J Mol Evol (1996) 42:281–293

JOURNAL OF MOLECULAR EVOLUTION

Relationships Between Bacterial Drug Resistance Pumps and Other Transport Proteins

J.H. Parish,[1] J. Bentley[2]

[1] Department of Biochemistry and Molecular Biology, The University of Leeds, Leeds LS2 9JT, United Kingdom
[2] Department of Biology, The University of York, York, Y01 5YW, United Kingdom

Received: 25 August 1995 / Accepted: 21 September 1995

Abstract. We have used three reference sequences representative of bacterial drug resistance pumps and sugar transport proteins to collect the 91 most closely related sequences from a composite, nonredundant protein sequence database. Having eliminated certain very close relatives, the remainder were subjected to analysis and alignment by using two different similarity matrices: one of these was a matrix based on structural conservation of amino acid residues in proteins of known conformation and the other was based on the more familiar mutational matrix. Unrooted similarity trees for these proteins were constructed for each matrix and compared. A systematic analysis of the differences between these trees was undertaken and the sequences were analyzed for the presence or absence of certain sequence motifs. The results show that the clades created by the two methods are broadly comparable but that there are some clusters of sequences that are significantly different. Further analysis confirmed that (1) the sequences collected by this objective method are all known or putative 12-helix (in some cases reported as 14-helix) transmembrane proteins, (2) there is evidence for few cases of an origin based on gene duplication, (3) the bacterial drug resistance pumps are distributed in more than one clade and cannot be regarded as a definitive subset of these proteins, and that (4) the diversity is such that there is no evidence of a single ancestral protein. The possible extension of the methods to other cases of divergent protein sequences is discussed.

Correspondence to: J.H. Parish

Key words: Transmembrane proteins — Drug resistance — Sugar transport — Similarity matrix

Introduction

Bacterial resistance to drugs including antibiotics and antiseptics can be mediated by mechanisms involving the efflux of these compounds. Such mechanisms have great practical importance as, unlike drug detoxification processes (such as penicillin hydrolysis and aminoglycoside adenylylation), the proteins involved in drug efflux can often confer resistance to a variety of chemically unrelated drugs (Lewis 1994). Also, this group of transporters includes several examples of proteins whose genes are present on transferable genetic elements, potentially allowing the spread of such resistance among members of heterogeneous bacterial populations. The biochemistry of such resistance is reviewed by Nikaido (1994).

The proteins involved in this study include many transport proteins that employ the proton motive force generated by energy-yielding metabolism (Mitchell et al. 1981) to effect the movement of substrate against a concentration gradient. We follow Mitchell and refer to proteins that translate the proton and substrate in the same direction as substrate-H^+ symporters and those in which these move in opposed directions as substrate/H^+ antiporters.

The antibiotic resistance proteins whose mechanism of action is known, or by analogy predicted, to involve a drug/H^+ antiport mechanism are proposed to belong to a larger group of 12- to 14-helical transmembrane proteins.

A tree based on sequence similarity has been constructed for some of these transmembrane proteins (Lewis 1994). It is known that these proteins show sequence relationships with sugar transport proteins including both facilitated diffusion transporters and sugar-H^+ symporters (Griffith et al. 1992), and the evidence for 12-transmembrane helices comes largely from side-directed mutagenesis experiments with members of this group (Baldwin, 1994). Several sequence analyses have been published which predict structural relationships between some of these membrane transporter proteins (Griffith et al. 1992; Marger and Saier 1993). However, now over 80 membrane transport proteins, including those involved in the transport of antibiotics, sugars, oligopeptides and amines, have been reported. All these proteins are predicted to have between 12 and 14 transmembrane domains and all share sequence similarity although there is, as yet, no comprehensive large-scale sequence analysis of these proteins.

Therefore, the aim of this study was to use a number of computer alignment and clustering algorithms to establish whether this large group of proteins indeed constituted a superfamily and whether the bacterial drug resistance pumps reviewed (Lewis 1994) were in any sense clustered within this larger superfamily. The approach is complementary with that of a recent study (Le Novère and Changeaux 1995) of the nicotinic acetylcholine receptors: these authors contrasted the use of a cladistic method based on the mutation matrix PAM250 with a parsimony-based phenetic analysis: we have not used the parsimony method but have contrasted the clades generated by using the same matrix with those generated by using a matrix based on amino acid conservation in proteins of known structure. We follow Le Novère and Changeaux in representing the findings as unrooted trees because, like them, we are dealing with a group of proteins with no obvious archaic progenitor.

Materials and Methods

The protein sequences used for alignments were abstracted from the OWL composite, nonredundant protein sequence database (Bleasby and Wootton 1990; Akrigg et al. 1992) by using three "reference sequences." They were (1) the human erythrocyte facilitated glucose transport protein, (2) the *Escherichia coli* lactose-H^+ symport protein, and (3) the predicted *E. coli* bicyclomycin resistance protein. The selected sequences were aligned by using CLUSTALV (Higgins et al. 1992) with two modifications. Firstly, the source code of CLUSTALV was changed to handle larger data sets: Secondly, alternative scoring matrices, Risler and PAM250, were employed. Two utilities were written for this purpose, the first converts the Risler matrix into FASTA format while the other, named "CLUTR", performs two passes on the nearest-joins output from CLUSTALV and generates files suitable for direct loading into the generic graph plotting program, GNUplot (Williams and Kelley 1992). The source code for CLUTR and other utilities specific to this work can be obtained by e-mail to either author (J.H.Parish@leeds.ac.uk or JB30@york.ac.uk).

We constructed cladograms of these selected proteins by using two alternative similarity matrices. The first matrix measures amino acid similarities by using a scoring system derived from the pattern recognition of amino acid substitutions of Risler et al. (1988). This is referred to as the "Risler matrix"; this matrix is objective and based on the known substitutions of amino acids in proteins of known isomorphic structure. The matrix was transformed to FASTA format required by CLUSTALV. This was achieved by using a utility RTOS. The Risler matrix itself uses the score of 0 for homology and increasing scores (up to 100) for unfavorable replacements. RTOS maps the scores of 0–100 to a range of two values, MAX and MIN, set to 20 and –20 in the calculations used in this paper. The cladogram was constructed from the nearest-joins output (Saitou and Nei 1987) calculated by the bootstrap algorithm of Felsenstein (1985) The second alignment and analysis was performed using the more familiar PAM250 matrix (Dayhoff et al. 1978).

Analysis of the internal homology in these proteins were performed by predicting the point of division between the left and right domains of these sequences by using either the PHD neural network program (Rost and Sander 1994) or the PEPPLOT program (Devereux et al. 1984) and then constructing a cladogram with these domains with the Risler matrix as described above.

Motifs were abstracted in part from the PROSITE database (Bairoch and Bucher 1994) and converted to the "CREGEX" (regular expression) format of Kolakowski et al. (1992). A program, PROTEST, was used to interrogate such REGEX files with the 82 protein sequence files.

Results

We set out to construct cladograms for the group of membrane transporter proteins predicted to belong to a large superfamily. Three reference sequences—(1) the human erythrocyte-facilitated glucose transport protein, (2) the *Escherichia coli* lactose-H^+ symport protein, and (3) the predicted *E. coli* bicyclomycin resistance protein were chosen to scan the database. It was considered that the human glucose transporter and the *E. coli* lactose symporter represented two transporter types previously thought to be distinct. The bicyclomycin resistance gene *bcr,* putative protein Bcr, is of interest to us as it is predicted to confer resistance to two unrelated drugs, bicyclomycin and sulphathiazole (Bentley et al. 1993; Lewis 1994), and might be another example of a protein containing a naturally suppressed opal codon (Kopelowitz 1992).

Ninety-one different proteins were selected from the OWL database by the three reference sequences as described in Materials and Methods. Eighty-two of these proteins are listed in Table 1. In order to reduce the complexity of the output we grouped certain very closely related proteins and selected one representative. An example of this is the group comprising the glucose transporters GTR1 from human, bovine, mouse, pig, rabbit, and rat, of which the human sequence was chosen as the representative. The *bcr* gene product was intentionally included twice where the first product is the full-length Bcr protein and the second is the product predicted from the nucleotide sequence distal to the TGA opal codon. If the TGA codon is read through in the *bcr* gene the predicted protein sequence is similar to that of the

Table 1. Proteins selected from the OWL database for further sequence analysis[a]

1	FUCP_ECOLI	B-S-s L-FUCOSE PERMEASE. - ESCHERICHIA COLI. (Lu and Lin 1989)
2	CSCB_ECOLI	B-S-s SUCROSE TRANSPORT PROTEIN (SUCROSE PERMEASE). - ESCHERICHIA COLI. (Bockmann et al. 1992)
3	LACY_ECOLI	B-S-s LACTOSE PERMEASE (LACTOSE-PROTON SYMPORT). - ESCHERICHIA COLI. (Kaback 1990)
4	LACY_KLEPN	B-S-s LACTOSE PERMEASE (LACTOSE-PROTON SYMPORT). - KLEBSIELLA (McMorrow et al. 1988)
5	RAFB_ECOLI	B-S-s RAFFINOSE PERMEASE. - ESCHERICHIA COLI. (Aslinidis et al. 1989)
6	ORF_BCNR	B-O-? extended orf of BCR_ECOLI [9] (Bentley et al. 1993)
7	ARAE_ECOLI	B-S-s ARABINOSE-PROTON SYMPORT (ARABINOSE TRANSPORTER). - ESCHERICHIA COLI (Maiden et al. 1987)
8	ATR1_YEAST	F-A-? AMINOTRIAZOLE RESISTANCE PROTEIN. - SACCHAROMYCES CEREVISIAE (BAKER'S YEAST). (Goempel-Klein and Brendel 1990)
9	BCR_ECOLI	B-A-a BICYCLOMYCIN RESISTANCE PROTEIN. - ESCHERICHIA COLI. (Bentley et al. 1993)
10	EMRD_ECOLI	B-A-? MULTIDRUG RESISTANCE PROTEIN D. - ESCHERICHIA COLI. (Burland et al. 1993)
11	BMRP_CANAL	F-A-? BENOMYL/METHOTREXATE RESISTANCE PROTEIN. - CANDIDA ALBICANS (YEAST). (Fling et al. 1991)
12	YIDY_ECOLI	B-O-? HYPOTHETICAL 41.5 KD PROTEIN IN TNAB 3′ REGION. - ESCHERICHIA COLI. (Burland et al. 1993)
13	CIT1_ECOLI	B-M-s CITRATE-PROTON SYMPORT (CITRATE TRANSPORTER) (CITRATE UTILIZATION DETERMINANT). - ESCHERICHIA COLI. (Sasatsu et al. 1985)
14	CIT2_ECOLI	B-M-s CITRATE-PROTON SYMPORT (CITRATE TRANSPORTER) (CITRATE UTILIZATION DETERMINANT). - ESCHERICHIA COLI. (Ishiguro and Sato 1985)
15	CITA_SALTY	B-M-s CITRATE-PROTON SYMPORT (CITRATE TRANSPORTER) (CITRATE CARRIER PROTEIN). - SALMONELLA TYPHIMURIUM. (Shimamoto et al. 1991)
16	CIT_KLEPN	B-M-s CITRATE-PROTON SYMPORT (CITRATE TRANSPORTER) (CITRATE CARRIER PROTEIN). - KLEBSIELLA PNEUMONIAE. (van der Rest et al. 1990)
17	CMLR_STRLI	B-A-a CHLORAMPHENICOL RESISTANCE PROTEIN. - STREPTOMYCES LIVIDANS. (Dittrich et al. 1991)
18	GAL2_YEAST	F-S-? GALACTOSE TRANSPORTER (GALACTOSE PERMEASE). - SACCHAROMYCES CEREVISIAE (BAKER'S YEAST). (Nehlin et al. 1989)
19	GLCP_SYNY3	F-S-? GLUCOSE TRANSPORT PROTEIN. - SYNECHOCYSTIS SP. (STRAIN PCC 6803). (Zhang et al. 1989)
20	GLF_ZYMMO	B-S-f GLUCOSE FACILITATED DIFFUSION PROTEIN. - ZYMOMONAS MOBILIS. (Barnell et al. 1990)
21	GTR1_HUMAN	M-S-f GLUCOSE TRANSPORTER TYPE 1, ERYTHROCYTE/BRAIN. - HOMO SAPIENS (HUMAN). (Mueckler et al. 1985) close relatives: GTR1_BOVIN glucose transporter type 1, erythrocyte/brain. - Bos taurus (bovine). (Boado and Pardridge 1991) GTR1_MOUSE glucose transporter type 1, erythrocyte/brain (gt1). - Mus musculus (mouse). (Kaestner et al. 1989) GTR1_PIG glucose transporter type 1, erythrocyte/brain (fragment). - Sus scrofa (pig). (Weiler-Güttler et al. 1989) GTR1_RABBIT glucose transporter type 1, erythrocyte/brain. - Oryctolagus cuniculus (rabbit). (Asano et al. 1988) GTR1_RAT glucose transporter type 1, erythrocyte/brain. - Rattus norvegicus (rat). (Williams and Birnbaum 1988) MUSGLUTRN mouse facilitated glucose transport protein mRNA, complete cds. - Mus musculus
22	GTR2_HUMAN	M-S-f GLUCOSE TRANSPORTER TYPE 2, LIVER. - HOMO SAPIENS (HUMAN). (Fukumoto et al. 1988)
23	GTR2_MOUSE	M-S-f GLUCOSE TRANSPORTER TYPE 2, LIVER. - MUS MUSCULUS (MOUSE). (Suzue et al. 1989)
24	GTR2_RAT	M-S-f GLUCOSE TRANSPORTER TYPE 2, LIVER. - RATTUS NORVEGICUS (RAT). (Thorens et al. 1988)
25	GTR3_CHICK	A-S-f GLUCOSE TRANSPORTER TYPE 3 (CEF-GT3). - GALLUS GALLUS (CHICKEN). (White et al. 1991)
26	GTR3_HUMAN	M-S-f GLUCOSE TRANSPORTER TYPE 3, BRAIN. - HOMO SAPIENS (HUMAN). (Kayano et al. 1988)
27	GTR3_MOUSE	M-S-f GLUCOSE TRANSPORTER TYPE 3, BRAIN. - MUS MUSCULUS (MOUSE). (Nagamatsu et al. 1992)
28	GTR4_HUMAN	M-S-f GLUCOSE TRANSPORTER TYPE 4, INSULIN-RESPONSIVE. - HOMO SAPIENS (HUMAN). (Fukumoto et al. 1989)
29	GTR4_MOUSE	M-S-f GLUCOSE TRANSPORTER TYPE 4, INSULIN-RESPONSIVE (GT2) - MUS MUSCULUS (MOUSE). (Kaestner et al. 1989)
30	GTR4_RAT	M-S-f GLUCOSE TRANSPORTER TYPE 4, INSULIN-RESPONSIVE. - RATTUS NORVEGICUS (RAT). (Birnbaum 1989)
31	GTR5_HUMAN	M-S-f GLUCOSE TRANSPORTER TYPE 5, SMALL INTESTINE. - HOMO SAPIENS (HUMAN). (Kayano et al. 1990)
32	GTR7_RAT	M-S-f GLUCOSE TRANSPORTER TYPE 7, HEPATIC MICROSOMAL. - RATTUS NORVEGICUS (RAT). (Waddell et al. 1992) close relative: MMGLUTRA Mouse mRNA for liver-type glucose transporter protein - Mus musculus Eukaryota (Asano et al. 1989)
33	HUP1_CHLKE	P-S-s H(+)/HEXOSE COTRANSPORTER. - CHLORELLA KESSLERI (CHLORELLA VULGARIS). (Sauer and Tanner 1989)
34	HXT2_YEAST	F-S-? HIGH-AFFINITY GLUCOSE TRANSPORTER HXT2. - SACCHAROMYCES CEREVISIAE (BAKER'S YEAST). (Kruckeberg and Bisson 1990)
35	VMT1_RAT	M-M-? CHROMAFFIN GRANULE AMINE TRANSPORTER. - RATTUS NORVEGICUS (RAT). (Liu et al. 1992)
36	VMT2_RAT	M-M-f SYNAPTIC VESICLE AMINE TRANSPORTER (MONOAMINE TRANSPORTER). - RATTUS NORVEGICUS (RAT). (Liu et al. 1992)
37	MMR_BACSU	B-A-a METHYLENOMYCIN A RESISTANCE PROTEIN (MMR PEPTIDE). - BACILLUS SUBTILIS. (Putzer et al. 1992)

Table 1. Continued

38	MMR_STRCO	B-A-a METHYLENOMYCIN A RESISTANCE PROTEIN (MMR PEPTIDE). - STREPTOMYCES COELICOLOR. (Neal and Chater 1987)
39	NORA_STAAU	B-A-a QUINOLONE RESISTANCE NORA PROTEIN. - STAPHYLOCOCCUS AUREUS. (Yoshida et al. 1990)
40	PRO1_LEIEN	Z-O-? PROBABLE TRANSPORT PROTEIN (LTP). - LEISHMANIA ENRIETTII. (Cairns et al. 1989)
41	QACA_STAAU	B-A-a ANTISEPTIC RESISTANCE PROTEIN. - STAPHYLOCOCCUS AUREUS. (Rouch et al. 1990)
42	QAY_NEUCR	F-M-? QUINATE TRANSPORTER. - NEUROSPORA CRASSA. (Geever et al. 1989)
43	QUTD_EMENI	F-M-? QUINATE PERMEASE. - EMERICELLA NIDULANS (ASPERGILLUS NIDULANS). (Hawkins et al. 1988)
44	RAG1_KLULA	F-S-? LOW-AFFINITY GLUCOSE TRANSPORTER. - KLUYVEROMYCES LACTIS (YEAST). (Wesolowski-Louvel et al. 1992)
45	SNF3_YEAST	F-S-? HIGH-AFFINITY GLUCOSE TRANSPORTER SNF3. - SACCHAROMYCES CEREVISIAE (BAKER'S YEAST). (Marshall-Carson et al. 1990)
46	STP1_ARATH	P-S-? GLUCOSE TRANSPORTER (SUGAR CARRIER). - ARABIDOPSIS THALIANA (MOUSE-EAR CRESS). (Sauer et al. 1990)
47	TCRB_BACSU	B-A-a TETRACYCLINE RESISTANCE PROTEIN. - BACILLUS SUBTILIS. (Sakaguchi et al. 1988)
48	TCR1_ECOLI	B-A-a TETRACYCLINE RESISTANCE PROTEIN (TRANSPOSON TN10). - ESCHERICHIA COLI. (Nguyen et al. 1983)
49	TCR2_BACSU	B-A-a TETRACYCLINE RESISTANCE PROTEIN. - BACILLUS SUBTILIS. (Noguchi et al. 1986) close relative: TCR_STRPN tetracycline resistance protein. - Streptococcus pneumoniae, Bacillus cereus and Bacillus subtilis. (Palva et al. 1990)
50	TCR2_ECOLI	B-A-a TETRACYCLINE RESISTANCE PROTEIN. - ESCHERICHIA COLI. (Preden 1983)
51	TCR3_ECOLI	B-A-a TETRACYCLINE RESISTANCE PROTEIN (TRANSPOSON TN1721). - ESCHERICHIA COLI. (Waters et al. 1983)
52	TCR_BACST	B-A-a TETRACYCLINE RESISTANCE PROTEIN. - BACILLUS STEAROTHERMOPHILUS. (Hoshino et al. 1985)
53	TCR_STAAU	B-A-a TETRACYCLINE RESISTANCE PROTEIN. - STAPHYLOCOCCUS AUREUS. (Mojumdar and Khan 1988)
54	TCR_STRAG	B-A-a TETRACYCLINE RESISTANCE PROTEIN. - STREPTOCOCCUS AGALACTIAE. (van der Lelie et al. 1989)
55	XYLE_ECOLI	B-S-s XYLOSE-PROTON SYMPORT (XYLOSE TRANSPORTER). - ESCHERICHIA COLI. (Maiden et al. 1987)
56	YIEO_ECOLI	B-O-? HYPOTHETICAL 51.5-KD PROTEIN IN RBSR 3' REGION. - ESCHERICHIA COLI. (Burland et al. 1993)
57	HXT1_YEAST	F-S-? HIGH-AFFINITY GLUCOSE TRANSPORTER HXT1. - SACCHAROMYCES CEREVISIAE (BAKER'S YEAST). (Ko et al. 1993)
58	JQ1479	B-A-a TETRACYCLINE RESISTANCE PROTEIN - ESCHERICHIA COLI TRANSPOSON TN1721 (Allmeier et al. 1992)
59	BMR1_BACSU	B-A-a MULTIDRUG RESISTANCE PROTEIN. - BACILLUS SUBTILIS. (Neyfakh et al. 1991)
60	JQ1201	B-A-a CMLA PROTEIN - PSEUDOMONAS SP. PLASMID R1033 TRANSPOSON TN1696 (Stokes and Hall 1991)
61	B40046	B-A-a TETRACYCLINE RESISTANCE PROTEIN HOMOLOG ACTII-2-STREPTOMYCES COELICOLOR (Fernandez-Moreno et al. 1991)
62	S24752	B-A-a LINCOMYCIN RESISTANCE PROTEIN LMRA - STREPTOMYCES LINCOLNENSIS (Zhang et al. 1992)
63	TCR_STAHY	B-A-a TETRACYCLINE RESISTANCE PROTEIN - STAPHYLOCOCCUS HYICUS (Schwartz et al. 1992)
64	S18539	B-A-? ACTVA-1 PROTEIN - STREPTOMYCES COELICOLOR (actinorhodin gene cluster) (Caballero et al. 1991)
65	TCMA_STRGA	B-A-? TCMA PROTEIN - STREPTOMYCES GLAUCESCENS TETRACENOMYCIN C RESISTANCE AND EXPORT PROTEIN. - (Guilfoile and Hutchinson 1992)
66	S25009	P-S-? SUGAR TRANSPORT PROTEIN STP4 - ARABIDOPSIS THALIANA (Sauer et al. 1992)
67	S25015	P-S-? MONOSACCHARIDE TRANSPORT PROTEIN MST1 - COMMON TOBACCO (Sauer and Stadler 1992)
68	A45611	Z-O-? PUTATIVE HEXOSE TRANSPORTER - TRYPANOSOMA BRUCEI (Bringaud and Baltz 1992)
69	B43319	M-M-? SYNAPTIC VESICLE AMINE TRANSPORTER, SVAT - RAT (Liu et al. 1992)
70	TCR4_ECOLI	TETRACYCLINE RESISTANCE PROTEIN, CLASS E. - ESCHERICHIA COLI. (Allard and Bertrand 1993)
71	ISTN10	B-A-a TRANSPOSON TN10 SEQUENCE ENCODING TETRACYCLINE RESISTANCE - ESCHERICHIA COLI PROKARYOTA (Hillen and Schollmeier 1983)
72	B48442	Z-S-? D2 = MEMBRANE TRANSPORT PROTEIN (CLONE D1.16.S) - LEISHMANIA DONOVANI (glucose transporter) (Langford et al. 1992)
73	TH11_TRYBB	Z-S-? GLUCOSE TRANSPORTER 1B/1C/1D/1F/2B. - TRYPANOSOMA BRUCEI BRUCEI. (Bringaud and Baltz 1993)
74	TH2A_TRYBB	Z-S-? GLUCOSE TRANSPORTER 2A. - TRYPANOSOMA BRUCEI TRYPANOSOMA BRUCEI (Bringaud and Baltz 1993) close relative: TBTHT3 T. brucei genes for hexose transporters - Trypanosoma brucei (Bringaud and Baltz 1993)
75	S14144	P-S-s C. KESSLERI HUP1 GENE FOR H(+)/HEXOSE-COTRANSPORTER - CHLORELLA KESSLERI EUKARYOTA (Sauer and Tanner 1989)

Table 1. Continued

76	S38453	P-S-s C. KESSLERI HUP2 M-RNA - CHORELLA KESSLERI EUKARYOTA (Sauer and Tanner 1989)
77	RCCSCP	P-S-? RICINUS COMMUNIS (CLONE PST293) SUGAR CARRIER PROTEIN (RCSTC) M-RNA, COMPLETE CDS. - RICINUS COMMUNIS EUKARYOTA (Weig et al. 1992)
78	RCCSCPS	P-S-? RICINUS COMMUNIS (CLONE PST29) SUGAR CARRIER PROTEIN (RCSTA) M-RNA, COMPLETE CDS. - RICINUS COMMUNIS EUKARYOTA (Weig et al. 1992)
79	GAL2_YEAST	F-S-? SACCHAROMYCES CEREVISIAE GALACTOSE TRANSPORTER (GAL2) GENE, COMPLETE CDS. - SACCHAROMYCES CEREVISIA EUKARYOTA (Nehlin et al. 1989)
80	GTR3_RAT	M-S-f RAT M-RNA FOR NEURONE GLUCOSE TRANSPORTER. - RATTUS NORVEGICUS EUKARYOTA (Nagamatsu et al. 1993)
81	RATGLUTV	M-S-f RATTUS NORVEGICUS FRUCTOSE TRANSPORTER (GLUT5) M-RNA, COMPLETE CDS. - RATTUS NORVEGICUS EUKARYOTA (Rand et al. 1993)
82	GALP_ECOLI	B-S-s GALACTOSE TRANSPORT PROTEIN (GALP) ESCHERICHIA COLI (Griffith et al. 1992)

[a] The sequences, codes, and brief descriptions were obtained by using the display facility of DELPHOS (see Materials and Methods) with the following exceptions. All descriptions from the database are in uppercase (certain older entries appear in lowercase). If erroneous biochemical descriptions (e.g., use of the word "permease") occur in the description these have not been corrected. Lowercase entries in the descriptions are for two purposes: (1) following the phrase "close relatives" as codes for proteins that showed very small differences in the cladograms and were eliminated to simplify the subsequent diagrams and (2) to cover ORF_BCNR (see text for discussion). Key: A avian; B bacterial; F fungal; M mammalian; P plant; Z protozoa; S sugar transporter; M metabolite (not sugar) transporter; A antibiotic resistance; O ORF; a antiport; s symport; f facilitated diffusion

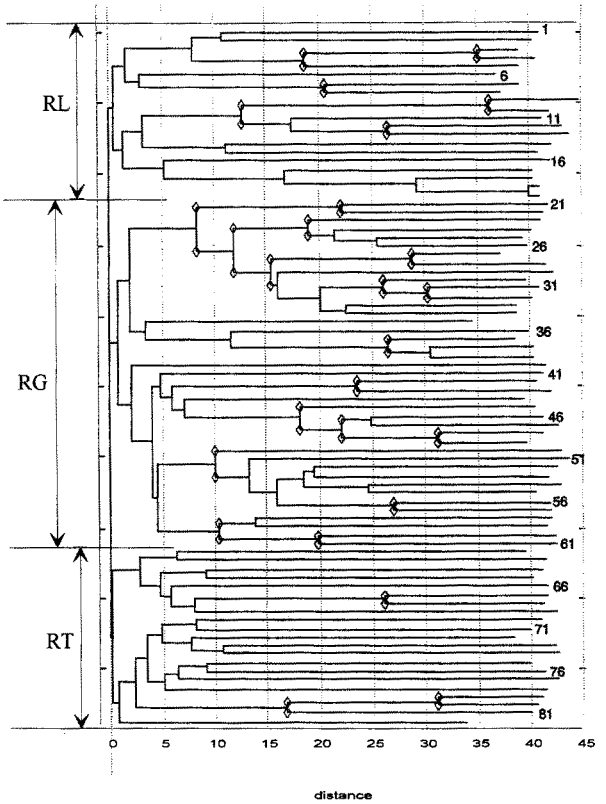

Fig. 1. Cladogram of members of the superfamily calculated by using a Risler matrix; 10,000 bootstrap attempts were made and nodes for which the confidence limit was 95% or more are marked with diamonds. Every fifth sequence is numbered using the convention noted in column (iii) of Table 3. The distance is the percent divergence (Kimura 1983).

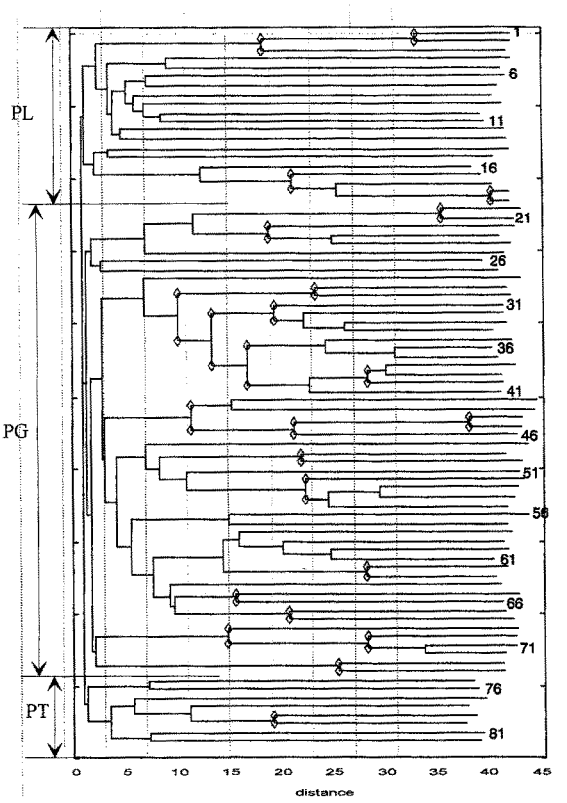

Fig. 2. Cladogram of members of the super family calculated by using a PAM matrix; 10,000 bootstrap attempts were made and nodes for which the confidence limit was 95% or more are marked with diamonds. Every fifth sequence is numbered using the convention noted in column (viii) of Table 3.

Pseudomonas spp. chloramphenicol resistance gene (*cmlA*) product CmlA, with highest sequence similarity predicted at the N-terminal (Lewis 1994; Bentley et al. 1993). The corresponding amino acid in CmlA is W (codon TGG).

As the proteins belong to several functional classes and come from widely divergent taxa we used two different matrices for constructing the cladograms. The results obtained by using the Risler matrix are shown in Fig. 1 and those by using the PAM250 matrix in Fig. 2.

Table 2. Summary of motifs used for analysis[a]

A. PB0001	[RK].GR[RK]	STRICT_12_MOTIF
B. PB0002	[RK]..[RK]	SLACK_12_MOTI1
C. PB0003	[RK]...[RK]	SLACK_12_MOTI2
D. PS00216	[LSTA][DE].[LFYA]GR[RK]....G	S_T_1
E. PS00217	[LF].G[LFA]..G.....[LIFY]..[EQ].....[RK]	S_T_2
F. PS00896	G[L][L].D[RK]LGL[RK][RK].[L][L]W	LACY_1
G. PS00897	P.[LF][LF]NR[L]G.KN[STA][L][L][L]	LACY_2
H. PS00873	Detail omitted	NA_ALANINE_SYMP
I. PS00211	Detail omitted	ABC_TRANSPORTER
J. PS00402	Detail omitted	BPD_TRANSP_INN_MEM

[a] Motifs A, B, and C represent the "strict" and general transporter motifs (Marger and Saier 1993), D and E are sugar transporter motifs while F and G are LacY motifs. H, I, and J represent typical motifs from other distinct transmembrane families (the details of these motifs have been omitted for clarity but can be accessed from PROSITE by using the motif code) (Bairoch and Bucher 1994). The listed sequences follow the REGEX convention (Materials and Methods). Thus motif A defines a sequence of five amino acids: either R or K; any residue; G; R; either R or K. Abbreviations used: [L] = [LIVM], S_T = SUGAR_TRANSPORTER, . = any residue

Both Figs. 1 and 2 are drawn as unrooted trees as there is no obvious outgroup for such a divergent set. In order to compare the two sets of data, we list the order in which these proteins appear in each cladogram in Table 3. We have also divided each of the two cladograms into the three large clades separated at the point of trichotomy. We have arbitrarily named these clades G, L, or T depending on the position of three characteristic proteins—namely, the facilitated glucose transport proteins GTR (G), the lactose/H^+ symport protein LACY (L), and the tetracycline resistance proteins TCR (T). We refer to the two matrices as PAM250 (P) and Risler (R); therefore, in Table 3 the PAM matrix L clade is termed PL and the corresponding Risler matrix clade is RL.

Part of our analysis relied on the prediction of motifs or signatures within the 82 proteins. For this we first abstracted all the PROSITE entries that matched the 82 sequences of Table 1; motifs that occurred on less than two occasions and non-discriminating motifs (such as leucine zipper, myristylation sites, etc.) were rejected. As a result, only four motifs survived (two sugar transporters and two LACY motifs). We also constructed three motifs based on the "strict" motif [RK]XG[RK] identified by Henderson (1990) and more general motifs [RK]$X_{2\text{-or-}3}$[RK] of Marger and Saier (1993). In these, X is any amino acid and [RK] means either R or K. These three motifs (termed A, B, and C) plus the PROSITE entries—sugar transport 1 and 2 (D and E); LACY 1 and 2 (F and G) along with three control motifs typical of other families of transmembrane proteins (H, I and J)—were used to scan the 82 sequences. These motifs are listed in Table 2. None of the proteins scored with the three control motifs—namely, the sodium/alanine symporter (H), the ABC superfamily motif (I), and mitochondrial inner membrane protein (J). Table 3 includes a summary of the motif searches.

Certain motifs do seem localized to particular clades or domains: the LACY motifs are restricted to a group of proteins within the domain RL/PL and sugar transport motifs are found in the PG/RG domain, although the sugar motif is also found in several proteins whose function is not known to involve the transport of sugars. The fact that all the sequences scored at least one of the motifs A, B, or C of Table 2 confirms that the objective method used to construct our set of sequences in Table 1 has not collected any spurious accidental entries, even though some of the sequences are open reading frames (ORFs) of no known function.

Although a superficial glance suggests differences between Figs. 1 and 2 and the corresponding parts of Table 3, the overall differences are not clear. In Fig. 3 we compare systematically the order of proteins in Figs. 1 and 2 by presenting them as a scatter plot. Following the clade terminology adopted above, LACY, for example, appears in both clades PL and RL, and this is referred to as the RL/PL "domain"; similarly, we refer to domains RG/PG and RT/PT. If the PAM250 and Risler cladograms were comparable, the points of Fig. 3 would lie on a diagonal and only the domains RL/PL, RG/PG, and RT/PT should be occupied. We have allocated the proteins that are clustered into "groups" (I–VIII). The small number of outlying proteins are given letters (*a–g*). In general, there is no significance in the fact that, for example, protein *e* (Table 3) appears as such an outlying sequence: it may be the only member of a larger cluster sequenced so far. The majority of sequences (subgroups III, IV, VI, and VII) do fall, more-or-less on the predicted diagonal. (Group II does not: it is a heterogeneous domain, RL/PG.) Subgroups IV and VI could probably be united; arguably *c* and *d* should be associated with subgroup IV and *a* with subgroup II; *f* seems to be intermediate between groups V and VI. Two of the heterogeneous domains (RL/PT and RG/PT) are unoccupied; in reality we suspect that RT/PG should be unoccupied because its sole member (*g*) is probably a redundant and erroneous sequence: the protein has the code JQ1479 (58 in Table 1) and is allegedly the same as TCR3_ECOLI (51 in Table 1). Protein 51 is a member of subgroup VII (RT/PT). Since discovering this anomaly, we have rechecked the sequences of these two proteins: the reported

Table 3. Summary of positions of 82 members of the superfamily within the two cladograms Fig. 1 (Risler—RIS) and Fig. 2 (PAM 250—PAM)[a]

(i) No.	(ii) Type	(iii) RIS	(iv) PAM	(v) domain	(vi) Group	Motifs ABC	Motifs DE	Motifs FG	(vii) PAM	(viii) No.
74	Z-S	1	43	RL, PG	I	−−+	+−	−−	1	69
72	Z-S	2	44	RL, PG	I	+++	−+	−−	2	36
73	Z-S	3	45	RL, PG	I	−++	−−	−−	3	35
68	Z-O	4	46	RL, PG	I	−++	−−	−−	4	65
40	Z-O	5	47	RL, PG	I	−++	−+	−−	5	64
12	B-O	6	28	RL, PG	a	+++	−−	−−	6	61
9	B-A	7	73	RL, PG	b	−++	+−	−−	7	60
6	B-O	8	74	RL, PG	b	−++	+−	−−	8	38
54	B-A	9	21	RL, PG	II	−++	−−	−−	9	56
52	B-A	10	22	RL, PG	II	−++	−−	−−	10	62
63	B-A	11	23	RL, PG	II	−++	−−	−−	11	37
53	B-A	12	24	RL, PG	II	−++	−−	−−	12	41
49	B-A	13	25	RL, PG	II	−++	−−	−−	13	8
47	B-A	14	26	RL, PG	II	−++	−−	−−	14	11
3	B-S	16	19	RL, PL	III	−++	−−	++	15	10
11	F-A	15	14	RL, PL	III	−++	−−	−−	16	2
2	B-S	17	16	RL, PL	III	−++	−−	++	17	5
4	B-S	18	18	RL, PL	III	−++	−−	++	18	4
5	B-S	19	17	RL, PL	III	−++	−−	++	19	3
1	B-S	20	20	RL, PL	III	−++	−−	−−	20	1
81	M-S	21	30	RG, PG	IV	−+−	−+	−−	21	54
31	M-S	22	31	RG, PG	IV	+++	−+	−−	22	52
32	M-S	23	32	RG, PG	IV	+++	−+	−−	23	63
23	M-S	24	33	RG, PG	IV	+++	−+	−−	24	53
24	M-S	25	34	RG, PG	IV	+++	−+	−−	25	49
22	M-S	26	35	RG, PG	IV	+++	−+	−−	26	47
30	M-S	27	37	RG, PG	IV	+++	−+	−−	27	17
29	M-S	28	36	RG, PG	c	+++	−+	−−	28	12
25	A-S	29	29	RG, PG	IV	+++	−+	−−	29	25
27	M-S	30	40	RG, PG	IV	+++	−+	−−	30	81
80	M-S	31	39	RG, PG	IV	+++	−+	−−	31	31
26	M-S	32	41	RG, PG	IV	+++	−+	−−	32	32
28	M-S	33	38	RG, PG	IV	+++	−+	−−	33	23
21	M-S	34	42	RG, PG	IV	+++	−+	−−	34	24
17	B-A	35	27	RG, PG	d	−++	−−	−−	35	22
16	B-M	36	69	RG, PG	V	+++	−+	−−	36	29
15	B-M	37	71	RG, PG	V	+++	++	−−	37	30
14	B-M	38	70	RG, PG	V	+++	++	−−	38	28
13	B-M	39	72	RG, PG	V	+++	++	−−	39	80
8	F-A	40	13	RG, PL	e	−++	++	−−	40	27
45	F-S	41	51	RG, PG	VI	−++	−+	−−	41	26
43	F-M	42	49	RG, PG	VI	−++	−+	−−	42	21
42	F-M	43	50	RG, PG	VI	−++	−+	−−	43	74
55	B-S	44	65	RG, PG	f	−++	−+	−−	44	72
57	F-S	45	48	RG, PG	VI	−++	−−	−−	45	73
44	F-S	46	54	RG, PG	VI	+++	−+	−−	46	68
34	F-S	47	52	RG, PG	VI	+++	++	−−	47	40
79	F-S	48	53	RG, PG	VI	−++	−+	−−	48	57
18	F-S	49	55	RG, PG	VI	−++	−+	−−	49	43
76	P-S	50	57	RG, PG	VI	+++	−+	−−	50	42
77	P-S	51	56	RG, PG	VI	+++	−+	−−	51	45
78	P-S	52	59	RG, PG	VI	+++	−+	−−	52	34
66	P-S	53	60	RG, PG	VI	+++	−+	−−	53	79
67	P-S	54	58	RG, PG	VI	+++	−+	−−	54	44
46	P-S	55	61	RG, PG	VI	+++	−+	−−	55	18
75	P-S	56	62	RG, PG	VI	+++	−+	−−	56	77
33	P-S	57	63	RG, PG	VI	+++	−+	−−	57	76
20	B-S	58	66	RG, PG	VI	+++	−+	−−	58	67
19	F-S	59	64	RG, PG	VI	+++	−+	−−	59	78
82	B-S	60	67	RG, PG	VI	+++	−+	−−	60	66
7	B-S	61	68	RG, PG	VI	+++	−+	−−	61	48
51	B-A	62	75	RT, PT	VII	+++	−−	−−	62	75

Table 3. Continued

(i) No.	(ii) Type	(iii) RIS	(iv) PAM	(v) domain	(vi) Group	Motifs ABC	Motifs DE	Motifs FG	(vii) PAM	(viii) No.
48	B-A	63	76	RT, PT	VII	+++	−−	−−	63	33
71	B-A	64	77	RT, PT	VII	+++	−−	−−	64	19
70	B-A	65	78	RT, PT	VII	+++	−−	−−	65	55
59	B-A	66	81	RT, PT	VII	+++	++	−−	66	20
58	B-A	67	70	RT, PT	g	+++	−−	−−	67	82
50	B-A	68	80	RT, PT	VII	+++	−−	−−	68	7
39	B-A	69	82	RT, PT	VII	−++	−−	−−	69	16
64	B-A	70	5	RT, PL	VIII	+++	−−	−−	70	14
61	B-A	71	6	RT, PL	VIII	+++	−−	−−	71	15
60	B-A	72	7	RT, PL	VIII	+++	+−	−−	72	13
62	B-A	73	10	RT, PL	VIII	−++	−−	−−	73	9
38	B-A	74	8	RT, PL	VIII	−++	−−	−−	74	6
65	B-A	75	4	RT, PL	VIII	−++	−−	−−	75	51
56	B-O	76	9	RT, PL	VIII	−++	−−	−−	76	48
41	B-A	77	12	RT, PL	VIII	−++	−−	−−	77	71
37	B-A	78	11	RT, PL	VIII	−++	−−	−−	78	70
69	M-M	79	1	RT, PL	VIII	−++	−−	−−	79	58
36	M-M	80	2	RT, PL	VIII	−++	−−	−−	80	50
35	M-M	81	3	RT, PL	VIII	−++	−−	−−	81	59
82	B-O	82	15	RT, PL	VIII	+++	−−	−−	82	39

[a] The first two columns of the table correspond to the number of each protein (Table 1) and type of transporter (as in Table 1 but with the mechanism omitted). Column (iii) lists the order in which these proteins appear in Fig. 1 while the corresponding position of the same proteins in the PAM matrix is shown in column (iv). Columns (v) and (vi) (domain and group) are discussed in the text. The column headed "motifs" is a summary of the scores of the motifs of Table 2 in that order. Proteins that do not fall into groups I–VIII are indicated by a letter (a–g). Thus, for example, the first entry (sequence 74 in Table 1) lacks motifs A and B, contains motifs C and D, and lacks motifs E, F, and G. Note that the entry (for the first three motifs) +− is impossible as motif A is a special case of motif C (Table 2). The last two columns repeat the data of columns (i) and (ii) ordered as in Fig. 2

differences are genuine but, as TCR3_ECOLI is the more recently deposited sequence, it is presumably the correct sequence. Although JQ1479 is probably an erroneous sequence we have left it in place for two reasons: first, our survey of the databases was objective and we avoid eliminating awkward sequences; second, the identification of this unique member of a domain in Fig. 3 and the recognition of the fact that other tetracycline resistance proteins occur elsewhere (see Discussion) provide another application of our method of analysis.

We consider briefly the members of the heterogeneous domains and others that are not within the "diagonal" subgroups II, IV, V, and VII. Within the domain RL/PG the five protozoal sugar transporters (proteins numbered 1–5 in the Risler order) form a distinct subgroup (I) of proteins which lie to the left of the diagonal. The two translations of the *bcr* gene (*b*) and an *E. coli* ORF (*a*) are outliers. Proteins of group II (also in domain RL/PG) are exclusively bacterial antibiotic transporters (numbers 9–14). Group V, although in the domain RG/PG, is set apart from the diagonal: this comprises bacterial metabolite transporters. Protein *e* is the aminotriazole resistance protein from *Saccharomyces cerevisiae,* ATR1_YEAST. Protein *f* is the xylose-H^+ symporter from *E. coli,* XYLE_ECOLI, which is known to have a larger cytoplasmic domain between helices 6 and 7 than that predicted for most other proteins within this superfamily (Griffith et al. 1992). Group VIII (the entire RT/PL domain) contains protein sequences that are not closely related (Figs. 1 and 2) and these comprise bacterial antibiotic resistance proteins, ORFs, and some mammalian metabolite transport proteins. This subgroup also includes the two proteins MMR_BACSU (37 in Table 1) and QACA_STAAU (41 in Table 1); there is evidence that these two contain 14 rather than 12 transmembrane helices.

Sequences in groups other than II, III, IV, VI, and VII (Fig. 3) deviate significantly from a linear relationship in the plot. We analyzed further several such sequences. It is possible that some of the abnormalities might arise if some of the proteins arose as a result of gene fusions. Our reason for making this supposition was that it has been suggested by other sources (including Baldwin 1994) that the genes for such proteins might have arisen from duplication of an archaic gene. The evidence comes from the observation that there are broad similarities between the "left part" of the protein sequence (containing predicted transmembrane helices 1–6) and the "right part" (containing predicted helices 7–12) for several proteins of this superfamily. We chose individual or representative sequences that lie remote from the diagonal in Fig. 3 and scanned the 82 original protein sequences

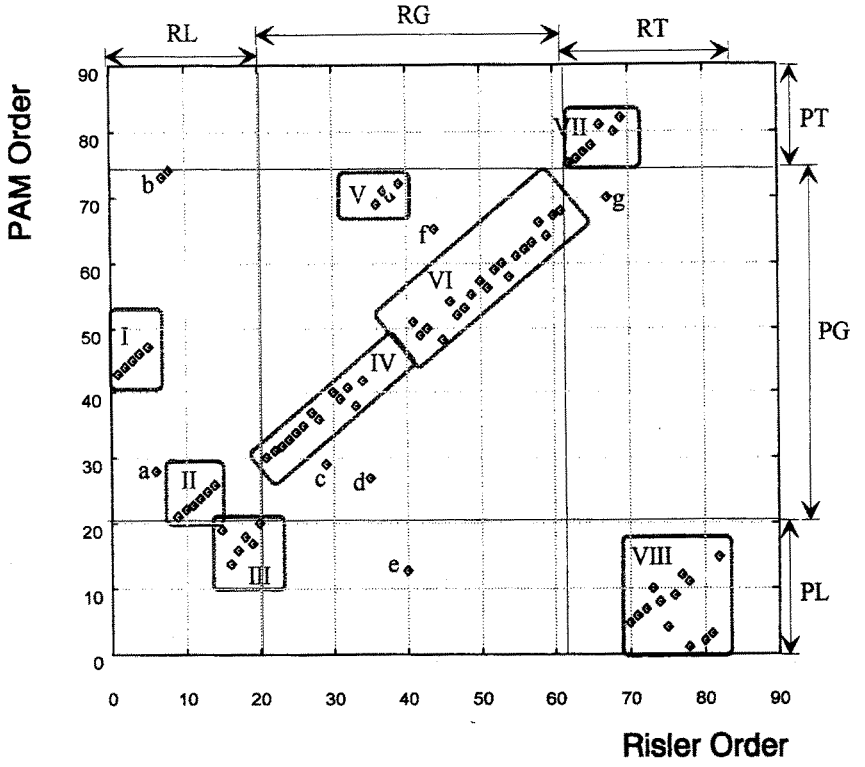

Fig. 3. Scatter diagram of the orders of entries in Figs. 1 and 2 and Table 3. Each point represents one sequence: the x-axis is the order for the Risler matrix (Fig. 1) and the y-axis that from the PAM matrix (Fig. 2). The protein can be referred to from Table 3. For example, the point at coordinates Risler 21, PAM 30 is RATGLUV (No. 81 in Table 1). The three clades (RL, RG, RT) for each of the cladograms based on the Risler matrix are delimited by vertical lines and those for the PAM matrix (PL, PG, PT) are delimited by horizontal lines. Therefore, the bottom left rectangle is the " domain" RL/PL (see text). The eight major groups are identified and numbered I–VIII.

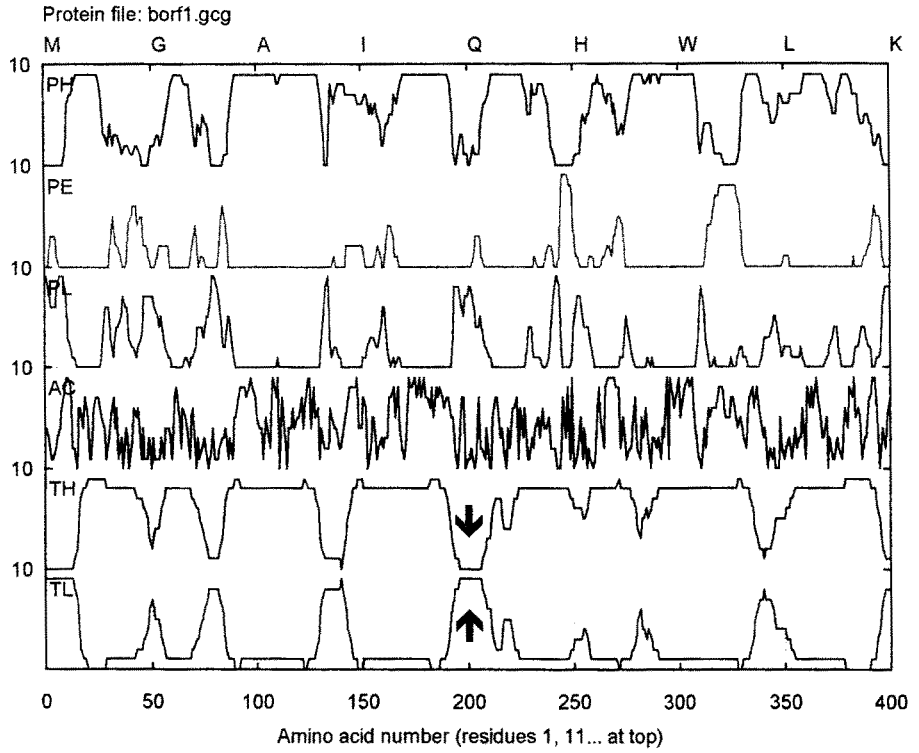

Fig. 4. The predicted secondary structure of ORF_BCNR (Table 1, entry 6). The output from the network was plotted in six horizontal panels using a score of 0 to 9 in each case. The abscissa is the protein sequence (every tenth amino acid residue is written along the top). The panels (from the top to the bottom) are as follows: *PH* is the helix propensity; *PE* is the beta-sheet propensity; *PL* is the loop propensity; *AC* is the predicted accessibility. In this case the neural network predicted that the protein was a transmembrane type and generated additional data plotted in the bottom two panels: *TH* is a measure of the likelihood of a transmembrane helix and *TL* is the likelihood of an interhelix loop. *TH* and *TL* should be (and indeed are) the inverse of one another. The arrows (↑↓) represent the gap between helices 6 and 7.

with the left and right parts of these representative sequences. First the point of division was identified by objective methods. A position between predicted helices 6 and 7 was identified by using either the PHD neural network program or the PEPPLOT program. As an example, output from PHD is shown in Fig. 4 for ORF_BCNR (No. 6 in Table 1).

The half sequences used are summarized in Table 4 and Fig. 5. Evidence for such internal homology in this set of "anomalous" sequences is partial: the two halves

Table 4. Half sequences in the order of the cladogram of Fig. 5[a]

Order (Fig. 5)	Code	No.	Side	Clade
1	MATS_RAT	36	Left	X
2	ATR1_YEAST	8	Right	
3	TCR_STAAU	54	Right	
4	LACY_ECOLI	3	Left	
5	ORF_BCNR	6	Right	
6	MATS_RAT	36	Right	
7	ATR1_YEAST	8	Left	
8	LACY_ECOLI	3	Right	
9	TBTH5	74	Left	
10	XYLE_ECOLI	55	Left	
11	HUP1_CHLKE	75	Left	
12	TCR1_ECOLI	48	Right	Y
13	JQ1201	60	Right	
14	TCR_STAAU	54	Left	
15	TCR1_ECOLI	3	Left	
16	CMLR_STRLI	17	Right	
17	CMLR_STRLI	17	Left	
18	JQ1201	60	Left	
19	ORF_BCNR	6	Left	
20	TBTH5	74	Right	Z
21	XYLE_ECOLI	55	Right	
22	HUP1_CHLKE	75	Right	

[a] The codes are those of the sequences from which these left and right halves were derived and the column headed No. cross-refers to the numbers in Table 1

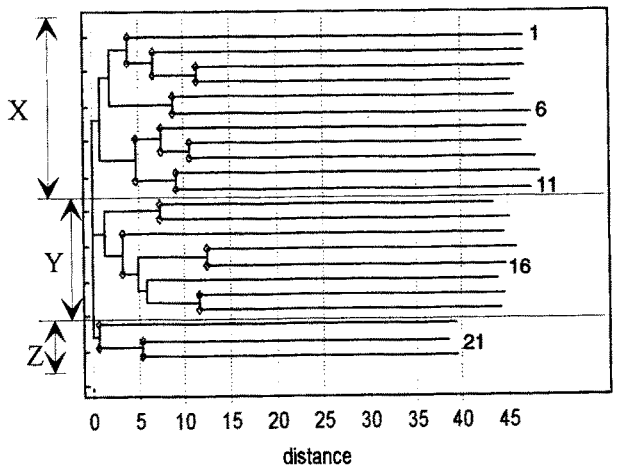

Fig. 5. Cladogram of half sequences of selected members of the superfamily calculated by using a Risler matrix; 10,000 bootstrap attempts were made and nodes for which the confidence limit was 80% or more are marked with diamonds. Every fifth sequence is indicated by its position in Table 4.

of CMLR_STRLI (16 and 17 in Fig. 5) are clearly related. Otherwise, of the three main clades (X, Y, and Z in Fig. 5), clade Z contains only right halves; ORF_BCNR, which occupies a highly anomalous position in Fig. 3 is confirmed as having its left half related to JQ1201 whereas its right half related to other proteins (not any of those that appear to be anomalous and hence are not represented in Fig. 5). We had previously observed this by qualitative examination of a 2-D plot (Bentley et al. 1993).

Discussion

If the 12-helix/14-helix transmembrane proteins are referred to as a superfamily, we propose that our 8 groups and the domains of Table 3 can be used as a working categorization. Given this, it is clear that the bacterial antibiotic resistance proteins are not clustered as Lewis (1994) predicted but occur in domains RL/PG, RG/PG, RT/PT, and RT/PL and in groups II, III, V, VIII, and XI. Moreover, tetracycline resistance can be mediated by proteins that are poorly related whether one chooses structural (''Risler matrix'') or mutational (''PAM matrix'') criteria. The qualitative picture emerges of the superfamily comprising members that are grouped in such a way that for the majority, mutational considerations are not much constrained by selective fitness (this majority is represented by sequences that lie on or near the diagonal of Fig. 3). This provides a partly quantitative justification for the case that the superfamily is representative of a versatile basic structure that has been recruited for a variety of transport purposes (see for example Griffith et al. 1992). However, we see no evidence of a candidate for a single ancestral protein of this type. There may be cases of divergent and convergent evolution in the superfamily: the use of the Risler matrix would not distinguish between these alternatives, and sequences that lie off the diagonal of Fig. 3 might include cases of convergent evolution However, it is at present difficult to speculate about this for two reasons: there is some, albeit weak, evidence for duplication and fusion referred to in our consideration of Fig. 5 and references such as Baldwin (1994). A second complicating factor is that as many of these are bacterial proteins and several are known products of plasmids, transposons, and integrons, the evolution of this superfamily will have included many cases of the lateral transmission of genetic information across large taxonomic gaps. We propose that one extension of the work might be to contrast codon usage in the genes for such proteins: to take one single example that we have considered briefly before (Bentley et al. 1993), the similarity of the N-terminal portion of Bcr and that of the *Pseudomonas* Cml protein suggests that a gene has been transferred from one organism to another in which a novel codon bias must represent a new selection pressure.

The methods developed for this analysis can obviously be applied with benefit to other cases of proteins of related function and/or structure (or presumed structure) but for which there is little apparent sequence homology. One example would be the lipocalins which have been the subject of independent approach searching for motifs by a statistical method (Lawrence et al. 1993). However, we should like to see two extensions of our method. First, data are lost in the representation of Fig. 3 because it deals only with the order of sequences in the cladograms: the points in such a plot should be distances although we recognize there is a problem of data repre-

sentation. Second, we believe that the method should be made generic: the principle of using two different similarity matrices for analyzing sequence relationships is, we hope, justified by this work. The receptor proteins of Le Novère and Changeaux (1995), which form another rather more conserved family, represent a good example of a set of proteins to which, we believe, our method could usefully be applied. More generally, the method could be applied to deducing the similarity matrices that actually apply and, by replacing the equivalent of Fig. 3 by an array in multidimensional space, the rules that determine amino acid replacement in a superfamily could objectively be determined.

Acknowledgments. J.B. was an SERC CASE student (with Smith-Kline Beecham) at the time of the experimental work on the *bcr* gene. The work was supported by an SERC research grant to J.H.P., P.J.F. Henderson, and S.A. Baldwin. We are grateful to A.J. Bleasby of DRAL (Daresbury) for helpful advice, and part of the computational work was performed using the SEQNET platform at DRAL.

References

Akrigg D, Attwood TK, Bleasby AJ, Findlay JBC, North ACT, Maughan NA, Parry-Smith DJ, Perkins DN, Wootton JC (1992) SERPENT—an information storage and analysis resource for protein sequences. CABIOS 8:295–296

Allard JD, Bertrand KP (1993) Sequence of a class E tetracycline resistance gene from *Escherichia coli* and comparison of related tetracycline efflux proteins. J Bacteriol 175:4554–4560

Allmeier H, Cresnar B, Greck M, Schmitt R (1992) Complete nucleotide sequence of Tn1721: gene organization and a novel gene product with features of a chemotaxis protein. Gene 111:11–20

Asano T, Shibasaki Y, Jasuga M, Kanazawa Y, Takaku F, Akanuma Y, Oka Y (1988) Cloning of a rabbit brain glucose transporter cDNA and alteration of glucose transport messenger RNA during tissue development. Biochem Biophys Res Commun 154:1204–1211

Asano T, Shibasaki Y, Lin JL, Akanuma Y, Takaku F (1989) The nucleotide sequence of cDNA for a mouse liver-type glucose transporter protein. Nucleic Acids Res 17:6386

Aslinidis C, Schmidt K, Schmidt R (1989) Nucleotide sequences and operon structure of plasmid-borne genes mediating uptake and utilization of rafinose in *Escherichia coli.* J Bacteriol 171:6753–6763

Bairoch A, Bucher P (1994) Prosite—recent developments Nucleic Acids Res 22:3583–3589

Baldwin SA (1994) Mammalian passive glucose transporters: members of an ubiquitous family of active and passive transport proteins. Biochim Biophys Acta 1154:17–49

Barnell WO, Yi KC, Conway T (1990) Sequence and genetic organization of a *Zymomonas mobilis* gene cluster that encodes several enzymes of glucose metabolism. J Bacteriol 172:7227–7240.

Bentley J, Hyatt LS, Ainley K, Parish JH, Herbert RB, White GR (1993) Cloning and sequence analysis of an *Escherichia coli* gene conferring bicyclomycin resistance. Gene 127:117–120

Birnbaum MJ (1989) Identification of a novel gene encoding an insulin responsive glucose transporter protein. Cell 57:305–315

Bleasby AJ, Wootton JC (1990) Construction of validated, non-redundant protein sequence databases. Protein Eng 3:153–159

Boado RJ, Pardridge WM (1991) Molecular cloning of the bovine blood-brain barrier glucose transporter cDNA—phylogenetic conservation of the 5′-untranslated region of the GLUT-1 isoform. Clin Res 39:A38

Bockmann J, Heuel H, Lengeler J (1992) Characterization of a chromosonally encoded, non-PTS metabolic pathway for sucrose utilization in *Escherichia coli.* Mol Gen Genet 235:22–32

Bringaud F, Baltz T (1992) A potential hexose transporter gene expressed predominantly in the bloodstream form of *Trypanosoma brucei.* Mol Biochem Parasitol 52:111–121

Bringaud F, Baltz T (1993) Differential regulation of 2 distinct families of glucose transporter genes in *Trypanosoma brucei.* Mol Cell Biol 13:1146–1154

Burland V, Plunkett G, Daniels DL, Blattner FR (1993) DNA sequence and analysis of 136 kilobases of the *Escherichia coli* geneome—organizational symmetry around the origin of replication. Genomics 16:551–561

Caballero JL, Martinez E, Malpartida F, Hopwood DA (1991) Organisation and functions of the *actVA* region of the actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor.* Mol Gen Genet 230:401–412

Cairns BR, Collard MW, Landfears SM (1989) Developmentally regulated transporter in *Leishmania* encodes a putative membrane transport protein (parasitic protozoan gene expression glucose transporter). Proc Natl Acad Sci USA 86:7682–7686

Dayhoff MO, Schwartz RM, Orcutt BC (1978) In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, supplement 3. NBRF, Washington, DC, p 345

Devereux J, Haeberli P, Smithies D (1984) A comprehensive set of sequence-analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Dittrich W, Betzler M, Schrempf H (1991) The unstable tetracycline resistance gene of *Streptomyces lividans* encodes a putative transmembrane protein. Mol Microbiol 5:2789–2797

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fernandez-Moreno MA, Caballero JL, Hopwood DA, Malpartida F (1991) The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces.* Cell 66:769–780

Fling ME, Kopf J, Tamarkin A, Gorman JA, Smith HA, Koltin Y (1991) Analysis of a *Candida albicans* gene that encodes a novel mechanism for resistance to benomyl and methotrexate. Mol Gen Genet 227:318–329

Fukumoto H, Seino S, Imura H, Seino Y, Eddy RL, Fukushima Y, Byers MG, Shows TB, Bell GI (1988) Sequence, tissue distribution and chromosomal localization of messenger RNA encoding a human glucose transporter-like protein. Proc Natl Acad Sci USA 85: 5434–5438

Fukumoto H, Kayano T, Buse JB, Edwards Y, Pilch PF, Bell GI (1989) Cloning and characterization of the major insulin-responsive glucose transporter expressed in human skeletal muscle and other insulin-responsive tissues. J Biol Chem 264:7776–7779

Geever RF, Huiet L, Baum JA, Tyler BM, Patel VB, Rutledge BJ, Case ME, Giles NH (1989) DNA sequence, organization and regulation of the *qa* gene cluster of *Neurospora crassa.* J Mol Biol 207:15–34

Goempel-Klein P, Brendel M (1990) Allelism of SNQ1 and ATR1, genes of the yeast *Saccharomyces cerevisiae* required for controlling sensitivity to 4-nitroquinoline N-oxide and aminotriazole. Curr Genet 18:93–96

Griffith JF, Baker ME, Rouch DA, Page MGP, Skurray RA, Paulsen IT, Chater KF, Baldwin SA, Henderson PJF (1992) Membrane transport proteins: implications of proteins sequences. Curr Opin Cell Biol 4:684–695

Guilfoile PG, Hutchinson CR (1992) The *Streptomyces glaucescens* TcmR protein represses transcription of the divergently oriented *tcmR* and *tcmA* genes by binding to an intergenic operator region. J Bacteriol 174:3650–3666

Hawkins AR, Lamb HK, Smith M, Keyte JW, Roberts CF (1988) Molecular organization of the quinic acid utilization *(qut)* gene cluster in *Aspergillus midulans.* Mol Gen Genet 214:224–231

Henderson PJF (1990) The homologous glucose-transport proteins of prokaryotes and eukaryotes. Res Microbiol 141:384–395

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Hillen W, Schollmeier K (1983) Nucleotide sequence of the Tn10 encoded tetracycline resistance gene. Nucleic Acids Res. 11:525–539

Hoshino T, Ikeda T, Tomizuka N, Furukawa K (1985) Nucleotide sequence of the tetracycline resistance gene of pTHT15, a thermophilic *Bacillus* plasmid—comparison with staphylococcal TCR controls. Gene 37:131–138

Ishiguro N, Sato G (1985) Nucleotide sequence of the gene determining plasmid mediated citrate utilization. J Bacteriol 164:977–982

Kaback HR (1990) The Lac permease of *Escherichia coli*—a prototypic energy-transducing membrane protein. Biochim Biophys Acta 1018:160–162

Kaestner KH, Christy RJ, McLenithan JC, Braiterman LT, Cornelius P, Pekala PH, Lane MD (1989) Sequence, tissue distribution and differential expression of messenger RNA for a putative insulin-responsive glucose transporter in mouse 3T3-L1 adipocytes. Proc Natl Acad Sci USA 86:3150–3154

Kayano T, Burant CF, Fukumoto H, Gould GW, Fan Y-S, Eddy RL, Byers MG, Shows TB, Seino S, Bell GI (1990) Human facilitative glucose transporters—isolation, functional characterization, and gene localization of cDNAs encoding an isoform (GLUT5) expressed in small intestine, muscle, and adipose tissue and an unusual glucose transporter pseudogene-like sequence (GLUT6). J Biol Chem 265:13276–13282

Kayano T, Fukumoto H, Eddy RL, Fan Y-S, Byers MG, Shows TB, Bell GI (1988) Evidence for a family of human glucose transporter-like proteins—sequence and gene localization of a protein expressed in fetal skeletal muscle and other tissues. J Biol Chem 263:15245–15248

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England

Ko CH, Liang H, Gaber RF (1993) Roles of multiple glucose transporters in *Saccharomyces cerevisiae*. Mol Cell Biol 13:638–648

Kolakowski LF, Leunissen JAM, Smith JE (1992) Prosearch—fast searching of protein sequences with regular expression patterns related to protein structure and function. Biotechniques 13:919–921

Kopelowitz J, Hampe C, Goldman R, Reches M, Engelberg-Kulka H (1992) Influence of codon context on UGA supression and readthrough. J Mol Biol 225:261–269

Kruckeberg AL, Bisson LF (1990) The HXT2 gene of *Saccharomyces cerevisiae* is required for high affinity glucose transport. Mol Cell Biol 10:5903–5913

Langford CK, Ewbank SA, Hanson SS, Ullman B, Landfear SM (1992) Molecular characterization of two genes encoding members of the glucose transporter superfamily in the parasitic protozoan *Leishmania donovani*. Mol Biochem Parasitol 55:51–64

Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald F, Wootton JC (1993) Detecting subtle sequence signals—a Gibbs sampling strategy for multiple alignment. Science 262:208–214

Le Novère N, Changeaux J-P (1995) Molecular evolution of the nicotinic acetylcholine receptor: an example of multigene family in excitable cells. J Mol Evol 40:155–172

Lewis K (1994) Multidrug resistance pumps in bacteria: variations on a theme. Trends Biochem Sci 19:119–123

Liu Y, Peter D, Roghani A, Schuldiner S, Prive GG, Eisenberg D, Brecha N, Edwards RH (1992) A cDNA that suppresses MPP^+ toxicity encodes a vesicular amine transporter. Cell 70:539–551

Lu Z, Lin CC (1989) The nucleotide sequence of *Escherichia coli* genes for L-fucose assimilation. Nucleic Acids Res 17:4883–4884

McMorrow I, Chin DT, Fiebig K, Pierce JL, Wilson DM, Reeve ECR, Wilson TH (1988) The lactose carrier of *Klebsiella pneumoniae* M5A1—the physiology of transport and the nucleotiude sequence of the *lacY* gene. Biochim Biophys Acta 945:315–323

Maiden MCJ, Davis EO, Baldwin SA, Moore DCM, Henderson PJF (1987) Mammalian and bacterial sugar-transport proteins are homologues. Nature 325:641–643

Marger MD, Saier MH (1993) A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. Trends Biochem Sci 18:13–20

Marshall-Carlson L, Celenza JL, Laurent BC, Carlson M (1990) Mutational analysis of the SNF3 glucose transporter of *Saccharomyces cerevisiae*. Mol Cell Biol 10:1105–1115

Mitchell PD, Skulachev VP, Hinkle PC (1981) Chemiosmotic protein circuits: in honour of Peter Mitchell. Addison-Wesley, Reading, MA

Mojumdar M, Khan SAJ (1988) Characterization of the tetracycline resistance gene of plasmid pT181 of *Staphylococcus aureus*. J Bact 170:5522–5528

Mueckler M, Caruso C, Baldwin SA, Panico M, Blenchi I, Morris HR, Allard WJ, Lienhard GE, Lodish HF (1985) Sequence and structure of a human glucose transporter. Science 229:941–945

Nagamatsu S, Kornhauser JM, Seino S, Mayo KE, Steiner DF, Bell GI (1992) Glucose transporter expression in brain—cDNA sequence of mouse GLUT3, the brain facilitative glucose transporter isoform, and idenfication of sites by in situ hybridization. J Biol Chem 267:467–472

Nagamatsu S, Sawa H, Kamada K, Nakamichi Y, Yoshimoto K, Hoshino T (1993) Neuron-specific glucose transporter (NSGT)-CNS distribution of GLUT3 rat glucose transporter (RGT3) in rat central neurons. FEBS Lett 334:289–295

Neal RJ, Chater KF (1987) Nucleotide sequence analysis reveals similarities between proteins determining methylenomycin resistance in *Streptomyces* and tetracycline resistance in eubacteria. Gene 58: 229–241

Neyfakh AA, Bidenko VE, Chen LB (1991) Efflux-mediated multidrug resistance in *Bacillus subtilis*—similarities and dissimilarities with the mammalian system. Proc Natl Acad Sci USA 88:4781–4785

Nehlin JO, Carlberg M, Ronne H (1989) Yeast glactose permease is related to yeast and mammalian glucose transporters. Gene 85:313–319

Nguyen TT, Postle K, Bertrand KP (1983) Sequence homology between the tetracycline resistance determinants of pBR322 and Tn10. Gene 25:83–92

Nikaido H (1994) Prevention of drug access to bacterial targets: permeability barriers and active efflux. Science 164:382–388

Noguchi N, Aoki T, Sasatsu M, Kono M, Shishido K, Ando T (1986) Determination of the complete nucleotide sequence of pNS1, a staphylococcal tetracycline resistance plasmid propagated in *Bacillus subtilis*. FEMS Microbiol Lett 37:283–288

Palva A, Vidgren G, Simonen M, Rintala H, Laamaen P (1990) Nucleotide sequence of the tetracycline resistance gene pBC16 from *Bacillus cereus*. Nucleic Acids Res 18:1635

Peden KWC (1983) Revised sequence of the tetracycline resistance gene of pBR322. Gene 22:277–280

Putzer H, Gendron N, Grunberg-Manago M (1992) EMBL/GenBank submission

Rand EB, Depaoli AM, Davidson NO, Bell GI, Burant CF (1993) Sequence, tissue distribution, and functional characterization of the rat fructose transporter GLUT5. Am J Physiol 264:G1169–G1176

Risler JL, Delorme MO, Delacroix A (1988) Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. J Mol Biol 24: 1019–1029

Rost B, Sander M (1994) Combining evolutionary information and neural networks to predict protein secondary structure. Proteins Struct Funct Genet 19:55–72

Rouch DA, Cram DS, DiBernadino D, Littlejohn TG, Skurray RA (1990) Efflux mediated antiseptic resistance gene qacA from Staphylococcus aureus—common ancestry with tetracycline-transport and sugar-transport proteins. Mol Microbiol 4:2051–2062

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sakaguchi R, Amano H, Shishido K (1988) Nucleotide sequence homology of the tetracycline resistance determinant naturally maintained in *Bacillus subtilis* Marburg-168 and the tetracycline resistance gene of *B. subtilis* plasmid pNS1981. Biochim Biophys Acta 950:441–444

Sasastsu M, Mishra TK, Chu L, Laddaga R, Silvers S (1985) Cloning and DNA sequence of a plasmid determined citrate utilization system in *Escherichia coli*. J Bacteriol 164:983–993

Sauer N, Friedländer K, Gräml-Wicke U (1990) Primary structure, genomic organization and heterologous expression of a glucose transporter from *Arabidopsis thaliana*. EMBO J 9:3045–3050

Sauer N, Stadler R (1992) EMBL data library submission

Sauer N, Illig J, Baier K, Stadler R (1992) EMBL data library submission

Sauer N, Tanner W (1989) The hexose carrier from *Chlorella*—cDNA cloning of a eukaryotic H^+-cotransporter. FEBS Lett 259:43–46

Schwarz S, Cardoso M, Wegener HC (1992) Nucleotide sequence and phylogeny of the *tet*(L) tetracycline resistance determinant encoded by plasmid pSTE1 from *Staphylococcus hyicus*. Antimicrob Agents Chemother 36:580–588

Shimamoto T, Izawa H, Daimon H, Ishiguro N, Shinagawa M, Sakano Y, Tsudo M, Tsuchiya T (1991) Cloning and nucleotide sequence analys of the gene (*citA*) encoding a citrate carrier from *Salmonella typhimurium*. J Biochem 110:22–28

Stokes HW, Hall RM (1991) Sequence analysis of the inducible chloramphenicol resistance determinant in the Tn1696 integron suggests regulation by translational attenuation. Plasmid 26:10–19

Suzue Z, Lodish HF, Thorens B (1989) Sequence of the mouse liver glucose transporter. Nucleic Acids Res 17:10099–10099

Thorens B, Sarkar HK, Kaback HR, Lodish HF (1988) Cloning and functional expression in bacteria of a novel glucose transporter present in liver, intestine, kidney, and beta-pancreatic islet cells. Cell 55:281–290

van der Lelie D, Bron S, Venema G, Oskam L (1989) Similarity of minus origins of replication and flanking open reading frames of plasmids pUB110, pTB913 and pMV158. Nucleic Acids Res 17: 7283–7294

van der Rest ME, Schwartz E, Oesterhelt D, Konings WN (1990) DNA sequence of a citrate carrier of *Klebsiella pneumoniae*. Eur J Biochem 189:401–407

Waddell ID, Zomerschoe AG, Voice MW, Burchell A (1992) Cloning and expression of a hepatic microsomoal glucose transport protein—comparison with liver plasma membrane glucose transport protein GLUT-2. Biochem J 286:173–177

Waters SH, Rogowsky P, Grinsted J, Altenbuchner J, Schmitt R (1983) The tetracycline resistance determinants of RP1 and Tn1721—nucleotide sequence analysis. Nucleic Acids Res 11:6089–6105

Weig A, Franz J, Sauer N, Komor E (1992) EMBL data library submission

Weiler-Güttler H, Zinke H, Moeckel B, Frey A, Gassen HG (1989) cDNA cloning and sequence analysis of the glucose transporter from porcine blood-brain barrier. Biol Chem Hoppe Seyler 370: 467–473

Wesolowski-Louvel M, Giffrini P, Ferrero I, Fukuhara H (1992) Glucose transport in the yeast *Kluyveromyces lactis*. 1. Properties of an inducible low-affinity glucose transporter gene. Mol Gen Genet 233:89–96

White MK, Rall TB, Weber M (1991) Differential expression of glucose transporter isoforms by the src oncogene in chicken-embryo fibroblasts. J Mol Cell Biol 11:4448–4454

Williams SA, Birnbaum MJ (1988) The rat facilitated glucose transporter gene—transformation and serum-stimulated transcription initiate from identical sites. J Biol Chem 263:19513–19518

Williams T, Kelley C (1992) GNUplot software. email: info-gnuplot-request@ames.arc.nasa.gov

Yoshida H, Bogaki M, Nakamura S, Ubukata K, Konno M (1990) Nucleotide sequence and characterization of the *Staphylococcus aureus norA* gene, which confers resistance to quinolones. J Bacteriol 172:6942–6949

Zhang CC, Durand MC, Jeanjean R, Joset F (1989) Molecular and genetic analysis of the fructose-glucose transport system in the cyanobacterium *Synechocystis* PCC6803. Mol Microbiol 3:1221–1229

Zhang HZ, Schmidt H, Piepersberg W (1992) Molecular cloning and characterization of two lincomycin-resistance genes, *ImrA* and *ImrB*, from *Streptomyces lincolnensis* 78–11. Mol Microbiol 6: 2147–2157