bacsu	F.SLSKDKIG	IFISHRLVAA	KLADRIIVMD	KGEIVGIGTH	EELLKTCPLY

Table I. Continued.

	201				244
cftr_human	SSKLMGCDSF	DQFSAERRNS	ILTETLHRFS	LEGDAPVSWT	ETKK
yhd5_yeast	KEKYVQLSSR	DSINEKNANR	LKAP.....	.....	....
mrp1_human	AEFLRTYAST	EQEQDAEENG	V.....	.....	....
ycfi_yeast	LWKLLNngKS	NEFGDSSESS	VRESSI....	.....	....
mdr1_enthi	.....	.....	.....	.....	....
yawb_schpo	SKKDTASSTG	ADTPLSRSQS	VITSSTDVTS	SASRSSDTVS	NYPK
yfib_bacsu	ESQLYKRIYE	SQFGREGSES	.....	.....	....
mdr_leita	RGELKGSKDV	ESCSSDVDTE	SATAETAPYV	AKAKGLNAEQ	ET..
cydd_haein	.....GFF	AELLAQRQQD	I.....	.....	....
sur_cricr	FEHWKTLMNR	QDQemERKAS	EPSQGLPRAM	SSRDGLLLDE	EEEE
yfic_bacsu	YSDL.....	.....	.....	.....	....
yor1_yeast	INLLQFSSQN	SEKEDEEQEA	VVAGELGQLK	YESEVK.ELT	ELKK
heta_anasp	W.....KY	HQMQHESGQT	.....	.....	....
sur_rat	FeeLEKETVM	ERKAPEPSQG	LPRAMSSRDG	LLLDEDEEEE	EAAE
sfuc_serma	.....Y	LRPVDEPTAS	FLGETL....	.....	....
y015_mycge	AR.....	.....	.....	.....	....
yk83_yeast	LRESINNDSK	NTTHNQIDLK	RSTTSKKTKN	GDPEGGNSQD	E...
cydc_ecoli	YQFKQG....	.....	.....	.....	....
coma_strpn	YAHLV.....	.....	.....	.....	....
prtd_erwch	QARMNPTAAM	PQ.....	.....	.....	....
y099_haein	SPNHLETAKF	MGESIVLPAN	LLDENTAQCQ	L.....	....
nata_bacsu	MSKLV.....	.....	.....	.....	....
mesd_leume	ASLF.....	.....	.....	.....	....
ybba_haein	.....	.....	.....	.....	....
mbpx_marpo	FVGIF.....	.....	.....	.....	....
ndva_rhime	AALL.....	.....RASG	ILTDEDVRKS	LT.....	....
uvra_serma	.....	.....	.....	.....	....
ywja_bacsu	SR.....	.....	.....	.....	....
potg_ecoli	SAEFIGsnVF	EGVLKERqgL	VLDSPGLVHP	LKVDADASVV	D...
nist_lacla	QELYYSEQYE	DN.....	.....	.....	....
md12_yeast	LSQLLNEKAA	PGPSDQQLQ.	.....	.....	....
hlyb_actac	YSYL.....	.....	.....	.....	....
spab_bacsu	KKMDESENYM	NPLEEEGSK.	.....	.....	....

sentative set of nucleotide binding proteins (Table 3) shows a strong degree of conservation of topological motifs. With the increased number of reported structures of NB proteins it is now possible to cluster these structures in at least three topologically related types: (1) mitochondrial F_1-ATPase/recA, (2) ADK and (3) p21/G-proteins (Figures 3a,b,c). They all share an α/β motif (known as the Rossman fold; [Rossman et al. (1974)]), consisting of 5 to 7, predominantly parallel β-strands, surrounded by 3 to 7 helices. The topology of these proteins (Figure 3a,b,c) shows strong conservation of a central motif, composed of a β-strand followed by an α-helix, a β-strand that can be far away in the sequence, and an additional α-helix and β-strand (β1αA . . . β4 . . . αFβ6, boxed in Figure 3a). Topological insertions or changes are concentrated in the regions between helix αA and strand β4, and between β4 and αF. This conserved core motif contains two important consensus sequences $GX_4GK[T/S]$, motif A (also known as the "P-loop"), and $RX_{6-8}h_4D$, motif B. The P-loop forms the binding site for triphosphate (or diphosphate) nucleotides and for Mg^{++}. Strand β1, loop (GX_4G) and helix αA (Table 3 and Figures 3) define the motif containing the P-loop.

Motif B is located in strand β4 and ends with a conserved negative charged residue (aspartate, D) . The third consensus signature, motif C, is not equally general; for example it is not present in the F_1/recA NBD. In the case of motifs A and B, not only the secondary structure is conserved, but also their spatial arrangement (Figure 4). Clearly, any structural model of a protein containing a nucleotide binding cassette must account for the facts described above.

We have chosen a member of the first topological class (Figure 3a) as a template for modeling NBDs. F_1-ATPase is a five subunit protein with stoichiometry

Table II. Automatic alignment [Rost (1996)] of 250 residues of NBD2 sequence of CFTR with 25 other sequences of non-CFTR proteins (listed below) found in the SwissProt sequence database. The amino acids at the beginning and the end of an insertion are indicated with lower case letters, the intervening amino acids are omitted

Code in Table 1	Identity %	No aa	No. of Ins. plus del.	Size of Ins.	Prot. Size	Swissprot Database Code	Probable Transport ATP-Binding Protein
cftr_human	100	250	0	0	1480	P13569	Dependent Chloride Channel (CFTR)
mrp1_human	38	221	1	1	1531	P33527	Multidrug Resistance-Associate
yawb_schpo	38	221	1	1	1478	Q10185	Probable ATP-Dependent Permease
ycfi_yeast	37	222	3	3	1515	P39109	Metal Resistance Protein
sur_human	36	220	1	1	395	Q09428	Sulfonylurea Receptor
yor1_yeast	36	220	4	23	1477	P53049	Oligomycin Resistance ATP
sur_rat	36	220	1	1	1580	Q09429	Sulfonylurea Receptor.
sur_cricr	35	220	1	1	1581	Q09427	Sulfonylurea Receptor.
yk84_yeast	35	230	3	22	306	P36171	Probable ATP-Dependent Transporter
yhd5_yeast	34	231	4	19	1592	P38735	Probable ATP-Dependent Permease
msba_haein	33	216	2	3	587	P44407	Probable Transport ATP-Binding Protein
lcnc_lacla	33	211	3	5	715	Q00564	Lactococcin A Transport
yfib_bacsu	32	225	4	9	573	P54718	Hypothetical ABC Transporter
ste6_yeast	32	217	3	5	1290	P12866	Protein Homolog) (P-Glycoprotein)
cydc_ecoli	32	211	3	4	573	P23886	Transport ATP-Binding Protein
ywja_bacsu	32	218	2	2	575	P45861	Hypothetical ABC Transporter
cydd_haein	32	203	3	7	586	P45082	Transport ATP-Binding Protein
mdr_leita	32	241	3	10	1548	P21441	Multidrug Resistance Protein
cyab_borpe	31	213	2	2	712	P18770	Cyclolysin Secretion ATP-Binding
msba_ecoli	31	213	2	3	582	P27299	Probable Transport ATP-Binding
heta_anasp	31	213	2	2	607	P22638	Heterocyst Differentiation
nist_lacla	31	219	3	7	600	Q03203	Nisin Transport ATP-Binding
cydc_haein	30	218	2	3	576	P45081	Transport ATP-Binding Protein
yabj_haein	30	182	7	25	229	P44986	Hypothetical ABC Transporter
hst6_canal	30	213	3	51	1323	P53706	ATP-Dependent Permease HS
yfic_bacsu	30	218	2	2	604	P54719	Hypothetical ABC Transporter

	1		|-----A-------|		50
cftr_human	LENISFSISP	GQRVGLLGRT	GSGKSTLLSA	FLRLLNTEGE	IQIDGVSWDS
mrp1_human	LRHINVTING	GEKVGIVGRT	GAGKSSLTLG	LFRINEseGE	IIIDGINIAK
yawb_schpo	LNDISVNIKP	QEKIGIVGRT	GAGKSTLTLA	LFRLIetSGD	IQLDDINITS
ycfi_yeast	LKHINIHIKP	NEKVGIVGRT	GAGKSSLTLA	LFRMIEaeGN	IVIDNIAINE
sur_human	LKHVNALISP	GQKIGICGRT	GSGKSSFSLA	FFRMVDteGH	IIIDGIDIAK
yor1_yeast	LKNLNLNIKS	GEKIGICGRT	GAGKSTIMSA	LYRLNetAGK	ILIDNVDISQ
sur_rat	LKHVNALISP	GQKIGICGRT	GSGKSSFSLA	FFRMVDmeGR	IIIDGIDIAK
sur_cricr	LKHVNTLISP	GQKIGICGRT	GSGKSSFSLA	FFRMVDmeGR	IIIDGIDIAK
yk84_yeast	LDNVSFKVKA	GTKVGIVGRT	GAGKSSIIAA	IYRLSDwnGT	ITIDNKDIKH
yhd5_yeast	IRNVSFKVDP	QSKIGIVGRT	GAGKSTIITA	LFRLLEptGC	IKIDGQDISK
msba_haein	LNNISFSVPA	GKTVALVGRS	GSGKSTIANL	VTRFYDiqGE	ILLDGVNIQD
lcnc_lacla	LSEIELSIKE	NEKLTIVGMS	GSGKSTLVKL	LVNFFqtSGT	ITLGGIDLQQ
yfib_bacsu	LRNVSFSAKP	RETIAILGAT	GSGKSTLFQL	IPRLYQpsGR	IYIDEKPVQD
ste6_yeast	.KNMNFDMFC	GQTLGIIGES	GTGKSTLVLL	LTKLYNCegK	IKIDGTDVND
cydc_ecoli	LKGISLQVNA	GEHIAILGRT	GCGKSTLLQQ	LTRAWDpqGE	ILLNDSPIAS
ywja_bacsu	LNDINLSIQA	GETVAFVGPS	GAGKSTLCSL	LPRFYEaeGD	ITIDGISIKD
cydd_haein	...LNFQIPA	NHNVALVGQS	GAGKTSLMNV	ILGFLPYEGS	LKINGQELRE
mdr_leita	LRGVSFQIAP	REKVGIVGRT	GSGKSTLLLT	FMRMVEvgGV	IHVNGREMSA
cyab_borpe	LRNVSLRIAP	GEVVGVVGRS	GSGKSTLTRL	IQRMFVarGR	VLIDGHDIGI
msba_ecoli	LRNINLKIPA	GKTVALVGRS	GSGKSTIASL	ITRFYDieGE	ILMDGHDLRE
heta_anasp	LNNITLTIER	GKTTALVGAS	GAGKTTLADL	IPRFYdtEGQ	ILVDGLDVQY
nist_lacla	LKNINLSFEK	GELTAIVGKN	GSGKSTLVKI	ISGLYqtMGI	IQYDKMRSSL
cydc_haein	LKNLTLDLEQ	GKKIAILGKT	GSGKSSLLQL	LVRNYDaqGE	LLLAEKPISA
yabj_haein	....NLSVNA	GERVAIIGES	GAGKSTLLNL	IAGFEFpqGE	IWLNDKNH..

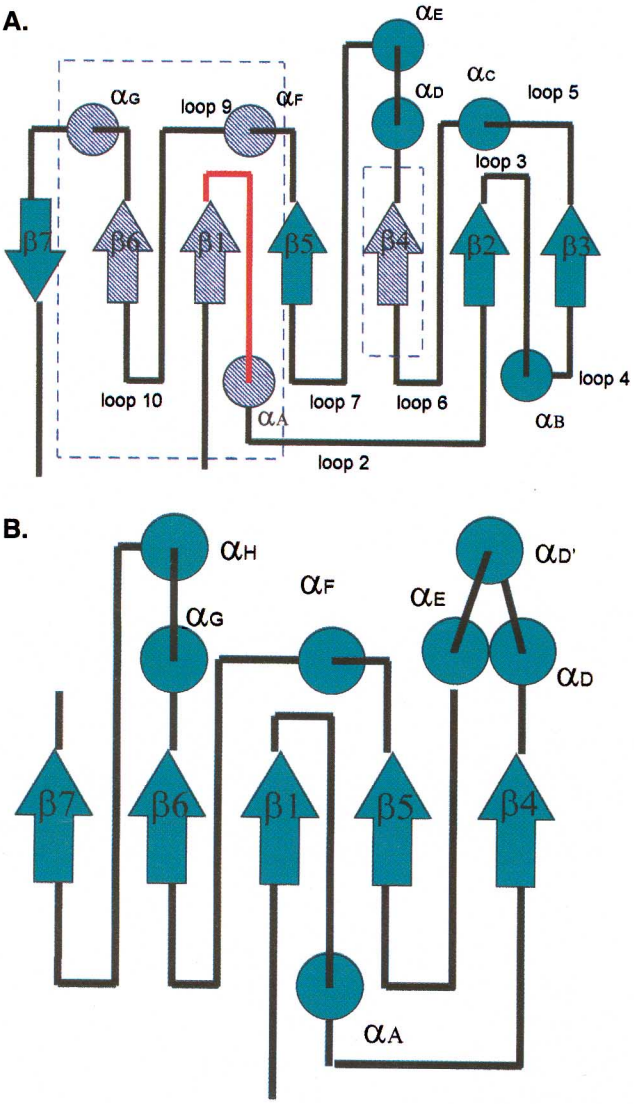
Table II. Continued.

hst6_canal	LKSISLDVKK	FTTIGIVGQS	GSGKSTILKI	LFRLYDidQT	VKIFNQNLYL
yfic_bacsu	LKHLQFTVPA	GQSIAFVGPT	GAGKTTVTNL	LARFYEpdGK	ILIDGTDIKT
	51	|----D----|			100
cftr_human	ITLQQWRKAF	GVIPQKVFIF	SGTFRKNLDP	YEQWSDQEIW	KVADEVGLRS
mrp1_human	IGLHDLRFKI	TIIPQDPVLF	SGSLRMNLDP	FSQYSDEEVW	TSLELAHLKD
yawb_schpo	IGLHDLRSRL	AIIPQENQAF	EGTIRENLDP	NANATDEEIW	HALEAASLKQ
ycfi_yeast	IGLYDLRHKL	SIIPQDSQVF	EGTVRENIDP	INQYTDEAIW	RALELSHLKe
sur_human	LPLHTLRSRL	SIILQDPVLF	SGTIRFNLDP	ERKCSDSTLW	EALEIAQLKL
yor1_yeast	LGLFDLRRKL	AIIPQDPVLF	RGTIRKNLDP	FNERTDDELw	qKPDENGTHG
sur_rat	LPLHTLRSRL	SIILQDPVLF	SGTIRFNLDP	EKKCSDSTLW	EALEIAQLKL
sur_cricr	LPLHTLRSRL	SIILQDPVLF	SGTIRFNLDP	EKKCSDSTLW	EALEIAQLKL
yk84_yeast	IPLERLRNSI	SCIPQDPTLF	DGTVRSNLDP	FDRYSDVQIY	GVLSKVGLIe
yhd5_yeast	IDLVTLRRSI	TIIPQDPILF	AGTIKSNVDP	YDEYDEKKIF	KALSQVNLIS
msba_haein	YRLSNLRENC	AVVSQQVHLF	NDTIANNIAY	AAqySREEII	AAAKAAYALE
lcnc_lacla	FDKHQLRRLI	NYLPQQPYIF	TGSILDNLLl	nENASQEEIL	KAVELAEIRA
yfib_bacsu	IPAEGLRRQI	GYVPQEVLLF	SGTIKENIAw	kENASLDEIM	DAAKLAQIHE
ste6_yeast	WNLTSLRKEI	SVVEQKPLLF	NGTIRDNLTy	qDEILEIEMY	DALKYVGIHD
cydc_ecoli	LNEAALRQTI	SVVPQRVHLF	SATLRDNLLl	sPGSSDEALS	EILRRVGLEK
ywja_bacsu	MTLSSLRGQI	GVVQQDVFLF	SGTLRENIAY	GRlaSEEDIW	QAVKQAHLEE
cydd_haein	SNLADWRKHI	AWVGQNPLLL	QGTIKENLLL	GdqANDEEI.	...NQALMRS
mdr_leita	YGLRDVRRHF	SMIPQDPVLF	DGTVRQNVDP	FLEASSAEVW	AALELVGLRE
cyab_borpe	VDSASLRRQL	GVVLQESTLF	NRSVRDNIAl	rPGASMHEVV	AAARLAGAHE
msba_ecoli	YTLASLRNQV	ALVSQNVHLF	NDTVANNIAY	aeQYSREQIE	EAARMAYAMD
heta_anasp	FEINSLRRKM	AVVSQDTFIF	NTSIRDNIAY	GTsaSEAEIR	EVARLANALQ
nist_lacla	MPEEFYQKNI	SVLFQDFVKY	ELTIRENIGL	ssQWEDEKII	KVLDNLGLDF
cydc_haein	YSEETLRHQI	CFLTQRVHVF	SDTLRQNLQF	adKISDEQMI	EMLHQVGLSK
yabj_haein	TRSAPYERPV	SMLFQENNLF	plTVQQNLAl	lTALEQEKIE	QVACSVGLGD
hst6_canal	INSGLLCQTI	AIVPQFPKFF	SGTIYDNLTy	sSSVSDSEII	KILKLVNLHQ
yfic_bacsu	LTRASLRKNM	GFVLQDSFLF	QGTIRENIRy	rLDASDQEVE	AAAKTANAHS
	101	|--C--------|		|-----B-------------|	150
cftr_human	VIEQFPGKLD	FVLVDGGCVL	SHGHKQLMCL	ARSVLSKAKI	LLLDEPSAHL
mrp1_human	FVSALPDKLD	HECAEGGENL	SVGQRQLVCL	ARALLRKTKI	LVLDEATAAV
yawb_schpo	FIQTLDGGLY	SRVTEGGANL	SSGQRQLMCL	TRALLTPTRV	LLLDEATAAV
ycfi_yeast	vLSMSNDGLD	AQLTEGGGNL	SVGQRQLLCL	ARAMLVPSKI	LVLDEATAAV
sur_human	VVKALPGGLD	AIITEGGENF	SQGQRQLFCL	ARAFVRKTSI	FIMDEATASI
yor1_yeast	KMHKF..HLD	QAVEEEGSNF	SLGERQLLAL	TRALVRQSKI	LILDEATSSV
sur_rat	VVKALPGGLD	AIITEGGENF	SQGQRQLFCL	ARAFVRKTSI	FIMDEATASI
sur_cricr	VVKALPGGLD	AIITEGGENF	SQGQRQLFCL	ARAFVRKTSI	FIMDEATASI
yk84_yeast	kLRNRFIDLN	TVVKSGGSNL	SQGQRQLLCL	ARSMLGARNI	MLIDEATASI
yhd5_yeast	SHEFEEvnLH	TEIAEGGLNL	SQGERQLLFI	ARSLLREPKI	ILLDEATSSI
msba_haein	FIEKLPQVFD	TVIGENGTSL	SGGQRQRLAI	ARALLRNSPV	LILDEATSAL
lcnc_lacla	DIEQMQLGYQ	TELSSDASSL	SGGQKQRIAL	ARALLSPAKI	LILDEATSNL
yfib_bacsu	TILKLPNGYD	TVLGQRGVNL	SGGQKQRISI	ARALIRKPAI	LLLDDSTSAL
ste6_yeast	FVISSPQGLD	TRI..DTTLL	SGGQAQRLCI	ARALLRKSKI	LILDECTSAL
cydc_ecoli	LLE..DAGLN	SWLGEGGRQL	SGGELRRLAI	ARALLHDAPL	VLLDEPTEGL
ywja_bacsu	LVHNMPDGLD	TMIGERGVKL	SGGQKQRLSI	ARMFLKNPSI	LILDEATSAL
cydd_haein	QAKEFTDKLG	leIKDGGLGI	SVGQAQRLAI	ARALLRKGDL	LLLDEPTASL
mdr_leita	RVASESEGID	SRVLEGGSNY	SVGQRQLMCM	ARALLKrsGF	ILMDEATANI
cyab_borpe	FICQLPEGYD	TMLGENGVGL	SGGQRQRIGI	ARALIHRPRV	LILDEATSAL
msba_ecoli	FINKMDNGLD	TVIGENGVLL	SGGQRQRIAI	ARALLRDSPI	LILDEATSAL
heta_anasp	FIEEMPEGFD	TKLGDRGVRL	SGGQRQRIAI	ARALLRDPEI	LILDEATSAL
nist_lacla	LKTNNQYVLD	TQlfQEGHQL	SGGQWQKIAL	ARTFFKKASI	YILDEPSAAL
cydc_haein	LLEQEGKGLN	LWLGDGGRPL	SGGEQRRLGL	ARILLNNASI	LLLDEPTEGL
yabj_haein	YLERLP....	.......NSL	SGGQKQRVAL	ARCLLRDKPI	LLLDEPFSAL
hst6_canal	FIVSLPQGLL	TIMNDSDntF	SGGQLQLLAI	ARALLRNPKI	LLLDECTSNL
yfic_bacsu	FIERLPKGYD	TVLTQNGSGI	SQGQKQLISI	ARAVLADPVL	LILDEATSNI

Table II. Continued.

	151				200
cftr_human	DPVTYQIIRR	TLKQAFADCT	VILCEHRIEA	MLECQQFLVI	EENKVRQYDS
mrp1_human	DLETDDLIQS	TIRTQFEDCT	VLTIAHRLNT	IMDYTRVIVL	DKGEIQEYGA
yawb_schpo	DVETDAIVQR	TIRERFNDRT	ILTIAHRINT	VMDSNRILVL	DHGKVVEFDS
ycfi_yeast	DVETDKVVQE	TIRTAFKDRT	ILTIAHRLNT	IMDSDRIIVL	DNGKVAEFDS
sur_human	DMATENILQK	VVMTAFADRT	VVTIAHRVHT	ILSADLVIVL	KRGAILEFDK
yor1_yeast	DYETDGKIQT	RIVEEFGDCT	ILCIAHRLKT	IVNYDRILVL	EKGEVAEFDT
sur_rat	DMATENILQK	VVMTAFADRT	VVTIAHRVHT	ILSADLVMVL	KRGAILEFDK
sur_cricr	DMATENILQK	VVMTAFADRT	VVTIAHRVHT	ILSADLVMVL	KRGAILEFDK
yk84_yeast	DYISDAKIQK	TIRETMKNTT	ILTIAHRLRS	VIDYDKILVM	EMGRVKEYDH
yhd5_yeast	DYDSDHLIQG	IIRSEFNKST	ILTIAHRLRS	VIDYDRIIVM	DAGEVKEYDR
msba_haein	DTESERAIQS	ALEELKKDRT	VVVIAHRLST	IENADEILVI	DHGEIRERGN
lcnc_lacla	DMITEKKILK	NLLP..LDKT	IIFIAHRLSV	AEMSHRIIVV	DQGKVIESGS
yfib_bacsu	DLQTEAKLLE	AIST..YHCT	TLIITQKITT	AMKADQILLL	EDGELIEKGT
ste6_yeast	DSVSSSIINE	IVKKGPPALL	TMVITHSEQM	MRSCNSIAVL	KDGKVVERGN
cydc_ecoli	DATTESQILE	LLAEMMREKT	VLMVTHRLRG	LSRFQQIIVM	DNGQIIEQGT
ywja_bacsu	DTETEAAIQK	ALQELSEGRT	TLVIAHRLAT	IKDADRIVVV	TNNGIEEQGR
cydd_haein	DAQSENLVLQ	ALNEASQHQT	TLMITHRIED	LKQCDQIFVM	QRGEIVQQGK
mdr_leita	DPALDRQIQA	TVMSAFSAYT	VITIAHRLHT	VAQYDKIIVM	DHGVVAEMGS
cyab_borpe	DYESEHIIQR	NMRDICDGRT	VIIIAHRLSA	VRCADRIVVM	EGGEVAECGS
msba_ecoli	DTESERAIQA	ALDELQKNRT	SLVIAHRLST	IEKADEIVVV	EDGVIVERGT
heta_anasp	DSVSERLIQE	SIEKLSVGRT	VIAIAHRLST	IAKADKVVVM	EQGRIVEQGN
nist_lacla	DPVAEKEIFD	YFVALSENNI	SIFISHSLNA	ARKANKIVVM	KDGQVEDVGS
cydc_haein	DRETERQILR	LILQHAENKT	LIIVTHRLSS	IEQFDKICVI	DNGRLIEEGD
yabj_haein	DQKLRVEMLA	LIAKleKDLT	LLLVTHQPSE	LisIDQVLVV	ENGQISQLQ.
hst6_canal	DPITTKIIIN	VIKSLHGKLT	ILFVTHDKEL	MRIADNLIVM	KDGQIVEQGD
yfic_bacsu	DTVTEVNIQE	ALARLMEGRT	SVIIAHRLNT	IQRADQIVVL	KNGEMIEKGS

	201				250
cftr_human	IQKLLNERSL	FRQAISPSDR	VKLFPHRNSS	KCKSKPQIAA	LKEETEEEVQ
mrp1_human	PSDLLQQRGL	FYSMAKDAGL	V.....	.....	.....
yawb_schpo	TKKLLENKAS	LFYSLAKESG	L.....	.....	.....
ycfi_yeast	PGQLLSdkSL	FYSLCMEAGL	VN.....	.....	.....
sur_human	PEKLLSRKDS	VFASFVRADK	.....	.....	.....
yor1_yeast	PWTLFSQesI	FRSMCSRSGI	VE.....	.....	.....
sur_rat	PEKLLSQKDS	VFASFVRADK	.....	.....	.....
sur_cricr	PETLLSQKDS	VFASFVRADK	.....	.....	.....
yk84_yeast	PYTLISDRnf	YRLCRQSGEF	ENLFELAKVS	.....	.....
yhd5_yeast	PSELLkeRGI	FYSMCRDSGG	LELLKqkQSS	K.....	.....
msba_haein	HKTLLEQNGA	YKQLHS....	.....	.....	.....
lcnc_lacla	HVDLLAQNGF	YEQ.....	.....	.....	.....
yfib_bacsu	HSELLSESQL	YKRIYE....	.SQFGREGSE	SC.....	.....
ste6_yeast	FDTLYNNRGE	LFQIVSNQSS	.....	.....	.....
cydc_ecoli	HAELLARQGR	YYQ.....	.....	.....	.....
ywja_bacsu	HQDLIEAGGL	YSRLHQAQ..	.....	.....	.....
cydd_haein	FTELQHEGFF	.....	.....	.....	.....
mdr_leita	PRELVMNHQS	MFHS.....	..MVESLGSR	GSKDFYELLM	GRRIVQPAV.
cyab_borpe	HETLLAAGGL	YAR.....	.....	.....	.....
msba_ecoli	HNDLLEHRGV	YAQ.....	.....	.....	.....
heta_anasp	YQELLEQRGK	LWK.....	.....	.....	.....
nist_lacla	HDVLLRRCQY	YQELYYSEQ.	.....	.....	.....
cydc_haein	YNSLITKENG	FFKRLIER..	.....	.....	.....
yabj_haein	.....	.....	.....	.....	.....
hst6_canal	FQQLISNDGE	FTK.....	.....	.....	.....
yfic_bacsu	HDELIRQKGF	YSDLYESQ..	.....	.....	.....



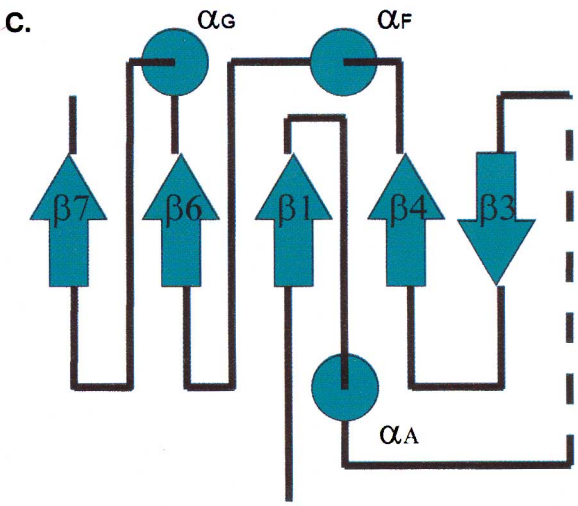


Fig. 3. *Diagrams of the three different types of topologies found in Nucleotide Binding Cassette proteins.* a) Mitochondrial F_1-ATPases/recA fold. b) Adenylate Kinase fold; c) Ras P21/EF-Tu/G-Protein fold. In part a of the figure, conserved structural elements common to all three proteins are boxed and striped filled and enclosed in the box; also, the P-loop is highlighted in red. In part c of the figure the dashed line shows a zone where insertions are present; examples of such insertions are: the large α-helical domain (6 helices) in heterotrimeric G-Proteins, or a "downward" α-helix in myosin.

$\alpha_3\beta_3\gamma\delta\epsilon$. The larger subunits α and β have sequence similarities, and share the same overall fold. The fold can be divided into three domains: a N-terminal beta barrel domain, a nucleotide binding domain, and a carboxy-terminal helical domain. The nucleotide binding domain of this protein was used as a template.

Sequence Alignment and Modeling of the CFTR Nucleotide Binding Domains

The first step (and probably the most important) in successful modeling is the correct alignment of the target and templates sequences. Figure 5 shows an alignment of the sequences of CFTR NBD1 and NBD2, with $F_1\alpha$- and $F_1\beta$-subunits respectively. This alignment takes into account conserved structural features in the NBC present in the proteins used as templates. Insertions and deletions in the NBD sequences with respect to the templates must be placed in regions where the F_1/recA fold can accommodate such changes with little structural disturbance. The first 40 residues in the alignment of both NBDs, which includes motif A, shows strong similarity with the same region of the NBDs of F_1. The first large insertion appears after this region, at the end of αA (Figure 3a; residues 482–494 in NBD1 and 1267–1280 in NBD2). This insertion correlates with the high frequency of insertions observed in this zone (positions from 44–48/38–40 in Table 1 and 2 respectively) in ABC transporters. This insertion is compatible with the F_1/recA fold which has a variable size loop 2 (Figure 3a) in that region. Strand β2 in the core of the fold, following the insertion, aligns well with the hydrophobic regions 495–503 of NBD1 (1284–1290 of NBD2). The last large insertion is observed in NBD2 before the region that aligns with the strand β3. This hydrophilic stretch of 14 residues forms a topological insertion which is probably exposed. The equivalent region in F_1 subunits is mainly hydrophobic, residing at the interface between

Table III. Alignment of the main secondary elements (α helices or β strands) found in a representative set of Nucleotide Binding proteins. The minus sign indicates anti-parallel β strands

Protein	Nucleotide Binding Domain Topology						Residue range
F_1-ATPase	β1αAβ2αB	β3αC	β4αD	αE β5	αFβ6αG	(−β7)	**150–336**
recA	β1αAβ2αB	β3αC	β4	αE β5	αFβ6	(−β7)	**149–326**
ADK	β1αA		β4 αDαD′	αE β5	αFβ6αGαH	β7	**9–194**
Ras p21	β1αA	(−β3)	β4		αFβ6αG	β7	**1–146**
EF-tu	β1αA	(−β3)	β4		αFβ6αG	β7	**99–173**
Transducin	β1αA (α)[6]	(−β3)	β4		αFβ6αG	β7	**59–323**
Myosin	β1αA (−αB)	(−β3)	β4 (α)[8]	αE β5	αFβ6αG		**172–680**

the NBD and the β-barrel domain that forms the N-terminal domain. Inspection of the sequence alignments of a representative group of ABC transporters proteins (Tables 1 & 2) suggests the presence at position 64 of the fold (504 of CFTR) of a highly conserved negatively charged residue preferentially a glutamic acid followed by a conserved asparagine (in region D), placed between motifs A and B. Mutations such as E504Q do cause CF. Sequence analysis (Fig. 5) aligns this region with an equivalent region of F_1: loop 3 and helix αB (Fig. 3a). This suggests that glutamic 504 of CFTR is equivalent to glutamic 188 of the β subunits of F_1-ATPase. This charged residue is thought to be the general base for the activation of the water molecule that attacks the γ-phosphate during ATP hydrolysis by F_1 [Abrahams et al. (1994)]. In the α-subunit of F_1, which is not active in ATP hydrolysis, a glutamine residue (Q208) is present in this position. An equivalent residue, Q1291, is present in NBD2 and highly conserved in the NBD2-like class (Table 2), suggesting a role for NBD2 similar to that of the F_1 α-subunit. Based on these and other considerations, NBD1 was modeled based on $F_1\beta$ and NBD2 based on $F_1\alpha$. The sequence of loop 5 of F_1 (EPP or DAA in α and β-subunits) can be aligned with conserved residues [E/D]XG, just before motif C, which is highly conserved in ABC transporters (Tables 1 & 2). A sequence analogous to motif C seems to be absent in F_1.

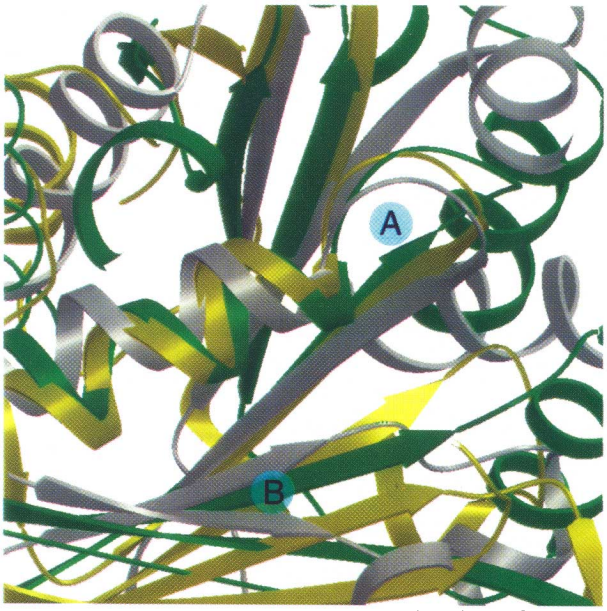


Fig. 4. *View of a spatial alignment of Nucleotide Binding (NB) cassettes of a representative set of NB proteins:* ADK is shown in silver, P21 in gold and recA in green.

Three-dimensional Modeling of the CFTR Nucleotide Binding Domains

The detailed three-dimensional models of the NBDs of CFTR are presented in Figures 6 (NBD1) and 7 (NBD2). Figures 8 and 9 show the secondary structure predictions [Chou & Fasman (1978)] in comparison with the suggested models based on the F_1/recA topology. Only two insertions, positions 44–53 in NBD1 and position 90 in NBD2, required careful adjustment. The structural alignment of F_1 subunits α and β to each other was based on the structure of rat liver F_1-ATPase [Bianchet et al. (1998)]. The two NBD structural models are similar to each other as expected from the similarity of their templates, the β and α subunits of F_1. The starting motif, N-terminus—β1-αA, was predicted to be similar in size to that of F_1. The insertion found in ABC transporters (after position 44 and 38 in Tables 1 and 2 respectively) was modeled as a longer αA, a loop and a short anti-parallel β-strand in the NBD1. A loop containing the D region

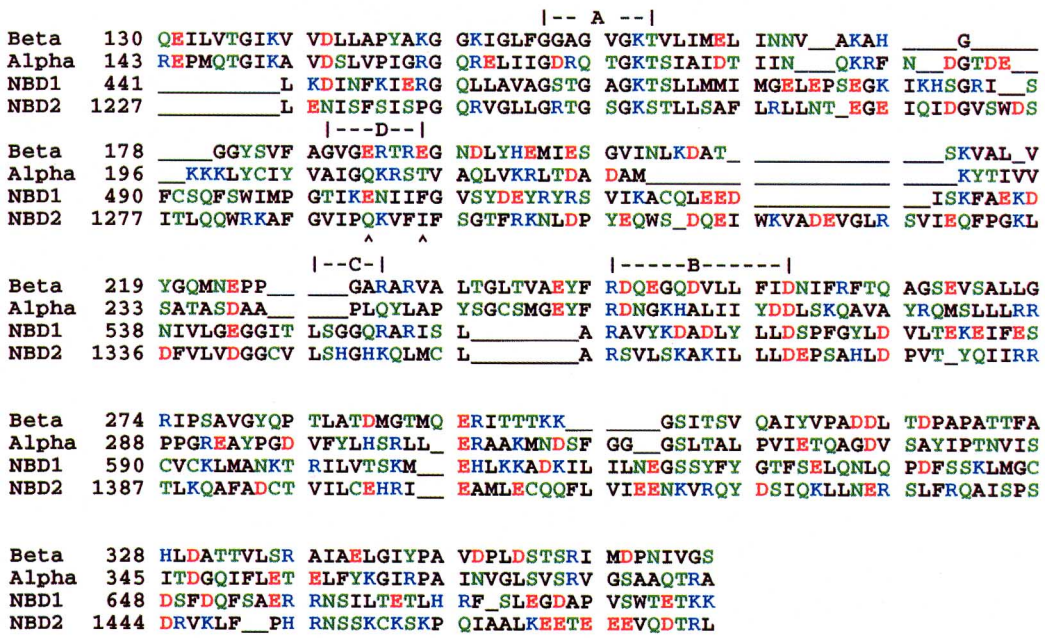


Fig. 5. *Alignment of the NBD of CFTR with α- and β-subunits of rat Liver F_1-ATPase.* Residues are represented by a single letter code, colored by hydropathy; black, hydrophobic; red, acidic hydrophilic; green, neutral hydrophilic; and blue, basic hydrophilic. Walker motifs A and B and regions C and D are indicated by brackets above the sequences. The positions of the proposed catalytic base and of F508 of CFTR are indicated by arrows below the sequences. The sequence alignment of both subunits of F_1 are based on the alignment of the three-dimensional structures of α- and β-subunits of rat Liver F_1-ATPase.

was placed close to the P-loop. The position of E504 in NBD1, and Q1291 in the NBD2 are, as expected, near the γ-phosphate binding pocket. Helix αB was modeled longer than the one observed in F_1 and recA. F508 is present at the beginning of αB, close to the adenosine binding pocket. The overall folds of the motifs A and B in the model presented here are similar to those proposed in previous studies [Hyde et al. (1990), Mimura et al. (1991), Carson et al. (1995)]. However, the sequence of region D (501-TIKENIIF-508 in CFTR NBD1 or 1287-GVIPQKVFIF-1296 in NBD2) is placed in a loop and in the first turn of helix αB, forming part of the binding site and probably interacting with the adenosine moiety. The end of region D in our model is associated with a hydrophobic patch close to the binding site and probably at an interface region. The model suggests that E504 has a function similar to that of E188 in F_1-ATPase. Certainly, the disease causing mutation E504Q, and probably the ΔI507 and ΔF508 mutations have an effect on the function, position, and in the geometry of the loop that contains E504. Significantly, a synthetic peptide that includes the carboxyl end of region D can bind ATP (unpublished observation), suggesting the participation of this region in ATP binding, as predicted by our model.

In the present model of NBD1 the adenosine and the sugar moiety are close to F508. Deletion of one residue at the beginning of helix αB could produce a 100° rotation of the helix if the loop containing E504 remains fixed; this would produce a major change in a region on the surface of the model. This change suggests an explanation for why ΔI507 and ΔF508 manifest themselves as folding mutants (Thomas et al. (1993) and references within), with temperature dependent membrane traffic and targeting problems [Cheng et al. (1990), Rich et al. (1991)]. Our model places motif C near the loop containing E504 and region D. The G551D mutation, which occurs in this region, may have charge and steric conflicts with E504 which account for the effects of the mutation. It is also possible that the G551D mutation impairs critical NBD1 interactions with other CFTR domains.

In the second nucleotide domain the putative catalytic base is a glutamine (Q1291). Unlike E504 in NBD1, Q1291 is predicted to be a poor catalytic base. Therefore, NBD2 is predicted to have a much lower catalytic capacity than NBD1. It is possible also that

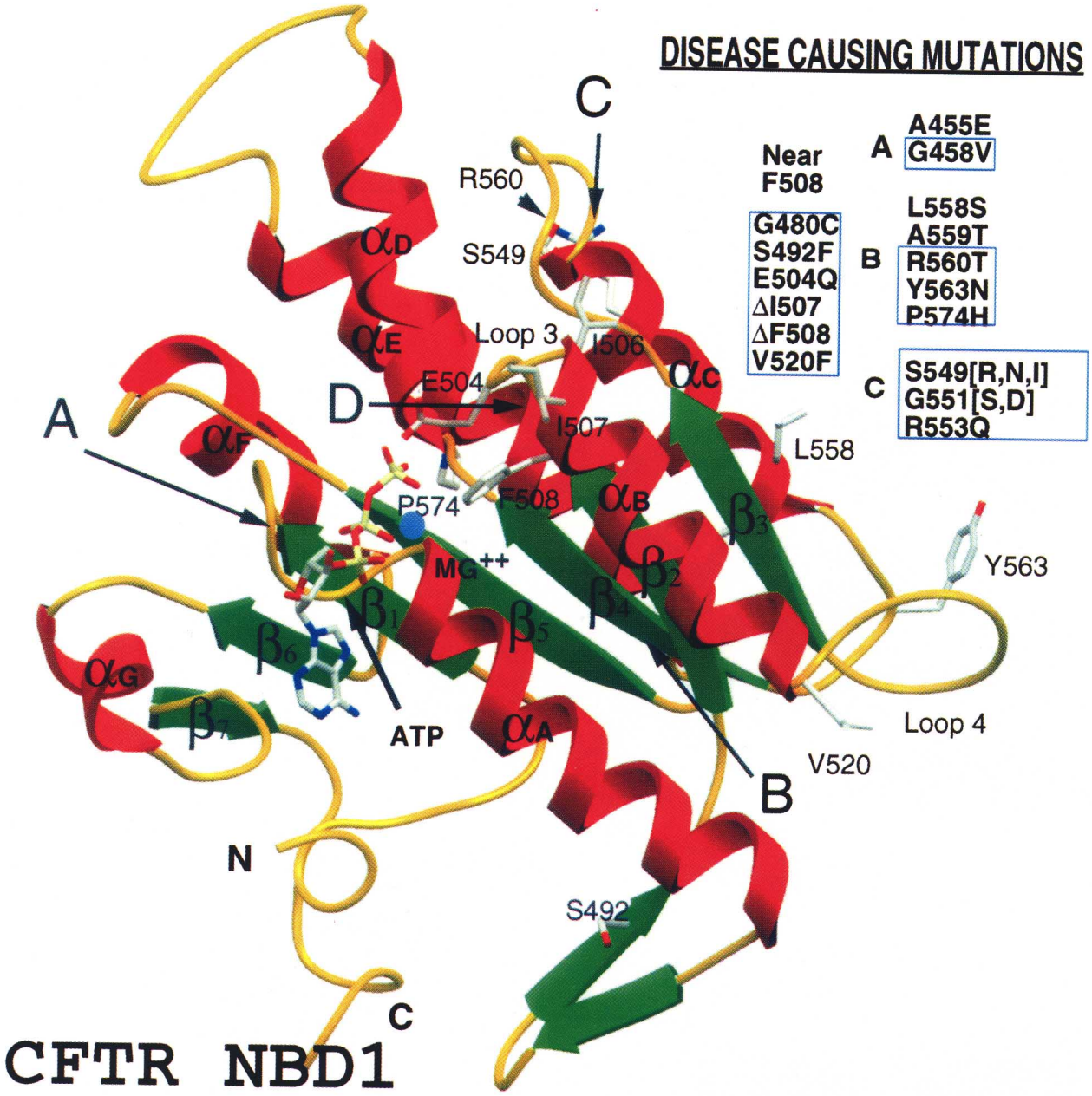


Fig. 6. *F_1/recA based model of the first Nucleotide Binding Domain (NBD1) of CFTR.* An ATP molecule is shown bound in the proposed binding site. Residues which when mutated cause CF are represented and tabulated in the Figure.

NBD2 has no catalytic capacity and only serves an ATP-dependent regulatory role. The same pattern is found in the alpha subunit of F_1-ATPase in which the equivalent residue is also a glutamine (Q208, F_1 numbering). To date, the intact α-subunit is believed to have no catalytic capacity. It should be noted, however, that in NBD2-like domains frequently the charged residue that follows the conserved glutamine of region D is an acidic one, usually aspartate (with lower frequency E), which may serve as catalytic base, e.g. MDR1 (mrp1_human in Table 2). This is not the case in all ABC transporters, e.g., CFTR and Ste6, where a basic residue (K and R respectively) follow the conserved glutamine.

Overall the three-dimensional models contain 35/37% α-helix and a 24/25% β-strand, for NBD1/NBD2.

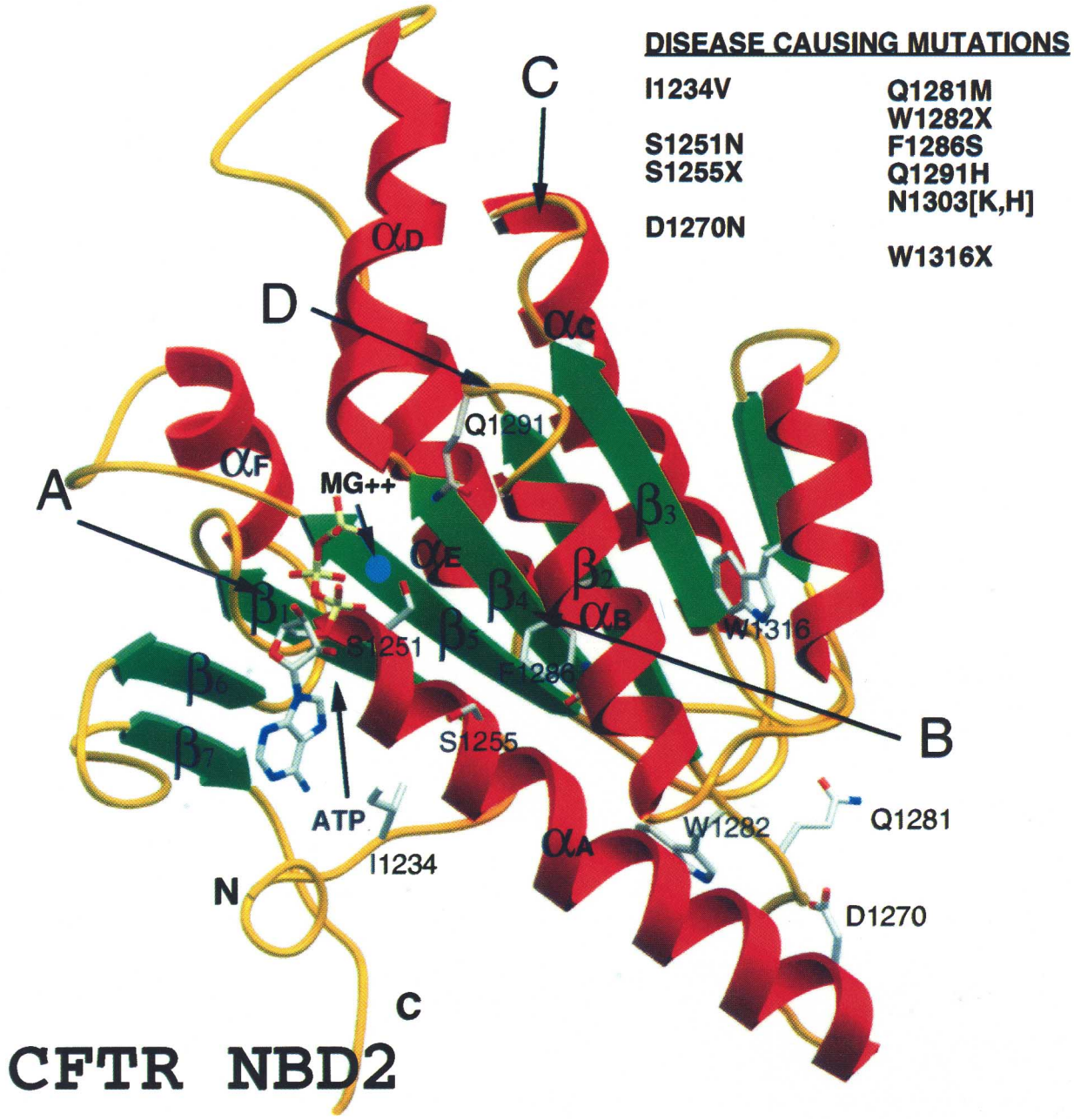


Fig. 7. *F_1/recA based model of the second Nucleotide Binding Domain (NBD2) of CFTR.* An ATP molecule is shown bound in the proposed binding site. Residues which when mutated cause CF are represented and tabulated in the Figure.

These values are in good agreement with values obtained using circular dichroism spectroscopy that show 40% α-helix and 24% β-strand using recombinant NBD1 of the traditional size (F443-S589) [Ko et al. (1994)]. The model of NBD1 presented here shows 39% α-helix and 24% β-strand in the same region.

Comparison with Previous Models

Several models of ABC transporter NBDs have been discussed in the literature, three of CFTR [Hyde et al. (1990), Carson et al. (1995), Annereau et al. (1997)] and one of the periplasmic histidine permeases

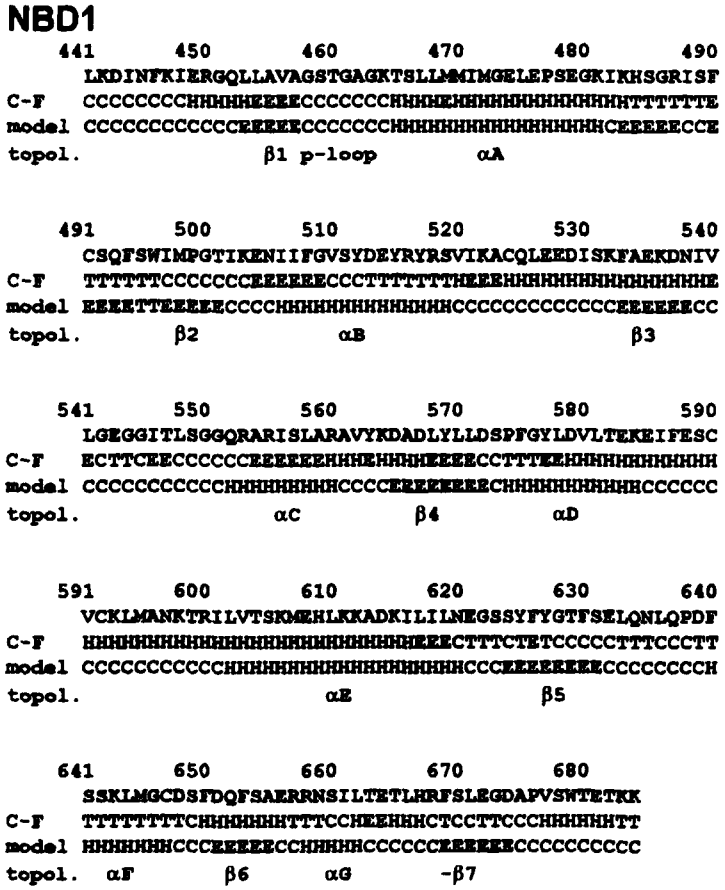


Fig. 8. *Secondary structure assignment of the first nucleotide binding domain of CFTR.* The sequence of NBD1 is aligned with the predicted secondary structure using a Chou and Fassman method (second line), and with the proposed secondary structure (third line). H, T, C and E are helix, turn, coil and beta strand respectively. The name of the secondary structural elements are shown in the last line.

[Mimura et al. (1991)]. All these models are based on the three folding motifs described previously (Fig. 3 a,b,c). The NBD of Hyde et al. (1990) and Mimura et al. (1991) were modeled using the ADK-fold and the Carson et al. (1995b) model was based on G-proteins. Recently, Annereau et al. (1997) reported a model based on F_1-ATPase. The NBDs in all the models have motifs A and B present in the secondary structural motif β1αA ... β4 (boxed in Figure 3a). Table 4 summarizes the similarities and differences among the different NBD models.

Hyde et al. (1990) and Mimura et al. (1991) made the first attempt to obtain a model of the nucleotide binding domain of ABC transporters. The model is based on the nucleotide binding domain of adenylate kinase, one of the few NBC proteins with a known three-dimensional structure at the time. This enzyme catalyzes the conversion of ATP and AMP to two ADP molecules and contains two nucleotide binding sites. An ADK-like fold, such as the one proposed by Hyde et al. (1990), assigns the region of F508 to an alpha-helical subdomain (AMP site in ADK), far away from the nucleotide site. Motif C (LSXGX[R/K]) has no role in this model. In contrast, motif C participates in ATP (or GTP) binding in models based on the G-protein-fold such as the one proposed by Carson et al. (1995b). It has been suggested that motif C is associated with nucleotide binding based on a proposed homology between CFTR and a heterotrimeric G-protein [Manavalan et al. (1995)]. The sequence DX[G/A]GQR, in heterotrimeric G-proteins, participates in binding of the GTP phosphates,[4] and Dearborn and Manavalan (1994) suggest that this sequence is equivalent to motif C of the ABC transporters. Supporting this

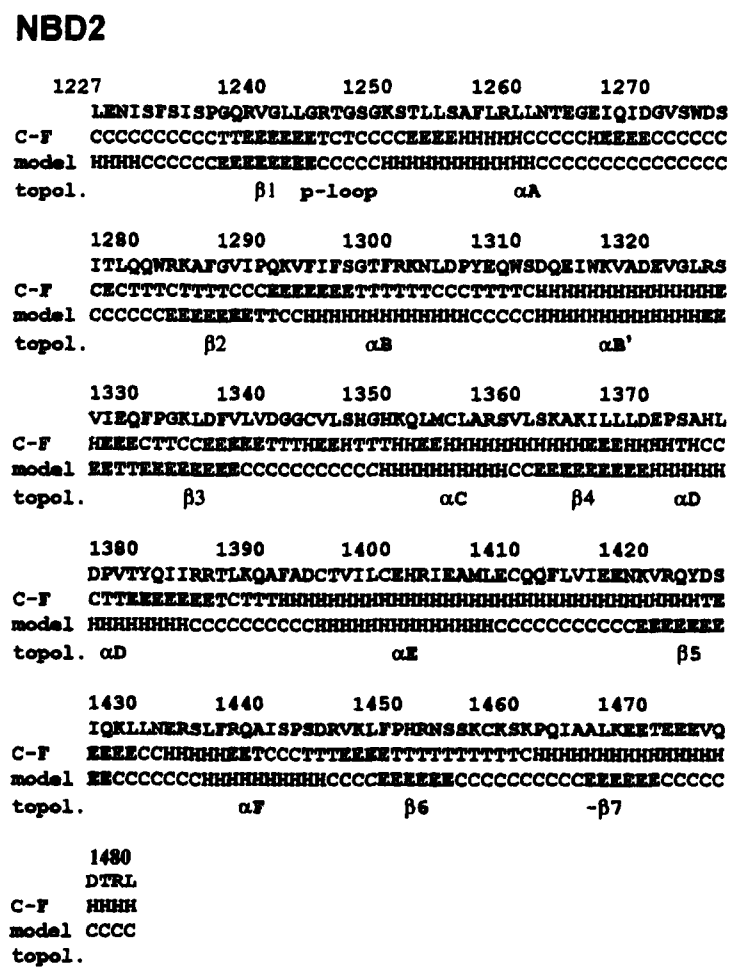


Fig. 9. *Secondary structure assignment of the second nucleotide binding domain of CFTR.* The sequence of the NBD2 of CFTR is aligned with the secondary structure predicted using a Chou and Fassman method (second line), and the proposed secondary structure (third line). H, T, C and E are helix, turn, coil and beta strand respectively. The name of the secondary structural elements are shown in the last line.

view, is the fact that GTP binding has been observed in a synthetic peptide of the NBD2 of CFTR [Randak et al. (1996)]. However, this is not compelling because GTP binding and even GTPase activity, are usually found in ATPases. Arguing against the hypothesis depicting the participation of motif C in nucleotide binding is the observation that a shorter synthetic peptide that does not include motif C can still bind ATP [Ko et al. 1994]. Also, the suggested equivalent of motif C, DX[G/A]GQR, occurs just after motif B in G-proteins, and not before as in the ABC transporters. Topologically, this sequence is inserted in the loop before β4 in ras P21/G-proteins (Fig. 3c). However, the G-protein based model suggests a direct explanation for why the mutation G551D, in the core of motif C, affects the ATP binding function. Supportive of these finding is the corrective mutation R555K. Nevertheless, a G-protein like fold for CFTR requires a large topological insertion between αA and β4 as is observed in the heterotrimeric G-proteins. The highly conserved feature of the motif B, located in this insertion, would be affected. Because of this, no explanation for other

[4] The proposed general base, E202 (G-Protein numbering), comes just after this sequence.

Table IV. Comparison of the predicted position of motif C and F508 in various models for CFTR

Model	Protein used in modeling	Position of motif C		Position of F508	
		Secondary structure	Relative to nucleotide binding site	Secondary structure	Relative to nucleotide binding site
This paper	F_1	loop	near	α-helix	within
Annereau et al. (1997)	F_1	loop	near	β-strand	outside
Hyde et al. (1990)	ADK	?	?	α-helix	outside
Carson et al. (1995)	G-proteins	loop	within	?	?

conserved features (such as the mutations in the region D) can be given based on this model.

Although, Annereau's et al. (1997) model also uses F_1-ATPase as template, their NBD sequence alignment differs with that presented here (Fig. 5). This model localizes F508 on a β-sheet buried in the core of the nucleotide binding fold.

Quaternary Association of NBD1 and NBD2 of CFTR

Modification of only one NBD in ABC transporters affects the total ATPase function. In particular, the observed behavior of CFTR in ATP hydrolysis [Carson et al. (1995)] suggests a close interaction between the two different NBDs. ATP hydrolysis in the second domain is believed to close the channel opened by the hydrolysis of ATP in the first domain. Also, mutation of lysines K458 in NBD1 and K1250 in NBD2 produce variations in the start and in the duration of the channel activity burst, respectively. These experiments assign a distinct function to each NBD and suggest a proximity relation between both domains. The assembly of recA into a recA polymer [Story et al. (1992), Yu and Egelman (1997)] and the $\alpha_3\beta_3$ association of subunits in F_1 show similarities and suggest a mode of association of the two NBDs in CFTR. The NBDs in F_1-ATPase, i.e. α- and β-subunits, have two interfaces with NBDs of adjacent subunits. Two different contact interfaces between molecules are present in the F_1 trimer of αβ pairs: α/β and β/α. Each type of interface has one of the ATP binding sites [Abrahams et al. (1994), Bianchet et al. (1997)]. This molecular quaternary association NBD1/NBD2 may also occur in CFTR, and two types of interactions are possible, one similar to the α/β interaction in F_1, in which the nucleotide binding site of NBD2 is at the interface and another similar to the β/α interface in F_1, in which the nucleotide binding site of NBD1 is at the interface. In both modes, the hydrolysis of ATP, as in F_1-ATPase could be associated with a conformational change [Abrahams et al. (1994)] easily transmitted to the second domain through the interface.

Helices αE and αF and strand β7 in NBD1 loop5 (motif C), helix αD and loop 3 (P1290-F1294, part of motif D) in NBD2, are placed in the buried interface in the α/β-like association (Fig. 10a). Symmetrically, αE, αF and strand β7 of NBD2 and loop 5 (motif C), αD and loop 3 (region D) of the NBD1 are in the buried interface in the β/α-like association (Fig. 10b); F508 is buried in this mode of association. Relevant regions such as motif C of NBD1, the locus for several of the critical disease causing mutations in CF, are localized at the proposed interfaces in the β/α-like association, and more "exposed" in the α/β-like interface. Residues at the onset of the motif B, where mutations cause CF, are in all the cases at the "exposed" face of helix αC in both NBDs. The effect of mutations in this region, are likely to affect the interaction of the NBDs with others portions of the protein.

The extent and the characteristics of the accessible surface area (ASA) buried in each mode of association (Table 5) can be used to rank the two possibilities. ASA calculations in both types of association shows the NBD2 buries more polar and less apolar area and the reciprocal for NBD1. Also, there are substantial differences in the amount of buried apolar and polar surfaces associated with the two possible interfaces. The putative α/β-like interface buries 1906 $Å^2$ of accessible apolar area and 475 $Å^2$ of polar accessible area, 245 $Å^2$ more of apolar and 304 $Å^2$ less of polar area than the β/α-like interface. From a thermodynamical point of view, the α/β-like association is favored. Figures 11a,b show the accessible surface of the α/β- and β/α-like quaternary association respectively, with

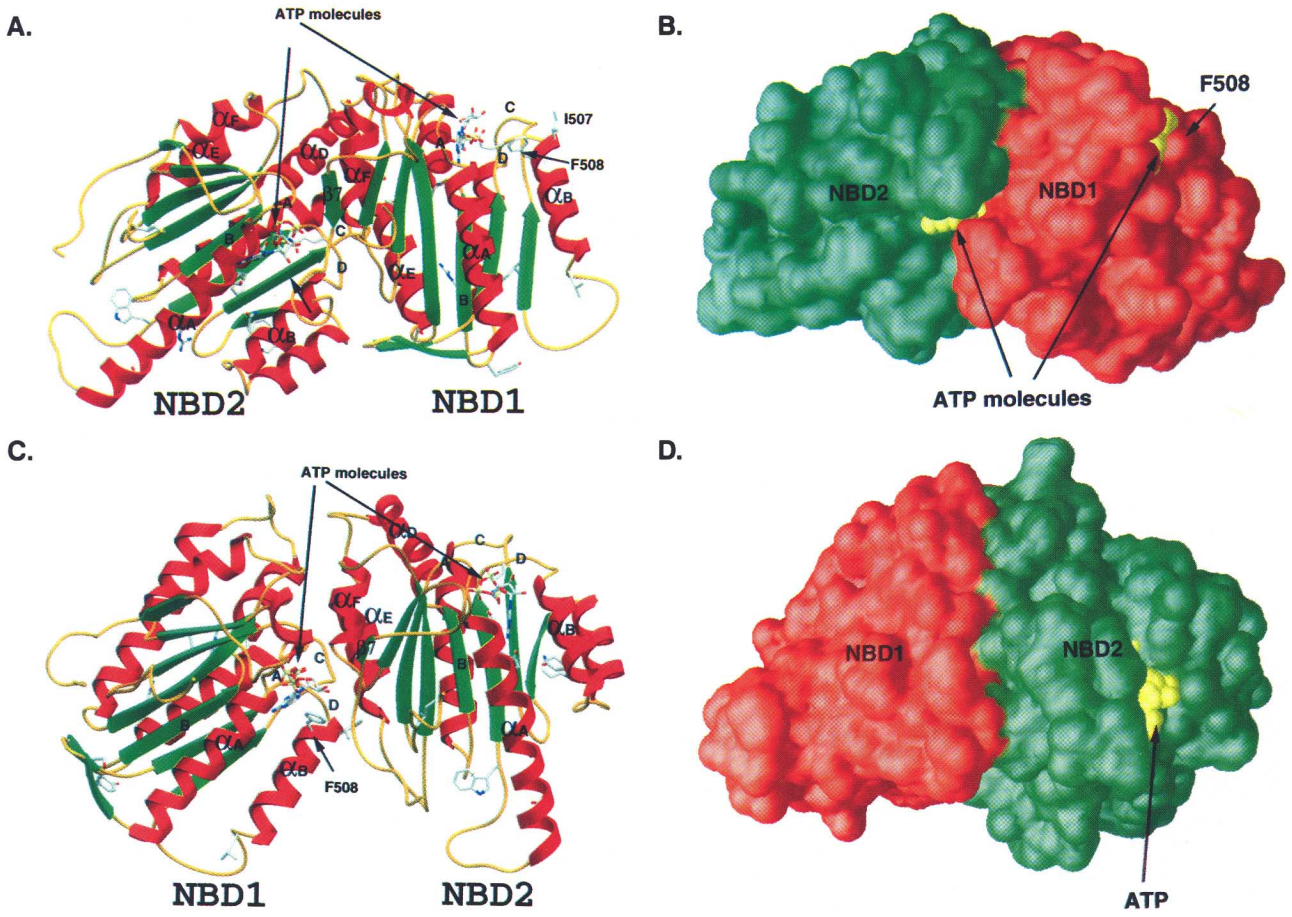


Fig. 10. *Modes of Quaternary association of the nucleotides Binding Domains.* a) Ribbon cartoon and b) space filling model of the α/β-like association, with each domain of different color, red NBD1 and green NBD2; c) ribbon cartoon and d) space filling model of the β/α-like association, with each domain of different color, red NBD1 and green NBD2.

the hydropathy of each surface atom color coded (green hydrophobic and red hydrophilic). The figures depict the hydropathy pattern of a typical soluble protein with no large hydrophobic patches. An electrostatic analysis (Figure 11c,d) displays a more charged NBD2 with concentrated positive charge (colored blue) near the nucleotide binding site.

The end of region D where the critical F508 residue lies, together with 4–6 hydrophobic residues upstream of motif C, are associated with a hydrophobic patch, which is relatively exposed in the favored α/β-like model. The concentration of mutations that cause CF in this zone and the intracellular traffic problems associated with these mutations, suggest that this hydrophobic patch may have a function in sorting and membrane trafficking of the CFTR protein to its final location in the plasma membrane.

Table V. Accessible surface area for the two different modes of association proposed. Contributions of each domain alone in each mode are also shown. All the areas are in Å^2

		NBD1	NBD2	Total
not associated	apolar	9047.6	8692.6	17740.2
	polar	4296.9	4582.0	8878.9
in the α/β-like	apolar	8096.4	7737.7	15834.1
association	polar	4022.1	4382.1	8404.2
in the β/α-like	apolar	8132.4	7946.8	16079.2
association	polar	3926.9	4173.7	8100.6

CONCLUSIONS

The structure of F_1 provides a useful template for modeling the nucleotide binding domains of ABC

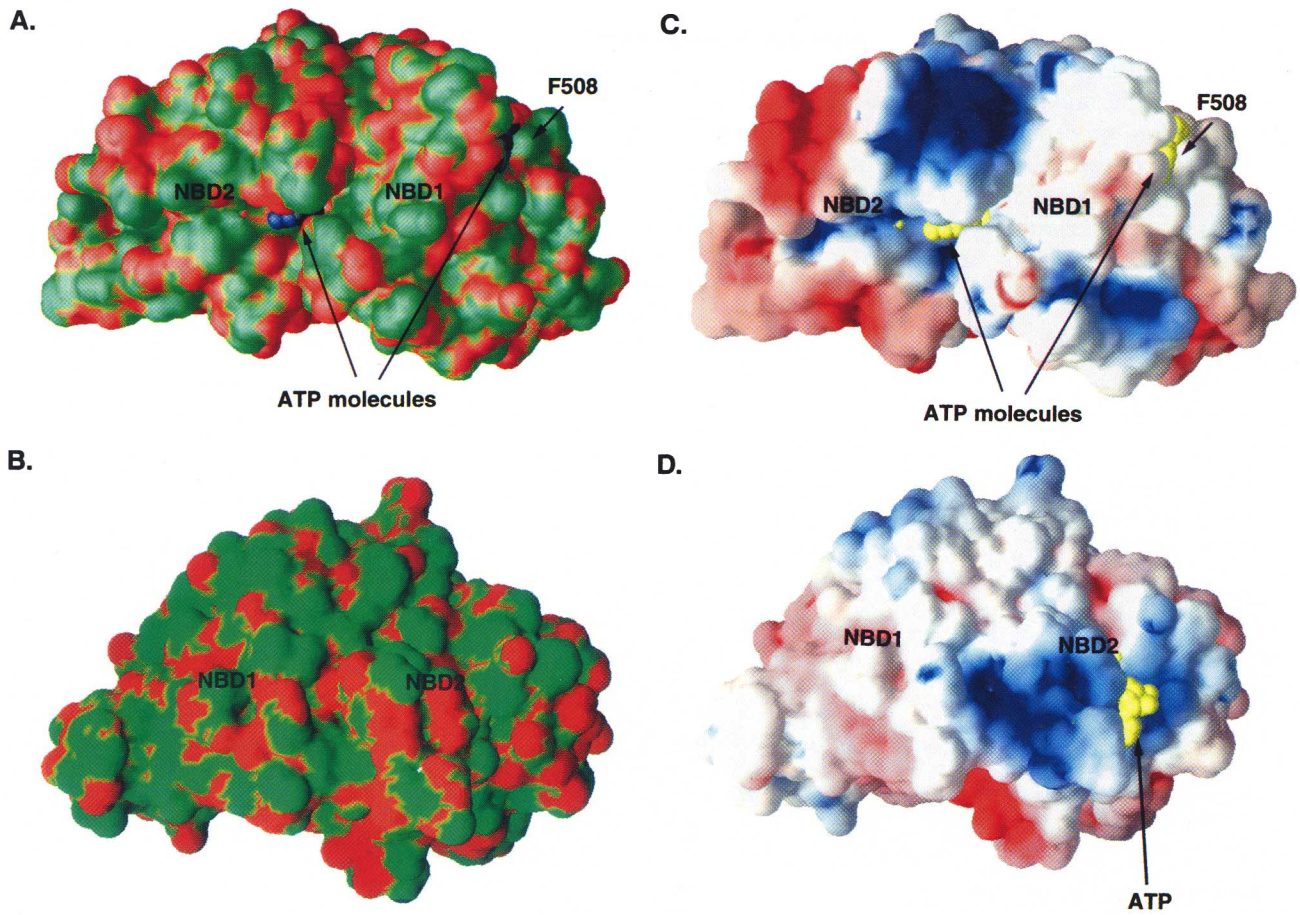


Fig. 11. *Views of the accessibility surface of the NBD complexes.* Carbon and sulfur atoms depicting hydrophobic surface are colored green and nitrogen and oxygen atoms depicting hydrophilic surface are in red. a) α/β-like, b) β/α-like, c) and d) their respective molecular surfaces colored with the calculated electrostatic potentials in KT units, blue 5KT (positive), red −5KT (negative).

transporters and in particular those of CFTR. The NBDs alignment, hydropathy profile, and alignment against the F_1 subunits, suggest NBDs longer than that traditionally accepted. The CFTR NBD1 model that results (Fig. 6) gathers the disease causing mutations in three different clusters: (1) mutations affecting the nucleotide binding pocket and the putative general base: A455E, G458V, E504Q ΔI507 ΔF508 P574H; (2) mutations in motif C which are probably related to an interaction with region D: S549[R,N,I] G551[S,D], R553Q; and (3) mutations within or near motif B, L558S, A559T, R560T, Y563N and mutations S492F and G480C. Sequence alignments suggest an association of the two domains similar to that of the α and β subunits of F_1. Deletions near the nucleotide binding pocket are certainly predicted to affect the proposed fold around the putative interface, and impair nucleotide associated functions. Changes in other regions predicted to be at the interface between CFTR domains might affect assembly of the protein. The proposed domain association also allows for a cooperative behavior between NBD1 and NBD2 as observed in intact CFTR [Carson et al. (1995a)]. Therefore, it seems possible, that the hydrolysis of ATP in NBD1 produces a conformational change that opens the channel and makes the second NBD active, until binding or hydrolysis of ATP in NBD2 reverts the opening and returns the NBD1 domain to its original state.

The models presented here indicate that NBD2 may exhibit a much lower ATP hydrolysis rate than NBD1 or perhaps no hydrolytic rate at all. Therefore, more ATP utilization would be predicted to be involved in channel opening than closing. It remains possible that channel closing requires preferentially binding of ATP to NBD2 rather than hydrolysis *per se*, and that

this ATP binding to NBD2 induces a conformational change in NBD1.

Although, this paper focuses on CFTR, most of the modeling procedures can be generalized to other ABC transporters. Here it should be noted that simply by examining the region D to establish whether it has a glutamic acid residue (E) or a glutamine residue (Q), one can predict which of the two NBDs (or both) may be functioning primarily in ATP hydrolysis.

ACKNOWLEDGMENTS

We thank Dr. Margarita Faig for reading the manuscript and for many helpful discussions during its preparation.

REFERENCES

Annereau J.P., Wulbrand U., Vankeerberghen A., Cuppens H., Bontems F., Tummler B., Cassiman J.J. and Stoven V. (1997) *FEBS Lett.* **407**, 303–308.

Abrahams J.P., Leslie A.G.W., Lutter R. and Walker J.E. (1994) *Nature* **370**, 621–628.

Bianchet M.A., Hullihen J., Pedersen P.L. and Amzel L.M. (1998) In Review.

Bianchet M., Ko Y.H. and Pedersen P.L. (1995a) Ped. Pulmon. Suppl. 12, abstract 32.

Bianchet M.A., Ko Y.H., Amzel L.M., Pedersen P.L. (1995b) *Biophysical J.* **70**(2):A214.

Carson M.R., Travis S.M. and Welsh M.J. (1995a) *J. Biol. Chem.* **270**(4):1711–1717.

Carson M.R., Travis S.M. and Welsh M.J. (1995b) *Biophysical J.* **69**, 2443–2448.

Cheng S.H., Gregory R.J., Marshall J., Paul S., Souza D.W., White G.A., O'Riordan C.R. and Smith A.E. (1990) *Cell* **63**, 827–829.

Cheng S.H., Rich D.P., Marshall J., Gregory R.J., Welsh M.J., Smith A.E. (1991) *Cell* **66**, 1027–1036.

Chou P.Y. and Fasman G.D. (1978) *Adv. Enzymol.* **47**, 45–148.

Denning G.M., Ostergard L.S., Cheng S.H., Smith A.E. and Welsh M.J. (1992) *J. Clin. Invest.* **89**, 339–349.

Dalemans W., Barbry P., Champigny G., Jallat S., Dott K., Dreyer D., Crystal R.G., Pavirani A., Lecocq J.P. and Lazduski M. (1991) *Nature* **354**, 526–528.

Doige C.A. and Ames G.F-L. (1993) *Annu. Rev. Microbiol.* **47**, 291–319.

Evans S.V. (1993) *J. Mol. Graphics* **11**, 134–138.

Higgins C.F. (1992) *Annu. Rev. Cell Biol.* **267**, 6455–6458.

Hyde S.C., P. Emsley, M.J. Harsthorn, M.M. Mimmack, U. Gileadi, S.R. Pearce, R.E. Hubbard & C.F. Higgins. (1990) *Nature* **346**, 362–365.

Kerem E., Corey M., Kerem B.S., Rommens J., Manneweiz D., Kobayashi K., Knowles M.R., Boucher R.C., O'Brien W.E., Beaudet A.L. (1990) *J. Hum. Genet.* **110**, 599–605.

Ko Y.H. and Pedersen P.L. (1995) *J. Biol. Chem.* **268**, 24330–24338.

Ko Y.H., Thomas P.J. and Pedersen P.L. (1994) *J. Biol Chem.* **269**, 14584–14588.

Ko Y.H., Delannoy M. and Pedersen P.L. (1997) *Biochem.* **36**, 5053–5064.

Levison H., Tsui L.C., Durie P. (1990) *N Engl. J. Med.* **323**, 1517–1522.

Li C.H., Ramjeesingh Wang W., Garami E., Hewryk M., Lee D., Rommens J.M., Galley K., Bear C.E. (1996) *J. Biol. Chem.* **271**, 28463–28468.

Manavalan P., Dearborn D.G., McPherson J.M. and Smith A.E. (1995) *FEBS Lett.* **366**, 87–91.

Morikawa K., la Cour T.F., Nyborg J., Rasmussen K.M., Miller D.L., Clark B.F. (1978) *J. Mol. Biol.* **5**;125, 325–338.

Nicholls A., Sharp K. and Honig B. (1991) *PROTEINS: Structure, Function and Genetics*, **11**, 281–296.

Noel J.P., Hamm H.E., Sigler P.B. (1993) *Nature*, **366**, 654–663.

Pai E.F., Krengel U., Petsko G.A., Goody R.S., Kabsch W., Wittinghofer A. (1990) *EMBO J.*, **9**, 2351–2359.

Pedersen P.L. and Amzel L.M. (1993) *J. Biol. Chem.* **268**, 9937–9940.

Qu B-H. and Thomas P. (1996) *J. Biol. Chem.* **271**, 7261–7264.

Riordan J.R., Rommens J.M., Kerem B., Alon N., Rozmahel R., Grzelczak Z., Zielenski J., Lok S., Plavsic N., Chou J., Drumm M.L., Iannuzzi C., Collins F.S. and Tsui L. (1989) *Science* **245**, 1066–1073.

Rich D.P., Anderson M.P., Gregory R.J., Cheng S.H., Paul S., Jefferson D.M., McCann J.D., Klinger K.W., Smith A. and Welsh M.J. (1991) *Nature* **347**, 358–363.

Rossmann M.G., Moras D. and Olsen K.W. (1974) *Nature* **250**, 194–199.

Rost B. (1996) *Meth. in Enzym.* **266**, 525–539.

Saraste M., Sibbald P.R. and Wittinghofer A. (1990) *Trends in Biochem. Sci.* **15**, 430–434.

Shirakihara Y., Leslie A., Abrahams J.P., Walker J., Ueda T., Sekimoto Y., Kambara M., Saiga K., Odaka M., Yoshida M. and Kagawa Y. (1997) *Structure* **5**, 825–836.

Story R.M., Weber I.T., Steitz T.A. (1992) *Nature* **355**, 567.

Schulz G.E., Elzinga M., Marx F., Schrimer R.H. (1974) *Nature* **250**, 120–123.

Taylor W.R. and Green N.M. (1989) *FEBS* **179**, 241–248.

Teem J.L., Berger L.S., Ostedgaard D.P., Rich D.P., Tsui L-C. and Welsh M.J. (1993) *Cell* **73**, 335–346.

Thomas P.J., Pedersen P.L. (1993) *J. Bioener. & Biomem.* **25**:11–20.

Tsui L-C. (1992) *Trends of Genetics* **8**, 392–398.

Walker J.E., Saraste M., Runswick M.J. and Gay N.J. (1982) *EMBO J.* **1**, 945–951.

Weber J. and Senior A.E. (1997) *Biochem. Biophys. Acta* **1319**, 19–58.

White S. (1994) Ed. S. White, Oxford University Press, New York, 97–124.

Yu X., Egelman E.H. (1997) *Nat. Struct. Biol.* **4**, 101.