The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations

Donald L. Jack, Ian T. Paulsen† and Milton H. Saier, Jr

Author for correspondence: Milton H. Saier, Jr. Tel: +1 858 534 4084. Fax: +1 858 534 7108.
e-mail: msaier@ucsd.edu

Department of Biology,
University of California
at San Diego, La Jolla,
CA 92093-0116, USA

In this paper an analysis of 175 currently sequenced transport proteins that comprise the amino acid/polyamine/organocation (APC) superfamily is reported. Members of this superfamily fall into 10 well-defined families that are either prokaryote specific, eukaryote specific or ubiquitous. Most of these proteins exhibit 12 probable transmembrane spanners (TMSs), but members of two of these families deviate from this pattern, exhibiting 10 and 14 TMSs. All members of these families are tabulated, their functional properties are reviewed and phylogenetic/sequence analyses define the evolutionary relationships of the proteins to each other. Evidence is presented that the APC superfamily may include two other currently recognized families that exhibit greater degrees of sequence divergence from APC superfamily members than do the proteins of the 10 established families from each other. At least some of the protein members of these two distantly related families exhibit 11 established TMSs. Altogether, the APC superfamily probably includes 12 currently recognized families with members that exhibit exclusive specificity for amino acids and their derivatives but which can possess 10, 11, 12 or 14 TMSs per polypeptide chain.

Keywords: amino acids, transport, evolution, superfamily, secondary carriers

INTRODUCTION

During the early evolution of life on Earth, before archaea and eukarya diverged from bacteria, life forms were undoubtedly much simpler than they are now. The primordial prokaryotic cell probably possessed few transporters and these served the essential functions of maintaining cytoplasmic ionic homeostasis and transmembrane electrical potentials, allowing entry and exit of plentiful nutrients by facilitation, and allowing the accumulation of scarce nutrients employing energy-

coupled transporters. These early transport systems provided essential 'housekeeping' functions without which the cells could not survive. Due to gene duplication and divergence as well as speciation, these early transporters developed into our current, large, ubiquitous superfamilies.

One of the half-dozen largest of these superfamilies is the amino acid/polyamine/organocation (APC) superfamily [TC (transporter classification) no. 2.A.3; Closs *et al.*, 1993; Reizer *et al.*, 1993; Saier, 1998, 1999a, b, c, 2000a, b; see also <http://www-biology.ucsd.edu/~msaier/transport/>]. This superfamily includes members that function as solute:cation symporters and solute:solute antiporters (Devés & Boyd, 1998; Isnard *et al.*, 1996; Kashiwagi *et al.*, 1997). Evidence suggests that in mammalian cells, simultaneous co- and counter-transport of ions with the amino acid can occur with a variety of complicated stoichiometries. For example, uptake of an amino acid can be accompanied by cotransport of three Na⁺ and one Cl⁻ while being countertransported against a K⁺ (Beckman & Quick, 1998).

† **Present address:** The Institute for Genomic Research (TIGR), 9712 Medical Center Drive, Rockville, MD 20850, USA.

Abbreviations: AAP, amino acid/auxin permease; AAT, amino acid transporter; ABT, archaeal/bacterial transporter; ACT, amino acid/choline transporter; APA, basic amino acid/polyamine antiporter; APC, amino acid/polyamine/organocation; CAT, cationic amino acid transporter; EAT, ethanolamine transporter; GGA, glutamate:GABA antiporter; HAAAP, hydroxy/aromatic amino acid permease; LAT, L-type amino acid transporter; SGP, spore germination protein; TC, transporter classification; TMS, transmembrane α -helical spanner; YAT, yeast amino acid transporter.

The majority of homologous integral membrane APC transport proteins appear to exhibit a uniform topology with twelve transmembrane α -helical spanners (TMSs) in a single polypeptide. This predicted topology has been experimentally verified for several bacterial homologues (Cosgriff & Pittard, 1997; Ellis *et al.*, 1995; Hu & King, 1998a). APC permease polypeptide chains vary in size from about 400 aminoacyl residues to about 800 residues (Sophianopoulou & Dhalluin, 1995). The larger proteins exhibit hydrophilic N- and C-terminal extensions as well as occasional enlarged inter-TMS loops. Proteins of one group, the LAT (L-type amino acid transporter) family (TC no. 2.A.3.8) derived from animals, are sometimes found in association with auxiliary proteins of the rBAT family (TC no. 8.A.9), and some of these same proteins, when defective, give rise to human diseases (Devés & Boyd, 1998; Estévez *et al.*, 1998; Markovich *et al.*, 1993; Mastroberardino *et al.*, 1998; Palacín *et al.*, 1998; Sato *et al.*, 1999; Torrents *et al.*, 1998; Verrey *et al.*, 1999). Some of the animal homologues of the CAT (cationic amino acid transporter) family (TC no. 2.A.3.3) have been found to serve as viral receptors (Kim *et al.*, 1991; Reizer *et al.*, 1993; Wang *et al.*, 1991).

The substrate specificities of several APC superfamily permeases have been carefully studied revealing that while some have exceptionally broad specificity, others are restricted to just one or a few amino acids or related compounds (Brechtel & King, 1998; Hu & King, 1998b, c; Isnard *et al.*, 1996; Kashiwagi *et al.*, 1997; Sato *et al.*, 1999). One of these permeases, a histidine permease of *Saccharomyces cerevisiae*, surprisingly, has been implicated in manganese transport (Farcasanu *et al.*, 1998). Their regulatory properties have been the subject of investigations (see, for example, Sanders *et al.*, 1998).

No comprehensive phylogenetic analysis of the APC superfamily has been conducted since 1993, when only 14 members of the family were analysed (Reizer *et al.*, 1993). Six years later, this superfamily has expanded to include 175 sequenced members that cluster in 10 defined families. Representation is found in all major domains of life (Paulsen *et al.*, 1998a, b and unpublished results; see <http://www-biology.ucsd.edu/~ipaulsen/transport/>). A recent study has led to the suggestion that this superfamily is related to other currently recognized families of amino acid transporters, but this suggestion has yet to be rigorously established (Young *et al.*, 1999).

In this paper we update our earlier reports, providing structural, functional and phylogenetic descriptions of the proteins of the APC superfamily. All currently sequenced members of each of the 10 families that comprise this superfamily are tabulated, and their substrate specificities and modes of action, when known, are described. Phylogenetic tree construction reveals their evolutionary relationships to each other. Motif analyses provide sequence fingerprints for each of the 10 constituent families of the APC superfamily. The members of each of these families are also shown to exhibit uniform or nearly uniform topological features that prove characteristic of the superfamily as a whole.

Distinctive properties that distinguish some of the different families are also noted. Thus, while all members of eight families appear to exhibit 12 TMSs, the members of one prokaryotic family [the SGP (spore germination protein) family] have 10 TMSs while the eukaryotic members of one of the ubiquitous families (the CAT family) have 14 TMSs. Further evidence is presented for a relationship of the APC superfamily to two other families of amino acid transporters.

METHODS

Computer methods. Sequences of the proteins that comprise the 10 families within the APC superfamily were obtained by recursive PSI-BLAST searches without iteration until all potential members had been retrieved from the GenBank, PIR and SWISS-PROT databases (P value $\geq 10^{-4}$) (Altschul *et al.*, 1997). Phylogenetic trees were constructed based on multiple alignments developed with the CLUSTAL X 8.1 (Thompson *et al.*, 1997) and the TREE (Feng & Doolittle, 1990) programs. The two programs gave very similar results (see Young *et al.*, 1999 for evaluation of these and other programs concerned with phylogenetic-tree construction). Family assignments were based on the phylogenetic results and on the statistical analyses obtained with the GAP program (Devereux *et al.*, 1984). The main features of the 10 families are presented in Table 1.

RESULTS

The APC family: an overview

Table 1 presents the 10 families of the APC superfamily. The names, abbreviations and TC numbers are provided for each family (columns 1 and 2). Column 3 tabulates the numbers of sequenced members of each family retrieved from the databases. Column 4 summarizes the source organisms from which the proteins of these families are derived. Column 5 presents the size ranges of the proteins in each family in numbers of aminoacyl residues per polypeptide chain. Column 6 lists the numbers of established or putative transmembrane α -helical spanners (TMSs) in the proteins of these families, and column 7 provides, as an example, one functionally well-characterized member of each family, when available.

The families vary in size from 4 members [the EAT (ethanolamine transporter) family; TC 2.A.3.5] to 36 members [the AAT (amino acid transporter) family; TC 2.A.3.1]. Other relatively large families include the YAT (yeast amino acid transporter) family with 29 members, the LAT family with 26 members, the CAT family with 21 members and the APA (basic amino acid/polyamine transporter) family with 20 members. All remaining families are relatively small, the largest being the ACT (amino acid/choline transporter) family with 13 members. For an introduction to the TC system of transport protein classification recently approved by the transport nomenclature panel of the International Union of Biochemistry and Molecular Biology, see Saier (1999c, 2000b).

Of the 10 families, half of them are represented only

Table 1. The 10 established families in the APC superfamily

Family TC no.	Family name and abbreviation	No. members	Source organism*	Size range (no. residues)	No. TMSs†	Well-characterized example‡
2.A.3.1	Amino acid transporter, AAT	35	B	433–556	12	PheP Eco
2.A.3.2	Basic amino acid/polyamine antiporter, APA	20	B	418–501	12	PotE Eco
2.A.3.3	Cationic amino acid transporter, CAT	21	B, Pl, An	462–657	14 (eukaryotes); 12 (prokaryotes)	Cat1 Hsa
2.A.3.4	Amino acid/choline transporter, ACT	13	Y, F, Pl	517–618	12	GabA Eni
2.A.3.5	Ethanolamine transporter, EAT	4	B	465–552	12	EntP Rer
2.A.3.6	Archaeal/bacterial transporter, ABT	8	B, Ar	424–736	12	–
2.A.3.7	Glutamate:GABA antiporter, GGA	9	B	466–540	12	GadC Lla
2.A.3.8	L-type amino acid transporter, LAT	26	An, Y	440–662	12	Lat1 Rno
2.A.3.9	Spore germination protein, SGP	10	<i>Bacillus</i> sp.	358–382	10	GerXB Ban
2.A.3.10	Yeast amino acid transporter, YAT	29	Y, F	521–852	12	Hip1 Sce

* B, bacteria; Ar, archaea; Y, yeast; F, fungi; An, animals; Pl, plants.

† Each family is believed to exhibit uniform topology except the CAT family, in which eukaryotic proteins have 14 TMSs while prokaryotic proteins have 12.

‡ See Table 3 for descriptions of these proteins.

Table 2. Interfamilial binary comparison scores

The GAP program with 100 random shuffles was used to generate the comparison scores reported, and these are expressed in standard deviations (SD). Percentage similarity values are also provided. The entire sequences of the proteins compared were used in these studies except for the comparisons presented below the dashed line, where scores were optimized by selection of the regions of highest similarity. These segments were always in excess of 100 residues in length.

TC no.*	Protein 1	TC no.	Protein 2	Comparison score (SD)	Identity (%)	Similarity (%)
2.A.3.1	AroP Eco	2.A.3.10	Can1 Sce	50	38	51
2.A.3.1	AroP Eco	2.A.3.6	Orf1 Mth	20	31	45
2.A.3.1	AroP Eco	2.A.3.3	AtrC1 Rpr	14	27	40
2.A.3.1	AroP Eco	2.A.3.2	ArcD Lsa	13	29	42
2.A.3.3	RocE Mtu	2.A.3.8	Lat1 Rno	23	30	39
2.A.3.3	RocE Mtu	2.A.3.5	EutP Rer	22	33	44
2.A.3.2	ArcD Lsa	2.A.3.7	XasA Eco	9	29	40
2.A.3.2	ArcD Lsa	2.A.3.9	GerAB Bsu	9	28	40
2.A.3.5	EutP Rer	2.A.3.4	Hnm1 Sce	10	27	40
2.A.42.2.1 (HAAAP family)	SdaC Eco	2.A.42.1.1 (HAAAP family)	TyrP Ctr	11	22	41
2.A.42.2.1 (HAAAP family)	SdaC Eco	2.A.3.1.2 (APC superfamily)	LysP Eco	8	23	39
2.A.42.1.1 (HAAAP family)	TyrP Ctr	2.A.3.9.2 (APC superfamily)	GerBB Bsu	7	22	39
2.A.42.1.2 (HAAAP family)	Mtr Eco	2.A.18.4.1 (AAAP superfamily)	Ynk1 Sce	8	20	39

* The TC numbers of the families within the APC superfamily are provided for all comparisons between proteins within the APC superfamily, presented above the dashed line. The individual proteins compared can be identified in Table 3. The TC numbers of the proteins selected for the interfamilial comparisons reported below the dashed line are provided. Using our standard criteria, the values allow establishment of homology for the 10 APC families (above the dashed line) but not for the AAAP and HAAAP families (Saier, 1994).

Table 3. Sequenced protein members of the 10 established families of the APC superfamily

Protein	Description in database	Source organism	No. residues	Accession no.*	NCBI gi no.
AAT family (TC 2.A.3.1)					
Orf1 Asp	Putative transporter	<i>Arthrobacter</i> sp.	466	gbAB007122	3116225
YtnA Bsu	Amino acid transporter	<i>B. subtilis</i>	463	gbAF008220	2293166
AapA Bsu	Amino acid permease	<i>B. subtilis</i>	459	spO06005	3023220
YdgF Bsu	Hypothetical transporter	<i>B. subtilis</i>	458	spP96704	3025135
YbxG Bsu	Hypothetical transporter	<i>B. subtilis</i>	462	spP54425	3123299
GabP Bsu	γ -Aminobutyrate permease	<i>B. subtilis</i>	469	spP46349	2828523
HutM Bsu	Putative histidine permease	<i>B. subtilis</i>	475	spP42087	1170426
YbgF Bsu	Similar to histidine permease	<i>B. subtilis</i>	470	gbZ99105	2632526
RocE Bsu†	Amino acid permease	<i>B. subtilis</i>	467	spP39137	730601
RocC Bsu	Amino acid permease	<i>B. subtilis</i>	470	spP39636	730600
YvbW Bsu	Similar to amino acid permease	<i>B. subtilis</i>	447	spO32257	2635914
AroP Cgl	Aromatic amino acid transporter	<i>Corynebacterium glutamicum</i>	463	spQ46065	2492533
ProY Eco†	Proline-specific permease	<i>E. coli</i>	457	spP77327	2492534
PheP Eco	Phenylalanine-specific permease	<i>E. coli</i>	458	spP24207	130068
CycA Eco†	D-serine/D-alanine/glycine transporter	<i>E. coli</i>	470	spP39312	732281
AroP Eco†	Aromatic amino acid transporter	<i>E. coli</i>	457	spP15993	2506093
YifK Eco	Probable transporter	<i>E. coli</i>	461	spP27837	2851572
AnsP Eco†	L-Asparagine permease	<i>E. coli</i>	499	spP77610	3183513
YkfD Eco	Hypothetical protein	<i>E. coli</i>	475	spQ47689	2492535
GabP Eco†	γ -Aminobutyrate permease	<i>E. coli</i>	466	spP25527	120778
LysP Eco†	Lysine-specific permease	<i>E. coli</i>	489	spP25737	547880
YeeF Eco	Hypothetical transporter	<i>E. coli</i>	454	spP33016	465570
YcjJ Eco	Hypothetical transporter	<i>E. coli</i>	461	spP76037	3915976
Orf1 Eco	Lysine-specific permease	<i>E. coli</i>	433	gbD90767	1742117
RocE Hpy	Amino acid permease	<i>Hel. pylori</i>	519	gbAE000610	2314162
Orf1 Hpy	Amino acid permease	<i>Hel. pylori</i>	475	gbAE001474	4154932
AnsP Mle	Putative L-asparagine permease	<i>Mycobacterium leprae</i>	505	gbAL035310	4200260
CycA Mtu	Putative transporter	<i>Myc. tuberculosis</i>	556	gbZ98268	2326747
AnsP Mtu	Putative transporter	<i>Myc. tuberculosis</i>	489	gbZ97559	2253066
Yao6 Mtu	Hypothetical protein	<i>Myc. tuberculosis</i>	487	spO06297	3024959
HutT Ppu	Inducible histidine transporter	<i>Pseudomonas putida</i>	467	gbAF032970	2642339
ProY Sty	Proline-specific permease	<i>Salmonella typhimurium</i>	456	spP37460	2506094
YifK Sty	Probable transporter	<i>Sal. typhimurium</i>	461	spP37456	586738
AnsP Sty	L-Asparagine permease	<i>Sal. typhimurium</i>	497	spP40812	728862
Orf48 Sty	Amino acid permease YeeF-like protein	<i>Sal. typhimurium</i>	448	gbX99945	1526981
Orf1 Sco	Putative L-asparagine permease	<i>Streptomyces coelicolor</i>	489	gbAL049587	4584481
APA family (TC 2.A.3.2)					
Orf1 Asa	Polytopic cytoplasmic membrane protein	<i>Aer. salmonicida</i>	426	pirJC5086	2126237
ArcD Bli	Arginine:ornithine antiporter	<i>B. licheniformis</i>	468	gbY17554	3687418
YvsH Bsu	Amino acid permease	<i>B. subtilis</i>	469	gbZ99121	2635847
YkbA Bsu	Hypothetical transporter	<i>B. subtilis</i>	438	gbAJ002571	2632007
ArcD Ctr	Arginine:ornithine antiporter	<i>Chl. trachomatis</i>	483	gbAE001310	3328798
ArcD Cpn	Arginine:ornithine antiporter	<i>Chlamydia pneumoniae</i>	485	gbAE001683	4377363
ArcD Cpe	Arginine:ornithine antiporter	<i>Clo. perfringens</i>	476	spQ46170	3913092
LysI Cgl	L-Lysine transport protein	<i>Cor. glutamicum</i>	501	spP35865	547879
PotE Eco†	Putrescine:ornithine antiporter	<i>E. coli</i>	439	spP24170	130692
YjdE Eco	Hypothetical protein	<i>E. coli</i>	445	spP39269	2851627
CadB Eco†	Probable cadaverine:lysine antiporter	<i>E. coli</i>	444	spP23891	115415
ArcD Eco†	Arginine:ornithine antiporter	<i>E. coli</i>	460	spP77429	3913086

Table 3 (cont.)

Protein	Description in database	Source organism	No. residues	Accession no.*	NCBI gi no.
YjeH Eco	Hypothetical transporter	<i>E. coli</i>	418	spP39277	731999
YhfM Eco	Hypothetical transporter	<i>E. coli</i>	445	spP45539	2851642
Orf1 Eco	Hypothetical transporter	<i>E. coli</i>	462	gbU18997	606304
PotE Hin	Putrescine:ornithine antiporter	<i>Hae. influenzae</i>	435	spP44768	1172565
ArcD Lsa	Arginine:ornithine antiporter	<i>Lactobacillus sakei</i>	475	spO53092	3913083
ArcD Pae†	Arginine:ornithine antiporter	<i>Pseudomonas aeruginosa</i>	482	spP18275	114116
PotE Rpr†	Putrescine:ornithine antiporter	<i>Rickettsia prowazekii</i>	429	gbAJ235272	3861036
OrfA Sco	Putative integral membrane transporter	<i>Str. coelicolor</i>	474	gbAL031515	3560003
CAT family (TC 2.A.3.3)					
Aat1 Ath†	Amino acid transporter AAT1	<i>Ara. thaliana</i>	533	pirS51171	1076291
Orf1 Ath	Putative amino acid transporter	<i>Ara. thaliana</i>	569	gbAC004238	3033391
YfnA Bsu†	YfnA protein	<i>B. subtilis</i>	462	gbD86418	2116759
YhdG Bsu	Hypothetical protein	<i>B. subtilis</i>	465	gbY14082	2226202
Orf7 Cel	Similar to amino acid permease	<i>Cae. elegans</i>	589	gbAF040642	2746784
Orf8 Cel	Similar to amino acid permease	<i>Cae. elegans</i>	583	gbAF024493	2394432
Orf9 Cel	Similar to amino acid permease	<i>Cae. elegans</i>	585	gbAF025452	2429392
CtrA Ddi	Similar to cationic amino acid transporter	<i>Dictyostelium discoideum</i>	556	gbAF015566	2353183
Cat1 Hsa†	Cationic amino acid transporter 1	<i>Homo sapiens</i>	629	refNP_003036	4507047
Cat2 Hsa†	Cationic amino acid transporter 2	<i>Homo sapiens</i>	657	refNP_003037	4507049
Cat4 Hsa	Cationic amino acid transporter 4	<i>Homo sapiens</i>	636	refNP_004164	4759138
Err Mau	Ecotropic murine retrovirus receptor	<i>Mesocricetus auratus</i>	628	gbU26454	4096268
Cat1 Mmu	Cationic amino acid transporter	<i>Mus musculus</i>	618	gbU70859	1575776
Cat2 Mmu	Cationic amino acid transporter 2	<i>Mus musculus</i>	657	pirA46073	1083250
RocE Mtu	RocE protein	<i>Myc. tuberculosis</i>	476	gbZ79702	1524280
Orf1 Mtu	Hypothetical protein	<i>Myc. tuberculosis</i>	495	gbZ95121	2072720
Cat1 Rno	Cationic amino acid transporter 1	<i>Rattus norvegicus</i>	624	gbU70476	1589917
Cat2 Rno†	Cationic amino acid transporter 2 (fragment)	<i>Rat. norvegicus</i>	261	gbU53927	1297194
Cat3 Rno	Cationic amino acid transporter 3	<i>Rat. norvegicus</i>	619	gbAB000113	2116552
AtrC1 Rpr	Cationic amino acid transporter 1	<i>Ric. prowazekii</i>	470	gbAJ23527	3860868
Orf1 Sco	Putative amino acid transporter	<i>Str. coelicolor</i>	503	gbAL023702	3169028
ACT family (TC 2.A.3.4)					
Orf3 Ath	Putative amino acid or GABA permease	<i>Ara. thaliana</i>	517	gbAC006200	4262224
Gpt1 Cal†	Putative polyamine transporter	<i>Candida albicans</i>	553	gbAF080132	3415007
GabA Eni†	GABA permease	<i>Emericella nidulans</i>	517	gbAJ131668	4972245
Aap2 Ncr†	Amino acid permease	<i>Neurospora crassa</i>	551	spO59942	3929337
Hnm1 Sce†	Choline transporter	<i>Sac. cerevisiae</i>	563	spP19807	117619
Bio5 Sce	Transmembrane regulator of KAPA/DAPA transporter	<i>Sac. cerevisiae</i>	561	spP53744	1730687
Uga4 Sce†	GABA-specific permease	<i>Sac. cerevisiae</i>	571	spP32837	418150
Ykr4 Sce	Hypothetical amino acid permease	<i>Sac. cerevisiae</i>	618	spP36029	549738
YahB Spo	Hypothetical amino acid permease	<i>Schizosaccharomyces pombe</i>	544	spQ09887	1351645
Yao8 Spo	Hypothetical amino acid permease	<i>Schiz. pombe</i>	550	spQ10087	1351705
Orf3 Spo	Putative amino acid permease	<i>Schiz. pombe</i>	557	gbAL031543	3581896
Orf4 Spo	Amino acid permease	<i>Schiz. pombe</i>	554	gbAL023595	3150116
Orf5 Spo	Amino acid permease	<i>Schiz. pombe</i>	542	gbAL023290	3116147
EAT family (TC 2.A.3.5)					
Orf1 Mba†	Putative monomethylamine permease	<i>Methanosarcina barkeri</i>	552	gbAF013713	2746072
EutP Rer†	Ethanolamine permease	<i>Rhodococcus erythropolis</i>	488	gbL24492	403170
Orf1 Sme†	Hypothetical protein	<i>Sinorhizobium meliloti</i>	465	gbAF055582	3687686
EutP Zmo†	Ethanolamine permease	<i>Zymomonas mobilis</i>	491	gbAF124349	4378853

Table 3. (cont.)

Protein	Description in database	Source organism	No. residues	Accession no.*	NCBI gi no.
ABT family (TC 2.A.3.6)					
Cat1 Afu†	Putative cationic amino acid transporter 1	<i>Arc. fulgidus</i>	736	gbAE000971	2648945
Cat2 Afu†	Putative cationic amino acid transporter 2	<i>Arc. fulgidus</i>	531	gbAE000980	2648775
YbeC Bsu	Similar to amino acid transporter	<i>B. subtilis</i>	539	gbZ99105	2632498
YveA Bsu	Hypothetical protein	<i>B. subtilis</i>	520	gbZ94043	1945680
YbaT Eco†	Hypothetical protein	<i>E. coli</i>	430	spP77400	2495533
Y609 Mja†	Hypothetical protein	<i>Met. jannaschii</i>	435	spQ58026	2496073
Yw19 Mtu	Hypothetical protein	<i>Myc. tuberculosis</i>	440	spQ10858	1731248
Orf1 Mth	Hypothetical amino acid transporter	<i>Met. thermoautotrophicum</i>	424	gbAE000837	2621621
GGA family (TC 2.A.3.7)					
XasA Cpn†	Amino acid transporter	<i>Chl. pneumoniae</i>	468	gbAE001613	4376556
XasA Ctr	Amino acid transporter	<i>Chl. trachomatis</i>	466	gbAE001295	3328624
Orf54 Cpe	Membrane-spanning transport protein	<i>Clo. perfringens</i>	495	gbD49537	1402515
XasA Eco†	Amino acid antiporter	<i>E. coli</i>	511	spP39183	2507158
YjeM Eco	Hypothetical protein	<i>E. coli</i>	514	spP39282	732004
YgiI Eco	Hypothetical protein	<i>E. coli</i>	477	spP42590	1176115
Orf1 Eco	Putative transporter	<i>E. coli</i>	540	gbAE000192	1787127
GadC Lla	Putative glutamate:GABA antiporter	<i>Lco. lactis</i>	503	gbAF005098	2352486
XasS Sfl§	Amino acid antiporter	<i>Shigella flexneri</i>	511	spQ54152	2499135
LAT family (TC 2.A.3.8)					
Orf1 Cel	Similarity to <i>Schistosoma</i> amino acid permease	<i>Cae. elegans</i>	464	gbU50308	1226279
Orf2 Cel	Similar to amino acid permeases	<i>Cae. elegans</i>	563	gbAF026214	2435610
Orf3 Cel	Similar to amino acid permeases	<i>Cae. elegans</i>	566	gbAF003144	2088757
Orf4 Cel	Similar to amino acid permeases	<i>Cae. elegans</i>	562	gbU56963	1293821
Orf5 Cel	Similarity to <i>Schistosoma</i> amino acid permease	<i>Cae. elegans</i>	440	gbZ74042	3879689
Orf6 Cel	Similarity to amino acid permeases	<i>Cae. elegans</i>	662	gbAF040647	2746835
Orf7 Cel	Similarity to human membrane protein E16	<i>Cae. elegans</i>	493	gbZ68216	3876345
Mnd Dme	Minidiscs/similar to amino acid transporter	<i>Drosophila melanogaster</i>	499	gbAF139834	5081597
y ⁺ Lat1 Hsa†	Unknown protein in locus pSLC7A6	<i>Homo sapiens</i>	515	refNP_003974	4507053
y ⁺ Lat2 Hsa	Unknown protein in locus pSLC7A7	<i>Homo sapiens</i>	511	refNP_003973	4507055
Orf1 Hsa	Membrane protein E16	<i>Homo sapiens</i>	507	refNP_003477	4507051
Orf2 Hsa	4F2 light chain	<i>Homo sapiens</i>	507	gbAB017908	4519803
hbO, + AT Hsa	Putative glycoprotein-associated amino acid transporter	<i>Homo sapiens</i>	487	NA	NA
hLAT2 Hsa†	LAT2 (L-type amino acid transporter 2)	<i>Homo sapiens</i>	534	NA	NA
hLAT3 Hsa	LAT3 putative amino acid transporter	<i>Homo sapiens</i>	535	gbY18483	4581470
y ⁺ Lat1 Mmu	Putative amino acid transporter	<i>Mus musculus</i>	510	gbAJ012754	3970791
Orf1 Mmu	4F2/CD98 light chain	<i>Mus musculus</i>	512	gbAB017189	4519256
Xct Mmu	Cystine/glutamate transporter	<i>Mus musculus</i>	502	gbAB022345	4689081
mbO, + AT Mmu	Putative glycoprotein-associated amino acid transporter	<i>Mus musculus</i>	487	NA	NA
mLAT2 Mmu	LAT2 (L-type amino acid transporter 2)	<i>Mus musculus</i>	531	NA	NA
Lat1 Rno†	LAT1 (L-type amino acid transporter 1)	<i>Rat. norvegicus</i>	512	gbAB015432	3582136
Mup1 Sce†	High-affinity methionine permease	<i>Sac. cerevisiae</i>	574	spP50276	1709181
Mup3 Sce	Hypothetical 60·6 kDa protein in CBP2-SSBR1 intergenic region	<i>Sac. cerevisiae</i>	546	spP38734	731613
SPRM1 Sma†	Amino acid permease	<i>Schistosoma mansoni</i>	503	gbL25068	407047
ASUR4 Xla†	Amino acid transporter chain (AmAt-L-lc) ASUR4	<i>Xenopus laevis</i>	507	gbY12716	3114983
Lat-1 Xla	L-Amino acid transporter 1	<i>X. laevis</i>	507	gbAF019906	2731774

Table 3 (cont.)

Protein	Description in database	Source organism	No. residues	Accession no.*	NCBI gi no.
SGP family (TC 2.A.3.9)					
GerXB Ban†	Spore germination protein GerXB	<i>Bacillus anthracis</i>	359	gbAF108144	4092083
GerIB Bsu†	Spore germination protein	<i>Bacillus cereus</i>	364	gbAF067645	3290176
YndE Bha‡	Spore germination protein (gene 728-1876)	<i>Bacillus halodurans</i>	382	gbAB013370	4514335
GerAB Bme§	Germination protein (gene 3086-4189)	<i>Bacillus megaterium</i>	367	gbU61380	2204286
GerAB Bsu†	Spore germination protein	<i>B. subtilis</i>	365	spP07869	1708033
GerBB Bsu	Spore germination protein	<i>B. subtilis</i>	368	spP39570	729628
Orf1 Bsu	Spore germination protein	<i>B. subtilis</i>	363	gbZ99113	2634160
GrkB Bsu†	Spore germination protein	<i>B. subtilis</i>	373	spP49940	1708044
YfkT Bsu	Putative germination protein	<i>B. subtilis</i>	358	gbD86417	2443222
YAT family (TC 2.A.3.10)					
Aat1 Amu	Amino acid transporter	<i>Amanita muscaria</i>	538	gbAJ223504	4468097
Can1 Cal	Lysine/arginine permease	<i>Can. albicans</i>	571	spP43059	1168755
PutX Eni	Proline transport protein	<i>Eme. nidulans</i>	570	pirS42708	630383
Naap1 Ncr	Amino acid permease	<i>N. crassa</i>	572	gbAF001032	2197048
Lyp1 Sce†	Lysine-specific permease	<i>Sac. cerevisiae</i>	611	spP32487	1708894
Can1 Sce	Arginine permease	<i>Sac. cerevisiae</i>	590	spP04817	729014
Dip5 Sce	Dicarboxylic amino acid permease	<i>Sac. cerevisiae</i>	608	spP53388	1706439
Alp1 Sce	Basic amino acid permease	<i>Sac. cerevisiae</i>	573	spP38971	1703255
Gap1 Sce	γ -Aminobutyrate permease	<i>Sac. cerevisiae</i>	601	spP19145	544369
Tat2 Sce†	Tryptophan permease	<i>Sac. cerevisiae</i>	592	spP38967	730893
Put4 Sce†	Proline-specific permease	<i>Sac. cerevisiae</i>	627	spP15380	1709941
Agp1 Sce	Asparagine/glutamine permease	<i>Sac. cerevisiae</i>	633	spP25376	140359
Hip1 Sce*	Histidine permease	<i>Sac. cerevisiae</i>	603	spP06775	729714
Bap3 Sce	Valine amino acid permease	<i>Sac. cerevisiae</i>	604	spP41815	1709572
Gnp1 Sce	High-affinity glutamine permease	<i>Sac. cerevisiae</i>	663	spP48813	1708002
Agp3 Sce†	General amino acid permease	<i>Sac. cerevisiae</i>	558	spP43548	1175959
Bap2 Sce	Leucine/valine/isoleucine amino acid permease	<i>Sac. cerevisiae</i>	609	spP38084	586506
Val1 Sce	Valine/tyrosine/tryptophan amino acid permease	<i>Sac. cerevisiae</i>	619	spP38085	586507
Mmp1 Sce	S-methylmethionine transporter	<i>Sac. cerevisiae</i>	583	pirS50959	1077318
Sam3 Sce	S-adenosylmethionine transporter	<i>Sac. cerevisiae</i>	587	pirS65307	2133015
Agp2 Sce	General amino acid permease	<i>Sac. cerevisiae</i>	596	spP38090	586542
Ssy1 Sce	Putative amino acid permease	<i>Sac. cerevisiae</i>	852	spQ03770	3024669
Isp5 Spo	Putative amino acid permease/sexual differentiation	<i>Schiz. pombe</i>	543	spP40901	729860
Orf2 Spo	Putative amino acid permease	<i>Schiz. pombe</i>	521	gbAL031532	3560241
Orf1 Spo	Amino acid permease	<i>Schiz. pombe</i>	574	gbAL023594	3150139
Aap1 Spo†	Amino acid permease	<i>Schiz. pombe</i>	594	gbD87954	1620894
Orf3 Spo	Amino acid permease	<i>Schiz. pombe</i>	537	gbAL023590	3136040
Ina1 Tha†	Amino acid permease	<i>Trichoderma harzianum</i>	573	spP34054	462414
Orf1 Ufa	Putative permease	<i>Uromyces fabae</i>	561	gbU81794	1764098

* gb, GenBank; sp, SWISS-PROT.

† Included in the APC superfamily phylogenetic tree (Fig. 1).

‡ Excluded from phylogenetic analysis because the sequence is incomplete.

§ Excluded from phylogenetic analysis due to 100% identity with XasA Eco (2507158).

|| Not available; sequence obtained from Dr F. Verrey (Verrey *et al.*, 1999).

¶ Sequencing errors in GenBank entry have been corrected.

Fig. 1. Phylogenetic tree of the entire APC superfamily, showing segregation of the proteins into the 10 clusters that define the 10 constituent families. At least four representative proteins from each of the 10 families were included in the analysis with the sole exception of the GGA family, where only two proteins were included. The TREE program (Feng & Doolittle, 1990) was used to generate this and other trees reported in this paper, as well as the multiple alignments upon which the hydropathy profiles (Fig. 2) and signature sequences (Table 4) were based. Groups: 1, AAT family; 2, APA family; 3, CAT family; 4, ACT family; 5, EAT family; 6, ABT family; 7, GGA family; 8, LAT family; 9, SGP family; 10, YAT family. The bar represents phylogenetic distance in arbitrary units and is the same relative length in all figures.

in bacteria. One family, the ABT (archaeal/bacterial transporter) family, is found only in bacteria and archaea but not eukaryotes. The CAT family is found in bacteria and eukaryotes but not archaea. Finally, three families are found only in eukaryotes (the ACT, LAT and YAT families). While YAT family members are derived exclusively from the fungal kingdom (yeast, fungi and mushrooms), ACT family members are derived from both plants and fungi but not animals, and LAT family proteins occur in both fungi and animals but not plants. Thus, all animal proteins are included in only two families, all plant proteins are similarly included in just two families, but fungal proteins are included in three families. Four of the five largest families are those which have representation in eukaryotes, but the largest family is specific to bacteria. The four smallest families are prokaryote specific.

Without exception, the five families specific to bacteria include members with the smallest polypeptide chains. The smallest of these proteins are restricted to the SGP family (size range 329–373 residues). No member of any other family is this small and this is the only family in which members have less than 12 putative TMSs. Proteins of the SGP family display just 10 TMSs. For the remaining four bacterial families, the protein size range is nearly the same (418–556 residues). Surprisingly, the ABT family shows a size range of 422–736 residues. Finally, the four families represented in eukaryotes show considerable size variation (440–852 residues). One of these families, the CAT family, exhibits the unique characteristic that all of the eukaryotic members display 14 putative TMSs although the prokaryotic members of this same family display the more usual 12 TMSs. Thus, eight families include members that exhibit 12 TMSs, one family consists of proteins with either 14 or 12 TMSs and one includes members with 10 TMSs. Size

and topological variation will be analysed in greater detail below.

Establishment of homology

Table 2 summarizes the results of the statistical analyses of interfamilial binary sequence comparisons. As discussed previously (Saier, 1994, 1996), a comparison score of 9 standard deviations (sd) for a stretch of at least 60 residues in comparable regions of two proteins is considered sufficient to establish homology. The odds that this degree of similarity could have arisen by chance is 1 in 10^{-19} . We and others have published the arguments leading to the conclusion that a value of 9 sd could not have resulted either by chance or by a process of evolutionary sequence convergence (Doolittle, 1986; Saier, 1994). This conclusion is corroborated by the related functions and structures of the members of the APC superfamily. The results summarized in Table 2 therefore establish by these criteria that all 10 families listed in Table 1 include homologous proteins that are constituents of a single superfamily.

We have presented preliminary evidence that the eukaryote-specific amino acid/auxin permease (AAAP) family (TC 2.A.18) is distantly related to the APC family and that a small bacterial family, the hydroxy/aromatic amino acid permease (HAAAP) family (TC 2.A.42) is also a distant constituent of this superfamily (Young *et al.*, 1999). In Table 2 we present statistical analyses that further support this suggestion.

The entries below the dashed line in Table 2 show that when SdaC of *Escherichia coli* was compared with TyrP of *Chlamydia trachomatis*, a comparison score of 11 sd was obtained. This value is sufficient to establish that these two proteins belong to a single family. Other

comparisons reported in Table 2, where members of the HAAAP family are compared with members of the APC and AAP families, do not reach a value of 9 SD. The comparison scores therefore do not allow us to establish homology between these families. The values of 7–8 SD obtained, however, are sufficient to strongly suggest homology. We therefore suggest that the AAP and HAAAP families are distant constituents of the APC superfamily, although we cannot prove this contention statistically.

Sequenced members of the APC superfamily

Table 3 shows the identified, sequenced members of the 10 established families that comprise the APC superfamily. The table presents the abbreviations of the proteins to be used in this study, the database descriptions of these proteins, their source organisms, their sizes, and their database accession numbers as well as their GenBank identifier (gi) numbers. Analyses of the sequences of the listed proteins provide the basis for the conclusions presented in this paper.

APC superfamily tree

Fig. 1 presents a phylogenetic tree for the entire APC superfamily, where representative members of each family are included. Deep-rooted branching provides the phylogenetic basis for dividing the superfamily into its 10 constituent families. Thus, each family stems from a point near the centre of the unrooted tree. Further, all of the members of a family are more closely related to each other than any one of these members is related to a member of any other family. The family designations (see Table 1) were based on the available functional data as well as on the source organisms.

Topological predictions

Fig. 2 presents average hydropathy plots for all 10 families of the APC superfamily. These plots were based on multiple alignments generated for the full-length proteins of each family presented in Table 1. Several features of these plots are worth mentioning. (1) The plots reveal that eight of the families display 12 putative TMSs with relative positions and relative sizes of the peaks being fairly similar. This fact shows that the sizes of both the TMSs and the inter-TMS loops have in general been conserved during evolution of the superfamily. (2) The CAT family (TC 2.A.3.3) displays 14 hydrophobic peaks due to the presence of two additional putative TMSs localized to the extreme C termini of the eukaryotic proteins of this family. This trait proved to be a characteristic only of the eukaryotic proteins, as all prokaryotic members of the CAT family exhibited the more usual 12 putative TMSs. No member of any other family within the APC superfamily exhibited this 14-TMS characteristic. (3) The SGP family (TC 2.A.3.9) displayed 10 putative TMSs rather than 12. This proved to be due to a C-terminal truncation, resulting in the elimination of TMSs 11 and 12, present in all other

members of the APC superfamily. The 10-TMS topology is therefore characteristic of the SGP family. (4) The hydropathy plots revealed that one member of the ABT family (TC 2.A.3.6; Cat1 Afu, see Table 3) exhibits a long C-terminal hydrophilic extension, while some members of the YAT family (TC 2.A.3.10) exhibit long N-terminal hydrophilic extensions. The C-terminal extension of 200 aminoacyl residues in Cat1 Afu proved to be a conserved, duplicated, hydrophilic domain found in numerous archaeal, bacterial and eukaryotic proteins. Similarly, the N-terminal hydrophilic extensions of the yeast permeases of the YAT family proved to exhibit sequence similarity with conserved domains in other proteins. These hydrophilic N- and C-terminal domains undoubtedly serve important functions, possibly in promoting protein–protein interactions. Additionally, members of the CAT and GGA (glutamate:GABA transporter) families (TC 2.A.3.3 and 2.A.3.7, respectively) have short N-terminal extensions, members of the EAT family (TC 2.A.3.5) have short C-terminal extensions and members of the AAT and ACT families (TC 2.A.3.1 and 2.A.3.4, respectively) have short N- and C-terminal extensions. Whether or not these hydrophilic extensions serve a specific function has yet to be examined. Members of remaining families (APA, TC 2.A.3.2; LAT, TC 2.A.3.8; and SGP, TC 2.A.3.9) lack or have only very short hydrophilic extensions at both ends of the proteins. (5) According to biochemically established topological features of representative prokaryotic APC superfamily members, both the N and C termini of the proteins are located in the cytoplasm with a 12 TMS topology (Cosgriff & Pittard, 1997; Ellis *et al.*, 1995; Hu & King, 1998a). Loops connecting TMSs 1 and 2, 4 and 5, 9 and 10, and 11 and 12 are relatively short while remaining loops are longer. Three of these four short loops are putative extracytoplasmic loops while one of them is localized to the cytoplasm. The remaining loops appear to be of variable sizes and may be characteristic of the individual families. Nevertheless, extracytoplasmic loops are, on average, substantially shorter than cytoplasmic loops. In summary, both the degrees of relative hydrophobicity of the TMSs and the relative loop sizes are generally characteristic of the APC superfamily as a whole, with greatest constancy with respect to relative degrees of hydrophobicity of the various TMSs. This feature serves as a ‘footprint’ characteristic of the APC superfamily.

The complete multiple alignments of the proteins that comprise each of the 10 families of the APC superfamily, upon which the average hydropathy plots shown in Fig. 2 were based, can be found at <http://www-biology.ucsd.edu/~msaier/transport/phylo.html>. These alignments were generated with the TREE program (Feng & Doolittle, 1990). Examination of these multiple alignments reveals that in general, well-conserved regions in all of these alignments can be found in which some residues are fully conserved and many more are largely conserved. It is clear that the program used to generate these alignments has aligned the sequences correctly.

Fig. 2. Average hydropathy profiles for the proteins that comprise each of the 10 families of the APC superfamily. The program used was based on the hydropathy values for the amino acids reported by Kyte & Doolittle (1982). A sliding window of 21 residues was used. In all of these plots, putative TMS no. 1 were aligned vertically to facilitate comparison of the results obtained for the 10 families. Individual proteins from all 10 families were examined using two additional hydropathy programs: TopPred and TMPred (Claros & von Heijne, 1994; Hofmann & Stoffel, 1993; von Heijne, 1992). The results confirmed the topological predictions shown here.

APC superfamily signature sequences

Table 4 presents signature sequences for the proteins of the 10 families within the APC superfamily. These sequences serve as identification motifs for these families. They were screened against the SWISS-PROT and TREMBL databases and retrieved only members of the represented family. By this criterion, they are *bona*

fide signature sequences specific to these families. They can be used to identify additional members of these families as they are sequenced.

Several of the larger families in the APC superfamily are included within the Pfam (<http://www.sanger.ac.uk/Software/Pfam/>) database of protein families as the 'amino acid permease' family (accession no. PF00324,

Table 4. Signature sequences for the 10 families of the APC superfamily

TC no.	Family	Signature sequence*
2.A.3.1	AAT	[PASDE]-X-[WH]-[LIVMY]-X(2)-[LIVA]-[LIVATSG]-X(2)-[LIVFGAT]-[LIVFA]-[LIVGMF]-X-[LIVGAST]-X-N-[LIVFYMA]-X-[STANG]-[LIVT]-[KREQN]-X-[FYVT]-[GA]-[ENK]-X-[END]
2.A.3.2	APA	[FMVIL]-X(2)-[VCTSAG]-[LIVMA]-[LIVMSAN]-X-[TCLM]-[MFSVIL]-[FW]-[VCGAS]-[FY]-[TDVIL]-G-[WMVIL]-[EAI]-[LBCAS]-[AFVIL]
2.A.3.3	CAT	[FMLIV]-[AS]-[GALIV]-[FLIVP]-[CS]-Y-[AGT]-E-[FLIV]-[GAS]-[SALIV]-[REAST]-[LIVF]-P-[LIVGKR]-[AST]-G-[GS]-[ASLIV]-[FY]
2.A.3.4	ACT	[GPLST]-[WLIVFY]-[LIFV]-[LIVT]-[GASLIV]-[GACLIVM]-[ASPLIVF]-[LIVFCG]-X(2)-[LIVMFASC]-[LIVTSC]-[GAS]-X-[GST]-[LIVM]-[GA]-[ED]-[LIVFH]-[LIVASC]-[AS]-X-[MFLIVWY]-P
2.A.3.5	EAT	V-[LIV]-[SV]-G-[ED]-Y-[FAM]-G-W-[SN]-X-[STG]-[LIVW]-[GA]-X-[AG]-G
2.A.3.6	ABT	[FSY]-[AST]-[KER]-[AL]-[ASG]-[SAL]-X-[LIVFMY]-[PV]-X(2)-[AG]-G-X-[LIVY]-X-[NFY]-X(5)-G
2.A.3.7	GGA	G-[LIV]-X-[LIVAST]-W-[LIVATM]-[RSKN]-[DKNQSH]-[AST]-[LIVM]-G-X(2)-[WFLIV]-[GA]-[FY]
2.A.3.8	LAT	G-[WFY]-[DNFS]-X-[LIV]-[NH]-[FYT]-[LIVAGS]-[TALIV]-[EGPS]-E-[LIVM]-X-[NDE]-P-X-[RK]-[NT]-[LIVM]-[PK]
2.A.3.9	SGP	[FY]-[LIVF]-[ED]-[RH]-[LIVFS]-[ED]-[LYVFSI]-[LIVFM]-[LIVFMAS]-X-[LIVTFSA Y]-X-[WLMY]-[LIVAT]-[LIVMAP]-X(2)-[FYIM]-X(2)-[LIVFM]-[LIVAS]-[LIVMG]
2.A.3.10	YAT	E-X(2)-[FY]-[WACLIVFY]-[FLIVMAS]-[GASN]-X-[LIVMFCS]-K-[LIVA]-[LIVMGA]-[LIVSTAGF]-[LIVMTA]-[LIVMTFCA]-[LIVMTGA]-[LIFYWM]-[LIFVTY]-[LIVF]-[LIVFTYC]-[LASGTCM]-[LIVF]-[LIVC]-[LIVMTY]-X-[LIVFWCAS]-G

* Alternative residues at any one position are provided in square brackets. X indicates that any residue can occur at that position. Numbers in parentheses indicate numbers of positions. The one-letter abbreviations of the amino acids are used.

Fig. 3. Phylogenetic tree for the AAT (TC 2.A.3.1) family.

prosite no. PDOC00191). However, the SGP family has been assigned to a distinct family (Pfam accession no. PB004126). Several of the smaller APC families are not included in the Pfam database (Bateman *et al.*, 2000; Hofmann *et al.*, 1999).

Phylogeny of the 10 families within the APC superfamily

Figs 3–12 present phylogenetic trees for the proteins of the 10 families within the APC superfamily. Evaluation of these trees provides evidence regarding paralogous

and orthologous relationships. Thus, if divergence has arisen solely due to speciation, the genes are orthologues, but if they arose as a result of gene duplication events within a single organism, they are paralogues. The expected relative distances for orthologues can be estimated based on relative phylogenetic distances observed for the 16S rRNAs from the organisms under consideration. It should be noted, however, that while most homologues in a single species can be assumed to be paralogues, it is more problematic to assign orthologous relationships for proteins from different species unless functional data are available. The reader should

Fig. 4. Phylogenetic tree for the APA (TC 2.A.3.2) family.

therefore be cognizant of this problem when considering conclusions based on the phylogenetic trees presented below.

In Fig. 3 (the bacterial AAT family), there are five principal branches, and *E. coli* paralogues, of which there are 12, are found in all of these clusters. This fact suggests that gene duplication events occurred during the early evolution of this family. The nine *Bacillus subtilis* paralogues are found in three clusters and the three *Mycobacterium tuberculosis* paralogues appear in two of them. A few late gene-duplication events evidently gave rise to paralogues of very similar sequence. For example, PheP and AroP in cluster 3, the phenylalanine and aromatic amino acid permeases of *E. coli*, respectively, appear to have arisen by a gene-duplication event that probably occurred somewhat before Gram-positive bacteria diverged from Gram-negative bacteria (judging from the distances observed for putative *E. coli* and *B. subtilis* orthologues). The four *B. subtilis* paralogues in cluster 1 probably arose after or at about the same time Gram-positive bacteria diverged from Gram-negative bacteria. The gene-duplication events giving rise to RocE and Orf1 of *Helicobacter pylori* (cluster 1), and YdgF and AapA of *B. subtilis* (cluster 3) must have occurred much more recently. Likely candidates for *E. coli*–*B. subtilis* orthologous pairs include YkfD Eco and RocC Bsu and its three close paralogues (cluster 1), GabP Eco and GabP Bsu (cluster 2), ProY Eco and YtnA Bsu, YifK Eco and YbxG Bsu, and CycA Eco and either AapA Bsu or YdgF Bsu (all in cluster 3). Only in the case of the two GabP orthologues (cluster 2) are functional data available to substantiate the conclusion of common function (Brechtel *et al.*, 1996; Brechtel & King, 1998; King *et al.*, 1995).

Fig. 4 shows the tree for the bacterial APA family. Four clusters (1–4) and two long branches (5 and 6) can be seen. The seven *E. coli* paralogues appear in all four

Fig. 5. Phylogenetic tree for the CAT (TC 2.A.3.3) family.

primary clusters. Thus, three *E. coli* paralogues (PotE, YjdE and CadB) as well as a *Haemophilus influenzae* orthologue (PotE Hin), similar to PotE Eco, are found in cluster 1. In the large cluster 2, no two members are from the same organism. Thus, no paralogous pairs are found in this cluster. Nevertheless, these proteins are clearly not all orthologues. For example, the great distance between Yvs Bsu and ArcD Bli, derived from *B. subtilis* and *Bacillus licheniformis*, respectively, suggests that these two proteins diverged following an early gene-duplication event rather than as a result of speciation. Similarly, although seven of the nine proteins in cluster 2 have been annotated as ArcD proteins, the relative distances of these proteins do not generally correlate with the phylogenetic distances of the organisms. This fact suggests that they are not all orthologues of similar function. Cluster 3 consists of a probable pair of orthologues from *E. coli* and *Aeromonas salmonicida*, and cluster 4 consists of a possible orthologous pair from *E. coli* and *B. subtilis* (YkbA Bsu with either YhfM or Orf1 of *E. coli*). The two *E. coli* proteins in this cluster probably arose from a very recent gene-duplication event. Orf1 Sco (5) and PotE Rpr (6) are found at the ends of long branches. The great distance separating PotE Eco from PotE Rpr suggests that these two proteins are not orthologues. PotE Rpr may have been incorrectly annotated.

The tree shown in Fig. 5 for the CAT family shows clustering in accordance with the source organism. Thus, cluster 1 is specific for Gram-positive bacteria, clusters 2 and 3 include only mammalian members, cluster 4 is a group of three paralogues from the nematode *Caenorhabditis elegans* and cluster 5 includes

Fig. 6. Phylogenetic tree for the ACT (TC 2.A.3.4) family.

Fig. 7. Phylogenetic tree for the EAT (TC 2.A.3.5) family.

two paralogues from the plant *Arabidopsis thaliana*. Fig. 6 (the ACT family) shows a group of yeast and fungal proteins related to one plant protein. Little clustering is observed except that Aap2 Ncr and Orf5 Spo are likely to be orthologues, and the clustering of YahB Spo, Orf3 Spo and Gpt1 Cal suggests similar functions for these proteins. The EAT family tree (Fig. 7) is consistent with orthologous relationships between EntP Rer and EutP Zmo as well as between Orf1 Sme and Orf1 Mba.

The ABT family tree (Fig. 8) resembles the EAT family tree in that little clustering is observed. Neither the archaeal nor the bacterial proteins cluster together, suggesting that early gene duplication occurred before divergence of these two domains of prokaryotes. Alternatively, one could invoke lateral gene transfer to

Fig. 8. Phylogenetic tree for the ABT (TC 2.A.3.6) family.

account for the results. The two *B. subtilis* paralogues probably arose by a relatively recent gene-duplication event.

In the GGA family tree (Fig. 9), the four *E. coli* paralogues are each found on distant branches. The XasA proteins of *Chlamydia pneumoniae* and *Chl. trachomatis* are undoubtedly orthologues, but the close clustering of XasA Eco and GadC Lla is surprising. Because *E. coli* and *Lactococcus lactis* are Gram-negative and Gram-positive bacteria, respectively, such close clustering would not be anticipated. We considered that this protein pair might provide an example of a recent lateral gene-transfer event between these two bacterial kingdoms. The same seemed possible for Orf54

Fig. 9. Phylogenetic tree for the GGA (TC 2.A.3.7) family.

Cpe and YgjI Eco which, however, are more distant from each other. Indeed, examination of the G + C contents of these genes revealed that the Gram-positive bacterial homologues exhibit the G + C values expected for the source organisms (35.0 mol % for the *gadC* gene of *L. lactis* compared to 35.4 mol % expected for *L. lactis* genes, and 30.8 mol % for the *orf54* gene of *Clostridium perfringens* compared to 31.0 mol % expected for *Clo. perfringens* genes), but the *E. coli* genes deviated significantly from expectation based on the G + C content for *E. coli* genes as a whole (46.0 mol % for the *xasA* gene and 45.5 mol % for the *ygjI* gene compared to 51.4 mol % expected for *E. coli* genes). In fact, the G + C contents of the two *E. coli* genes are nearly the same at each of the three nucleotide positions of the codons in the *E. coli xasA* and *ygjI* genes, and they differed considerably from those expected for native *E. coli* genes, suggesting that they were obtained horizontally from the same organism or two very similar organisms.

Fig. 11. Phylogenetic tree for the SGP (TC 2.A.3.9) family.

Codon usages for these two genes proved also to be more similar to each other than to *E. coli* genes (data not shown). These results suggest that *E. coli* acquired both of these genes from a Gram-positive bacterium long after Gram-positive and Gram-negative bacteria diverged from each other.

The phylogenetic tree for the large LAT family (Fig. 10), which includes animal and yeast proteins, reveals clustering patterns that in many cases reflect the phylogenies of the source organisms. Thus, the two yeast proteins, Mup1 and Mup3, cluster loosely together as do four ORFs from *Cae. elegans*. However, *Cae. elegans* proteins are found on two additional branches that cluster very loosely with the mammalian proteins. A number of mammalian orthologous and close paralogous relationships can be proposed on the basis of the tree configuration (see Fig. 10).

The bacterial SGP tree (Fig. 11) reveals some early and late gene-duplication events as well as one probable orthologous relationship for the various established and putative germination proteins. Thus, GrkB Bsu and

Fig. 10. Phylogenetic tree for the LAT (TC 2.A.3.8) family.

Fig. 12. Phylogenetic tree for the YAT (TC 2.A.3.10) family.

GerAB Bme are probable orthologues, but the cluster of three *B. subtilis* proteins at the bottom of the tree (GerAB, Orf1 and GerBB) represent paralogues that probably arose by recent gene-duplication events (Corfe *et al.*, 1994; Zuberi *et al.*, 1987). On the other hand, YfkT, GrkB and GerXB represent *B. subtilis* proteins that presumably arose by early gene-duplication events, maybe during the early origin of sporulation in the *Bacillus* line. This tree illustrates the tremendous amount of gene duplication that must have occurred in response to pressures arising during development of a program of prokaryotic differentiation. All of these proteins presumably arose by gene-duplication events in order to render spore germination responsive to the presence of amino acids in the growth medium.

Finally the YAT family (Fig. 12), where the vast majority of the proteins are from one species (*Sac. cerevisiae*), reveals a number of interesting orthologous and paralogous relationships. Particularly worthy of note is the huge cluster of 10 *Sac. cerevisiae* paralogues in a single cluster at the bottom of the tree. The occurrence of additional *Sac. cerevisiae* paralogues scattered throughout the tree is also noteworthy. Many of these yeast proteins are functionally characterized and they exhibit differing specificities. Thus, like bacilli, yeast has proliferated paralogues within the APC superfamily. Although many of these proteins are known to be active in the vegetative yeast cell, it will be interesting to determine if any of them function in the regulation of spore germination as has been demonstrated for the SGP family in *Bacillus*.

DISCUSSION

In this report we provide a detailed and up-to-date phylogenetic characterization of the APC superfamily. Based exclusively on phylogeny (Fig. 1), we identified 10

distinct families, five families specific to bacteria, one bacterial and archaeal family, one bacterial and eukaryotic family with representation only in animals and plants of the eukaryotic domain, and three eukaryote-specific families. Of the three eukaryote-specific families, one is fungi specific (including yeast and mushroom proteins), one includes members from both fungi and plants but not animals, and one includes members from both fungi and animals but not plants. Summarizing these observations, bacterial APC superfamily members are found in 7 of the 10 families, archaeal members are found in only one family, fungal proteins are found in three families, and both plant and animal members are each found in just two families.

Protein size and topological comparisons

Size and topological analyses showed that the SGP family (TC 2.A.3.9) includes the smallest proteins in the APC superfamily and these proteins display only 10 putative TMSs instead of the usual 12. The deficiency is at the C termini of these proteins. Interestingly, no member of this family has been shown to be a transporter, and the possibility that they serve as receptors rather than transporters can be entertained. Thus, loss of TMSs 11 and 12 may result in alteration of function. If, however, a transport function is demonstrated for one of these proteins, one can suggest that TMSs 11 and 12 in other APC superfamily permeases are not essential for transport.

One family, the CAT family, the only family to have representation in both prokaryotes and eukaryotes, shows a 12 TMS topology for the prokaryotic members but a 14 TMS topology for the eukaryotic members. Why these eukaryotic proteins, but no other members of the APC superfamily, have two extra putative TMSs is not known. Because the bacterial 12 TMS proteins

cluster separately from the eukaryotic proteins on the CAT family tree (Fig. 5), it is reasonable to propose that the addition of two extra TMSs to the C termini of these proteins occurred only once during the evolution of the APC superfamily. It is possible that this C-terminal extension is related to the viral-receptor function of several of these proteins, although it could not have evolved to serve this function (see Reizer *et al.*, 1993).

Functionally uncharacterized families

No functional data are available for members of one of the APC superfamily families, the ABT family (TC 2.A.3.6). This family is represented in three archaea (*Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum* and *Archaeoglobus fulgidus*) and three bacteria (*E. coli*, *B. subtilis* and *Myc. tuberculosis*). We can assume that these proteins transport amino acids and/or their derivatives, but direct experimentation will be necessary to establish this assumption. Similarly, although proteins of the SGP family (TC 2.A.3.9) have clearly been implicated in spore germination, there is no evidence as to whether these proteins function in transport or as transmembrane receptors, signalling the presence of amino acids to cytoplasmic regulatory proteins. Determination of the biochemical functions of these proteins will prove of interest for a more complete understanding of the mechanisms regulating endospore germination in *Bacillus* species.

Distant phylogenetic relationships

Although we were able to establish homology for all of the members of the APC superfamily included in this study, we also extended an earlier suggestion (Young *et al.*, 1999) that two other recognized amino acid transporter families are distantly related to the APC superfamily (Table 2; Saier, 2000a). Thus, the eukaryotic AAAP family and the prokaryotic HAAAP family are most likely distant constituents of the APC superfamily although maximal comparison scores of only 7–8 SD were obtained when interfamilial comparisons with any established member of the APC superfamily were done. The PSI-BLAST program with iterations (Altschul *et al.*, 1997), which uses a matrix/motif algorithm to detect distant phylogenetic relationships, also revealed similarities between members of these three families (unpublished results). It is therefore highly likely that the APC superfamily includes at least 12 recognized families instead of just 10, and because the AAAP family tree is complex, with several major sequence-divergent phylogenetic clusters (Young *et al.*, 1999), the AAAP family may appropriately be subdivided into more than one subfamily. However, because we prefer to adhere to a rigorous standard for establishment of homology, we will retain the AAAP and HAAAP families in the TC system separate from the APC superfamily. No evidence for a phylogenetic relationship between the APC superfamily and the many other families of transporters

capable of transporting amino acids and their derivatives has been forthcoming (Saier, 2000a).

It is interesting to note that members of the AAAP and HAAAP families exhibit 11 established or putative TMSs (see Young *et al.*, 1999 for a summary of the biochemical and computational evidence). This number is different from that observed for any one of the established members of the APC superfamily. This fact suggests that the former two families may be more closely related to each other than to established members of the APC superfamily. Assuming all of these families to be related, we conclude that members of the extended APC superfamily may exhibit 10, 11, 12 or 14 TMSs. The functional implications of this finding have yet to be elucidated.

Concluding remarks

Although the reported studies have revealed that the APC superfamily is one of the largest superfamilies found in nature, it has apparently not diversified greatly in substrate specificity. All of its functionally characterized members transport amino acids and/or their derivatives. In view of the broad specificities of permeases in certain other superfamilies such as the larger major facilitator superfamily (TC 2.A.1) (Pao *et al.*, 1998; Saier *et al.*, 1999a) or the smaller solute: sodium symporter family (TC 2.A.21), it is surprising that the APC superfamily has not diverged more extensively with respect to substrate recognition. An explanation for this observation is not currently at hand, but protein architectural constraints could be responsible (see Saier 1994, 1996, 1998).

Future DNA sequencing projects and functional analyses of the revealed genes will undoubtedly result in both numerical and functional expansion of the 10 currently established families of the APC superfamily. Moreover, we can expect that completely novel families within the APC superfamily will be discovered. Evolutionary links to other families such as the AAAP family may also become unequivocally established as 'missing-link' sequences become available. Additionally, more refined *in silico* tools for establishing homology as well as high resolution three-dimensional structural data for related secondary carriers may provide bases for establishing evolutionary links. We anticipate that the rules governing the evolutionary expansion of ancient superfamilies will soon become more clearly defined. Such advances will undoubtedly shed light on the most basic structural restrictions that determine the propensity for functional diversification of homologous transmembrane solute-transport systems during evolution.

ACKNOWLEDGEMENTS

We wish to thank Donna Yun, Monica Mistry, Milda Simonaitis and Mary Beth Hiller for their assistance in the preparation of this manuscript. Work in the authors' laboratory was supported by NIH grants 2R01 AI14176 from The National Institute of Allergy and Infectious Diseases and 9R01 GM55434 from the National Institute of General

Medical Sciences, as well as by the M. H. Saier, Sr. Memorial Research Fund.

REFERENCES

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000). The Pfam protein families database. *Nucleic Acids Res* **28**, 263–266.
- Beckman, M. L. & Quick, M. W. (1998). Neurotransmitter transporters: regulators of function and functional regulation. *J Membr Biol* **164**, 1–10.
- Brechtel, C. E. & King, S. C. (1998). 4-Aminobutyrate (GABA) transporters from the amino acid-polyamine-choline superfamily: substrate specificity and ligand recognition profile of the 4-aminobutyrate permease from *Bacillus subtilis*. *Biochem J* **333**, 565–571.
- Brechtel, C. E., Hu, L. & King, S. C. (1996). Substrate specificity of the *Escherichia coli* 4-aminobutyrate carrier encoded by *gabP*: uptake and counterflow of structurally diverse molecules. *J Biol Chem* **271**, 783–788.
- Claros, M. G. & von Heijne, G. (1994). TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* **10**, 685–686.
- Closs, E. I., Albritton, L. M., Kim, J. W. & Cunningham, J. M. (1993). Identification of a low affinity, high capacity transporter of cationic amino acids in mouse liver. *J Biol Chem* **268**, 7538–7544.
- Corfe, B. M., Moir, A., Popham, D. & Setlow, P. (1994). Analysis of the expression and regulation of the *gerB* spore germination operon of *Bacillus subtilis* 168. *Microbiology* **140**, 3079–3083.
- Cosgriff, A. J. & Pittard, A. J. (1997). A topological model for the general aromatic amino-acid permease, AroP, of *Escherichia coli*. *J Bacteriol* **179**, 3317–3323.
- Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387–395.
- Devés, R. & Boyd, C. A. R. (1998). Transporters for cationic amino acids in animal cells: discovery, structure and function. *Physiol Rev* **78**, 487–545.
- Doolittle, R. F. (1986). *Of Urfs and Orfs: a Primer on How to Analyze Derived Amino Acid Sequences*. Mill Valley, CA: University Science Books.
- Ellis, J., Carlin, A., Steffes, C., Wu, J., Liu, J. & Rosen, B. P. (1995). Topological analysis of the lysine-specific permease of *Escherichia coli*. *Microbiology* **141**, 1927–1935.
- Estévez, R., Camps, M., Rojas, A. M., Testar, X., Devés, R., Hediger, M. A., Zorzano, A. & Palacín, M. (1998). The amino acid transport system $\gamma^+L/4F2hc$ is a heteromultimeric complex. *FASEB J* **12**, 1319–1329.
- Farcasanu, I. C., Mizunuma, M., Hirata, D. & Miyakawa, T. (1998). Involvement of histidine permease (Hip1p) in manganese transport in *Saccharomyces cerevisiae*. *Mol Gen Genet* **259**, 541–548.
- Feng, D.-F. & Doolittle, R. F. (1990). Progressive alignment and phylogenetic tree construction of protein sequences. *Methods Enzymol* **183**, 375–387.
- Hofmann, K. & Stoffel, W. (1993). TMbase – a database of membrane spanning proteins segments. *Biol Chem Hoppe-Seyler* **347**, 166.
- Hofmann, K., Bucher, P., Falquez, L. & Bairoch, A. (1999). The prosite database, its status in 1999. *Nucleic Acids Res* **27**, 215–219.
- Hu, L. A. & King, S. C. (1998a). Membrane topology of the *Escherichia coli* γ -aminobutyrate transporter: implications on the topology and mechanism of prokaryotic and eukaryotic transporters from the APC superfamily. *Biochem J* **336**, 69–76.
- Hu, L. A. & King, S. C. (1998b). Functional significance of the ‘signature cysteine’ in helix 8 of the *Escherichia coli* 4-aminobutyrate transporter from the amino-acid-polyamine-choline superfamily. *J Biol Chem* **273**, 20162–20167.
- Hu, L. A. & King, S. C. (1998c). Functional sensitivity of polar surfaces on transmembrane helix 8 and cytoplasmic loop 8–9 of the *Escherichia coli* GABA (4-aminobutyrate) transporter encoded by *gabP*: mutagenic analysis of a consensus amphipathic region found in transporters from bacteria to mammals. *Biochem J* **330**, 771–776.
- Isnard, A. D., Thomas, D. & Surdin-Kerjan, Y. (1996). The study of methionine uptake in *Saccharomyces cerevisiae* reveals a new family of amino acid permeases. *J Mol Biol* **262**, 473–484.
- Kashiwagi, K., Shibuya, S., Tomitori, H., Kuraishi, A. & Igaragshi, K. (1997). Excretion and uptake of putrescine by the PotE protein in *Escherichia coli*. *J Biol Chem* **272**, 6318–6323.
- Kim, J. W., Closs, E. I., Albritton, L. M. & Cunningham, J. M. (1991). Transport of cationic amino acids by the mouse ecotropic retrovirus receptor. *Nature* **352**, 725–728.
- King, S. C., Fleming, S. R. & Brechtel, C. E. (1995). Ligand recognition properties of the *Escherichia coli* 4-aminobutyrate transporter encoded by *gabP*: specificity of Gab permease for heterocyclic inhibitors. *J Biol Chem* **270**, 19893–19897.
- Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.
- Markovich, D., Stange, G., Bertran, J., Palacin, M., Werner, A., Biber, J. & Murer, H. (1993). Two mRNA transcripts (rBAT-1 and rBAT-2) are involved in system b^{0,+}-related amino acid transport. *J Biol Chem* **268**, 1362–1367.
- Mastroberardino, L., Spindler, B., Pfeiffer, R., Skelly, P. J., Loffing, J., Shoemaker, C. B. & Verrey, F. (1998). Amino-acid transport by heterodimers of 4F2hc/CD98 and members of a permease family. *Nature* **395**, 288–291.
- Palacín, M., Estévez, R., Bertran, J. & Zorzano, A. (1998). Molecular biology of mammalian plasma membrane amino acid transporters. *Physiol Rev* **78**, 969–1054.
- Pao, S. S., Paulsen, I. T. & Saier, M. H., Jr (1998). The major facilitator superfamily. *Microbiol Mol Biol Rev* **62**, 1–32.
- Paulsen, I. T., Sliwinski, M. K. & Saier, M. H., Jr (1998a). Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. *J Mol Biol* **277**, 573–592.
- Paulsen, I. T., Sliwinski, M. K., Nelissen, B., Goffeau, A. & Saier, M. H., Jr (1998b). Unified inventory of established and putative transporters encoded within the complete genome of *Saccharomyces cerevisiae*. *FEBS Lett* **430**, 116–125.
- Reizer, J., Finley, K., Kakuda, D., MacLeod, C. L., Reizer, A. & Saier, M. H., Jr (1993). Mammalian integral membrane receptors are homologous to facilitators and antiporters of yeast, fungi, and eubacteria. *Protein Sci* **2**, 20–30.
- Saier, M. H., Jr (1994). Computer-aided analyses of transport protein sequences: gleaned evidence concerning function, structure, biogenesis, and evolution. *Microbiol Rev* **58**, 71–93.
- Saier, M. H., Jr (1996). Phylogenetic approaches to the identification and characterization of protein families and super-families. *Microb Comp Genomics* **1**, 129–150.

- Saier, M. H., Jr (1998). Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. In *Advances in Microbial Physiology*, vol. 40, pp. 81–136. Edited by R. K. Poole. San Diego, CA: Academic Press.
- Saier, M. H., Jr (1999a). Classification of transmembrane transport systems in living organisms. In *Biomembrane Transport*, pp. 265–276. Edited by L. Van Winkle. San Diego, CA: Academic Press.
- Saier, M. H., Jr (1999b). Eukaryotic transmembrane solute transport systems. In *International Review of Cytology: a Survey of Cell Biology*, pp. 61–136. Edited by K. W. Jeon. San Diego, CA: Academic Press.
- Saier, M. H., Jr (1999c). Genome archeology leading to the characterization and classification of transport proteins. *Curr Opin Microbiol* 2, 555–561.
- Saier, M. H., Jr (2000a). Families of transmembrane transporters selective for amino acids and their derivatives. *Microbiology* 146, 1775–1795.
- Saier, M. H., Jr (2000b). A functional/phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev* 64, 354–411.
- Saier, M. H., Jr, Beatty, J. T., Goffeau, A. & 11 other authors (1999a). The major facilitator superfamily. *J Mol Microbiol Biotechnol* 1, 257–279.
- Saier, M. H., Jr, Eng, B. H., Fard, S. & 15 other authors (1999b). Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochim Biophys Acta* 1422, 1–56.
- Sanders, J. W., Leenhouts, K., Burghoorn, J., Brands, J. R., Venema, G. & Kok, J. (1998). A chloride-inducible acid resistance mechanism in *Lactococcus lactis* and its regulation. *Mol Microbiol* 27, 299–310.
- Sato, H., Tamba, M., Ishii, T. & Bannai, S. (1999). Cloning and expression of a plasma membrane cystine/glutamate exchange transporter composed of two distinct proteins. *J Biol Chem* 274, 11455–11458.
- Sophianopoulou, V. & Dhalluin, G. (1995). Amino acid transporters of lower eukaryotes: regulation, structure and topogenesis. *FEMS Microbiol Rev* 16, 53–75.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25, 4876–4882.
- Torrents, D., Estévez, R., Pineda, M., Fernández, E., Lloberas, J., Shi, Y.-B., Zorzano, A. & Palacín, M. (1998). Identification and characterization of a membrane protein (y⁺L-amino acid transporter-1) that associates with 4F2hc to encode the amino acid transport activity y⁺L: a candidate gene for lysinuric protein intolerance. *J Biol Chem* 273, 32437–32445.
- Verrey, F., Jack, D. L., Paulsen, I. T., Saier, M. H., Jr & Pfeiffer, R. (1999). New glycoprotein-associated amino acid transporters. *J Membr Biol* 172, 181–192.
- Von Heijne, G. (1992). Membrane protein structure prediction, hydrophobicity analysis and positive-inside rule. *J Mol Biol* 225, 487–494.
- Wang, H., Kavanaugh, M. P., North, R. A. & Kabat, D. (1991). Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter. *Nature* 352, 729–731.
- Young, G. B., Jack, D. L., Smith, D. W. & Saier, M. H., Jr (1999). The amino acid/auxin:proton symport permease family. *Biochim Biophys Acta* 1415, 306–322.
- Zuberi, A. R., Moir, A. & Feavers, I. M. (1987). The nucleotide sequence and gene organization of the *gerA* spore germination operon of *Bacillus subtilis* 168. *Gene* 51, 1–11.

Received 19 October 1999; revised 28 March 2000; accepted 24 May 2000.