A Bacterial System for Investigating Transport Effects of Cystic Fibrosis–Associated Mutations

AMY L. GIBSON, LOIS M. WAGNER, FRANCIS S. COLLINS, DALE L. OXENDER*

LIV-I, a high-affinity system that transports neutral, branched-chain amino acids into *Escherichia coli*, has two components, LivG and LivF, that are homologous to the cystic fibrosis (CF) transmembrane conductance regulator (CFTR). CF-associated mutations of human CFTR were introduced into corresponding regions of LivG, and their effects on leucine transport could be grouped into three classes. Mutations were found that (i) abolished LIV-I–directed transport, (ii) retained about a quarter of wild-type activity at the Michaelis-Menten constant (K_M), and (iii) had minimal activity at the K_M. A mutation equivalent to a benign polymorphism had no effect on transport. The correlation of these mutational phenotypes in LivG and CFTR suggests that the LIV-I prokaryotic transporter is functionally similar to the CF protein and that this similarity can be exploited to clarify the properties of the nucleotide-binding fold in this superfamily of proteins.

THE CLONING AND SEQUENCING OF the *CF* (cystic fibrosis) gene (*1–3*), together with recent complementation studies (*4*), provide the opportunity to study the molecular basis for this common genetic disease. The *CF* gene encodes a large protein, CFTR (1480 amino acids), which includes two NBF (nucleotide-binding fold) domains. These structural elements are common to members of a superfamily of related proteins from mammals, yeast, and prokaryotes (*1, 5*), most of which are known to mediate transport processes. They share a conserved motif involving amino acid similarity over 200 residues, which suggests a conservation of three-dimensional protein structure. CFTR is involved in ion transport (*6*) and may be a chloride channel (7), but the precise functions of the protein and nature of the *CF* defect are unknown.

Even in the absence of a crystallographic structure determination, the wide representation of the common structural motif of the superfamily can be exploited to learn more about the structure and function of both CFTR and related proteins. Similar approaches have permitted inferences about the structures and functions of G proteins on the basis of the similarities to EF-Tu and *ras* p21 (*8*) and have been used in the design of mutagenesis studies of the F_1–adenosine triphosphatase (ATPase) of *E. coli* (*9*) and mammalian multidrug resistance protein (*10*) and in studies dissecting the multiple functions of a yeast DNA repair enzyme (RAD3) (*11*). A growing list of CFTR mutations with associated CF phenotypes (*12–21*) suggests the possibility of investigating specific structure-function relationships by introducing similar mutations into other members of the superfamily in regions of common sequence. This approach is particularly appealing in prokaryotic systems, in which powerful methods of gene expression, mutagenesis, and selection are readily available. In this study, we introduced CF-associated mutations into a CFTR-homologous component of LIV-I of *E. coli*.

The LIV-I operon of *E. coli* encodes six proteins that transport leucine, isoleucine, and valine into the cell (*22*). Two periplasmic binding proteins, LivJ and LivK, deliver their amino acid substrates to a multicomponent transporter complex in the membrane. By analogy with other systems (*23*), two integral membrane components, LivH and LivM, are believed to form a transmembrane channel, whereas two associated peripheral membrane components, LivG and LivF, provide energy for transport from hydrolysis of nucleoside triphosphates.

Alignments of the amino acid sequences of CFTR NBF domains with comparable components of LIV-I and other transporters show the conserved A and B motifs noted by Walker *et al.* (*24*) for ATPases and kinases as well as additional regions of similarity characteristic of the protein superfamily (*1, 3*). The inferred protein sequences of the two NBF domains of CFTR were aligned with the complete sequences of LivG and LivF by the method of cluster linkage analysis described by Higgins and Sharp (*25*) and implemented as the CLUSTAL package of programs. The alignment (Fig. 1) was calculated with gap penalties set to the default values and has not been adjusted manually. The mutation ΔF508, reported to be present in about 70% of human cystic fibrosis patients, is marked by the rightmost Δ symbol in the figure. This phenylalanine residue is conserved as residue 92 in both LivF and LivG. Leu^{91} of LivG and Val^{91} of LivF are treated as conservative substitutions for Ile^{507} of CFTR.

The alignment permitted the quantitation of the degrees of relatedness of the four protein domains. Paired residues of the two CFTR NBF domains are 29% identical. LivF and LivG share 31% identical residues. The CFTR NH_2-terminal NBF is 21% identical to LivF and 19% identical to LivG. The CFTR COOH-terminal NBF is 25% identical to LivF and 20% identical to LivG. We compared each pair of sequences using the logarithm-of-odds scoring matrix of Dayhoff *et al.* (*26*), which estimates the probability of a specific amino acid replacement on the basis of comparisons of amino acid sequences of closely related proteins. The odds of relatedness of the most distantly related pair of domains (CFTR-N and LivG) is 1×10^{18} (*27*).

Four benign polymorphisms of the NBF domains of human CFTR have been discovered on normal chromosomes: M470V (*16*), I506V (*28*), V562I, and G576A (*29*). Two of these allowable substitutions are seen at the same positions in LivG (V89, analogous to I506V of CFTR, and A182, analogous to G576A of CFTR).

Differences between the amino acid sequences of human and murine CFTR (*30*) are also allowable substitutions that may be treated as polymorphisms. Comparison of the mouse sequence to the multiple sequence alignment reveals additional residues conserved in the LIV-I proteins: three found in LivG, six found in LivF, and three found in both LivF and LivG (*31*).

Fourteen deletion or missense mutations of the NBF domains of CFTR have been described: ΔF508 (*3*); G551D and A559T (*13*); G551S (*21*); ΔI507, A455E, S549I, S549R, S549N, R560T, Y563N, and P574H (*16*); G458V (*17*); and N1303K (*18*). Twelve of these fourteen human mutations are alterations of residues that are conserved in LivG or LivF or both. (A455E and Y563N are alterations of nonconserved residues.) With the exception of residue 507, all analogous residues are identical, suggesting that these CFTR mutations highlight residues that are essential for the function of the common structural motif.

We have introduced eight CF-associated mutations and one benign polymorphism of human CFTR (at positions boxed in Fig. 1) into conserved regions of LivG by site-directed mutagenesis (*32*). Mutagenesis was

A. L. Gibson and L. M. Wagner, Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109.
F. S. Collins, Howard Hughes Medical Institute and Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI 48109.
D. L. Oxender, Department of Biotechnology of the Parke Davis Research Division of Warner Lambert, Ann Arbor, MI 48106.

*To whom correspondence should be addressed.

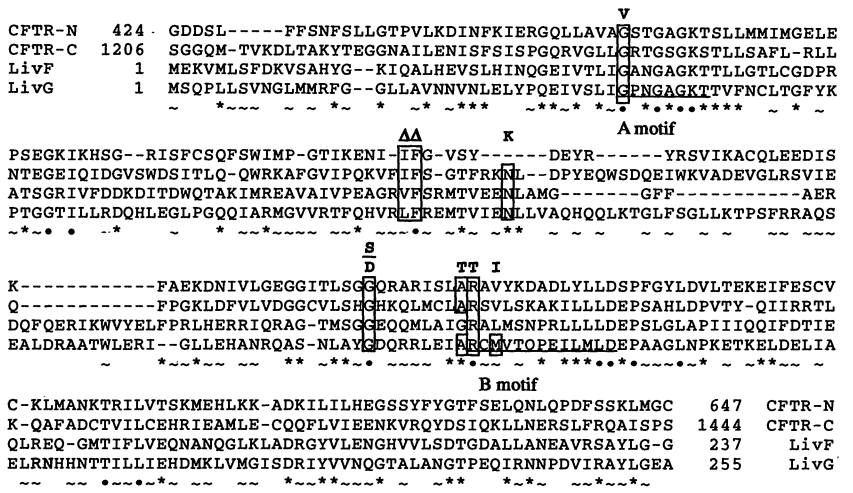

Fig. 1. Alignment of the nucleotide-binding portions of human CFTR with homologous regions of LivF and LivG. Numbers indicate beginning and ending residues. Dash (-) indicates a gap introduced for alignment. Numbers of identical aligned residues at each position are denoted by • (a match across all sequences), * (three identical residues or two pairs of identical residues), and ~ (one pair of identical residues). Reported changes in the published sequences of CFTR, LivG, and LivF are incorporated (*46*). Approximate Walker A and B motifs (*24*) are underlined, and aligned residues altered in LivG for this study are boxed. Residues deleted from CFTR at positions 507 and 508 are marked with the Δ symbol. Missense mutations are indicated with V (Gly^{458} of CFTR is replaced by Val), K (Asn^{1303} is replaced by Lys), D and S (Gly^{551} of CFTR is replaced by Asp and Ser), T (Ala^{559} and Arg^{560} are replaced by Thr), and I (a normal allele of CFTR may have Ile rather than Val at position 562).

confirmed by DNA sequencing (*33*). The CFTR mutations consist of the following: the deletion of an isoleucine at residue 507 (ΔI507) (*16*) and of phenylalanine from position 508 (ΔF508), G458V, G551D, G551S, A559T, R560T, and N1303K. Comparable mutations in LivG are ΔL91, ΔF92, G38V, G157D, G157S, A165T, R166T, and N100K, respectively. One of the polymorphisms reported for human CFTR (isoleucine instead of valine at position 562) (*29*) was introduced at the analogous position in LivG, replacing methionine (M168I).

Mutant forms of LivG were subcloned by digestion with Hind III and Eco RI followed by ligation into pUC19. A positive control plasmid expressing wild-type LivG (pOX21) (*34*) consisted of the same Hind III–Eco RI fragment cloned into pUC9 (*35*). The pUC19 vector without the *livG* gene served as the negative control.

We tested the effects of these mutations in LivG by transforming a *livG* strain, AE405 (*34*), with plasmid-borne alleles of *livG* and observing leucine uptake (Fig. 2). About 17% of the total leucine uptake by the positive control strain (at a ligand concentration near the K_M of transport, 0.1 μM) is due to a low-affinity transporter, LIV-II (*36*), and a small amount of nonsaturable transport. The deletion mutations (comparable to ΔI507 and ΔF508) and most of the tested missense mutations (analogous to G458V, G551S, A559T, and N1303K) failed to complement the transport defect. Two missense mutations (comparable to G551D and R560T) resulted in strains with LIV-I–directed leucine uptake equal to 23% and 35% of the uptake seen for the wild-type strain, respectively. A control mutant (M168I, corresponding to the benign allele V562I) transported leucine at a level corresponding to that observed for wild-type LivG. Addition of isopropyl β-D-thiogalactopyranoside (IPTG), to induce increased expression from genes following the *lac* promoter on the pUC plasmids, had no effect on the level of transport. The number of LIV-I transporters in the membrane is limited by the levels of transmembrane components (LivH and LivM) expressed from chromosomal alleles and is insensitive to variations in expression of genes on the high-copy plasmids.

Kinetic studies of LIV-I–specific transport in the four transformed mutant strains at varying concentrations of leucine (Fig. 3) show that the two deletion mutations were completely defective in leucine transport activity, even at high substrate concentrations. Parameters of the transport kinetics for the missense mutants were derived from double-reciprocal plots. Wild-type LivG exhibited a K_M of 0.16 and maximum initial velocity of transport (V_{max}) of 2.18. Four CF-associated missense mutations primarily caused decreases in the V_{max} of transport. LIV-I complexes incorporating a LivG corresponding to G551D had a K_M of 0.16 and V_{max} of 0.34 (16% of wild-type capacity). LivG incorporating the change analogous to R560T showed a K_M of 0.29 and V_{max} of 0.97 (50% of wild-type capacity). Similar experiments with LivGs analogous to G551S and A559T showed that these changes caused V_{max} defects as well (*37*).

Thus, eight CF-associated mutations of CFTR produced defective transport phenotypes in the context of a related bacterial transporter nucleotide-binding component, whereas a polymorphism of CFTR introduced nearby had no effect. These observations are consistent with the hypothesis that members of the protein superfamily containing this conserved amino acid sequence very likely share a conserved three-dimensional structure.

Two groups have proposed models of the structural motif common to the NBF domains of CFTR and LivG and LivF (*38*) on the basis of the known structure of adenylate kinase (ADK) (*39*) and secondary structure predictions for several proteins containing

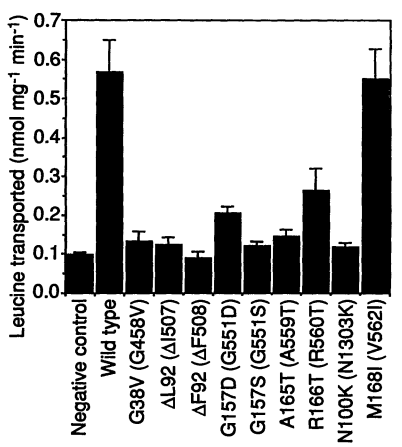

Fig. 2. Effects on leucine transport in *E. coli* due to the introduction of known CFTR mutations into homologous regions of *livG*. Plasmids carrying genes encoding wild-type LivG and LivGs bearing site-directed mutations corresponding to CF-associated changes (G38V, ΔL91, ΔF92, G157D, G157S, A165T, R166T, and N100K) or a benign polymorphism of CFTR (M168I) were used to transform AE405, a *livG* strain of *E. coli*. Corresponding mutations of CFTR are shown in parentheses. Initial-rate transport assays were performed as described (*36*). Plasmid-bearing strains were cultured in Vogel-Bonner medium supplemented with ampicillin, leucine, essential amino acids, thymine, and thiamine. Mid-log cells were chilled and washed three times in leucine-free medium. Cells (0.4 ml) were mixed with an equal volume of medium containing 0.2 μM leucine and ^{3}H-labeled leucine. After a 15-s incubation at 37°C, cells were isolated by filtration onto Millipore HA membranes. Excess radiolabeled medium was removed by washing twice with phosphate buffer. Leucine uptake was determined by scintillation counting. The range indicated on each bar represents the standard error of the mean (SEM) of triplicates from two independent cultures of each plasmid-bearing strain.

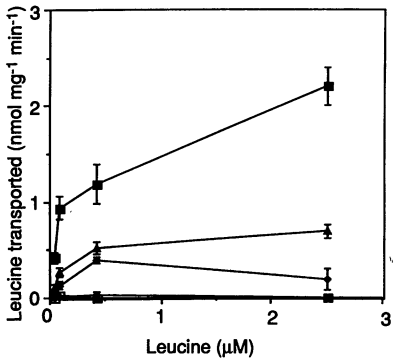

Fig. 3. Leucine transport kinetics of the strains described in Fig. 2. Symbols (from top to bottom on the plot) were used as follows: ■, wild-type LivG; ▲, R166T (R560T); ◇, G157D (G551D); ◆, ΔL91 (ΔI507); and □, ΔF92 (ΔF508). Transport assays were performed as described in Fig. 2 except that the final leucine concentration for each sample of cells was 0.05, 0.1, 0.5, or 2.5 μM. Uptake by four aliquots of a single culture of each plasmid-bearing strain was assayed at each ligand concentration. Transport measured for the negative control culture has been subtracted from total uptake by the other strains. Each point represents the average of triplicate samples from a single culture of each plasmid-bearing strain. Vertical bars show the range of expected value ± SEM. Corresponding mutations of CFTR are in parentheses.

the conserved motif. The G458V mutation appears as the first invariant glycine in the Walker A motif (GXXGXG). This feature, found in ATPases and kinases, is believed to form a glycine-rich flexible loop that changes conformation to admit triphosphate nucleosides to a catalytic site. Glycine is required at these positions for steric reasons, and replacement of glycine with the larger valine would impair mobility of the loop (*39, 40*). A cluster of tested residues (ΔI507, ΔF508, and N1303) appears in a poorly conserved region that is believed to have no counterpart in ADK (*39*). Our results are consistent with an earlier hypothesis (*28*) that residues 507 and 508 are found in an element (possibly in a β-sheet conformation) (*41*) where the relative length of the polypeptide is more important than the precise amino acid sequence.

Invariant residues at positions 551 (Gly) and 560 (Arg, part of the Walker B motif) and a small residue (Ala or Gly) at 559 may be important for positioning an essential aspartic acid residue (position 572 in CFTR) that is believed to coordinate with the metal ion of the Mg-ATP substrate (*42*). Our observation that two of the mutants of LivG altered at these residues retain some transport function was unexpected. Just as either isoleucine or valine are tolerated in the Walker B motif in CFTR (at position 562), the substitution of isoleucine for methionine is tolerated perfectly at the comparable position in LivG.

The conserved components of bacterial transporters provide energy from active transport by hydrolysis of triphosphate nucleosides (*23*). The functions of the corresponding NBF domains of CFTR are less clear. If CFTR is solely a chloride channel, three of the protein's five domains remain unexplained (*43*). The architecture of the conserved component might be capable of accomplishing either ATPase or kinase reactions. The possibility of an active transport function for the protein in mammalian cells remains, although substrate and direction of that transport are unknown. There are examples of proteins having nucleotide-binding folds with distinct functions (catalytic and regulatory) (*44*). Finally, there is genetic evidence that the conserved components of two bacterial transporters (the *spo0K* and *pst* operons of *Bacillus subtilis* and *E. coli*, respectively) (*45*) participate in signal transduction. In both cases, a conserved component similar to both LivG and the NBF domains of CFTR is thought to activate one or more histidine protein kinases, which in turn control expression of distant genes. CFTR may be a multifunctional molecule, but the geometry of the nucleotide-binding folds remains accessible through studies of related proteins of known function. It should be possible to further examine structure-function relationships by genetic selection of second-site revertants in *E. coli livG* mutants. A complementary approach to studies of the common structural motif of this protein superfamily can add to our understanding of fundamental mechanisms of transport in bacteria and the biochemical consequences of human CF-associated defects.

REFERENCES AND NOTES

1. J. R. Riordan *et al.*, *Science* **245**, 1066 (1989).
2. J. M. Rommens *et al.*, *ibid.*, p. 1059.
3. B.-S. Kerem *et al.*, *ibid.*, p. 1073.
4. M. L. Drumm *et al.*, *Cell* **62**, 1227 (1990); D. P. Rich *et al.*, *Nature* **347**, 358 (1990).
5. C. F. Higgins *et al.*, *J. Bioenerg. Biomembr.* **22**, 571 (1990); G. F.-L. Ames, *ibid.* **20**, 1 (1988).
6. M. R. Knowles *et al.*, *Science* **221**, 1067 (1983); P. M. Quinton and J. Bijman, *N. Engl. J. Med.* **308**, 1185 (1983); P. M. Quinton, *FASEB J.* **4**, 2709 (1990).
7. N. Kartner *et al.*, *Cell* **64**, 681 (1991); J. A. Wagner *et al.*, *Nature* **349**, 793 (1991).
8. F. Jurnak, S. Heffron, E. Bergmann, *Cell* **60**, 525 (1990).
9. D. Parsonage, S. Wilke-Mounts, A. E. Senior, *FEBS Lett.* **232**, 111 (1988).
10. M. Azzaria, E. Schurr, P. Gros, *Mol. Cell. Biol.* **9**, 5289 (1989).
11. P. Sung, D. Higgins, L. Prakash, S. Prakash, *EMBO J.* **7**, 3263 (1988).
12. L. Roberts, *Science* **247**, 1296 (1990).
13. G. R. Cutting *et al.*, *Nature* **346**, 366 (1990).
14. M. Vidaud *et al.*, *Hum. Genet.* **85**, 446 (1990).
15. M. B. White *et al.*, *Nature* **344**, 665 (1990).
16. B.-S. Kerem *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8447 (1990).
17. H. Cuppens, P. Marynen, C. Boeck, J. J. Cassiman, *Pediatr. Pulmonol. Suppl.* **5**, 203 (1990).
18. L. Osborne, R. Knight, G. Santis, M. Hodson, *Am. J. Hum. Genet.* **48**, 608 (1991).
19. M. Dean *et al.*, *Cell* **61**, 863 (1990).
20. K. Davies, *Nature* **348**, 110 (1990).
21. T. V. Strong *et al.*, in preparation.
22. M. D. Adams *et al.*, *J. Biol. Chem.* **265**, 11436 (1990).
23. L. Bishop, R. Agbayani, Jr., S. V. Ambudkar, P. C. Maloney, G. F.-L. Ames, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6953 (1989); M. L. Mimmack *et al.*, *ibid.*, p. 8257; D. A. Dean, A. L. Davidson, H. Nikaido, *ibid.*, p. 9134.
24. J. E. Walker, M. Saraste, M. J. Runswick, N. J. Gay, *EMBO J.* **1**, 945 (1982).
25. D. G. Higgins and P. M. Sharp, *Gene* **73**, 237 (1988); *CABIOS* **5**, 151 (1989).
26. M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt, in *Atlas of Protein Sequence and Structure*, M. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, DC, 1978), suppl. 3, vol. 5, pp. 345–352.
27. Odds of relatedness of the other paired domains are as follows: CFTR-N and CFTR-C, 7.9×10^{31}; LivF and LivG, 2.5×10^{41}; CFTR-N and LivF, 2.5×10^{19}; CFTR-C and LivF, 1.3×10^{25}; CFTR-C and LivG, 1.6×10^{21}.
28. K. Kobayashi, M. R. Knowles, R. C. Boucher, W. E. O'Brien, A. L. Beaudet, *Am. J. Hum. Gen.* **47**, 611 (1990).
29. M. Goossens, personal communication.
30. F. Tata *et al.*, *Genomics* **10**, 301 (1991).
31. The following additional residues from the NBF domains of murine CFTR align with identical residues in LivG, LivF, or both: E425, H432, L446, E452, A477, Q621, T623, K1307, I1375, G1390, R1408, and A1425.
32. T. A. Kunkel, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 488 (1985). No strains with mutant *livF* alleles are currently available.
33. F. Sanger, S. Nicklen, A. R. Coulson, *ibid.* **74**, 5463 (1977).
34. P. M. Nazos *et al.*, *J. Bacteriol.* **163**, 1196 (1985).
35. When LIV-I DNA is cloned as described, pUC9 and pUC19 are identical.
36. M. Rahmanian, D. R. Claus, D. L. Oxender, *J. Bacteriol.* **116**, 1258 (1973).
37. Data not shown. Activity of LivGs incorporating changes analogous to G458V and N1303K was not tested at higher leucine concentrations.
38. S. C. Hyde *et al.*, *Nature* **346**, 362 (1990); C. S. Mimura, S. R. Holbrook, G.F.-L. Ames, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 84 (1991).
39. D. Dreusicke, P. A. Karplus, G. E. Schulz, *J. Mol. Biol.* **199**, 359 (1988).
40. D. Dreusicke and G. E. Schulz, *FEBS Lett.* **208**, 301 (1986).
41. P. J. Thomas, P. Shenbagamurthi, X. Ysern, P. L. Pedersen, *Science* **251**, 555 (1991).
42. D. C. Fry, S. A. Kuby, A. S. Mildvan, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 907 (1986).
43. The existence of CF patients apparently carrying two null alleles [G. R. Cutting *et al.*, *N. Engl. J. Med.* **323**, 1685 (1990)] presents an intriguing problem. It is possible that multimeric complexes of truncated forms of CFTR could still provide functional ion channels and that the loss of other functional domains could be a much less severe defect than the presence NBF or R-domains with altered functions.
44. F. Jurnak, *Science* **230**, 32 (1985).
45. D. Z. Rudner, J. R. LeDeaux, K. Ireton, A. D. Grossman, *J. Bacteriol.* **173**, 1388 (1991); A. D. Grossman, *Cell* **65**, 5 (1991); N. N. Rao and A. Torriani, *Mol. Microbiol.* **4**, 1083 (1990).
46. J. Zielenski *et al.*, *Genomics* **10**, 214 (1991). Corrections to the original sequence of the LIV-I operon have been communicated to GenBank.

47. This work was supported by a grant to D.L.O. and F.S.C. from The Cystic Fibrosis Foundation. F.S.C. is an Investigator of the Howard Hughes Medical Institute. The authors acknowledge M. L. Drumm and R. M. Williamson for helpful discussions and M. D. Adams and S. A. Haney for critical reading of the manuscript.

19 February 1991; accepted 12 July 1991