Computational investigations of divalent heavy metal ion homeostasis

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

Mathematisch-Naturwissenschaftlich-Technischen Fakultät
der Martin-Luther-Universität Halle-Wittenberg

von Herrn Torsten von Rozycki

geboren am 4. Juli 1976 in Halle/Saale

Gutachter:
1. Prof. Dr. D.H. Nies
2. Prof. Dr. G.-J. Krauss
3. Prof. Dr. M.H. Saier Jr.

ZUSAMMENFASSUNG

So wie viele anderen *Burkholderiaceen* kann auch *Cupriavidus metallidurans* CH34 sehr vielfältige Habitate nutzen. Dies schliesst in diesem Fall auch Standorte mit ein, die stark schwermetallbelastet sind. Das ungewöhnlich grosse Genom dieses β-Proteobakteriums wurde vom Joined Genome Institute (JGI) entschlüsselt und konnte somit für die Identifikation und Charakterisierung wichtiger Schwermetallresistenzdeterminanten herangezogen werden. Für weiterführende bioinformatische, funktionelle und phylogenetische Untersuchungen wurden die Genomsequenzen mit denen der eng verwandten Proteobakterien *Cupriavidus eutrophus* H16, *Cupriavidus eutrophus* JMP134, *Ralstonia solanacearum* GMI1000 und mit zwei weiteren Burkholderiaceen verglichen. Die Ergebnisse bilden die Grundlage für eine lokale biologische Datenbank, die für weiterführende Untersuchungen herangezogen werden kann.

- Neben einem weitgehend chromosomal codierten orthologen Grundset an Proteinen für essentielle Stoffwechselleistungen besitzen die untersuchten Stämme auch ein gemeinsames Repertoire von Transportproteinen für die Aufnahme von Nährstoffen oder essentiellen Kationen. Sie besitzen dabei ein ungewöhnliches Übergewicht an Transportern für Di- und Tricarbonsäuren, Aminosäuren und Peptiden haben jedoch deutlich weniger zuckerspezifische Transporter als viele andere Bakterien. Viele essentielle Schwermetallentgiftungssysteme, insbesondere die CHR (TC 2.A.51), MIT (TC 1.A.35), CDF (TC 2.A.4) Transporterfamilien und die kupfer- und silberspezifischen CPX-ATPasen (TC 3.A.3) gehören ebenfalls zum gemeinsamen Repertoire dieser Proteobakterien

- Ausgehend von einem bereits mäßig schwermetallresistenten gemeinsamen Vorläuferstamm hat insbesondere in den untersuchten *Cupriavidus* Stämmen offenbar eine schrittweise evolutionäre Anpassung bestimmter Schwermetallresistenzdeterminanten stattgefunden. Dies hat mit *C. metallidurans* CH34 letztendlich zu einem Proteobakterium geführt, das durch viele zusätzliche Resistenzdeterminanten speziell an stark schwermetallbelastete Standorte angepasst ist. Viele dieser paralogen Schwermetalltransporter werden durch die Plasmide codiert und die zusätzlichen Replicons codieren auch viele ‚seltene' Orthologe, die in den untersuchten Stämmen weniger stark verbreitet sind. Stammspezifische Genduplikationen und auch horizontaler Gentransfer haben daher vermutlich wesentlich zur hohen genetischen Flexibilität von Schwermetallresistenzdeterminanten in *C. metallidurans* CH34 beigetragen. Von besonderer Bedeutung sind dabei die für Zink/Cadmium/Blei spezifischen paralogen CPX-ATPasen (TC 3.A.3) CadA, PbrA und CzcP. Die Calcium – spezifischen CPX-ATPasen, viele KUP (2.A.72) und TPS Transporter (TC 1.B.20) sowie zahlreiche MerT Proteine (TC 9.A.2) sind ebenfalls das Produkt einer Genduplikation.

- *C. metallidurans* CH34 besitzt auch viele zusätzliche stammspezifische paraloge RND (TC 2.A.6) Transporter für die Ausschleusung divalenter Schwermetallionen. Viele dieser HME (TC 2.A.6.1.) RND Proteine bilden mit OMF (1.B.17) Transportern und MFP (TC 8.A.1) Proteinen ein transperiplasmatisches CBA-Efflux System (Saier *et al*., 1994). In *C. metallidurans* CH34 sind viele dieser CBA-Komponenten in gemeinsamen Operons organisiert und weisen eine ähnliche evolutionäre Entwicklung auf. Jedoch codieren die entsprechenden plasmidständigen Operons in *C. metallidurans* CH34 auch viele zusätzliche Transporter von denen viele vermutlich einen anderen Ursprung haben. Die Expression vieler CBA-Transportkomplexe wird durch ECF (extracytoplasmic function) Sigmafaktoren kontrolliert (Lonetto, *et al.*, 1992) und viele dieser Regulatoren in *C. metallidurans* CH34 sind ebenfalls das Produkt einer Genduplikation. *C. metallidurans* CH34 besitzt hingegen kaum paraloge regulatorische Zwei-Komponenten-Systeme.

- Als integrales Membranprotein gehört YedZ (TC 9.B.43.) in *E. coli* einer noch weitgehend uncharakterisierten Transporterfamilie (TC-Klasse 9) an. Die Sequenzähnlichkeit von Transmembransegmenten (TMSs) in YedZ deutet auf eine intragenomische Gentriplikation hin, wobei ein 2 TMS codierendes Segment zu den insgesamt 6 Transmembransegmenten der YedZ Transporterfamilie geführt hat. Mehrere neuartige MFS (2.A.1) Transporter in magnetotaktischen Bakterien und auch einige Elektronentransportproteine in Cyanobacterien verfügen über eine C-terminale, zu YedZ ähnliche Proteindomäne die möglicherweise Redoxreaktionen oder den Elektronentransportprozess dieser Proteine kontrolliert.

- Die SbtA Transporterfamilie (TC 2.A.83) ist für die unspezifische Aufnahme von Hydrogencarbonat (HCO_3^-) in Cyanobacterien verantwortlich. Dabei arbeiten diese Transporter mit anderen Mechanismen zur Anreicherung von CO_2 (CCMs) als Substrat für die Ribulose Bisphosphate Carboxylase-Oxygenase (RuBisCO) zusammen (Badger *et al*., 2006). Die zehn transmembranen Segmente (TMSs) in SbtA resultieren vermutlich aus einer Genduplikation, wobei beide Hälften eine entgegengesetzte Orientierung in der Membran besitzen. Die Transporter der SbtA Familie lassen sich in zwei phylogenetisch unterschiedliche Gruppen mit möglicherweise verschiedenen Aufgaben unterteilen.

SUMMARY

The β-proteobacterium *C. metallidurans* CH34 and many related *Burkholderiacea* are able to use highly diverse ecological niches. Its remarkable ability to also survive in heavy metal polluted environments is however an almost unique feature of this organism. The extraordinary large genomic sequence of this strain has been deciphered and published by the Joined Genome Institute (JGI) and these genomic sequences have largely facilitated the identification and characterization of essential transition metal homeostasis determinants. Succeeding computational phylogenetic and functional investigations of these and other determinants have also relied on a comparison with closely related proteobacteria. This includes *C. eutrophus* H16, *C. eutrophus* JMP134, *R. solanacearum* GMI1000 and two additional strains of the *Burkholderia* group. The resulting information was gathered into a local database to facilitate further computational research.

- Particularly the chromosomal determinants of these strains encode many orthologs for essential housekeeping functions. Additionally has an orthologous core set of transporters for the uptake of nutrients or essential cations been revealed. All these strains encode many transporters for di-and tricarboxylates, amino acids and peptides but they have only few sugar specific transporters and they have thus probably largely shifted their diet from sugars to amino acids. Many proteins for heavy metal homeostasis are also widely distributed. This includes the putative CHR (TC 2.A.51), MIT (TC 1.A.35), CDF (TC 2.A.4) transporters and copper- and silver-specific CPX-ATPases (TC 3.A.3).

- The common ancestor of the *Cupriavidus* strains was probably already moderately heavy metal resistant. Particularly in *C. metallidurans* CH34 has a pursuing stepwise evolutionary upgrade of transition metal transporters probably broadened its adaptation to strongly polluted environments. The plasmids of these strains encode many paralogs and many sparsely distributed “rare” orthologs. The unusual genomic flexibility of the plasmid encoded metal ion transporters is thus probably largely based on horizontal gene transfer and extensive gene duplications. The duplication of the zinc/cadmium/lead specific P-type ATPases (TC 3.A.3) (CadA, PbrA and CzcP) in *C. metallidurans* CH34 and the duplication of its calcium specific P-type ATPases might reflect this adaptation. Many transporters of the TPS family (TC 1.B.20) and KUP family (2.A.72) are also paralogs and many paralogous MerT proteins (TC 9.A.2) have also been found in this strain.

- *C. metallidurans* CH34 contains many strain-specific paralogs of the RND (TC 2.A.6) transporter family for the extrusion of divalent heavy metal cations. Many HME RND (TC 2.A.6.1) transporters are combined with OMF (1.B.17) transporters and MFP (TC 8.A.1) proteins and they form a transperiplasmic CBA-type efflux complex (Saier *et al.*, 1994). In *C. metallidurans* CH34 encode the plasmid borne operons for CBA-type efflux systems many auxiliary transporters with a divergent evolutionary background. Many CBA-type efflux systems are controlled by ECF (extracytoplasmic function) sigma factors (Lonetto, *et al.*, 1992) and many of these regulators in *C. metallidurans* CH34 are the product of a gene duplication event. This contrasts to a largely conserved set of regulatory two-component-systems in the investigated strains.

- The integral membrane protein YedZ (TC 9.B.43.) of *E. coli* belongs to the group of poorly characterized transporters. Homology could be shown for several transmembrane segments (TMSs) of representatives of the YedZ family which suggests that an intragenic triplication event of a 2 TMS containing segment has resulted in the six transmembrane segments of this transport protein family. Several novel MFS (2.A.1) transporters in magnetotactic bacteria and some transmembrane electron carriers in cyanobacteria contain a C-terminal YedZ like domain that might facilitate or regulate oxidoreduction, transmembrane electron flow, and transport.

- The SbtA transporter Family (TC- 2.A.83) is responsible for low affinity sodium-dependent bicarbonate (HCO^{3-}) uptake in cyanobacteria. These transporters are combined with other specialized CO_2 concentrating mechanisms (CCMs) to supply the ribulose bisphosphate carboxylase-oxygenases (RuBisCO) with inorganic carbon. Sequence similarity analysis has suggested that the ten transmembrane segments (TMSs) of SbtA might have been formed by an intragenic duplication event and these two halves have an opposite orientation in the membrane. The SbtA family members can be grouped into two distinct phylogenetic clusters which may have multiple transport functions.

INDEX

ZUSAMMENFASSUNG I
SUMMARY III
INDEX V
ABBREVIATIONS VII
DEFINITIONS IX

1. INTRODUCTION 1

2. METHODS 5
   2.1. Sequence alignment 5
   2.2. Multiple sequence alignment 8
   2.3. Topology prediction 9
   2.4. Additional programs 10
   2.5. Protein structure analysis 11
   2.6. Databases 13
   2.7. Gap closure 18

3. RESULTS 20
   3.1. Overview of publications 20
   3.1. The YedZ family: possible heme binding proteins that can be fused to transporters and electron carriers. 21
   3.2. Sequence analyses of cyanobacterial bicarbonate transporters and their homologues. 24
   3.3. Genomic analyses of transport proteins in *Ralstonia metallidurans* CH34. 27
      3.3.1. Summary 27
      3.3.2. Results of succeeding investigations 27
   3.4. *Cupriavidus metallidurans*: evolution of a metal-resistant bacterium 35
   3.5. Unpublished results of supplementary investigations 40
      3.5.1. Ordering contigs of *C. metallidurans* CH34 40
      3.5.2. Annotation of *C. metallidurans* CH34 42
      3.5.3. Characterisation of metalloproteins 43
      3.5.4. The protein structure of CueO 44

4. DISCUSSION 47
4.1. Genomic changes can effect specific adaptations 47
4.2. The proteomic repertoires of proteobacteria 47
4.3. The proteomic repertoire of *C. metallidurans* CH34..48
4.4. The plasmids of proteobacteria are hotspots for new unique functions........................48
4.5. The transport protein repertoires of proteobacteria 50
4.6. Mechanisms of metal ion homeostasis..52
4.7. Transporter families for metal ion homeostasis 54
4.8. Metal ion transport proteins in *C. metallidurans* CH34 and related bacteria 55
4.9. The HME RND subfamily is overrepresented in *C. metallidurans* CH34 57
4.10. The composition of HME-CBA-type exporter clusters 57
4.11. The assembly of the CBA-type efflux complex 59
4.12. The conservation of regulatory proteins 62
4.13. YedZ contains fused domains for transport and electron transfer..............................64
4.14. SbtA combines similarity to secondary carriers and ABC transporters......................67
4.15. Outlook...69

5. REFERENCES..71
6. OVERVIEW OF PUBLICATIONS 1

ABBREVIATIONS

aa	amino acids
bp	basepairs
ABC-transporter	ATP-binding cassette-transporter
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
CDF	cation diffusion facilitator
C-terminal	carboxyterminal
Da	dalton
DOE	department of energy
DNA	desoxy ribonucleic acid
ECF	extracellular function
e.g.	for instance
EMBL	european molecular biology laboratory
Glc	glucose
HAE	hydrophobe/amphiphile efflux
HME	heavy metal efflux
HMM	hidden markov model
IUBMB	International Union of Biochemistry and Molecular Biology
IUPAG	international union of pure and applied chemistry
JGI	joined genome institute
Kbp	kilobases
Mbp	megabases
MFP	membrane fusion protein
MIC	minimal inhibitory concentration
MIT	metal inorganic transporter
MSA	multiple sequence alignment
MW	molecular weight
NCBI	national centre of biotechnology information
N-terminal	aminoterminal
NRAMP	natural-resistance-associated macrophage protein
pH	negative decadic logarithm of proton concentration
PP	periplasm
PEP	phosphoenolpyruvate
PMF	proton motive force
RDP	ribosomal database project
rmsd	relative mean square deviation
OMP	outer membrane protein (transport protein family)
OMF	outer membrane factor
Orf	open reading frame
RNA	ribonucleic acid
RND	resistance, nodulation, cell division protein family
Sec	seconds
Tab.	table
TC-System	transporter classification-system
TMS	transmembrane segment
URR	upstream regulatory region
UTP	uridine triphosphate
Vol.	volume
ZIP	ZRT, IRT-ähnliche Proteine
ZnT	zinc transporter

addresses of used bioinformatic internet sites

Site	Description	Address
PEDANT	Protein Extraction, Description and Analysis Tool	http://pedant.gsf.de/
EBI	European Bioinformatics Institute	http://www.ebi.ac.uk/genomes/
NCBI	National Center for Biotechnology Information	http://www.ncbi.nlm.nih.gov
TIGR	The Institute for Genomic Research	http://www.tigr.org
GOLD	Genomes Online Database	http://www.genomesonline.org
DDBJ	DNA Database of Japan	http://gib.genes.nig.ac.jp/
KEGG:	Kyoto Encyclopedia of Genes and Genomes	http://www.genome.jp/kegg/
CMR	Comprehensive Microbial Resource	http://www.tigr.org/tigr-scripts/CMR2/
IMG	Integrated Microbial Genomes	http://img.jgi.doe.gov/
COGS	Clusters of Orthologous Genes	http://www.ncbi.nlm.nih.gov/COG/
FUSIONDB	database of bacterial and archaeal gene fusion events	http://igs-server.cnrs-mrs.fr/FusionDB/
TCDB	The transport protein classification system	http://www.tcdb.org/
DOE JGI	Joined Genome Institute (US department of energy)	http://genome.jgi-psf.org/mic_home.html
SWISS-MODEL	comparative protein modelling server	http://www.expacy.ch/swissmod/SWISS-MODEL:htm
SWISS-PDB-VIEWER	protein structure analysis tool	http://www.expacy.ch/spdbv
RASMOL	protein structure analysis tool	http://www.umass.edu/microbio/rasmol
PDB (RCSB)	protein data bank (protein structures)	http://www.pdb.org
CHIME	protein structure analysis	http://www.umass.edu/microbio/chime
MDB	MetalloproteinDatabase	http://www.metallo.scripps.edu/
RDP-II	The ribosomal database project	http://rdp.cme.msu.edu/treebuilder/treeing.spr
SIGNALP	signal peptides cleavage sites prediction	http://www.cbs.dtu.dk/services/SignalP-2.0/
APPLESCRIPT	macintosh scripting language	http://www.apple.com/applescript/
Umweltbundesamt	Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit	http://www.umweltbundesamt.de

DEFINITIONS

Alignment
Juxtaposition of nucleotide or protein sequences with the aim to establish homology with maximal similarity and minimal inferred changes among the sequences

Annotation
An annotation is a note added by way of explanation or commentary. In a biological sense this includes the process of identifying all of the coding regions in a genome and determining what those genes do.

Bioinformatics
Use of computer programs for the analysis and storage of DNA and protein sequences

Bootstrapping
Bootstrapping or jackknifing are statistical methods that randomly sample or delete columns in sequence data and they allow to estimate sample variance

Competence
The ability of bacteria to take up extracellular DNA

Contig
The result of joining an overlapping collection of sequence reads

Conservation
Changes at a specific position of an amino acid or DNA sequence that preserve the physico-chemical properties of the original residue.

Coordination number
The total number of sites of the central metal atom or ion which are occupied by ligands

Dendrogram
Branching tree like phylogenetic diagram

Distance matrix
Approach for the construction of phylogenetic trees that minimizes the difference between the realized tree and measured distances

Draft sequence
Collection of contigs of various sizes with unknown order and orientation that may contain sequencing errors and possible misassemblies

Evolutionary distance
The sum of the physical distances that separate organisms in evolutionary trees

Evolutionary relatedness
Inverse proportional to evolutionary distance in evolutionary trees

Finished sequence:
Contiguously sequenced clone insert with high quality standard that do not contain unresolved gaps

G+C ratio
Percentage of the total nucleic acid consisting guanine plus cytosine bases in DNA or RNA sequences (as mol % GC)

Genomics
Discipline of mapping, sequencing and analysing genomes

Genomic Island
Large mobile or potentially mobile DNA segments that are integrated into the chromosome or other replicons

Heavy metals
Heavy metals are metals with a density of more than 5 g/cm^3.

Heuristic method
Analysis procedure that does not guarantee to find the optimal solution to a problem – usually used to increase the speed over exact methods

Homology
Similarity attributed to descent from a common ancestor

Homologs
Genes sharing a common origin

Hydrogen bond
Weak chemical bond between electronegative elements (oxygen/nitrogen atoms) and hydrogen atoms
Identity
The extent to which two (nucleotide or amino acid) sequences are invariant.
Outparalogs
Paralogous genes resulting from a duplication(s) preceding a given speciation event
Inparalogs
Paralogous genes resulting from a lineage-specific duplication(s) subsequent to a given speciation event
Ingroup
In phylogenetics an assumed monophyletic group, usually comprising the taxa of interest
IS-element
Insertion sequences are special types of transposable elements
k-tuple
The term k-tuple denotes a contiguous sequence of DNA bases that is k bases long.
Magnetosome
Organelle in magnetotactic bacteria that consists of a magnetic crystal surrounded by a lipid bilayer membrane
Maximum likelihood
Criterion to estimate a parameter from observed data under an explicit model. In phylogenetic analysis the optimal tree under the maximum likelihood criterion is the tree that is the most likely to have occurred given the observed data and the assumed model of evolution.
Maximum parsimony
Criterion for estimating a parameter from observed data based on the principle of minimizing the number of events needed to explain the data
Metagenomics
The isolation and sequencing of DNA derived from diverse and mixed microbial communities (environmental genomics or ecogenomics)
Metal Chelates
Polydentate ring forming ligand groups working as chelating agents
Metalloproteins
Proteins that bind tightly (with an association constant greater than 10^8 M^{-1}) to one or more metal ions or metal-containing cofactors
Metallomes
Sum of biological molecules containing biometals
MIC
Minimal inhibitory concentration – the minimum concentration of a substance necessary to prevent microbial growth
Monophyletic group
A group or taxa that contain their common ancestor and all of its descendants
Multidrug resistance
The simultaneous acquisition of resistance to many chemically unrelated compounds to which the cell has never been exposed.
Mutation
Inheritable change in base sequence in the genome of an organism
Neighbour joining
Heuristic search algorithm to find a minimum evolution tree
Neural network
Artificial neural networks or machine learning algorithms can learn non-linear mappings from even noisy labelled data sets and have the potential to analyze complex data structures in highdimensional spaces.
Open Reading Frame (ORF)
DNA sequence leading to a protein of known length and composition
Operon
Cluster of genes whose expression is controlled by a single common operator

Orthologs
Homologous sequences in different species that arose from a common ancestral gene during speciation
Outparalogs
Genes that derive from a duplication event before a speciation event of interest – they are thus not orthologs according to definition.
Outgroup:
One or more species, that are phylogenetically distant to the taxonomic group of interest (the ingroup).
Paralogs
Homologous genes, that are related by a duplication event. They might be in the same or in different genome.
Parsimony
In general when judging hypotheses parsimony is the preference for the least complex explanation for an observation. In phylogenetic analyses under maximum parsimony, the preferred phylogenetic tree is the tree that requires the least number of evolutionary changes.
Phylogenetic tree.
Representation of the historical relationships among lineages of organisms or their parts (e.g. genes)
PHB
Poly-beta-hydroxybutyrate (PHB) is a storage material in prokaryotes being composed of beta-alkanoic acids (PHAs)
Porins
Protein channels in the outer membrane of Gram-negative bacteria for small and medium sized molecules
Primary structure
Sequence of monomeric subunits of macromolecule
Promoter
Site of DNA for RNA polymerase binding and begin of RNA transcription
Prosthetic group
Nonprotein constituent of an enzyme
Proteome
Total complement of proteins present in an organism at one time
PMF
Proton motive force (PMF) energized state of a membrane created usually by an electron transport chain
Quarternary structure
Number and arrangement of individual polypeptides in a final protein molecule
Regulon
Set of operons that are all controlled by a single regulator
Repression
Inhibition of the synthesis of a protein by an external substance (repressor)
Response Regulator
Members of two-component-systems with a regulatory protein being phosphorylated by a sensor kinase
Rooted Tree
Tree in with a known position of the ancestor
RuBisCO
Ribulose bisphosphate carboxylase-oxygenases - primary carboxylating enzyme in photosynthetic organisms
Scaffold
A consistent scaffold is a unidirectional path that puts contigs in a definite order
Screening
Sorting of organisms by phenotype or genotype by allowing growth of only particular types of organisms
Secondary metabolite
A product excreted by an organism at the end of the growth phase or during the stationary phase

Secondary structure
Initial folding pattern of polypeptides or polynucleotides as determined by hydrogen bonding
Sensor Kinase
Members of two-component-systems located within the cell wall that phosphorylates itself and transferes the phosphoryl group to a response regulator protein
Siderophore
low molecular weight iron chelator for survival at low iron concentrations
Similarity
The extent to which nucleotide or protein sequences are related. It is based upon identity plus conservation.
Species
Collection of close related strains (>97% sequence homology of 16SrRNA and >70% genomic hybridization)
Synteny
Conservation in gene order
Synteny maps
These analyses produce sets of true, one-to-one orthologs, and this presentation incorporates a view of their relative physical positions across multiple genomes.
Taxonomy
Study of scientific classification and nomenclature
Topology
The branching pattern of a phylogenetic tree
Transition metals
Generally the elements existing between the Alkali metals and earths (Groups 1 and 2) and the non-metal elements (Groups 13 - 18), that is they are the elements where the d-orbitals are being filled
Two-component-system
Regulatory system containing of response regulators and sensor kinases
Unrooted Tree
Tree with no ancestral node
Xenobiotics
Compounds that are released in any compartment of the environment by the action of man and thereby occur in a concentration that is higher than natural

one / three letter abbreviations of amino acids

abbreviation		amino acid	abbreviation		amino acid
A	Ala	alanine	M	Met	methionine
C	Cys	cysteine	N	Asn	asparagine
D	Asp	aspartate	P	Pro	proline
E	Glu	glutamate	Q	Gln	glutamine
F	Phe	phenylalanin	R	Arg	arginine
G	Gly	glycine	S	Ser	serine
H	His	histidine	T	Thr	threonine
I	Ile	isoleucine	V	Val	valine
K	Lys	lysine	W	Trp	tryptophan
L	Leu	leucine	Y	Tyr	tyrosine

1. INTRODUCTION

The eastern heart of Germany comprises an extensive accumulation of diverse industrial facilities for the production of diverse chemical products. This ‘chemical triangle’ spans a large area around the cities Bitterfeld Halle/Saale and Leipzig (figure 1 B). Major chemical facilities are the BASF-Leuna-Werke which were founded in 1916 for ammonia synthesis and the BUNA-Werke for synthetic rubber production. The build up of an extensive chemical industry in this region was continued thought the entire communist decades. Not much care was taken to assess the detrimental effects of these industrial activities to the environment. The landscape, rivers and lakes were thus largely toxified by industrial wastes of nearly any kind. According to information’s of the Bundesumweltministerium were in 1990 only three percent of the east- German rivers and only one percent of the lakes ecologically intact (http://www.umweltbundesamt.de). Continuous mining and metallurgic activity in the Mansfeld district of this region has also produced large deposits of metalliferous sludge (Schreck, 1997; Krauss *et al.,* 2001). Figure 1 shows the resulting elevated heavy metal concentrations (1990) in sediments of certain East German and West German rivers (http://www.umweltbundesamt.de). The DFG-graduiertenkolleg ‘adaptive physiological biochemical reactions to ecological relevant effectors’ has coordinated investigations to uncover mechanisms of heavy metal homeostasis in plants and bacteria and to thus facilitate the development of new approaches for an efficient biodegradation of toxic wastes.

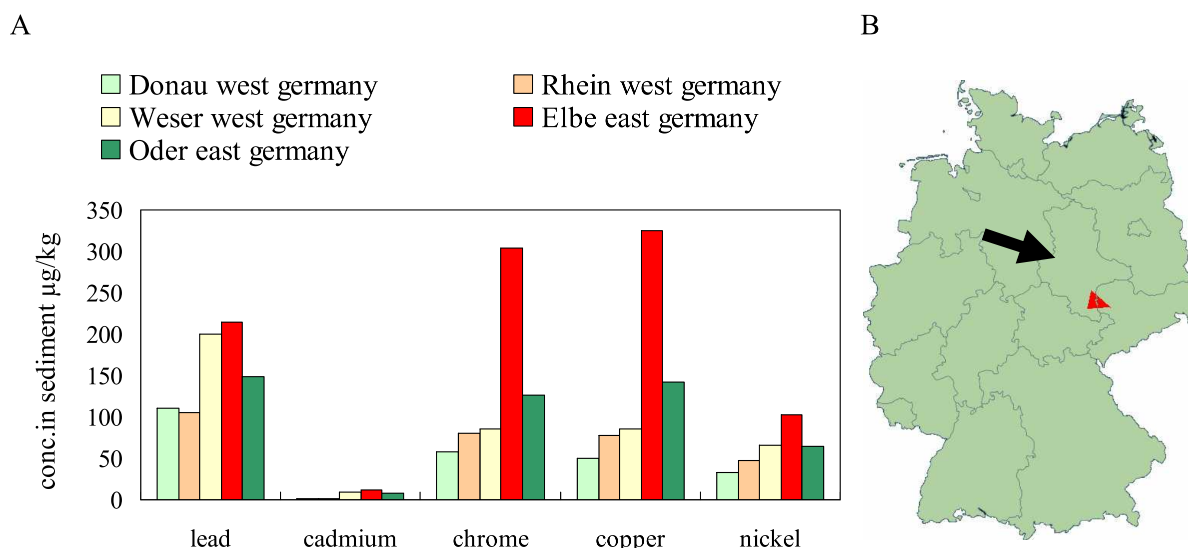
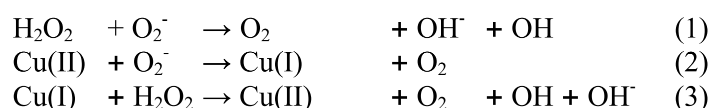


Fig. 1: Heavy metals pollution of major german rivers (http://www.umweltbundesamt.de)
The Elbe River passes nearby a region that exhibits an intensive accumulation of chemical industry. This ‘chemical triangle’ around the cities Bitterfeld Halle/Saale and Leipzig is shown as red triangle in map B. The red bars (diagram A) indicate the heavy metal ion concentrations (1990) of the Elbe compared to those in other large German rivers.

Not only major bioelements (C, O, H, N, P, S, Cl), but also many metal ions (e.g. Na^+, K^+, Mg^{2+}, Ca^{2+}, Fe^{3+}) are ubiquously distributed and widely used in bacterial cells and many enzymes take advantage of their unique chemical features (Gadd 1992; Wackett *et al.*, 2004). All divalent heavy metal cations are however potentially toxic and their toxicity increases within each group of the periodic table from top to down (Nies 2004b). Particularly the heavy metal ions mercury, lead, cadmium, and silver can bind to structural or functional important thiol groups of proteins (Nies, 2003). Heavy metal ions can also replace catalytical essential metal ions and chromate and copper can be implicated in the production of highly reactive oxygen species (Fenton's/ Haber Weiss reactions are shown in figure 2) (Fenton 1876; Koppenol, 2001; da Silva and Williams, 2001). The cellular metal ion homeostasis via uptake and efflux keeps the balance between a possible metal ion shortage and overflow (Rosen, 2002).



$$H_2O_2 + O_2^- \rightarrow O_2 + OH^- + OH \quad (1)$$
$$Cu(II) + O_2^- \rightarrow Cu(I) + O_2 \quad (2)$$
$$Cu(I) + H_2O_2 \rightarrow Cu(II) + O_2 + OH + OH^- \quad (3)$$

Fig. 2: Fenton-like reactions (Liochev and Fridovich, 2002; Santo *et al.*, 2008).
The Fenton's reaction is the reaction of ferrous ions with hydrogen peroxide that results in the production of highly reactive oxygen species. The rate of the reaction of dihydrogen peroxide with superoxide (equation 1), is strongly accelerated in the presence of copper (or iron) (equations 2 and 3): Cu (II) is initially reduced by superoxide (equation 2), followed by reoxidation by dihydrogen peroxide (equation 3), resulting in a net production of the hydroxyl radical

Many proteobacteria can use highly diverse ecological niches and their taxonomic name reflects their versatility. The godfather for the group of proteobacteria is the Greek God Proteus who was able to change his shape at will (Mergeay 2000; Mergeay *et al.*, 2003). Some proteobacteria are also specifically adapted to highly polluted environments and they are thus often used as model organisms for heavy metal homeostasis and for specialized biodegradation pathways (Grover *et al.*, 2006; Valls and de Lorenzo, 2002). A novel rather unusual and potentially problematic environmental niche is spacecraft. Not only pathogenic strains may cause problems in these enclosed environments. Proteobacteria might also inflict allergies, water- or food spoilage and even material deterioration. The latter effect may be caused by organic polymer degradation and metal oxidation (Novikova *et al.*, 2006). The β proteobacterium *C. metallidurans* CH34 was used to investigate these and other potential microbe-related problems in a recent project onboard of the international space station ISS (Pierson, 2001).

Cupriavidus (formerly called *Alcaligenes, Ralstonia, Wautersia*) *metallidurans* CH34 (Vaneechoutte *et al.,* 2004; Valls and de Lorenzo, 2002; Goris *et al.,* 2001; Vandamme and Coenye, 2004) is a mesophilic heavy metal resistant bacterium. It was isolated from a zinc decantation tank but it also exhibits a high degree of persistence (figure 3) to many other transition metal ions (Mergeay *et al.,* 1978; Mergeay *et al.,* 1985; Mergeay, 2000).

Tab. 1: Transition metal ion homeostasis determinants in *C. metallidurans* CH34 (Mergeay *et al*., 2003)

pMOL30		pMOL28	
determinant	substrates	determinant	substrates
czc	Zn^{2+}; Co^{2+}; Cd^{2+}	*cnr*	Co^{2+}; Ni^{2+}
cop	Cu^{2+}	*chr*	Chromate
mer	Hg^{2+}	*mer*	Hg^{2+}
pbr	Pb^{2+}		
ncc	Ni^{2+}; Co^{2+}; Cd^{2+}		

Many important heavy metal resistance determinants (table 1) are cumulated at its megaplasmids pMOL28 (180 Kbp; Taghavi *et al*., 1997) and pMOL30 (238 Kbp; Mergeay *et al*., 1985). The megaplasmid pMOL28 contains important determinants for the resistance to chromate (*chr* - Nies *et al*., 1990), mercury (*mer* - Diels *et al*., 1985) and for the export of cobalt and nickel (*cnr* - Siddiqui *et al*., 1988). Major resistance systems for zinc, cobalt and cadmium (*czc* - Nies *et al*., 1987) as well as lead (*pbr* - Corbisier *et al*., 1999, Mergeay 2000), silver (*sil* - Mergeay *et al*., 2003; Monchy *et al*., 2006), copper (*cop* - Dressler *et al*., 1991) and mercury (*mer* - Diels *et al*., 1985) can be found at plasmid pMOL30 (table 1) (Monchy *et al*., 2007). *C. metallidurans* CH34 containing biofilms were also detected on gold grains from two Australian sites (Reith *et al.,* 2007). Moreover has selenium (Se^{4+}) and gadolinium (Ga^{3+}) fixation been reported for this β-proteobacterium (Andres *et al*. 2000; Roux *et al*. 2001; Ledrich *et al*., 2005).

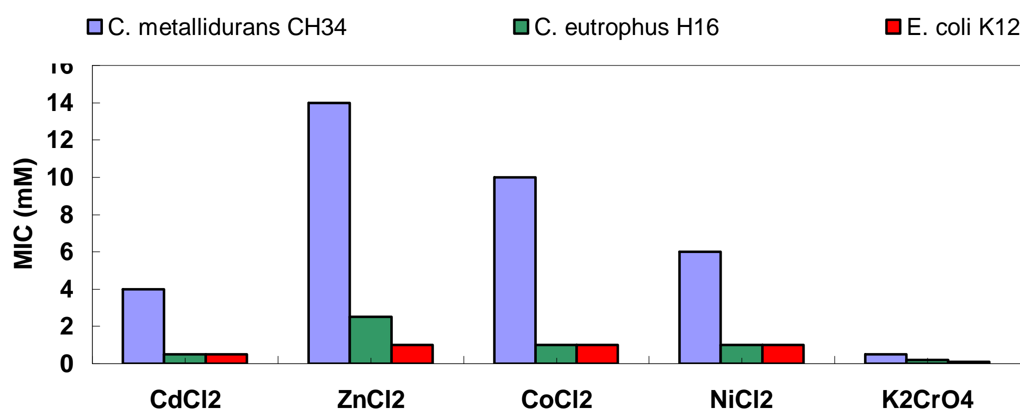


Fig. 3: MIC concentrations for metal ions in *C. metallidurans* CH34 and *E. coli* (Dressler *et al.,* 1991)
The potential of substances to be toxic for the cells is reflected by the MIC-concentration (minimal inhibitory concentration). The table shows the minimum concentration of certain heavy metals ions that prevents microbial growth of *C. metallidurans* CH34 (blue); *C. eutrophus* H16 (green) and *E. coli* K12 (red).

These determinants increase the MIC to zinc fifty fold, that to cobalt about thirty fold and that to cadmium about seven fold (Nies, 2003). A comparison of the MIC values to heavy metal ions of other proteobacteria is shown in figure 3. Computational investigations of its genomic sequences have been used to give further insights into its transition metal ion homeostasis. The transporter protein database TCDB (at http://www.tcdb.org/) was used to create an inventory of the total transport protein repertoire of *C. metallidurans* CH34. Pursuing investigations were focussed at the genomic flexibility of heavy metal specific transporters and their corresponding regulators. A comparison to the proteomic

sequences of closely related *Burkholderiacea* has supported these investigations. The evolution and function of the uncharacterized YedZ transporter family (TC 9.B.43.) in *E. coli* was also investigated. Similarily were computational phylogenetic and functional investigations also focussed at the emergence and function of the SbtA bicarbonate transporter family (TC 2.A.83) of cyanobacteria.

2. METHODS

2.1. Sequence alignment

Most computational phylogenetic and functional investigations rely basically on the sequence similarity (proof of orthology) to proteins of known function (Tatusov *et al.*, 1997). As a general assumption, have the sequences being compared a common evolutionary ancestral precursor sequence and the best guess at the actual path of evolution is thus the path that requires the fewest evolutionary events. Pairwise sequence alignments thus attempt to find the best match between two sequences with minimal replacements, inserts and deletions being necessary. Substitutions insertions and deletions are not equally likely and they are accordingly weighted. In a biological sense are longer inserts or a deletion more realistic, which justifies the use of affine gap penalties that makes the beginning of a gap more expensive than their extension this slows down the local alignment algorithm. Global alignments (Needleman-Wunsch) extend from one end of each sequence to the other. In contrast to this are local alignment (Smith and Waterman) intended to find optimal matching regions (“subsequences”) within two sequences.

2.1.1. Needleman and Wunsch (Needleman and Wunsch, 1970)

The global Needleman and Wunsch (NW) algorithm is guaranteed to find optimal alignments for the entire sequence length. Two sequences are compared in a matrix along x- and y-axes and if they are identical, a path along a diagonal is drawn. The process also includes the addition of gaps and conservative substitutions when needed.

2.1.2. Smith and Waterman (Smith and Waterman, 1981)

The Smith and Waterman (SW) algorithm is the most sensitive approach to find weakly related sequences and it performs optimal local sequence alignments. Other methods (BLAST, FASTA) are faster but less thorough. Smith and Waterman alignments (SW) and BLAST can be used to make conclusions about homology and local alignment procedures are also almost always used for database searches. It is also useful to find protein domains or short regions of homology within sequences

2.1.3. Hidden Markov models (HMMs) (Sonnhammer *et al.,* 1998)

The HMMER program uses multiple sequence alignments for the creation of Hidden Markov models (HMMs) which largely improves the quality of the alignments. HMMs describe the probability of having a particular amino acid residue in a column of a multiple sequence alignment. The resulting probabilistic models for HMMs give more sensitive alignments than traditional techniques.

2.1.4. Substitution matrices

2.1.4.1. PAM (Henikoff and Henikoff, 2000)

The Point Accepted Mutations substitution matrix (PAM) is derived from global alignments of closely related sequences. PAM accepts a mutation, if it is found in at least one homolog of a known protein (>85% identity). The matrix number (PAM40, PAM100) refers to the evolutionary distance with greater numbers indicating higher distances.

2.1.4.2. BLOSUM (Henikoff and Henikoff, 2000)

The Block Substitution Matrices (BLOSUM) are derived from local, ungapped alignments of distantly related sequences. BLOSUM is based on a larger dataset than PAM. It uses blocks of local alignments and takes conserved protein domains into account. For local similarity searches perform BLOSUM matrices thus generally better than PAM matrices.

2.1.5. BLAST (Altschul *et al.,* 1990)

The gapped Basic Local Alignment Search Tool (BLAST) is intended for database searches and it identifies homologous sequences by gapped local alignments. The hits are extended in either direction and in a refinement of BLAST are two independent hits in close proximity required, which greatly improves the speed for the alignment. The BLAST algorithm (Holmes *et al.*, 1998) uses a word based heuristic to approximate a simplification of the Smith-Waterman algorithm that is known as the maximal segment pair's algorithm (Altschul *et al.,* 1997). This makes BLAST more sensitive for protein sequences than e.g. FASTA. The expectation value E is the number of alignments with scores higher than or equal to score S that are expected to occur by chance in a database search. The expectation value of E decreases exponentially with increasing score S and good alignments result in high scores and low E values.

2.1.6. Tatusov –criteria for finding orthologs (Tatusov *et al.*, 1997).

BLASTP runs can be used to identify putative orthologous and paralogous proteins in related organisms. A protein A in proteome a is a candidate ortholog of protein B in proteome b if protein B is the best match (within proteome b) and protein A in reverse is also the best match when sequence B is searched against proteome a (Mushegian *et al.*, 1998*;* Koonin *et al.*, 1996). Additionally, no homolog in a taxonomic outgroup (*E. coli* K12 in the present analysis) should be closer to the proteins A or B than their corresponding putative orthologs (Gehring *et al.*, 1994*;* Tatusov *et al.*, 1997). This reciprocal-best- BLAST-hits (RBH) approach however may incorrectly predict a paralog as an ortholog when incomplete genome sequences or gene loss are involved (Fulton *et al.*, 2006). Orthology

is not transitive meaning that two proteins in distinct species which are both orthologous to a protein in a third species are not necessarily orthologous to each other (Fitch 1970). The determination can be further complicated by the occurrence of horizontal gene transfer and gene conversion (Sonnhammer and Koonin, 2002). BLAST is fast and detects biologically relevant homologies reliably, but it reports local similarities. The detected probable orthologs should thus share a sequence similarity at least over the majority of their length. Forcing the matched area to be longer than 50% of the longer sequence can be used to avoid domain-level matches and false positives can be detected by adding out-groups (Remm *et al.*, 2001)

2.1.7. PSI- BLAST (Altschul *et al.*, 1997)

The Position Specific Iterated BLAST search tool is also used to identify proteins of similar sequences. PSI- BLAST performs initial sequence similarity searches through a database with the gapped BLAST algorithm and it uses a standard weight matrix (Henikoff and Henikoff, 2000). After this initial iteration, the program constructs a profile (Gribskov, *et al*., 1987; Tatusov *et al*., 1994, Gribskov and Veretnik**,** 1996) from closely related proteins. This procedure iterates until no new closely related proteins can be found, or until the number of iterations has reached a certain threshold. This results in a list of possible homologues, sorted by their E-value. A low expectation value (E-value) points to a high probability that the match is not randomly occur in the database, which then implies that the matches are homologous (Karwath and King, 2002). Highly conserved positions in these protein sequences receive high scores which results in an increased sensitivity of this approach.

2.1.8. PHI- BLAST (Zhang *et al.,* 1998)

The Pattern Hit Initiated BLAST (PHI- BLAST) can identify homologous protein sequences that contain a regular expression pattern and it carries out a position specific iterative protein search.

2.1.9. BESTFIT (Devereux, *et al*. 1984)

BESTFIT is another program that can be used to carry out Smith and Waterman sequence alignments

2.1.10. MUMmer (Delcher *et al*., 1999)

MUMmer and related programs align large stretches of genomic DNA sequences from multiple species. It can thus be used for global genome comparison, contig assembling and investigations of evolutionary constrains.

2.1.11. GAP (Devereux *et al.,* 1984)

The University of Wiscounsin genetics computer group (UWGCG) has developed a package of (GCG Wiscounsin Package) programs for biological sequence analysis. The GAP program of this package carries out Needleman-Wunsch alignments. Its main emphasis is to maximize the number of matches and minimize the number of gaps. A binary comparison score for the two amino acid sequences is expressed in standard deviations (SDs). The similarity between the two sequences is compared with a

large number of random shuffles of these two sequences (thus eliminating discrepancies due to unusual amino acid compositions) to establish significance. The cut-off for establishing homology using the GAP program between two proteins is usually set to 9 standard deviations (SDs) for regions of at least 60 residues, using 500 random shuffles with a gap opening penalty of 8 and a gap extension penalty of 2 (Saier *et al.*, 1994).

2.1.12. IC program (Zhai and Saier, 2002)

Needleman-Wunsch alignments between multiple homologues can be performed using the IC program of the GCG Wiscounsin Package.

2.2. Multiple sequence alignment

For related proteins with a sequence identity of 20–30%, only one-half of the relationships can be detected by pairwise sequence comparisons, and for related proteins with lower identities, the proportion is much smaller (Brenner *et al.*, 1998). A way to overcome this limitation is the use of multiple sequence alignments, which can be three times as effective as pairwise comparisons (Park *et al.*, 1998). Conserved features, such as cysteines as part of important disulfide bridges, may be highly conserved. Other conserved motifs are transmembrane domains and important elements of the secondary structure. Many regulatory regions of genes may also have consensus sequences identifiable by multiple sequence alignments (MSA). For multiple sequence alignments a collection of three or more protein (or nucleic acid) sequences are partially or completely aligned. Homologous residues are aligned in columns across the length of the sequences. The progressive alignment of Feng-Doolittle starts with the two most closely related sequences and then adds the next closely related sequence. This is continued until all sequences are added to the multiple sequence alignment. Global alignment algorithms outperform local algorithms for multiple sequence alignment. A single query can however also be searched against a precomputed database of multiple sequence alignments (MSAs).

2.2.1. MSA methods

The most common programs available for the creation of progressive multiple alignments are PILEUP CLUSTAL X and CLUSTAL W.

2.2.1.1. PILEUP (Devereux *et al.*, 1984)

The PILEUP program as part of the GCG package calculates multiple sequence alignments based on a GLOBAL dynamic programming alignments of all sequence pairs. The closest pair is selected to begin the alignment and then each successive sequence is added according to its distance.

2.2.1.2. BIOEDIT (North Carolina state university)

BIOEDIT is a graphical alignment editor for single and multiple sequence alignments (CLUSTALW). It can also be used for comparative analysis, restriction mapping and it contains many other useful tools for sequence analysis.

2.2.1.3. CLUSTALW (Thompson *et al.*, 1994)

CLUSTALW uses the neighbour joining method to create a progressive alignment. The program uses position dependent gap penalties and multiple scoring tables to match sequences. As a drawback, can errors that are introduced early in the alignment procedure not be fixed and the program needs many (heuristic) parameters. A phylogenetic tree can be build based on final alignment with bootstrapping and omitting of gapped regions.

2.2.1.4. CLUSTALX (Thompson *et al.*, 1997)

The CLUSTALX program is also used for multiple alignments of homologous sequences

2.2.2. Phylogenetic trees

Phylogenetic trees are built from multiple sequence alignments. Tree construction seeks to understand the evolutionary relationship between certain taxa (groups of organisms). The branching pattern and the (additive) lengths of the peripheral and internal branches connecting two terminal nodes indicate the probable path of evolution (Ludwig and Klenk, 2000). Important methods used to create phylogenetic trees are distance matrix, parsimony and maximum likelihood.

Unrooted trees do not show where the ancestral node lies. The most common procedure to find the root of a tree is to use an outgroup organism, i.e. a taxon that is guaranteed to be more distant from all of the taxa of interest than any of them are from each other. The 16S rDNA sequences exhibit a high degree of sequence conservation and this allows using them for the investigation of the evolutionary relatedness between organisms. Global alignments are usually used for this purpose. For a comparison of the phylogenetic relationships between distinct protein families are local alignments more suitable.

2.2.3. TREE (Feng and Doolittle, 1990)

The TREE program can be used for the construction of a phylogenetic tree and the TREEVIEW drawing program can be used to visualize the results.

2.3. Topology prediction

2.3.1. TMS SPLIT program (Zhou *et al.*, 2003)

The TMS SPLIT program can be used to generate fragmented protein sequences which facilitates investigations of internal gene duplication events.

2.3.2. TMHMM (Sonnhammer *et al.*, 1998b), HMMTOP (Tusnady and Simon, 1998) and WHAT (Zhai and Saier, 2001)

TMHMM, HMMTOP and WHAT are programs for the estimation of the probable topologies of membrane proteins. The methods are based on a hidden Markov Model. The AVEHAS program can be used to visualize the corresponding results.

2.3.3. AVEHAS (Zhai and Saier, 2001)

AVEHAS is a program for the analysis of the average hydropathy, average amphipathicity and average similarity of protein sequences.

2.4. Additional programs

2.4.1. PSORT (Nakai and Kanehisa, 1991)

PSORT is used to predict the probable protein localisation based on protein sorting signal sequences.

2.4.2. FSED (Fichant and Quentin, 1995)

FSED is used to detect frameshift errors in genomic sequences by investigating the k-tuple distribution. The method is based on the assumption, that coding sequences exhibit a statistical bias in the distribution of 3-tuples and 6-tuples in the three frames of an open reading frame. The bias is used to create discriminate functions on a learning set of sequences (correspondence analysis)

2.4.3. SIGNALP (Bendtsen *et al*., 2004)

The SIGNALP server predicts putative signal sequences and signal peptide cleavage sites in amino acid sequences. This prediction is based on a combination of several neural networks and hidden Markov models.

2.4.4. STADEN (Bonfield *et al.*, 1995)

The STADEN package contains biological programs for the investigation of genomic sequences. A crucial point lies in computational gap closure which allows to finish a genome assembling.

2.4.5. GAP4 (Bonfield *et al*., 1995)

As part of the STADEN package is GAP4 a primary tool for sequence navigation and editing.

2.4.6. BIOTECHNIX3D (http://www.biotechnix3d.com)

BIOTECHNIX3D is a simple interface for DNA- and protein sequence analysis. It integrates many tools and useful links to molecular databases and enables an easy storage of annotation efforts.

2.4.7. APPLESCRIPT(http://www.apple.com/applescript/)

APPLESCRIPT is a dynamic, object-oriented scripting language implemented into OSX and OS9. APPLESCRIPT creates sets of written instructions to automate repetitive tasks, customize applications and it can even control complex workflows. A script can send instructions to one application, get the resulting data, and then pass the data on to one or more additional applications.

2.5. Protein structure analysis

2.5.1. Homology modelling

Homology modelling approximates the protein structure for a protein of unresolved three dimensional structure based on the resolved protein structure models of a homologous protein. This is based on pairwise sequence comparisons, fold recognition and the predictions based on statistical rules derived from protein structure models (e.g. secondary protein structure, transmembrane helices and coiled-coils) (Teichmann *et al.*, 1999). To allow this calculation in a recenable time very approximate models are used and the sequences must have sufficient similarity (80%). Theoretical protein modelling can be used to build up at least a ‘low resolution’ model for the protein of interrest. It can also guide the design of site directed mutagenesis studies for structural investigations.

Superpositioning procedures for protein structural models will align the C-alpha atoms and then optimize the fit by minimizing their relative mean square deviation (RMSD). A theoretical model is wrong, if part of its structure is missplaced relatively to the rest of the model. This can be e.g. the result of a erroneous sequence alignment or deviation from the ideal stereochemical value for bond lenghts or angles.

2.5.2. Hydrogen bonds

The probability for the presence of hydrogen bonds is commonly assessed based on both the distance between a hydrogen-bond donor/acceptor pair. The donor-hydrogen bonds usually range between 2.6 and 3.3 Å and the donor-H-acceptor angle is about 180°).

2.5.3. Protein structure analysis programs

2.5.3.1. CHIME and RASMOL (http://www.umass.edu/microbio/chime)

RASMOL is an interface for the exploration of molecule structure models. CHIME is more effective than RASMOL for the representation of chemical structure information.

2.5.3.2. PROTEIN EXPLORER (http://www.umass.edu/microbio/chime/explorer/pe_tut.htm)

The PROTEIN EXPLORER (PE) is a RasMol-like interface implemented in Chime. It exhibits a much greater assistance and automation and is more convenient and easier to use than the RASMOL program. However, all RASMOL commands can also be used in the PROTEIN EXPLORER program. The noncovalent bond finder is integrated into this system. It allows selecting any target moiety within a protein structure and finds the closest atoms in its proteomic neighbourhood.
PE can also report the distances between atoms, bond angles or salt bridges. The calculation of contact surfaces is another powerful feature of this program that allows visualizing the contacts to a single atom, to ligands or between residues of proteins.

2.5.3.3. SPDBV – 'Deep View' (Guex and Peitsch, 1997)

The SWISSPDB-VIEWER (http://www.expacy.ch/spdbv) has been renamed 'Deep View' (SPDBV) and is a multiplatform interface for the viewing and analysing of protein and nucleic acid structures. SPDBV also provides some advanced features including an interface for theoretical structure modelling and for the visualization of electron density maps or electrostatic surfaces. The measurement of bonds, angles and distances between atoms is also implemented. The PDB-files of protein structure models can be retrieved automatically by their PDB-ID. Deep View can attempt to superimpose molecules. This is facilitated by an implemented interface to the SWISS-MODEL server. Deep View can calculate the electrostatic potential induced by charged groups in the protein structure, and visualizes it a colour gradient at the molecular surface. With SPDBV the amino acid sequence of the structure can also be searched for the occurrence of specific PROSITE pattern e.g. glycosylation sites etc.

2.5.3.4. MSMS (Sanner *et al.*, 1996)

MSMS is a molecular surface modelling server

2.5.3.5. YASARA (Krieger *et al.*, 2002)

YASARA is a new commercially available program for comprehensive protein structure visualization.

2.5.3.6. GRASP (Nicholls *et al.*, 1991)

GRASP computes molecular surfaces

2.5.3.7. POV-Ray

POV-Ray is a multi-platform, free Ray-tracer that generates very high quality images. It is also implemented in SPDBV.

2.5.3.8. NONCOVALENT BOND FINDER (http://www.umass.edu/microbio/chime)

The NONCOVALENT BOND FINDER allows finding and visualizing probable noncovalent bonds. The initial cut-off radius of 2.5Å can be freely adjusted.

2.5.3.9. MSA3D (http://www.umass.edu/microbio/chime/explorer/pe_tut.htm)

MSA3D is a program implemented in the PROTEIN EXPLORER that allows colouring a molecule according to the conservation/mutation revealed from multiple sequence alignments

2.5.3.10. CASTP (Edelsbrunner 98).

Pockets are empty concavities on a protein surface into which solvent can gain access. A cavity is an interior empty space that is not accessible to the outside solvent. The CASTP program is based on recent theoretical and algorithmic results of computational geometry and can be used to find cavities and pockets.

2.5.3.11. CHARM, AMBER or GROMOS.

Idealization of the bond geometries and the removal of unfavourable non-bonded contacts can be performed by energy minimalization with force field.

2.6. Databases

Biological databases can gather nucleotide or protein sequences or describe common structural, evolutionary and functional aspects of proteins or protein families.

2.6.1. The NON-REDUNDANT NUCLEOTIDE DATABASE (NR) (Pruitt *et al.,* 2005)

The non-redundant nucleotide database is maintained by NCBI as a composite of GENBANK and EMBL with identical sequences being merged. It is usually used for BLAST searches

2.6.2. SWISS-MODEL server

The SWISS-MODEL server provides several tools for comparative protein modelling

2.6.3. GENBANK /EMBL / DDBJ (http://www.ncbi.nlm.nih.gov)

GENBANK, EMBL and the DNA Databank of Japan (DDBJ) nucleotide databases are principally just different names for the same database. Slight differences only occur by time lags for the propagation of new data into these databases.

2.6.4. JGI (http://genome.jgi-psf.org/mic_home.html)

The Joined Genome Institute combines several genomic databases of the US Department of Energy (DOE) and includes data from high throughput genomic sequencing and annotation. The main focus lies at bacteria of extreme habitats in regard to pH, heavy metal pollution or radioactivity.

2.6.5. SWISSPROT (Bairoch *et al.*, 1997)

The Swiss Protein Database (SWISS-PROT) is a database for protein sequences derived from DNA translations. It is maintained by the University of Geneva and the EMBL Data Library. Cross references to EMBL,PROSITE and PDB are included into this system.

2.6.6. TrEMBL (Bairoch *et al.*, 1997)

The TrEMBL database contains all sequences that are not yet integrated into the SWISSPROT protein database but that are available from translations of the TrEMBL Nucleotide Sequence database. This database is maintained by the European molecular biology laboratory (EMBL).

2.6.7. FUSIONDB (Suhre and Claverie, 2004)

FusionDB (http://igs-server.cnrs-mrs.fr/Fusion- DB/) is a database of probable bacterial and archaeal gene fusion events. A gene fusion event might have occurred if two not homologous reference genes within a genome both match a third protein (target) in the genome as their highest scoring BLAST hit. Also when split in the middle between the two BLAST hits, the two halves of the target ORF must match back to the original two reference genes as their best BLAST hit. Each putative fusion event is subjected to a scoring scheme based on different evaluations of its pair-wise and multiple alignments.

2.6.8. PDB (Bernstein, *et al.* 1977)

The Protein Databank of the Brookhaven National Library contains all available solved protein structures. Currently about 4000 protein chains of known three dimensional structures are available at PDB and it also includes the structure models of many transport proteins.

2.6.9. PIR (Sidman *et al.*, 1988)

The Protein Identification Resource is a combination of no redundant (PIR1) and redundant (PIR1/PIR2/PIR3) protein sequence databases.

2.6.10. PFAM (Sonnhammer *et al.*, 1998)

The protein families database (PFAM) is a large collection of multiple sequence alignments (MSAs) and hidden markov models (HMMs) of protein domains and families. It allows viewing domain architecture and contains links to sites for protein structure investigations.

2.6.11. PROSITE (Falquet *et al*., 2002)

PROSITE is a database of protein motifs at the University of Geneva.

2.6.12. BLOCKS (Henikoff *et al*. 1999)

BLOCKS is a nonredundant database of protein alignment blocks. A block is a gap free multiple alignment of sequences based on PROSITE motifs.

2.6.13. RDPII (Maidak *et al.*, 2001)

The ribosomal database project (RDPII) contains sequences from high throughput rRNA sequence analysis. It includes aligned and annotated rRNA sequences for over 100.000 bacterial small subunit rRNA genes. The Tree Builder (RDPII) can be used to create a phylogenetic tree of 16SrRNA (with bootstrap) with the Neighbour Joining method and with an additional organism as outgroup root

2.6.14. The PEDANT genome database (Riley *et al.*, 2005)

A comprehensive automatic annotation of genomes is available in the PEDANT database.

2.6.15. IUBMB (http://www.chem.qmul.ac.uk/iubmb/)

The International Union of Biochemistry and Molecular Biology (IUBMB) provides a functionally oriented classification of proteins.

2.6.16. transport protein classification – TC system (Zhou *et al*., 2003)

The transporter protein classification system (TC) has been adopted by IUBMB as the international standard for categorizing proteins concerned with the transport of substrates across cellular membranes. The breakdown of transporters in this classification is based on mode of transport, bioenergetics, topology, molecular phylogeny and substrate specifities (outlined in table 2) because several of those features have proven to be conserved among certain transport protein families. The assignment according to the TC-system also reveals the substrate range of the identified transport proteins.

Tab. 2: Organisation of the transporter protein classification system (TC) (Saier 2000).
The first digit of the TC classifier reflects the mode of transport and energy coupling mechanism. The second and third term indicate the phylogenetic family and phylogenetic cluster within the family. The last digit reflects the substrate specifities and polarity.

transporters are subdivided and classified with a five-character designation > D1.L1.D2.D3.D4. D1	
•D1	transporter class (i.e. channel, carrier, primary active transporter, group translocator or transmembrane electron flow carrier).
•L1	subclass, that, e.g. refers to the energy source used to drive transport.
•D2	transporter family (sometimes actually a superfamily).
•D3	subfamily (or the family of a superfamily)
•D4	corresponds to the specific transport system (defined range)

2.6.17. TCDB (Saier 2000)

The transporter protein database TCDB (at http://www.tcdb.org/) describes and classifies putative transport protein families according to the TC system. It also contains useful additional information about the mode of transport, topology, phylogenetic constrains and substrate ranges that are conserved for the respective transport protein families (Saier 2000).

2.6.18. TRANSPORTDB (http://www.membranetransport.org/)

Is another transport protein database that reveals the specific transport protein repertoires of certain organisms.

2.6.19. ABCDB (Quentin and Fichant, 2000)

ABC DB is a database that contains ATP driven transport protein families.

2.6.20. Databases for the investigation of metal binding sites

2.6.20.1. PROMISE (Degtyarenko *et al.*, 1999)

The PROMISE database of the University of Leeds gathers structural, functional and bibliographic information for prosthetic centres of proteins. The properties of each corresponding domain is comprehensively described and summarized. Tools and links to a wide variety of websites are also integrated.

2.6.20.2. MDB (http:// metallo.scripps.edu/).

The METALLOPROTEIN DATABASE (MDB) is maintained by the Scripps Research Institute to facilitate investigations of metalloproteins. It offers comprehensive structural information of known putative metal binding sites which have been derived from structures of metalloproteins at the

Brookhaven Protein Data Bank. These PDB entries have been surveyed to investigate the trends in geometrical and stereo chemical parameters of the metal binding sites. MDB allows to search for probable metal binding sites by geometric concerns (e.g. residues with a distance shorter than 2.8 Å from the metal) when the three dimensional structure model is available for the protein of interest. This identification of potential metal binding sites is carried out by the integrated Metal-binding Site Evaluator (MSE). The amino acid sequences of putative metal binding motifs as well as important geometric properties are gathered and the entries are sorted according to the type of bound metal ion. MDB also gives access to specific parameters of these sites including the metal bond distances and side chain torsion angles. It's also possible to evaluate the distribution of ligand patterns for a particular metal ion. A PDBviewer is also integrated to give visual insights into specific structural properties of the metal binding sites of metalloproteins (see figure 4).

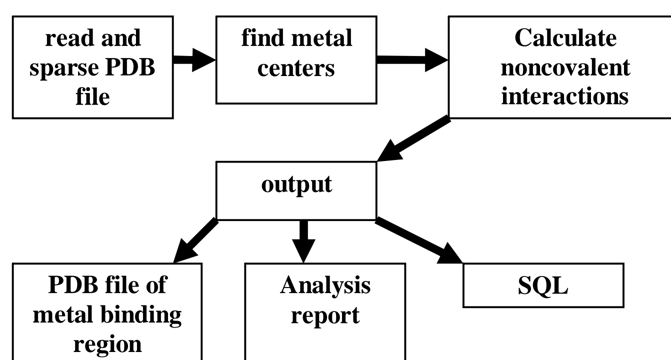


Fig. 4: the MetalloproteinDatabase MDB (http://www.metallo.scripps.edu/ simplified)
The picture shows the general workflow for the identification of putative metal binding sites with the MetalloproteinDatabase MDB

2.6.20.3. METAL-BINDING SITE EVALUATOR (MSE)

The Metal-binding Site Evaluator allows searching for metal-binding sites and it identifies first and second shell ligands. The search is based on specific properties of the protein geometry.

2.6.21 IMG (Markowitz *et al*., 2006)

The integrated microbial genomes (IMG) system is a new data management and analysis platform for microbial genomes provided by the Joint Genome Institute (JGI). IMG contains both Draft and complete genomes and facilitates individual or comparative functional investigations. IMG allows users to focus their analysis on subsets of genes and genomes of interest and to save the results of their analysis. IMG is available at http:// img.jgi.doe.gov.

2.6.22 Gene Ontology (GO) (Ashburner *et al*., 2000)

The Gene Ontology (GO) developed at the GO Consortium provides a suitable framework for functional annotation and allows categorization of genes in functional classes. This can be very useful

to understand the physiological meaning and to assess functional differences between subgroups of sequences.

2.6.23 The Kyoto encyclopaedia of genes and genomes KEGG (Kanehisa and Goto; 2000)

Pathways, reactions and compounds are summarized at KEGG. A gene can be examined in the context of its associated pathways.

2.6.24 The comprehensive microbial resource (CMR) (http://www.tigr.org/tigrscripts/CMR2)

The comprehensive microbial resource at TIGR contains more than a hundred sequenced prokaryotic genomes.

2.7. Gap closure (Staden *et al.,* 2000)

Computational gap closure can follow different strategies. One approach is to simply compute all possible overlaps between Contigs and to subsequently combine those strings, which exhibit the highest sequence similarity (figure 5). The final goal is to find the so called shortest common superstring of a set of sequences. Programs as MUMmer, CAP3, PHRAP, FINISHER, PROMER and the TIGRASSEMBLER (Sutton *et al*., 1995; Gordon *et al*., 1998; Huang and Madan, 1999) are available for this purpose. On the other hand it was shown that these kinds of scaffolding programs misassemble up to 19 % of the investigated contigs.

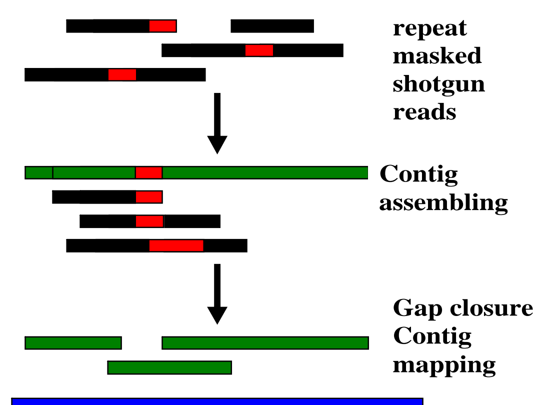


Fig. 5: General approach for computational gap closure (Wang *et al.*, 2002 simplified).
The process of genomic sequencing involves physical breaking of the target DNA and subsequent sequencing of the resulting random fragments which are then assembled according to overlapping reads. The contigs of unfinished gap containing Draft Versions are subsequently joined into larger scaffolds (gap closure) which reveal the probable order and orientation of the pieces.

2.7.1. PROJECTOR2 (Sacha *et al.,* 2003)

The PROJECTOR2 program uses a different strategy than other scaffolder programs. It has been developed to position contigs of an unfinished genome onto the genomic sequence of related organisms (see figure 6). The contigs ends of a genomic sequence assembly often contain unreliable

DNA sequences caused by Phage DNA IS elements and gene duplications. Unlike related scaffolder programs does the PROJECTOR program attempt to suppress those sequences prior of the scaffolding process. This is carried out by evaluating the G+C content and by masking double sequences. PROJECTOR also facilitates experimental gap closure by providing appropriate primer sequences for PCR experiments.

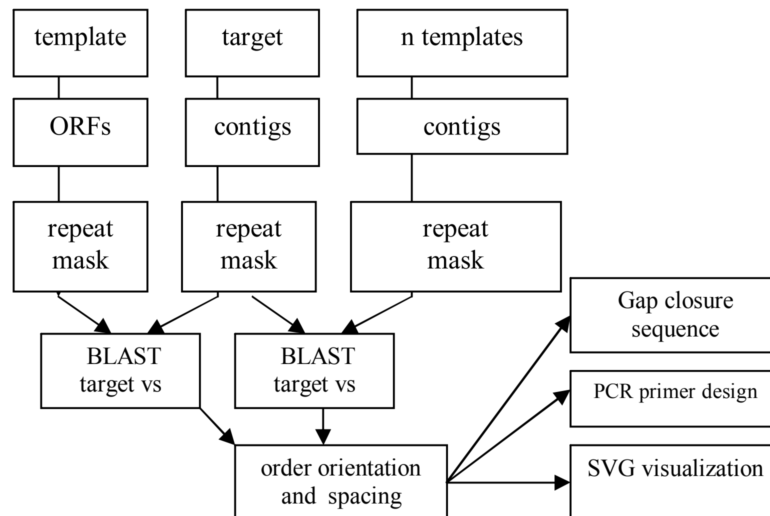


Fig. 6: PROJEKTOR2-approach for computational gap closure (Sacha *et al.,* 2005 simplified). The picture shows the general workflow for the s for contig mapping and Primer design with the PROJEKTOR2 programs, which allows experimental gap closure

3. RESULTS

3.1. Overview of publications

1:

von Rozycki, T., Yen, M.R., Lende, E.E., Saier, M. H. Jr. 2004

The YedZ family: possible heme binding proteins that can be fused to transporters and electron carriers.

J Mol Microbiol Biotechnol 8:129-40.

2:

von Rozycki, T., Schultzel, M.A., Saier, M. H. Jr. 2004

Sequence analyses of cyanobacterial bicarbonate transporters and their homologues.

J Mol Microbiol Biotechnol 7:102-8.

3:

von Rozycki, T., Nies, D.H., Saier, M. H. Jr. 2005

Genomic analyses of transport proteins in *Ralstonia metallidurans*.

Comp Func Genom. 6:17-56.

4:

von Rozycki, T., Nies, D.H. 2008

Cupriavidus metallidurans: evolution of a metal-resistant bacterium

(accepted *Antonie Van Leeuwenhoek*)

3.1. The YedZ family: possible heme binding proteins that can be fused to transporters and electron carriers.

(published in 2004 in *J Mol Microbiol Biotechnol*. 8(3):129-40.)

Summary

The YedZ (TC 9.B.43) transporter family belongs to the group of incompletely characterized transporters (TC-class 9). The *yed*YZ operon of *E. coli* encodes a two subunit putative oxidoreductase and YedZ anchors probably the catalytic subunit (YedY) to the inner membrane (Drew *et al.,* 2002; Brokx *et al.*, 2005). Six integral transmembrane spanning segments (TMS) can be found in most homologs of YedZ. They are represented as regions of high hydropathy in figures 8 and 9. Sequence comparisons with the GAP and IC programs (Devereux *et al.,* 1984) have been carried out to investigate the conservation of these transmembrane segments (figure 7). The GAP program was used for individual sequence alignments and a cut off of 9 standard deviations (SDs) was applied to establish homology (Saier *et al.,* 1994; Saier *et al.,* 1996). The IC-program was used to find homologies among multiple sequences. An extensive sequence similarity could be shown for several 2 TMSs containing segments of YedZ (10.7 SD). For example, align the TMSs 1 and 2 of a homolog from *Salmonella enterica* well with TMSs 3 and 4 of a homolog of YedZ from *Rhodospirillum rubrum* (as shown in figure 7). A GAP comparison score of 8.1 SD (31% sequence identity; 42% similarity) was observed. The TMSs 3 and 4 of a YedZ homolog from *Erwinia carotovora* aligns well with the TMSs 5 and 6 of a YedZ homolog from *Silicibacter* sp. These segments gave a GAP comparison score of 10.7 SD (37 % sequence identity; 48 % similarity). The similarities of these segments indicate that the six TMS containing YedZ homologs have probably been formed by the triplication of a primordial 2 TMS encoding element. Multiple sequence alignments with the CLUSTALX program (Thompson *et al.*, 1997) have also revealed that certain histidine residues are conserved in these segments. The TMS 5 in YedZ of *E. coli* contains two nearly, but not fully conserved histidyl residues and these motifs can be found in all but the most distant homologues of this protein. In several homologs in magnetotactic bacteria are YedZ-like domains fused to transporters of the MFS (2.A.1) superfamily. These proteins have a high sequence similarity to magnetosome transporters. Thus exhibit the 18 TMS containing YedZ orthologes in magnetotactic bacteria an extensive sequence similarity to the 12 TMS (428 aa) protein MamH (TC 2.A.1.43.1) of *Magnetospirillum gryphiswaldense* (39% identity; 59% similarity) (Schübbe *et al.,* 2003). A high sequence similarity to an MFS homologue from *Chlorobium tepidum* (gi 21674295, 25% identity; 44% similarity) has also been found. In some homologs of YedZ in cyanobacteria are YedZ-like domains fused to proteins for electron transfer e.g. the transmembrane electron carrier DsbD (TC #5.A.1) (Kimball *et al.,* 2003). Thus have the N-terminal regions of four cyanobacterial homologs of YedZ a high sequence similarity to an 80-residue segment of the DsbD protein of *E. coli* (TC 5.A.1.1.1). The homolog Npu1 in *Nostoc punctiforme* PCC exhibits 29% identity and 48% sequence similarity to the DsbD protein of *E. coli* and exhibits 32% identity and 42% similarity to a 180-residue segment in the DsbD and *Pasteurella multocida* (Q9CP40). The charged residues in transmembrane segments of the YedZ like domains of these proteins might be important for the binding of a prosthetic group and in the MFS homologs they may have a role in transmembrane proton conduction.

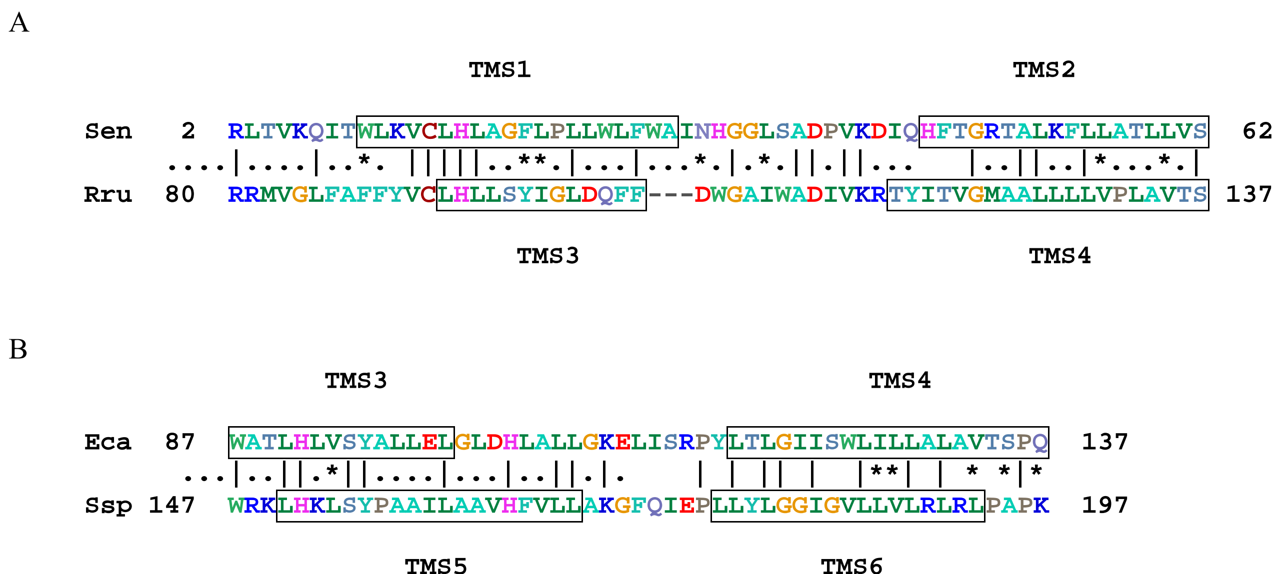


Fig. 7: Alignments of YedZ homologs with the IC and GAP programs.

A An alignments of YedZ homolog Sen (region with TMSs 1–2 shown) from *Salmonella enterica* (gi|29143626) with the YedZ homolog Rru (region around TMSs 3–4 shown) from *Rhodospirillum rubrum* (gi|48763376) revealed a sequence identity of 31% and a similarity of 42% (GAP comparison score of 8.1 SD) between these segments.

B An alignment of the YedZ homolog Eca (region with TMSs 3–4 shown) from *Erwinia carotovora* subsp. *Atroseptica* with the YedZ homolog Ssp (TMSs 5–6 shown) from *Silicibacter* sp. TM1040 was carried out. Only the first histidine in TMS 3 and the second histidine in TMS 5 are fully conserved in all of the eukaryotic homologues. The alignments were carried out with the GAP program (residue colours were manually changed) and gave a sequence identity of 37% and a similarity of 48% (GAP comparison score of 10.7 SD). The boxes indicate the positions of the putative TMSs; the residue numbers for the protein alignment are given at the beginning and end the sequence lines

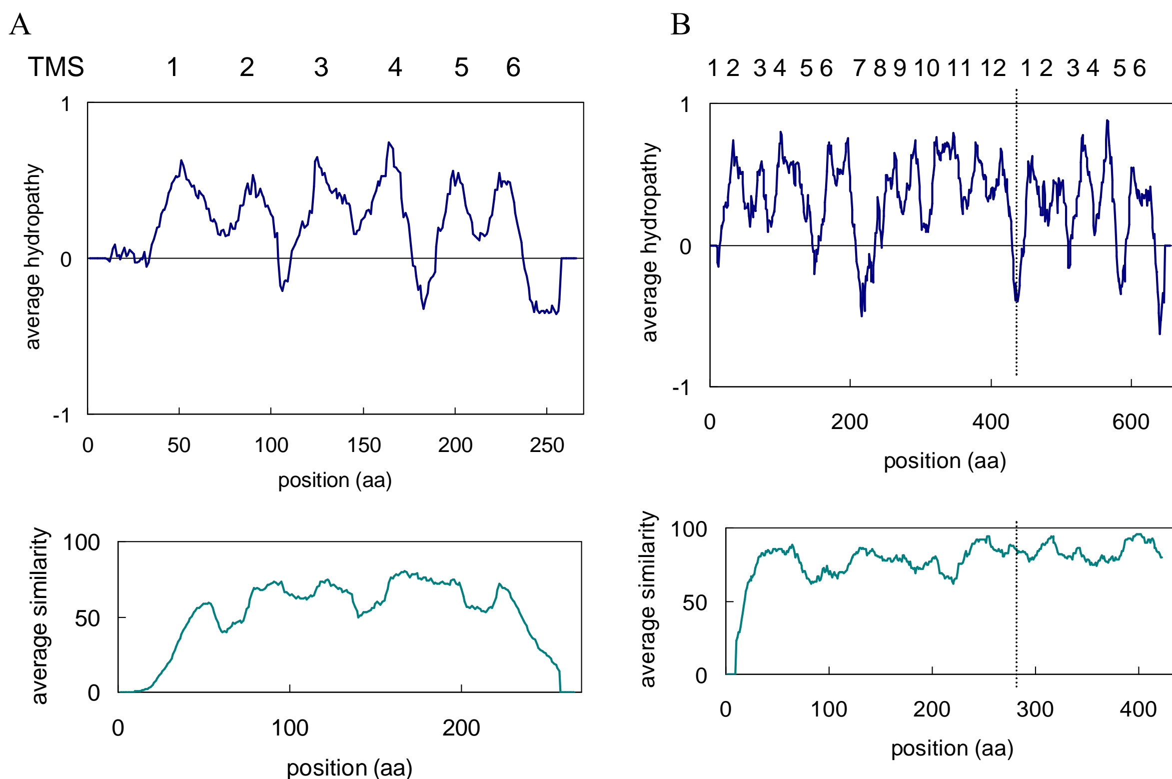


Fig. 8: Average hydropathy and average similarity plots of YedZ domains.

A The figure shows the average hydropathy and average similarity plot of prokaryotic 6 TMS containing homologs of YedZ

B The figure shows the average hydropathy and average similarity plot of proteins in magnetotactic bacteria. The vertical dotted line indicates the boundary between the MFS permease domain and the fused YedZ domain.

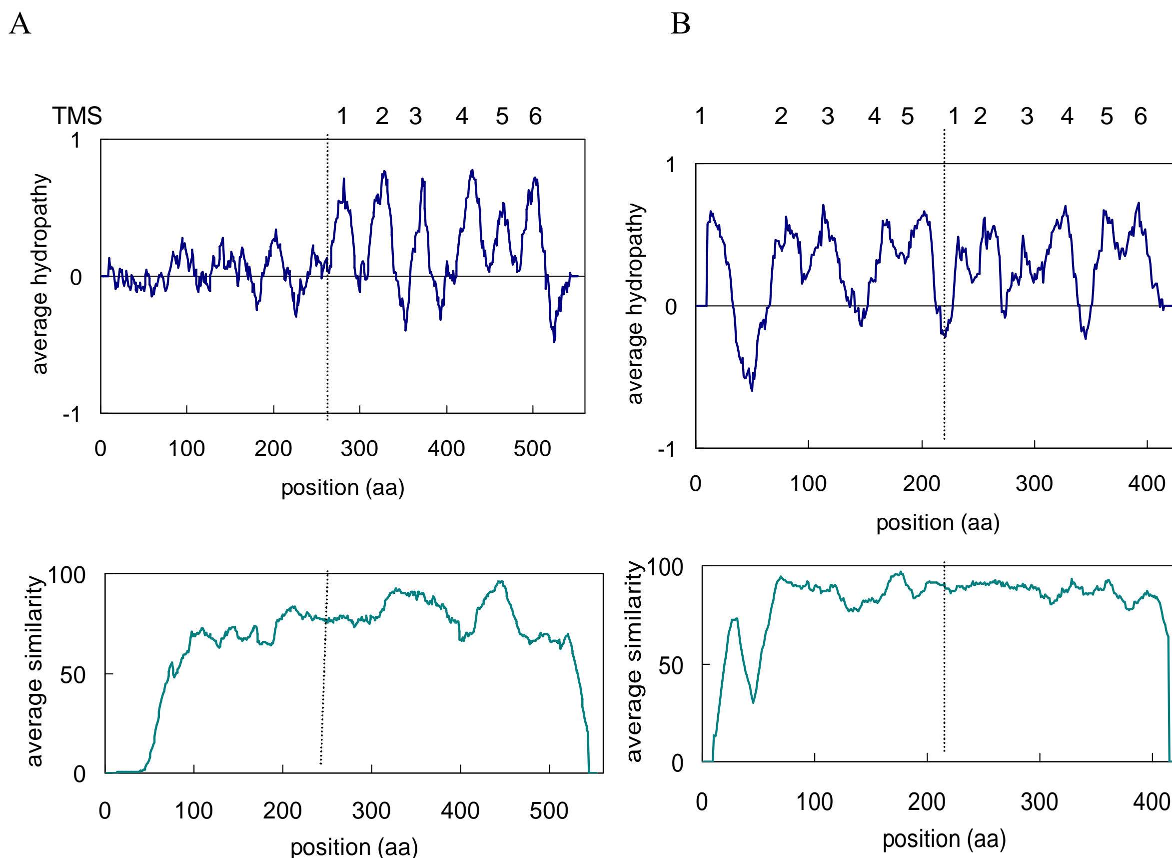


Fig 9: Average hydropathy and average similarity plots of YedZ domains.

A Average hydropathy and average similarity plot of the eukaryotic YedZ homologues. The AVEHAS program was used to generate a plot of the average hydropathy and the average similarity plot is based on a multiple sequence alignments that was generated with the CLUSTALX program.

B Average hydropathy and average similarity plot of cyanobacterial YedZ fusion proteins. The vertical dotted line indicates the boundary between the DsbD-like domain and the C-terminal YedZ domain. These plots were generated with the AVEHAS program and the multiple alignments were generated with the CLUSTALX program. The N-terminal transporter domains of homologs in magnetotactic bacteria and the oxidoreductase domains in homologues in cyanobacteria as well as nonhomologous regions in several eukaryotic (Gga3 and Gga4) homologs were removed before the plots were generated.

Publication 2

3.2. Sequence analyses of cyanobacterial bicarbonate transporters and their homologues.

(published in 2004 in *J Mol Microbiol Biotechnol*. 7(3):102-8.)

Summary

The SbtA (slr1512) protein of *Synechocystis* PCC 6803 is a sodium dependent bicarbonate uptake transporter of the SBT transport protein family (TC 2.A.83). The GAP and IC programs (Devereux *et al.,* 1984) have been used to investigate the conservation of transmembrane segments in homologs of this transporter family. A multiple sequence alignment of SbtA of *Synechococcus* sp and its homologues has been carried out with the CLUSTALX program (Thompson *et al.*, 1997). The phylogenies of the proteins do not follow the phylogenies of the source organisms and the tree topology suggests that SbtA family members might have multiple transport functions. The first half of a homolog of the SbtA transporter family (2.A.83) (Mtu) from *Mycobacterium tuberculosis* could be aligned (figure10) with the second half of another SbtA homolog (Pma3) from *Prochlorococcus marinus*. The two sequences gave a GAP comparison score of 9.0 SD (standard deviations) and exhibit 30% sequence identity. The positions of three putative TMSs in both sequences have shown to be conserved (figure 12). Thus are the first halves of SbtA family members homologous to their second halves and the ten transmembrane segments (TMSs) of SbtA have thus probably been formed by an intragenic duplication event. The two halves have opposite orientation in the membrane (Zhou *et al.,* 2003). An alignment of the N-terminal region of SbtA with a portion of the N-terminal sequence of an ABC-type L-arabinose transport protein (AraH- TC 3.A.1.2.2) gave a GAP comparison score of 9.8 SD (35% identity). The conserved positions correspond to TMSs 1–4 in SbtA which align well with TMSs 2–5 in AraH (figure11). These results reveal a novel close relationship between regions of ATP hydrolysis driven primary transporters (AraH- TC 3.A.1.2.2) on one side and electrochemically driven transporters (SbtA- TC 2.A.83.1.1) on the other side.

A

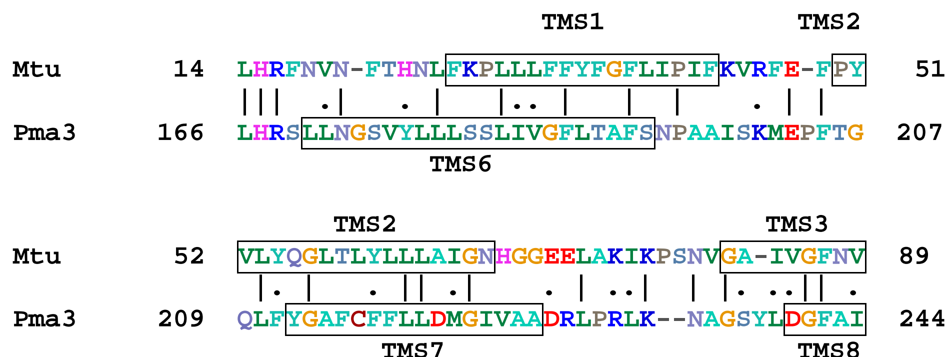


Fig. 10: Sequence Analyses of SbtA homologs
An alignment of the first half segment (a 75 aa segment) of the SbtA homologue Mtu in *Mycobacterium tuberculosis* with the second half (72 aa segment) of the SbtA homologue Pma3 from *Prochlorococcus marinus* was carried out. The two sequences exhibited 30% identity and gave a comparison score of 9.0 SD. Residue numbers are provided at the beginning and end of each line. The positions of putative TMSs in both proteins are shown as boxes.

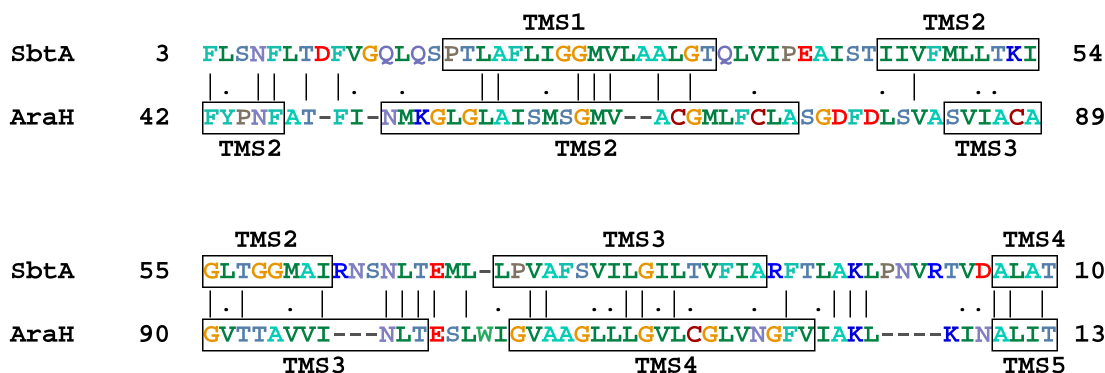


Fig. 11: Sequence Analyses of SbtA homologs
An alignment of the N-terminal region of SbtA with an N-terminal sequence segment of the ABC-type L-arabinose transport protein, AraH was carried out. The alignment gave 35% identity and a comparison score of 9.8 SD. Residue numbers are provided at the beginning and end of each line. The positions of putative TMSs in both proteins are shown as boxes.

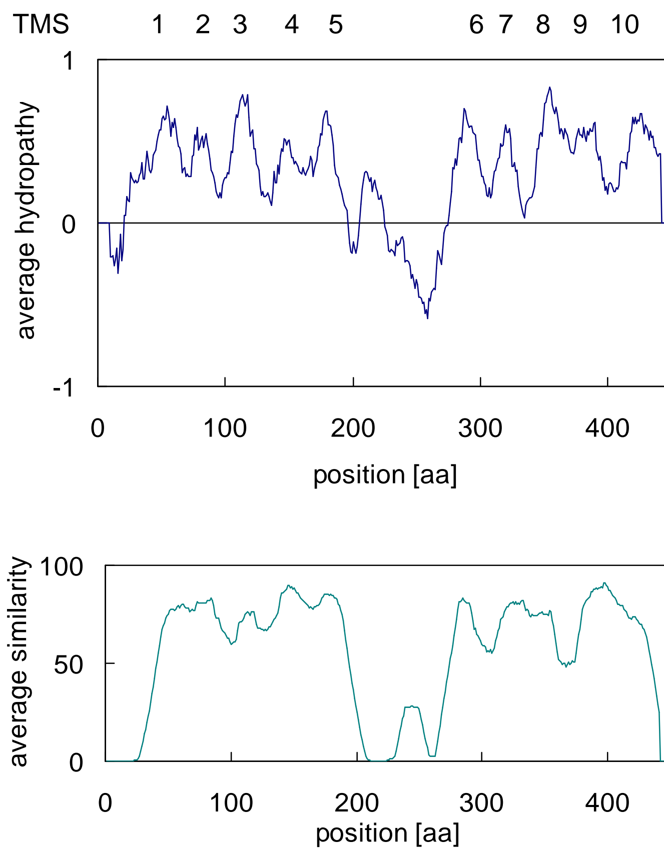


Fig. 12: Average hydropathy and average similarity plots for several SBT family members.
The average hydropathy plot was generated with the AveHas program. The central vertical line indicates where the protein sequences were cut in order to compare the two halves of the proteins. The multiple sequence alignments were generated with the CLUSTALX program.

3.3. Genomic analyses of transport proteins in *Ralstonia metallidurans* CH34.

(published in 2005 in *Comp Func Genom*. 6: 17-56.)

3.3.1. Summary

The predicted protein sequences (CDS translations) of *C. metallidurans* CH34 have been blasted against the protein sequences of the representatives of transport protein families at the transporter protein database TCDB (at http://www.tcdb.org/) (Saier, 2000). As cut off to establish homology to these sequences has an expectation value by chance of $1e^{-10}$ been applied (Saier *et al.*, 2006). A comparison of the membrane topologies, the protein sizes and functional assignments (SWISSPROT - Bairoch *et al.*, 1997) has been carried out to verify the assignment to transport protein families (Das *et al.*, 2007). The membrane topologies were determined by charged bias analysis with the TMHMM (Sonnhammer *et al.*, 1998b) program. The similarities to characterized protein domains have been identified using the protein pattern database PFAM (Sonnhammer *et al.,* 1998). Further protein functional analysis was based on the similarities to proteins at NCBI, SWISSPROT and TrEMBL (Bairoch *et al.,* 1997) databases. Typical substrate specifities for the transport protein families have been assigned according to specifications at TCDB and literature (e.g. Saier, 2000; Saier, 2000c; Saier, 2000d). Figure 13 gives a glance of the overall potential substrate transport capabilities in *C. metallidurans* CH34.

3.3.2. Results of succeeding investigations

3.3.2.1. The comparison of the transporter repertoires in proteobacteria

The transport protein repertoire of *C. metallidurans* CH34 has been compared to the predicted transport protein repertoires of several other proteobacteria which are available at the TRANSPORTDB (Ren *et al.,* 2004) database (http://www.membranetransport.org/). Orthologs of transporters in *C. eutrophus* JMP134, *B. cepacia* AMMD and *B. xenovorans* LB400, *R. solanacearum* GMI1000 and *C. eutrophus* H16 have also been investigated. A draft version of *R. solanacearum* UW551 (chromosomal sequences) was also included into this comparison. The genomic sequences of these strains were obtained from JGI (DOE Joined Genome Institute, http://www.jgi.doe.gov/) and NCBI (National Centre for Biotechnology Information, http://www.ncbi.nlm.nih.gov/). The species name "*Cupriavidus eutrophus*" (Tindall, 2008) was used for the strains H16 and JMP134, but the genomic sequence of strain H16 was published under the name "*Ralstonia eutropha*" (Pohlmann *et al.*, 2006) and these strains are also often termed "*Cupriavidus necator*".

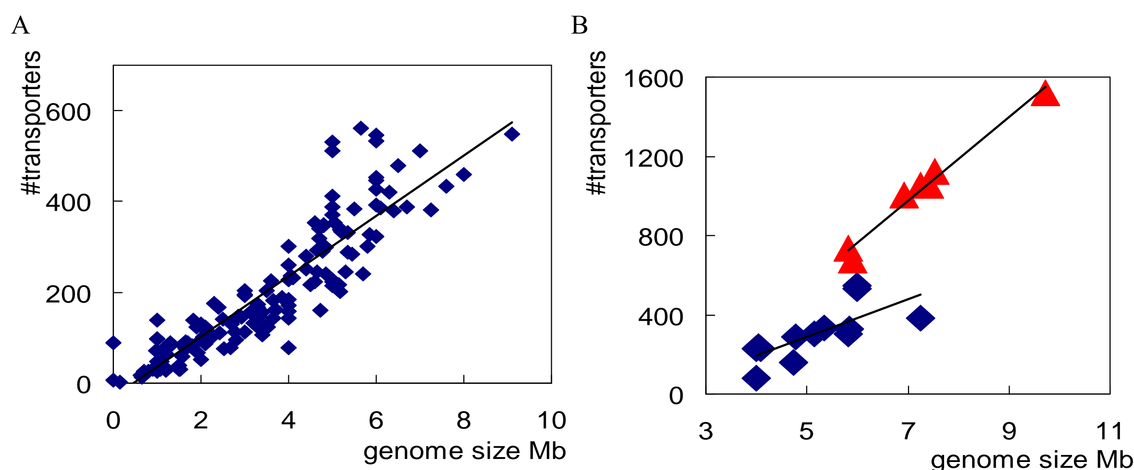


Fig. 13: Ratios of transport proteins compared to the genomic sizes

A The relationship between the number of transport proteins and the genome sizes of proteobacteria at TRANSPORTERDB was plotted.

B The relationship of transporter numbers and genome size is shown for the investigated *Burkholderiacea* (red triangle) is compared to this relationship of other proteobacteria at TRANSPORTERDB (blue dots) is shown

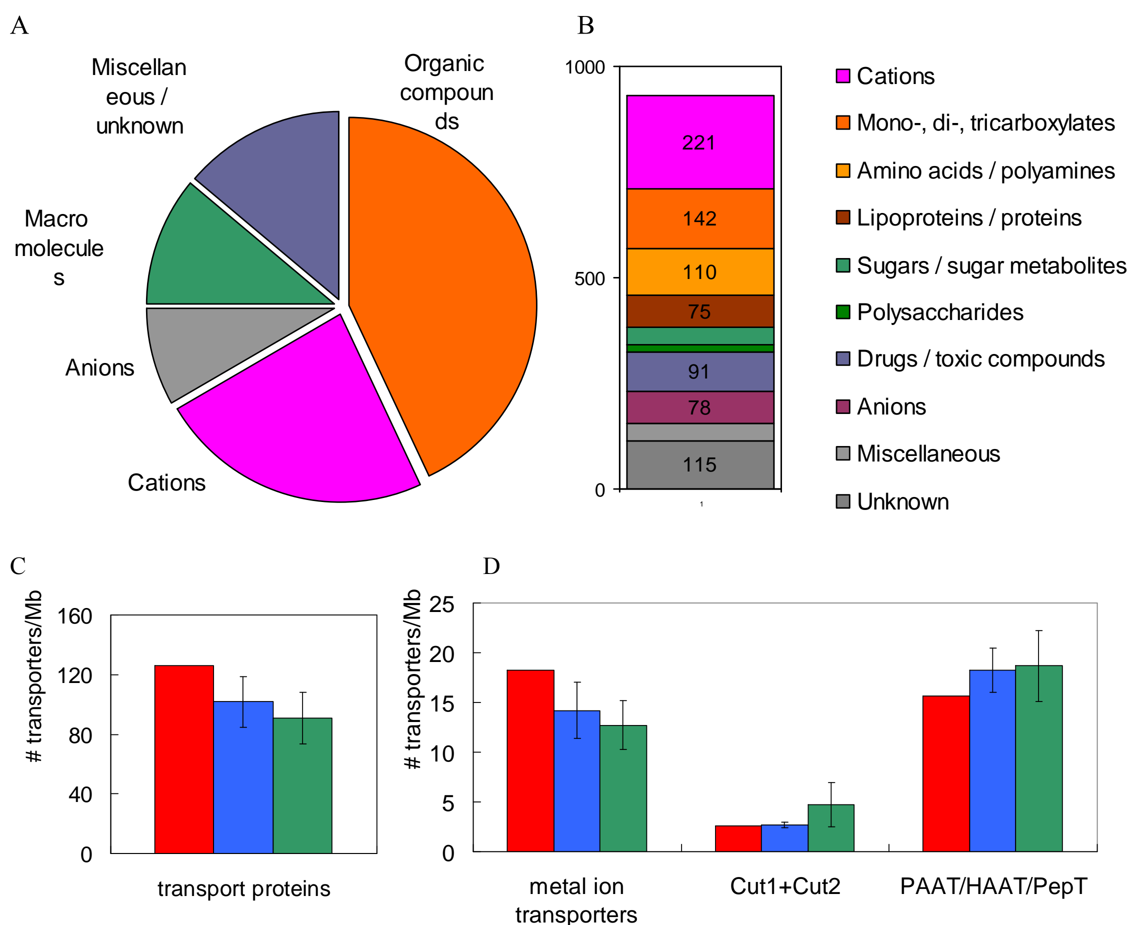


Fig. 14: Ratios of transport protein families and their substrate ranges

A The potential ratios of transported substrates according to the substrate specifities of transport protein families in *C. metallidurans* CH34 have been added up

B The total number of transport proteins with certain substrate ranges in *C. metallidurans* CH34 was added up

C The number of transport proteins in the investigated proteobacteria was related to their genome sizes. The red bars show the ratios for *C. metallidurans* CH34, the blue bars show the average ratios for the *Cupriavidus* cluster and green bars show the average ratios for all investigated *Burkholderiacea*

D The mean average of transporters/Mb with distinct substrate ranges are compared (CUT1and CUT2 are sugar specific transporters; PAAT and HAAT are amino acid transporters)

A common feature of these strains is a large genome. Larger genomes contain usually more transport proteins than smaller genomes and they contain often also more distinct transport protein families (Paulsen *et al.*, 1998). Particularily plant- or soil-associated organisms have often the largest variety and number of transporters (Ren and Paulsen, 2005).

The TRANSPORTDB (http://www.membranetransport.org/) is a quantitative inventory for the number of transporter families in distinct organisms. As outlined in figure 14 have the total transport protein numbers in the investigated *Burkholderiacea* been related to their respective genomic sizes and this relationship has also been compared to those of other proteobacteria at TRANSPORTDB. For the investigated proteobacteria has a nearly proportional relationship between their genome size and transport protein number been observed (figure 14) and this trend also holds true for *C. metallidurans* CH34 and the related *Burkholderiacea* (red triangles in figure 14). The larger *B. xenovorans* proteome has thus 50 % more and the smaller *R. solanacearum* has 30 % less transporters than *C. metallidurans* CH34. The ratios ranged between 167 transporters/Mbp and 125 transporters/Mbp and this corresponds to between 12.5 % and 16.7 % of all proteins.

3.3.2.2. The distribution of transporter families

The transporter classification (TC)-system distinguishes nine major classes of transport proteins which reflect distinct transport modes and energy coupling mechanisms (Saier *et al.*, 2006; Saier, 2000). The most identified putative transport proteins in the investigated proteobacteria fall into the class of primary active transporters (45±5%), followed by electrochemical potential-driven transporters (27±4%) and channels/ pores (11.5±2.5%) (shown in figure 15).

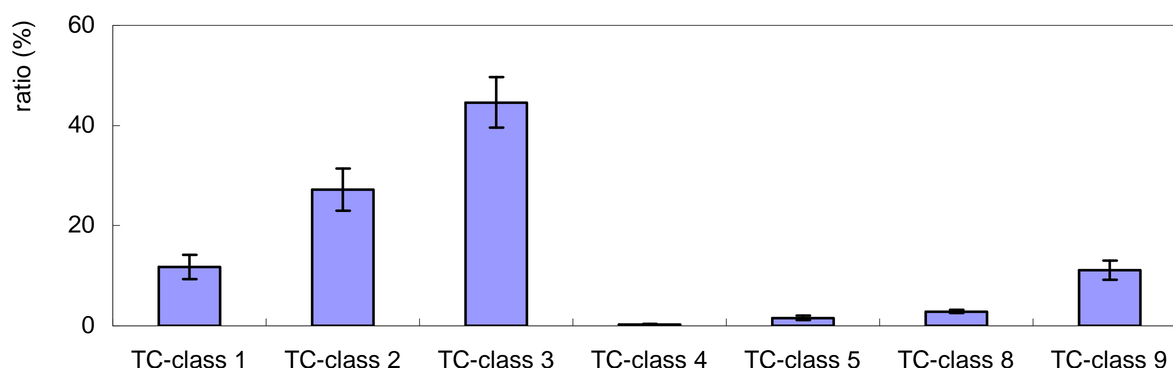


Fig. 15: Distribution of the major transport protein classes

The bars correspond to the fractions (ratio) of each TC-transporter class as part of the overall transport protein repertoires and indicate the mean percentage among all the investigated proteobacteria, the error bars indicate the corresponding variance.

TC-class 1 represents the group of channels and pores
TC-class 2 represents the group of electrochemical potential driven transporters
TC-class 3 represents the group of primary active transporters
TC-class 4 represents group translocators
TC-class 5 represents transport electron carriers
TC-class 8 represents accessory factors
TC-class 9 represents incompletely characterized transporters

The nine major transport protein classes in the investigated β-proteobacteria classes stand for different transport modes and energy coupling mechanisms (Saier 2000).

Primary transporters are less frequent in the *Cupriavidus* cluster (*C. metallidurans* CH34; *C. eutrophus* JMP134 and *C. eutrophus* H16) than in the other investigated proteobacteria. These strains contain also only few phosphotransfer-driven group translocators (PTS - Lengeler, 1990) for sugar import (TC 4.A.1) (1-6 representatives). Each of those organisms has however at least one (glucose specific) PTS system. The sugar specific transporters CUT1 (TC 3.A.1.1) and CUT2 (TC 3.A.1.2) are also underrepresented in this group, but instead more homologs of the amino acid transporters PAAT (TC 3.A.1.3), HAAT (3.A.1.4) and PepT (3.A.1.5) and di-/tricarboxylate transporters are prevalent in these bacteria.

3.3.2.3. The plasmids cumulate unique transporter determinants

Rare and unique transporters may indicate specific metabolic and functional capabilities (Paulsen *et al*., 1998). The distribution of the transporter families for the uptake of essential nutrients and essential cations as well as several plant pathogenesis factors (Type III Secretory Pathway (3.A.6) (Galan and Collmer, 1999) is similar in the investigated proteobacteria.

A rather unusual feature of these strains is an unusual high percentage of transporters for the transport of di-and tricarboxylates. They also have a preponderance of transporters for the utilization of amino acids and peptides. Sugar specific transporters are underrepresented in all these organisms and they contain only few group translocators (PTS systems) (Lengeler 1990; Reizer *et al.*, 1991) for the transport of carbohydrates. Some of the investigated β-proteobacteria are plant pathogens and they contain many proteins for the transfer of effector proteins into plant cells. These effector proteins modulate probably the host defence signalling pathways and promote disease inside the plant cell by interfering with host cell functions (Cunnac *et al*., 2004). The effector proteins can be transferred by the type III (Virulence-related) secretory pathway transporter family (TTSS) (TC 3.6) (Cornelis and Van Gijsegem, 2000; Büttner and Bonas, 2002; Szurek *et al.,* 2002). The TTSS transporter family is widely distributed among plant- and animal pathogenic bacteria and some of these type III proteins are located on mobile genomic islands (Tampakaki *et al.,* 2004; Boucher *et al.*, 2001; Genin and Boucher, 2004; Salanoubat *et al*., 2002; Toussaint *et al*., 2003). Many type III transporters have also been found in the herein investigated proteobacteria and *C. metallidurans* CH34 contains 12 additional paralogous TTSS secretory systems to enhance its capabilities for conjugal transfer.

C. metallidurans CH34 has also a much higher ratio (12 transporters /Mbp) of metal ion specific transporters than the other strains (average ratio 7.2±0.6 transporters /Mbp) (shown in figure 14 and figure 16). The ratio of metal ion transporters exceeds even largely those of the closely related strains in the *Cupriavidus* cluster (figure 16). The ratio of transition metal transporters increases from the plasmid of *C. eutrophus* H16 (4.4 transporters/Mbp) via chromosome 2 of *C. metallidurans* CH34 (12.0 /Mbp) and plasmid pJP4 of *C. eutrophus* JMP134 (22.2 /Mbp) to megaplasmid pMOL28 (29.4/Mbp) and megaplasmid pMOL30 (73.9 /Mbp) of *C. metallidurans* CH34 (figure 16). The chromosomes of all these strains encode between 6.6 and 9.9 transporters/Mbp for metal ions.

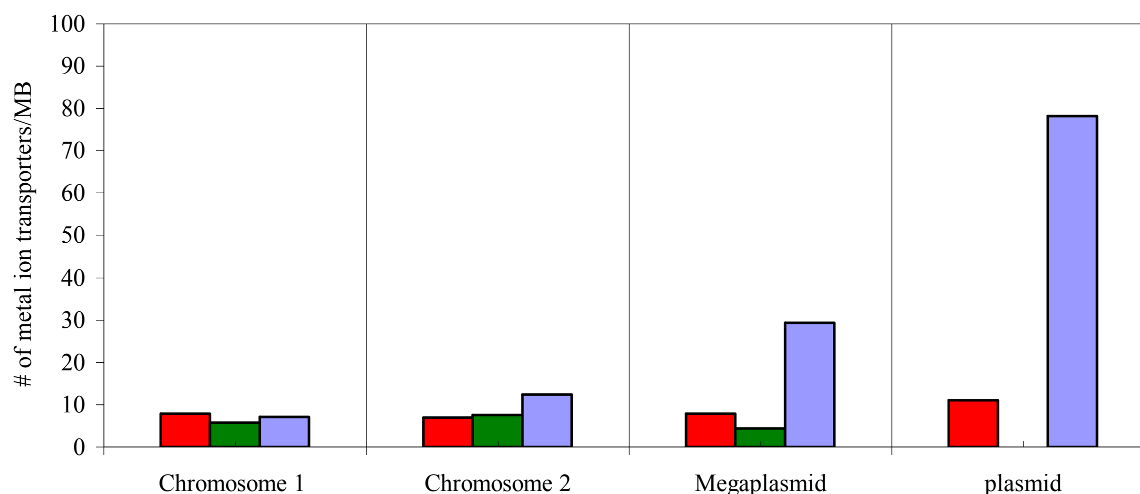


Fig. 16: Distribution of transition metal transport proteins in the *Cupriavidus* cluster.
The bars show the number of transition metal transporters per Mbp (replicon size) at distinct replicons for *C. eutrophus* JMP134 (red bars), *C. eutrophus* H16 (green bars) and *C. metallidurans* CH34 (blue bars).

C. eutrophus H16 uses the 'Knallgas reaction' to grow with H_2 as sole electron donor (Schwartz and Friedrich, 2001) and the major function of megaplasmid pHG1 might thus be molecular hydrogen oxidation. In plasmids *of C. metallidurans* CH34 are metal ion specific transporter determinants cumulated and about two third (18 of 29) of all transport proteins in plasmid pMOL30 are probably involved in transition metal transport.

3.3.2.4. The distribution of RND transporters

Particularly the group of RND (TC 2.A.6) transporters is overrepresented in *C. metallidurans* CH34. Characterized members of the RND (resistance nodulation and cell division) (TC 2.A.6) superfamily catalyze substrate efflux via an H^+ antiport mechanism (Saier *et al*., 1994; Dong and Mergeay, 1994). These proteins are found ubiquitously in bacteria, archaea and eukaryotes but three of its eight recognized phylogenetic families are largely restricted to Gram-negative bacteria (families 1-3) and these transporter families have differential substrate specificities (Paulsen *et al.*, 1998). The HAE1 subfamily (TC 2.A.6.2) of RND transporters can translocate multiple drugs (e.g. tetracycline, chloramphenicol, fluoroquinolones, β-lactams, etc.), whereas the HME subfamily of RND (2.A.6.1) transporters is implicated in heavy metal (Co^{2+}, Zn^{2+}, Cd^{2+}, Ni^{2+}, Cu^{+} and Ag^{+}) homeostasis (Tseng *et al.*, 1999). With 12 HME-RND transporters exhibits *C. metallidurans* CH34 an outstanding ratio of HME transporters but the number of its HAE1 RND transporters resembles those in other investigated strains.

3.3.2.5. The structure of the CzcCBA efflux complex

The RND (TC #2.A.6) transporters in the cytoplasmic membrane (Rensing *et al.* 1997, Anton *et al.,* 1999; Nies; 2003) can cooperate with membrane fusion proteins (OMF- TC 1.B.17) (Dong and Mergeay, 1994) in the outer membrane and the interaction of these transporters is probably supported

by adaptor proteins of the MFP (TC 8.A.1) family (Saier *et al*., 1994) (Rensing *et al*., 1997; Touze *et al.,* 2004; Zgurskaya and Nikaido, 1999; Lobedanz *et al*., 2007).

The protein structure model for a representative of the metal ion specific HME subfamily of RND transporters is not yet available. The protein structure model of the HAE1-RND drug exporter of *E. coli*, AcrB (TC #2.A.6.2.2) has been resolved in 2002 (Murakami *et al*., 2002). Specific structural features of AcrB have already helped to get insights into the probable transport mechanism of other HAE RND transporters (e.g. Murakami *et al.,* 2006; Hearn *et al*., 2006; Nehme and Poole, 2007). The protein structure model of AcrB (1IWG) has been obtained from the Brookhaven Protein Databank (PDB). A structure model for the OMF transporter TolC is also available (PDB-ID 1EK9) (Koronakis *et al*., 2000). The protein sequences of AcrB and the HME-RND transporter CzcA have been compared and the conservation of residues has been visualized with the protein structure analysis program DEEPVIEW (SPDBV) (Guex and Peitsch, 1997) and with the PROTEIN EXPLORER (MSA3D at http://www.umass.edu). Important HME1 RND transporters in *C. metallidurans* are the pMOL30-encoded CzcA, and a putative protein encoded by its chromosome 2 (Hme468, Rmet_4468). Both RND proteins share 80% sequence identity but have more closely related orthologs encoded by the two *C. eutrophus* strains (AcrB10 is 82% identical to CzcA; Reut_B3968 82% identical to Hme469). Multiple alignments (MSAs) were used to identify functionally important conserved residues of the RND transporters (table 3). The Deep View program has also been used to calculate the electrostatic potential of the proteins which are induced by charged groups in the protein structure model. The electrostatic potential has been visualized as a colour gradient at the molecular surface (electron density map EDM) (figure 18). Internal cavities have been calculated with CASTP. The transport of substrates by RND transporters is probably supported by an electrical field that is generated by a proton relay pathway (Mao *et al.,* 2002; Goldberg *et al*., 1999). The probable location of functional important charged residues (D402, D408 and E415 in AcrB) of this putative (DDE) charge relay network in the transmembrane segments of AcrB are shown in figure 17. The figure 18 also shows the structure of docking domains between the components of the AcrABTolC complex.

Tab. 3: Examples of homologs of the HAE-RND transporter AcrB of *E.coli*

Accession #	homolog
P31224	AcrB acriflavine resistance protein B of *E.coli* (TC 2.A.6.2.2)
P24177	AcrD (acriflavine resistance protein d) of *E.coli* (TC 2.A.6.2. 7)
P38054	CusA putative cation efflux system protein CusA of *E. coli* (TC 2.A.6.1.4)
P13511	CzcA cobalt-zinc-cadmium resistance protein of *C. metallidurans* CH34 (TC 2.A.6.1.2
Rmet_4468	Hme468 homolog of CzcA at the chromosome of *C. metallidurans* CH34
P37972	CnrA nickel and cobalt resistance protein CnrA of *C. metallidurans* CH34 (TC 2.A.6.1.1)
Q9ZHC9	SilA (silver ion specific efflux pump) of *Salmonella typhimurium* (TC 2.A.6.1.3)
B3968	homolog of CzcA in *C. eutrophus* JMP134

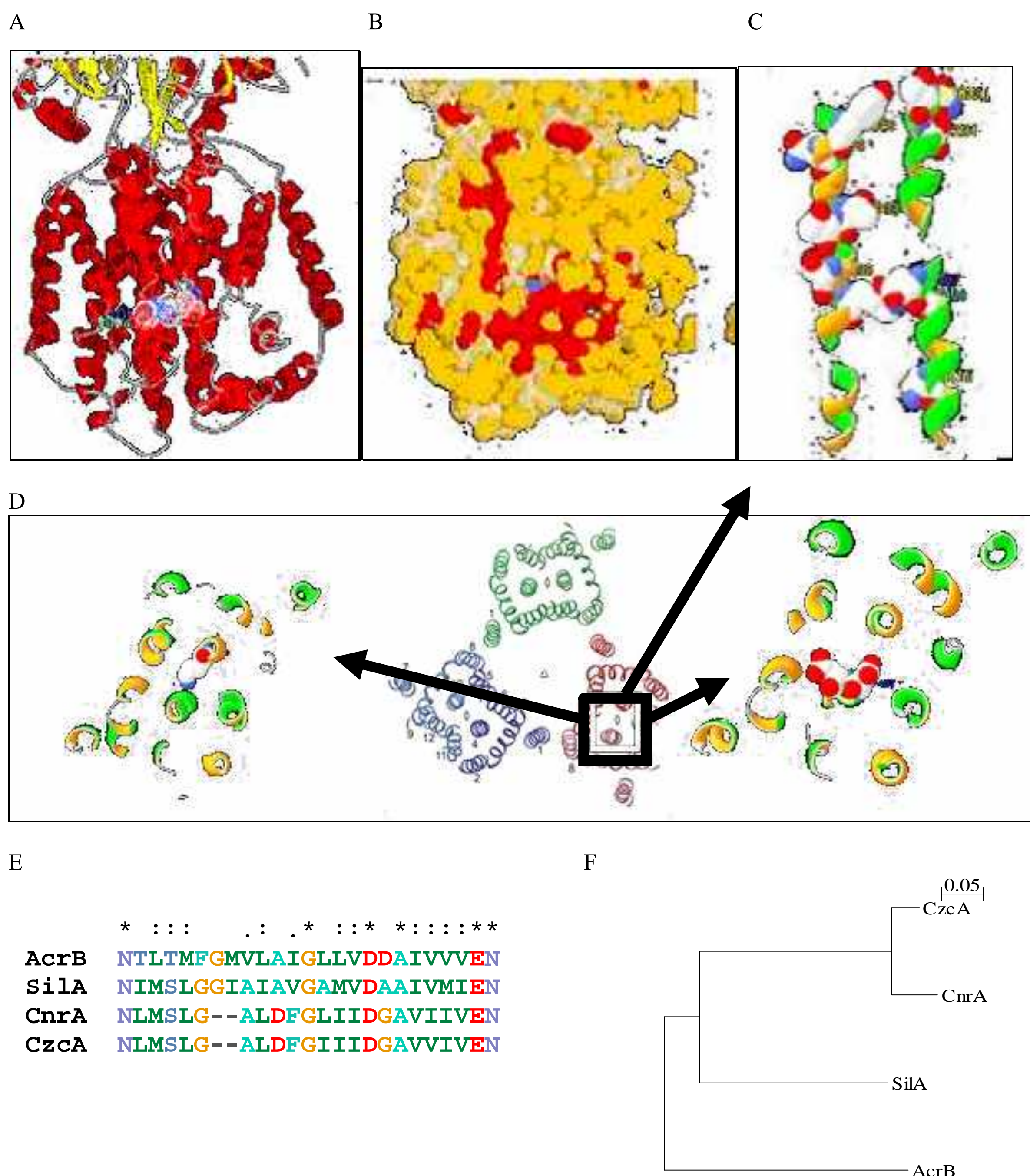


Fig. 17: The protein structure model of the RND transporters

A The picture shows the location of Asp407/Asp408/Lys940 as part of the charge relay pathway in TMS4 and TMS10 of AcrB (Seeger *et al.*, 2006). Putative internal cavities in the AcrB structure model have been calculated and visualized with CASTP and protein structure model viewers PE and YASARA (right side).

B The molecular surface and charge distribution of probable internal cavities of AcrB was calculated and visualized with PE in slab mode (AcrB structure model cut along the Cavities and the polar residues in TMS 4 and TMS 10. Shown in blue colour are charged residues, red are not charged residues as part of putative internal cavities

C The charged residues of the DDE charge relay network of AcrB in TMS4 are displayed in ribbon mode

D The geometry between Asp407/Asp408/Lys940 TMS4 (left) and TMS 10 (right side) in AcrB are shown as view from top in ribbon mode (created with SPDBV)

E Multiple sequence alignment of certain RND transporters (table 3) was carried out (only a segment of TMS4 with the putative DDE charge relay network of CzcA displayed) was carried out with CLUSTALW

F The phylogenetic tree for the sequences of the investigated RND transporters (table 3) was constructed with ClustalX

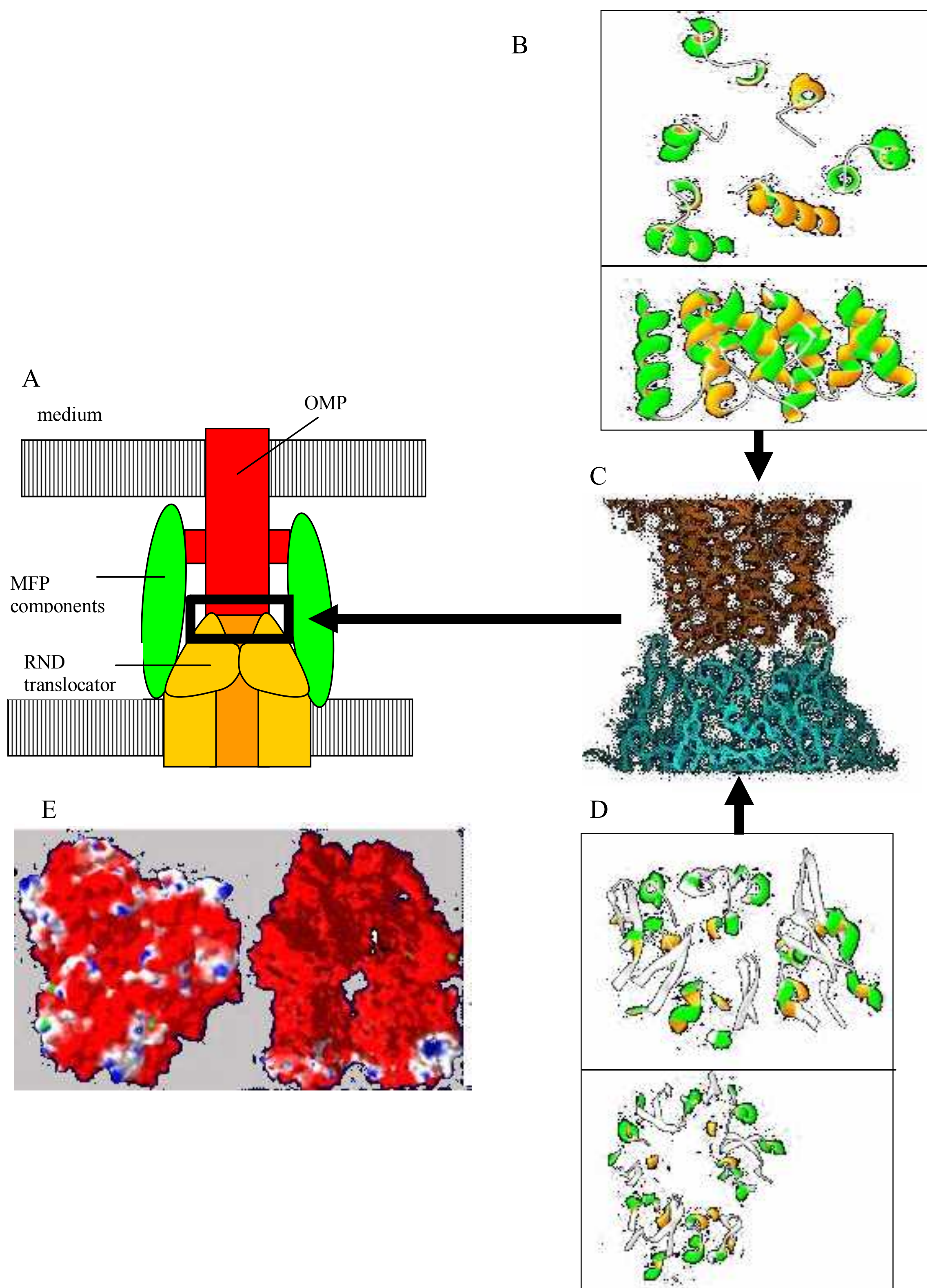


Fig. 18: The docking domains of the AcrABTolC complex
A The picture shows a schematic representation of the AcrABTolC efflux complex
B and C A periplasmic coiled-coil interface (Lobedanz *et al*., 2007) in TolC supports the recruitment by AcrB and facillitates the assembly of the CBA-type efflux complex (top view, side view)
D The picture shows the finger like structures of the AcrB headpiece (created with SPDBV) that bind the respective coiled-coil interface of TolC in close connection, (described as “periplasmic kissing”).
E Shows the charge distribution (top view, side view) of AcrB (created with PE)

3.4. *Cupriavidus metallidurans*: evolution of a metal-resistant bacterium

(accepted in 2008 *Antonie Van Leeuwenhoek*)

The evolution of genomic sequences is determined by vertical descent (speciation), gene duplication, gene loss, horizontal transfer, fusion events, or other genomic rearrangements (Koonin, 2005). Orthologs are proteins in different species, that go back to a single protein in the last common ancestor of these species and they are thus often responsible for a similar function (ortho= exact) (Fitch, 1970; Storm and Sonnhammer, 2002). Paralogs arose by a gene duplication event within an organism (para= in parallel) and they exhibit a higher sequence similarity to another protein within the same proteome than to any protein of another organism (Tatusov *et al*., 1997). A comparison of the BLASTP similarity score and the amino acid identities was carried out to subcategorize homologs into orthologs and paralogs in *Burkholderiacea*. The results were improved by only accepting reciprocally best matching protein pairs as orthologs. This reciprocal-best-BLAST-hits (RBH) approach (Tatusov *et al*., 1997) works usually reasonably well for the comparison of the proteomic sequences of bacteria (Mushegian *et al*., 1998; Alexeyenko *et al*., 2006; Alexeyenko *et al*., 2006b).
Preceding investigations revealed a high degree of conservation for essential regulatory proteins (EF-TU, RpoA B and C and RpoD) but an unusual high degree of evolutionary flexibility for the investigated metal ion transporters and certain sigma factors have also been observed (Dräger, 2005).
To pursue these investigations have all protein sequences of *C. metallidurans* CH34, *C. eutrophus* H16 and *C. eutrophus* JMP134, *R. solanacearum* GMI1000, *Burkholderia cepacia* AMMD and *Burkholderia xenovorans* LB400 been investigated for orthologous relationships (Tatusov *et al.*, 1997). Strain *R. solanacearum* UW551 with its unfinished genome was omitted. The protein sequences of all these investigated bacteria have been blasted vice versa. Reverse BLAST and a comparison to the sequence identities to proteins in *E. coli* K12 have been used to verify candidate orthologous relationships (Mushegian *et al.,* 1998).The results are outlined in figure 19.
Orthologs have been for most (76 to 82%) proteins in *C. metallidurans* CH34 and the most of these orthologs are encoded at the largest replicons. About one quarter (26%) of all proteins in *C. metallidurans* CH34 have orthologs in all these strains. Some 'rare' orthologs (9% of all its proteins) are only distributed among some of these strains (figure 19) and most of these proteins were encoded at the smaller replicons.

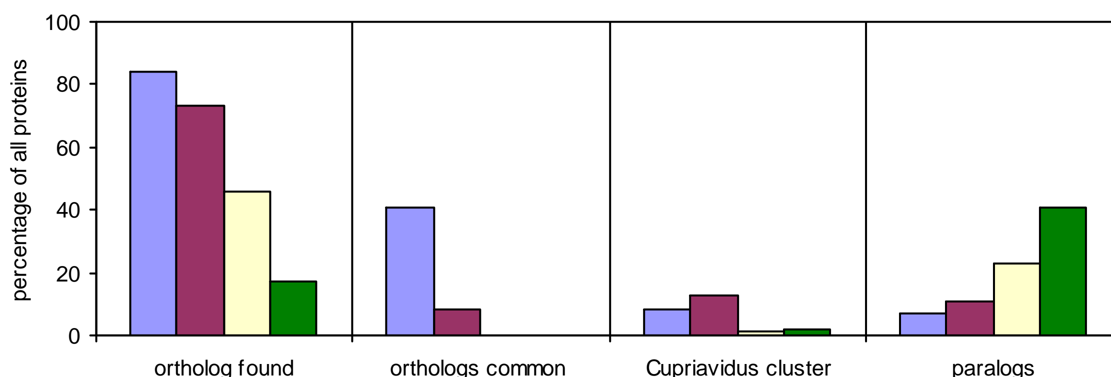


Fig. 19: Distributions of orthologs and paralogs on different replicons.

The bars show the ratios of orthologs and paralogs in *C. metallidurans* CH34. Blue bars represent the ratios for the chromosomally encoded proteins, The brown bars show the ratios for the megaplasmid encoded proteins and yellow and green bars show the ratios for plasmids pMOL28 and pMOL30 respectively.
The category on the left side ('orthologs found') show the ratios of proteins in *C. metallidurans* CH34, which have an ortholog in at least one of the investigated proteobacteria. The next category ('orthologs common') shows the ratios of proteins that have orthologs in all investigated strains. The rare orthologs in the *Cupriavidus* Cluster and the ratios of paralogs are also shown.

3.4.1. The distribution of paralogs

The total ratios of paralogs (9±3%) were nearly equal in all the investigated proteobacteria (figure 20). The total frequencies of putative paralogous transporters were also within this 9±3% range. Many paralogs were plasmid- or megaplasmid encoded (31% in *C. eutrophus* H16; 21% in *C. eutrophus* JMP134 megaplasmid). In *C. metallidurans* CH34 are nearly the half of the plasmid pMOL30 encoded proteins (41%) and one quarter of plasmid pMOL28 encoded proteins paralogs but only 10.25% of the chromosomal encoded proteins fall into this group.

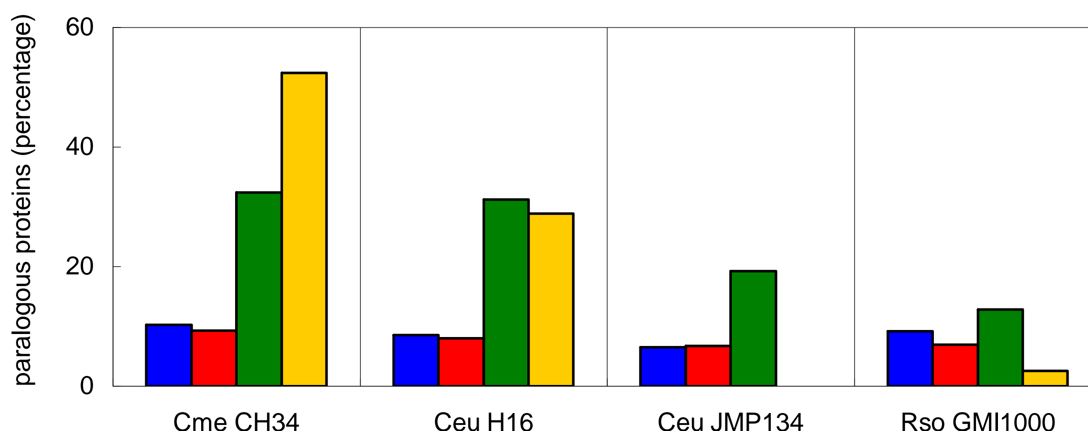


Fig. 20: Distribution of paralogs

The diagram shows the ratios of paralogous proteins in the *Cupriavidus* cluster (*C. metallidurans* CH34 (Cme CH34), *C. eutrophus* H16 (Ceu H16), *C. eutrophus* JMP134 (Ceu JMP134)) and in *R. solanacearum* GMI1000 (Rso GMI1000). The blue bars show the percentages of paralogs within the entire proteomes, red bars show only the fractions of paralogous putative transporters, green bars show only the fractions of paralogous proteins, that are encoded at the bacterial plasmids and yellow bars show the percentages of paralogs among transport proteins that are encoded at bacterial plasmids.

In *C. metallidurans* CH34 are not only one third of all plasmid encoded proteins but even about the half of all plasmid encoded putative transport proteins paralogs (28.9 % in *C. eutrophus* H16, 2 % in *C. eutrophus* JMP134). The plasmid encoded proteins of the *Cupriavidus* cluster exhibit thus an unusual genomic flexibility and in *C. metallidurans* CH34 is this flexibility specifically directed to the evolution of transport proteins.

The putative function for the bulk of paralogous transport proteins differs in these bacteria. In *C. eutrophus* H16 are paralogs more prevalent among transporters of the Major Facilitator Superfamily (MFS) (2.A.1) (Pao *et al*., 1998) the ATP-binding Cassette (ABC) Superfamily (3.A.1) (Saurin *et al*., 1999) and the Proton-translocating Transhydrogenase (PTH) Family (3.D.2) (Saier *et al*., 2000). In contrast to this have many paralogs in *C. metallidurans* CH34 been found among the transporters of the Type IV Secretory Pathway (IVSP) Family (3.A.7) (Gauthier *et al*., 2003) and its plasmids encode many paralogous heavy metal ion transporters.

In preceding investigations have sequence similarities between some of these putative heavy metal transport proteins and similarities between sigma factors in proteobacteria been correlated to the phylogenetic relatedness (16S-rRNA identities) of theses strains (Dräger, 2005).

3.4.2. The distribution of recent / middle aged and old paralogs

For further investigations of orthologous and paralogous relationships can the similarities of biological sequences be compared (reconciled) with the species tree (Sonnhammer and Koonin, 2002). Thus have recent paralogs (inparalogs) been formed by a lineage specific duplication which occurred subsequently to the speciation of an organism from related species. In contrast to this are middle aged and old paralogs (outparalogs) the result of a lineage specific duplication preceding this speciation event and this results in orthologous relationships that are restricted to this phylogenetic lineage (Koonin, 2005).

The outparalogs in the investigated proteobacteria have been subdivided into old paralogs which were duplicated before *R. solanacearum* GMI1000 has separated from other *Burholderiacea* and into middle aged paralogs which evolved during the divergence of the *Cupriavidus* cluster. To identify outparalogs and to shed light into the probable evolutionary timescale of the gene duplication were certain proteomes stepwise removed from consideration. When certain bacteria are taken out of consideration will the corresponding orthologous relationships to their proteins be ignored. This might then result in the occurrence of new paralogous relationships in the remaining organisms (outparalogs). However might not only gene duplication events, but also gene losses and horizontal gene transfer result in the occurrence of outparalogs. The iteration process started with the strain which exhibits the lowest degree of phylogenetic relatedness to *C. metallidurans* CH34 and was concluded with the most closely related strain. The phylogenetic relationships between the investigated proteobacteria were revealed by the Tree Builder (RDPII) program (Maidak *et al.,* 2001).

The phylogenetic tree (figure 21) was based on multiple sequence alignments of the corresponding 16S-rRNA sequences.

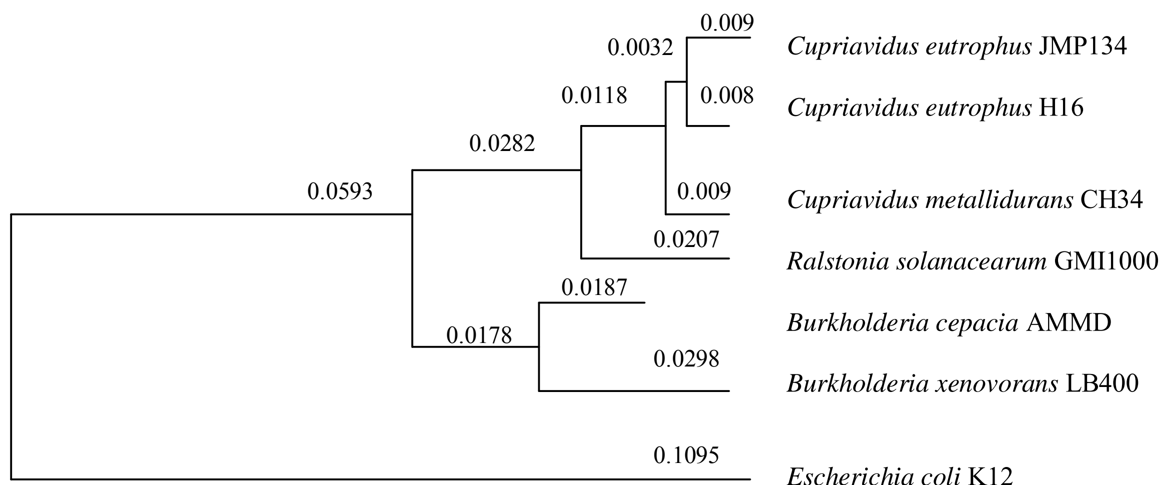


Fig. 21: Phylogenetic tree of the investigated proteobacteria
The Tree Builder (RDPII) program was used to create a phylogenetic tree that was derived from 16SrRNA sequences (Neighbour Joining with bootstrapping). It indicates the evolutionary relationship between the investigated proteobacteria. The 16SrRNA sequence of *E. coli* K12 was used as the taxonomic outgroup to find the root for this tree.

3.4.3. The evolution of metal ion transporters

The CDF (TC 2.A.4), MerTP (TC 9.A.2), MFP (TC 8.A.1), MIT (TC 1.A.35), NiCoT (TC 2.A.52), OMF (1.B.17), OMR (1.B.14), P-ATPase (TC 3.A.3), CHR (2.A.51), HME (RND) (TC 2.A.6.1) and ZIP (2.A.5) transport protein families are specific for heavy metal ions (outlined in table 4).

Tab 4: Important protein families for heavy-metal transport.

TC-Family	TC-Family	TC #	references
OMR	The outer membrane receptor Family	1.B.14	(Locher *et al*. 1998)
OMF	The Outer Membrane Factor Family	1.B.17	(Dong and Mergeay, 1994),
MIT	The CorA Metal Ion Transporter Family	1.A.35	(Smith and Maguire, 1998)
MFS	The Major Facilitator Superfamily	2.A.1	(Saier, 2000)
CDF	The Cation Diffusion Facilitator Family	2.A.4	(Nies *et al.,* 1995)
ZIP	The Zinc (Zn^{2+})-Iron (Fe^{2+}) Permease Family	2.A.5	(Grotz *et al*., 1998)
RND	Resistance, Nodulation, Cell division Family	2.A.6	(Dinh *et al*., 1994)
CHR	The Chromate Ion Transporter Family	2.A.51	(Nies and Brown, 1998)
NiCoT	The $Ni^{2+-}Co^{2+}$ Transporter Family	2.A.52	(Eitinger and Friedrich, 1991)
ABC	ATP-Binding Cassette Transporters	3.A.1	(Saurin *et al.,* 1999)
P-type ATPases	The P-type ATPase Family	3.A.3	(Solioz and Vulpe, 1996)
MFP	The Membrane Fusion Protein Family	8.A.1	(Dinh *et al*., 1994
MerTP	The MerTP Mercuric Ion (Hg^{2+}) Permease Family	9.A.2	Qian *et al*., 1998
FeoB	The Ferrous Iron Uptake Family	9.A.8	(Kammler *et al*., 1993).
ILT	The Iron/Lead Transporter Family	9.A.10	(Cao, *et al*., 2007)

Several putative transporters of the MerT, and KUP families and many lead, cadmium and calcium specific P-type ATPases (3.A.3) in *C. metallidurans* CH34 are probably the product of a recent gene duplication event. Only few (6 to 23%) of the components of CBA-type efflux systems (RND, MFP and OMF) in this organism have orthologs. The most important components of CBA-type-efflux systems are heavy metal efflux (HME) transporters (TC 2.A.6.1) of the RND superfamily (Saier *et al.*, 1994, Dinh *et al.*, 1994). These can be subcategorized into HME1 ($Zn^{2+}/Co^{2+}/Cd^{2+}$), HME2 (Ni^{2+}/Co^{2+}), HME3a (Zn^{2+}), HME3b (Co^{2+}) and HME4 (Cu^{+}/Ag^{+}) transporters (Nies, 2003). The proteins of HME3/HME4 subfamilies of RND transporters have not recently been duplicated in this organism. Not only the HME4 RND transporters SilA and CusB (Gupta *et al.*, 1999) but also the corresponding MFPs of their respective CBA-type efflux complex are middle aged paralogs and orthologs were most widely distributed in *C. eutrophus* JMP134 and *C. eutrophus* H16. These proteins may have thus been formed by a gene duplication event that occurred prior to the divergence of the *Cupriavidus* cluster. Components of the CBA-type efflux complex CzcCBA for cobalt zinc and cadmium resistance are however old paralogs. The periplasmic metal binding protein CzcE that is also encoded at the *czc*-operon is a recent paralog (inparalog). The two cobalt and nickel specific homologs of the HME2 subfamily of RND transporters are also recent paralogs. The high genomic flexibility of the HME RND families has obviously been an important factor for the unusual heavy metal resistance in *C. metallidurans* CH34. The CDF (CzcD, DmeF and FieF), CHR ($ChrA_1/ChrA_3$ and $ChrA_{4N}/ChrA_{4c}$), MIT transporters and P-ATPases of *C. metallidurans* CH34 have not extensively been duplicated and the most of these transporters are thus old paralogs.

3.4.4. The conservation of regulators

Many metal ion transporters extrude not only toxic, but also essential cations and their expression is often tightly regulated (van der Lelie *et al.*, 1997). Particularly the group of ECF (extracytoplasmatic function) sigma factors (Raivio and Silhavy, 2001) plays a crucial role in the response to transition metal ions and other cellular stressors (Grosse *et al.*, 2007). Putative regulatory proteins in *C. metallidurans* CH34 and the related proteobacteria where identified by applying keyword searches in NCBI and SWISSPROT (Bairoch *et al.,* 1997) databases. The information and sequences from the supplementary material of recent publications (e.g. Grosse *et al.,* 2007; Nies *et al.,* 2006) and the information about the proximity to determinants coding for the efflux system was also used.
An unusual high degree of evolutionary flexibility has already been reported for the sigma factor RpoN (Dräger, 2005). RpoN (nitrogen metabolism) controls operons that have to remain absolutely silent when not needed (Reitzer and Schneider 2001; Nies, 2004).
Several plasmid encoded ECF-sigma factors are unique to *C. metallidurans* CH34 (e.g. CnrH) and *C. metallidurans* has triplicated its main sigma factor for iron homeostasis namely RpoI (Sig680), RpoJ (Sig611) and RpoK (Sig697) (Große *et al*. 2007). RpoK and RpoJ are recent paralogs. This flexibility

of ECF sigmafactors contrasts to a largely conserved set of chromosomal encoded sigma factors of the Sigma70 and Sigma54 family. *C. metallidurans* CH34, *C. eutrophus* H16 and *C. eutrophus* JMP134 have a nearly equal number of putative two-component-systems. About two third of these regulators in the three organisms are orthologs. Extensive recent gene duplications have not occurred in this family.

3.5. Unpublished results of supplementary investigations

3.5.1. Ordering contigs of *C. metallidurans* CH34

C. metallidurans CH34 contains a primary chromosome (3.93 Mbp), a secondary chromosome (2.58 Mbp) and two megaplasmids pMOL28 (171 Kbp) and pMOL30 (234 Kbp) (Taghavi *et al.*, 1997; Mergeay *et al.*, 1985). Both of these low copy number plasmids are stably maintained even without selective pressure and they are self-transferable at low frequencies (Szpirer, *et al.* 1999). The genomic sequence assembly of this β-proteobacterium has been carried out by the Joined Genome Institute (JGI). The assembling progress of its genomic sequence has however long been hampered by misassembles e.g. on the junctions of insertion elements. Those are widely distributed in its genome and thus have for many years only incomplete (Draft) Versions for the genomic sequence of this organism been published by DOE JGI. These Draft versions contained many uncertainties about the order and position of genomic sequences. Previous Draft versions for *C. metallidurans* CH34 at JGI (finished state- 5/1/2006) and their GENBANK identifiers are summarised in table 5.

Tab. 5: Draft versions for *C. metallidurans* CH34
The table shows an overview of the actual genome sequence (version published in may 01 2006) and previous Draft versions for the *C. metallidurans* CH34 assembly at JGI (JGI project ID: 2351356)

finished assembly date		major contigs (>10 reads and >2Kbp)	GENBANK Accession #	Size (basepairs)
May 01 2006	chromosome1	Contig1808	CP000352	3928091
	chromosome2	Contig1743	CP000353	2580086
	pMOL30	Contig1731	CP000354	233722
	pMOL28	Contig1651	CP000355	171461
				691336 total

Draft Version (Phase III in prefinishing)

	total contigs	major contigs (>10 reads and >2Kbp)	Size (basepairs)
Dec 17 2004	29		6953150
Nov 17 2003	72	45	6800000
dec2000	783		5242141

Computational gap closure has been used to give insights into the probable organisation of the *C. metallidurans* CH34 genome before the completely finished genomic sequence version was published in 2006. The protein sequences of distinct Draft versions have been blasted (BLASTP, BLASTN) vice versa to keep track of the gene identifiers of all published Draft versions. Several other sequence alignment strategies were also applied. Particularly, the MUMmer program has facilitated these large scale genomic comparisons. A contig mapping of the genomic sequences of *C. metallidurans* CH34 against those of close relatives such as *C. eutrophus* JMP134 and *R. solanacearum* GMI1000 has been carried out (Fleischmann *et al*., 1995) with BLASTN and the PHRAP (Staden *et al*., 2000), CAP3 (Huang and Madan. 1999), PROJECTOR (Sacha *et al*., 2003) and CONSED (Gordon *et al.,* 1998), programs have also been used. The most of these assembler programs use stringent BLAST to find probable overlaps between contigs. Local alignments were carried out with the MUMmer program. In some of these programs is a repeat masking step carried out prior to the alignment procedure. The table 6 provides a summary of all these investigations.

Tab. 6: Conversion table of Draft versions for *C. metallidurans* CH34.
BLASTP was used to associate the contig identifiers of older Draft versions with contig identifiers of newer Draft versions of *C. metallidurans* CH34 (older contigs are ordered from left to right)

Draft 2006	Draft 2006	Draft 2004	Draft 2003 contigs	Draft 2000 contigs
Contig 1651	**pMOL 28**	58	335	783
Contig 1731	**pMOL 30**	70 72	203 279	567 686 509 457 429 617 718 573 724 557 645 756 475 230 468 699 407 441 500 584 482 570 679 665 685 596 447 656 720 526 690 740 675
Contig 1743	**Mega-plasmid**	49 51 52 54 56 57 59 62 64 66 67 69 70 71 72	222 185 187 222 187 197 222 204 202 201 204 197 222 197 202 204 222	668 681 633 535 489 510 634 508 647 417 517 589 613 562 693 655 586 721 619 755 576 502 660 560 611 433 444 656 767 726 659 735 723 568 700 648 695 157 706 587 635 386 494 762 626 398 380 588 448 608 585 775 778 774 532 661 676 537 571 432 652 108 594 689 578 480 746 771 748 572 664 777 730 434 431 719 205 453 730 768 591 506 455 438 423 420 737 590 530 743 749 646 524 744 765 776 707 758 499 741 552 616 721 553 605 650 490 609 711 703 734 641 412 546 629 51 472 755 745 701 782 773 781 757 521 396 603 716 622 654 559 387
Contig 1808	**Chromo-some1**	41 42 44 46 48 50 53 60 61 62 63 65 66 68 71 72	238 186 205 181 205 183 188 192 191 205 199 195 199 205 219 194 205 182 205 219	381 615 763 529 670 486 516 541 518 684 574 601 404 577 597 769 691 638 564 692 403 704 637 229 555 452 556 606 683 713 527 680 528 406 766 752 687 739 751 540 549 561 717 651 409 503 600 523 671 397 563 580 612 583 470 339 593 688 520 682 519 694 462 437 505 436 632 366 430 709 733 610 729 456 662 495 627 614 419 325 667 421 698 582 644 427 458 566 391 550 599 543 515 451 507 312 408 620 554 714 595 575 712 538 388 504 569 697 450 696 533 674 579 618 628 649 621 702 551 666 435 602 547 481 467 754 484 374 658 653 442 592 728 640 496 604 607 178 492 473 229 389 483 415 630 485 663 545 414 425 493 678 525 511 539 643 742 636 477 410 727

The lengths of contigs of distinct Draft versions were added up and compared to ensure that the results of the gap closure attempts are meaningful. These investigations were also supported by recent publications in which e.g. the location of several sigma factors and several resistance determinants were experimentally revealed (e.g. Grosse *et al*., 2007; Mergeay *et al.,* 2003). Thus have BLASTN runs between the corresponding nucleotide sequences of such determinants (submissions from NCBI) and the nucleotide sequences of contigs of *C. metallidurans* CH34 been carried out. These investigations have helped to reveal the probable orientation and order of genes and contigs of *C.*

metallidurans CH34 and they have supplied the annotation and transport protein classification for this organism with the identifiers of previous and novel Draft versions.

3.5.2. Annotation of *C. metallidurans* CH34

Pursuing investigations of the genomic sequences of *C. metallidurans* CH34 were combined with available annotation for this strain to support further investigations of its heavy metal homeostasis determinants. The annotation is an explanation or commentary that describes the potential function and location of proteins or genomic sequences and much of this information can be retrieved from DOE JGI. The genomic sequences of *C. metallidurans* CH34 have passed through an automatic genome analysis pipeline at the JGI's production facility (PGF) of the Oak Ridge National Laboratory (ORNL) to reveal this information (Hauser *et al*., 2004, Markowitz *et al.,* 2006).
The derived information at this site provides e.g. the gene identifier, predicted operon structures, sequence similarity to homologs at the Non-Redundant Nucleotide Database (Pruitt *et al.,* 2005) and SWISSPROT (Bairoch *et al.*, 1997), characteristic protein domains (PFAM -Sonnhammer *et al.,* 1998) as well as COG (cluster of orthologous groups) entries (Gene Ontology Consortium., 2004). More evidence for potential protein functions is given by the classification according to the GO (Ashburner *et al*., 2000) and KEGG (Kanehisa and Goto; 2000) databases. Websites at JGI also provide the nucleotide sequences and CDS translations (protein sequences). Several of these entries were obtained and stored and additional information was derived from its DNA or protein sequences to reveal specific capabilities of *C. metallidurans* CH34. The corresponding gene identifiers of all previous Draft versions have also been added. A multiple sequence alignment (CLUSTALW - Thompson *et al.,* 1994) with close homologs was carried out and phylogenetic trees were constructed with CLUSTALX (Thompson *et al.*, 1997). Corresponding protein sequences of related bacteria have been obtained from TIGR (CMR), NIH or JGI (Peterson *et al*., 2001). The probable protein localisations were revealed by searching for putative signal sequences (PSORT - Nakai and Kanehisa, 1991). The hydropathical profile (TMHMM - Sonnhammer *et al.,* 1998b) was determined and a detection of probable frameshift mutations (FSED - Fichant and Quentin, 1995) was carried out. An overview of the genomic neighbourhood of the genes according to the MIP-Protein Extraction Description and Analysis Tool (Riley *et al*., 2005) has also been integrated.
Genomic sequences and results of annotation efforts have subsequently been integrated into a local database which can be used with BIOTECHNIX3D. As shown in figure 22 allows this program an easy maintenance and manipulation of stored annotation and it provides tools and links for further computational investigations. These pursuing computational investigations were particularly focused on genetic regions which contain putative transport proteins and corresponding regulators. Particularly transporters being implicated in heavy metal ion homeostasis have been a central target for these bioinformatic investigations.

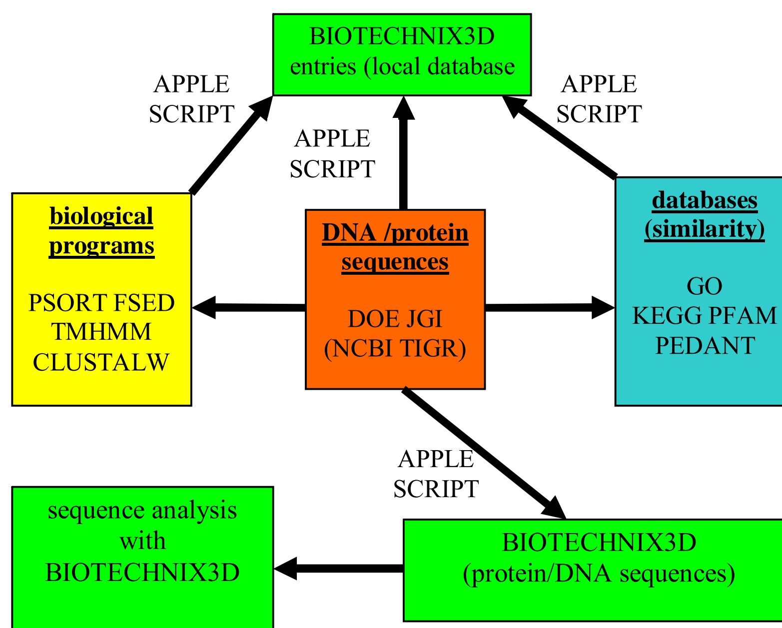


Fig. 22: Workflow for the annotation of *C. metallidurans* CH34.
The figure shows an overview of the computational investigations that were carried out with genomic sequences of *C. metallidurans* CH34. The genomic sequences were obtained from the Joined Genome Institute (JGI) (orange) and were investigated with external programs (yellow) or with programs, which are implemented in the BIOTECHNIX3D package (green). The similarity of these sequences to entries at certain biological databases (dark green) was also investigated. The process of assembling information and also the integration of this data into a database for Biotechnix3d was facilitated by using the APPLESCRIPT scripting language (http://www.apple.com/applescript/) that is implemented in the Macintosh OS (MacOSX and OS9 supported). Manipulations of HTML files with Javascript code have been used to automatically access databases and programs at the internet.

3.5.3. Characterisation of metalloproteins

The probable metal binding motifs of putative heavy metal specific transporters and other metalloproteins in *C. metallidurans* CH34 have been identified and stored. The metal-binding sites of transporters and metalloproteins can be very diverse, varying e.g. in their coordination numbers, geometries and metal preferences (Babor *et al.*, 2005; Opella *et al.*, 2002; Patel *et al.*, 2007). The corresponding signature sequences (e.g. metal binding sites) can facilitate functional investigations and these motifs can also be used to subcategorize metal ion transporter families according to their substrate specifities (Saier, 2000). Thus is e.g. an N-terminal Cys-X-X-Cys rich metal binding motif a common feature of CPX-ATPases, but this motif is modified in the lead specific P-type efflux ATPases of *C. metallidurans* CH34 (Cys replaced by Glu) and a modified (CXXEE) motif can also be found in another CPX-ATPase (CadA) of this organism (Tong *et al.*, 2002; Battistuzzi *et al.*, 1996, Borremans *et al.*, 2001; Mergeay *et al.*, 2003). A novel histidine-rich CPx-ATPase (Bxa1) has a

unique Cys-Cys (CC) sequence element that probably facilitates a resistance to both monovalent (Cu^+ and Ag^+) and divalent (Zn^{2+} and Cd^{2+}) metal ions (Tong *et al*., 2002).

The charged residues in transmembrane helices of metal ion specific (HME) RND transporters can similarly be used to predict the specificity of identified putative RND transporters and to assign them to subgroups of this transport protein family (Goldberg *et al.*, 1999; Nies; 2003; Guan and Nakae, 2001).

The potential metal binding motifs of metal ion transporters or metalloproteins are however usually combined of only four to six ligands (Alberts *et al*., 1998; Bock *et al*., 1995; Katz *et al*., 1996). Many of these sites bind diverse metal ions with similar affinities and the identification of metal binding sites is not easily predictable from standard sequence analysis alone (Andreini *et al.,* 2006; Andreini *et al.,* 2007; Katayama *et al*., 2002; Bertini and Rosato, 2007).

The metalloprotein database MDB (http:// metallo.scripps.edu/) is an inventory of metal-binding patterns (MBP). The protein sequences (CDS translations) of *C. metallidurans* CH34 were investigated for the occurrence of metal-binding pattern (MBP) as stated at MDB. The mining of regular expression patterns along the protein sequence was combined with an inspection of the conservation of amino acids near the putative metal binding sites (Andreini *et al*., 2004). The protein sequences of known metal binding proteins at MDB have been blasted (BLASTP) against the proteomic sequences of *C. metallidurans* CH34. The similarity of proteins to GENBANK, PROSITE (Falquet *et al*., 2002) and SWISSPROT entries and the functional patterns databases PROMISE and PFAM have also been evaluated (Andreini *et al.*, 2006; Andreini *et al.*, 2007; Bertini and Rosato, 2007). The published results of experimental investigations (e.g. Andersen *et al*., 2002; Battistuzzi *et al.* 1996; Eitinger *et al*., 1997) have also been used. The results have been used to supplement the transport protein classification of *C. metallidurans* CH34.

3.5.4. The protein structure of CueO

Charged residues may also contribute to the recognition and binding of siderophores to multicopper oxidases (MCOs) (Cao and Klebba, 2002). The multicopper oxidase (MCO) CueO of *E. coli* is an important player in copper homeostasis and it can probably oxidize Cu(I) to the less toxic Cu(II) (Outten *et al*., 2001b; Grass and Rensing**,** 2001; Grass *et al*., 2004). CueO is probably also involved in iron homeostasis, especially in the oxidative release of iron from the catechol iron siderophore enterobactin (Roberts *et al*., 2002; Kim *et al*., 2001; Sakurai and Kataoka, 2007). Enterobactin (Entercheline) (figure 23) is a cyclic trimer of 2,3 Dihydroxybenzoylserine and it is one of the most effective ferric iron chelating compounds (O'Brien and Gibson 1970; Pollack and Neilands, 1970 Raymond *et al*., 2003).

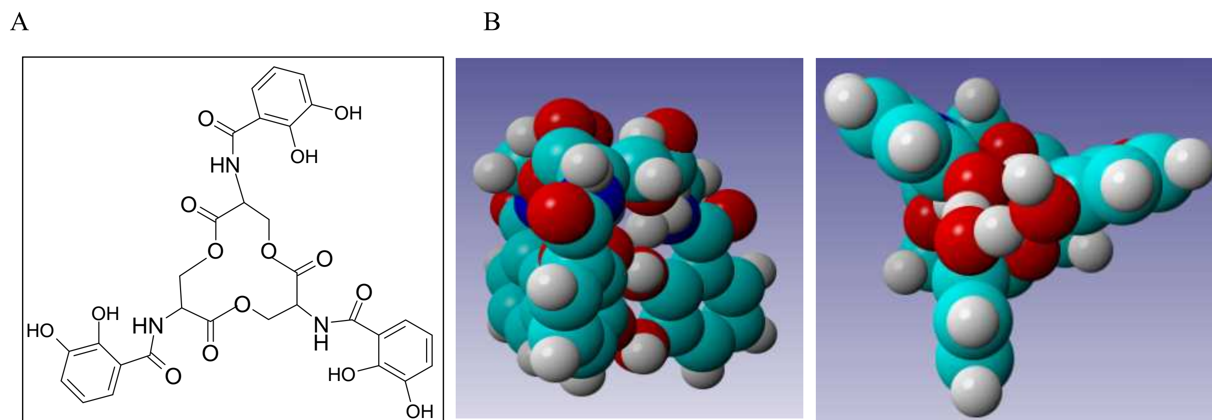


Fig. 23: Structure model of enterobactin (corresponds to Raymond *et al.,* 2003)
A chemical model and B structure model for enterobactin (side view and view from top). The chemical structure was created with the MDL IsisDraw program; the picture for the structure model was generated with YASARA

Initial attempts were carried out to fit enterobactin into identified cavities of the structure model of CueO. The putative substrate-binding site is probably located near the N - terminal - methionine-rich region. A shallow depression at the protein surface might allow substrates to approach the T1 copper site of this protein (Roberts *et al.*, 2002; Roberts *et al.*, 2003; Kataoka *et al.*, 2007). The protein structure model for CueO (PDB 1KV7) has been obtained from the Brookhaven protein databank. The structure of the catecholate enterobactin was build according to its chemical structure with the MDL IsisDraw (http://www.mdli.com) as shown in figure 23. Investigations of the protein structure model of CueO were carried out with PE, SPDBV and the YASARA programs (Krieger *et al.*, 2002) (figure 24). The cavities near the T1 copper site of CueO have been calculated with CASTP (Edelsbrunner 98). The size of the cavities and the distribution and orientation of aromatic residues have been determined and enterobactin as the probable substrate was fit into these regions (figure 24). High-affinity binding sites for Fe^{3+} siderophores are often lined with aromatic residues e.g. in the outer membrane iron transporters FepA and FhuA (Braun and Braun, 2002). The distribution of aromatic residues in CueO might support the notion of a probable enterobactin binding. The size, charge, hydrophobicity and hydrogen bonding may however also contribute to the recognition and binding of colicins (Cao and Klebba, 2002) and further investigations will be necessary.

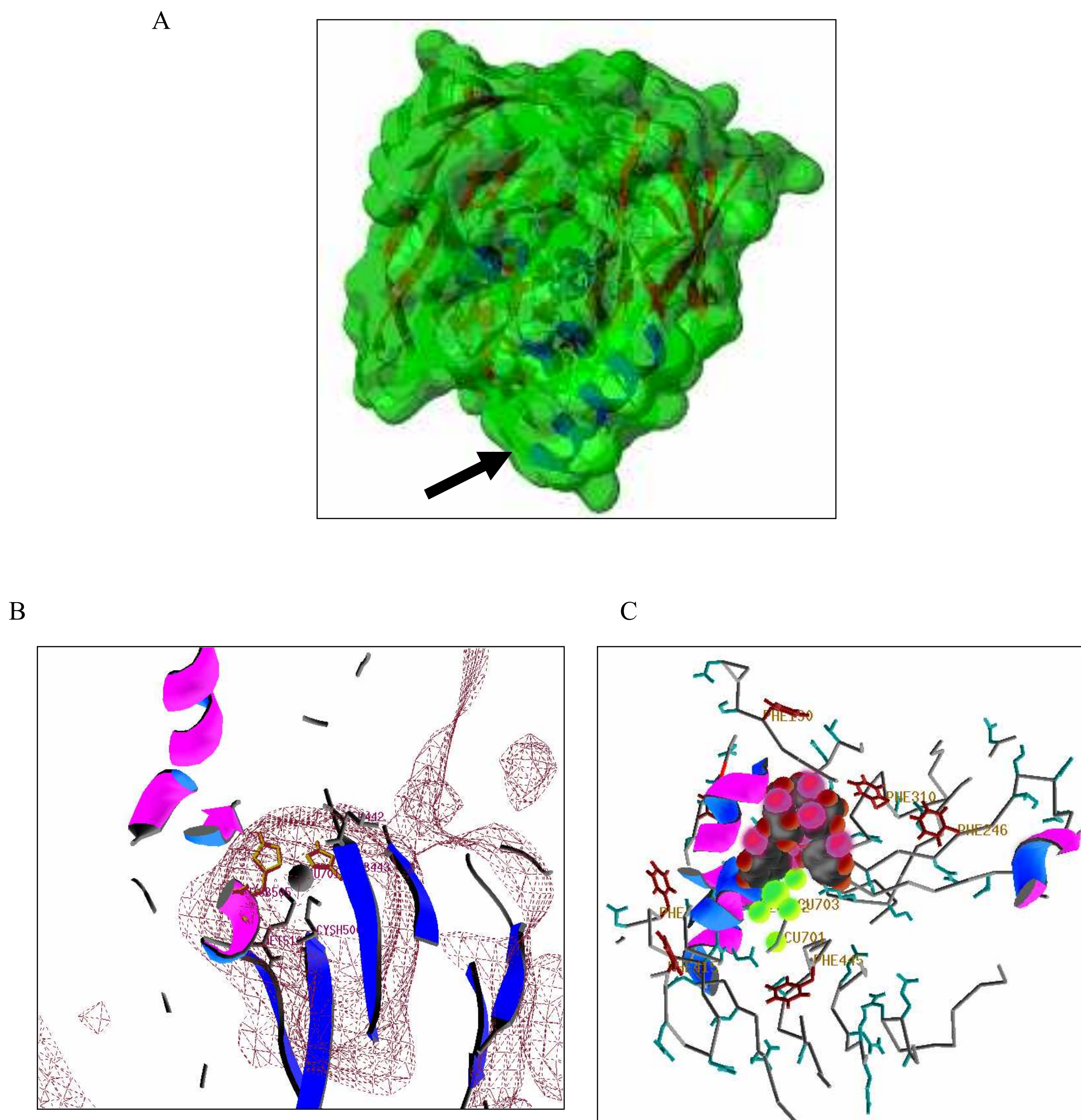


Fig. 24: Structure model of the MCO CueO in *E. coli*

A The figure shows a ribbon diagram of the 53.4 kDa large periplasmic MCO protein CueO. It exhibits three domains with a cupredoxin like fold (eight stranded β-barrel fold). The copper1 site (arrow) is positioned at the bottom of a widely opened substrate binding cleft containing a motif of several charged residues (His443, Cys500, His505 and Met510). Among these, Asp has shown not to be involved in the binding of phenolic substrates. (created with the YASARA program)

B Attempt to fit of enterobactin into the type1 copper site of CueO. The internal cavities near the type1 copper site of CueO have been calculated with CASTP (picture B) and examined and visualized with SPDBV

C The enterobactin structure model was fit into this region. The green dots represent the type1 copper site (created with SPDBV and PE)

4. DISCUSSION

4.1. Genomic changes can effect specific adaptations

Within the past decade has the number of available sequenced genomes nearly doubled every year (Overbeek, 2000). About half of all proteins of newly sequenced bacteria can usually be functionally classified with computational methods (Fraser *et al.,* 2002; Goesmann *et al.*, 2003). These methods rely on the bioinformatic proof of sequence similarity (proof of orthology) to proteins with experimentally verified cellular functions (Koonin *et al.*, 1996; Tatusov *et al.,* 1997, Dehal and Boore, 2006; Apatoff *et al.*, 2008). When the distribution of orthologs in different organisms does not follow the general pattern of evolutionary divergence can this be an indication for horizontal gene transfer (Dröge *et al.*, 1998; Ragan, 2001).

Bacteria can not only acquire foreign DNA from an exogenous source, but they can also shuffle their genomes to produce novel arrangements of genes (Ochman and Moran, 2001; Top and Springael, 2003). New functions or differential gene expression can also be achieved by gene duplication (Gogarten and Olendzenski, 1999). Gene duplication and gene loss are up to three times more frequent than horizontal gene transfer and the half of all *E. coli* proteins have probably evolved through gene duplication (Koonin *et al.*, 1995; Labedan and Riley**,** 1999; Moran, 2007; Kunin and Ouzounis**,** 2003). Since compactness is a governing principle of the bacterial genome evolution, have only those diverged genes duplicates persisted, that confer a selective advantage to these cells and a high genomic flexibility is of special importance for bacteria that invade problematic new environmental niches (Fitch and Ayala, 1994; van der Lelie *et al.*, 2005; Salanoubat *et al.*, 2002). The identification of evolutionary changes in the proteome can thus shed light into lineage specific adaptations and capabilities of bacteria (Metzgar and Wills, 2000; Salanoubat *et al.,* 2002; Koonin, 2005).

4.2. The proteomic repertoires of proteobacteria

The major habitat of *Burkholderiacea* is soil, but proteobacteria of this group can also often be found in the rhizosphere of plants (Tabacchioni *et al*., 2002; Berg *et al*., 2005; Payne *et al.,* 2006; Caballero-Mellado *et al.*, 2007; Alisi *et al.* 2005). The bacterial strains in this habitat and those in biofilms have often a higher conjugative gene transfer frequency than those in other environments and an overwhelming high fraction of genes in *Burkholderiacea* has been acquired from external sources (van Elsas 1992; Dröge *et al.,* 1998, 1999; Hausner and Wuertz 1999; Thomas and Nielsen; 2005). This was achieved by mobile genetic elements as plasmids, transposons, insertion (IS) elements and

genomic islands (Bennett, 2004; Hall, 1997; Overbeek *et al.*, 2003). The high genomic flexibility of *Burkholderiacea* has resulted in a broad diversity of nutritional pathways and large sets of uptake transporters (Coenye *et al.*, 1999, Salanoubat *et al.*, 2002; Genin and Boucher. 2004). Similarly is an unusual spontaneous genomic variability probably also responsible for their high saprophytic fitness and pathogenic flexibility (Fitch and Ayala, 1994; Metzgar and Wills, 2000; McCutcheon and Moran, 2007; Moxon *et al.*, 1994). One of the most important plant pathogens is *R. solanacearum* (Grover *et al.*, 2006). *R. solanacearum* is a soil-borne pathogen that causes lethal wilting diseases in more than 200 plant species (Genin and Boucher, 2002, 2004; Cunnac *et al.*, 2004). Its extraordinary nutritional versatility has enabled this γ-proteobacterium to switch between alternative, extremely different ecological niches (Valls *et al.*, 2006). The replicons of *R. solanacearum* GMI1000 (the 3.7-Mb chromosome and the 2.1-Mb megaplasmid) have a mosaic like structure and they contain many genetically mobile elements. Lateral gene transfer has thus probably been of special importance for this organism (Salanoubat *et al.,* 2002; Coenye and Vandamme, 2003; Grover *et al.*, 2006).

4.3. The proteomic repertoire of *C. metallidurans* CH34

The β-proteobacterium *C. metallidurans* CH34 is also a good recipient of foreign genes (Lejeune *et al.*, 1983). Its large genomic island (PAGI-2) is conserved at 100% in rather distantly related strains of *P. aeruginosa* (Larbig, *et al.* 2002). Two mercury transposons Tn4378 (at plasmid pMOL28) and Tn4380 (at plasmid pMOL30) can be found and major genes for heavy metal resistance in pMOL28 are grouped in a region, which is flanked by insertion elements (IS*1071*) (Diels *et al.,* 1985) (Monchy *et al.*, 2007). Its high genomic flexibility might have facilitated its extraordinary ability to withstand heavy metal ion concentrations that are lethal to other bacteria (Mergeay *et al.*, 1985; Taghavi *et al.* 1997b). *C. metallidurans* CH34 gains thus increasing interest as model organism for heavy metal homeostasis (Nies, 2003). The recent closure of its genome has now allowed evaluating the responsible determinants for their emergence and acquisition. Gene models, pre-computed sequence similarity relationships, functional annotation and pathway information for genetic determinants of *C. metallidurans* CH34 have been assembled and stored to support these investigations. The phylogenetic profiles, hydropacity profiles and a representation of the genomic neighbourhood of genes have also been included. This information and the protein- and nucleotide sequences were integrated into a local database which facilitates further computational investigations.

4.4. The plasmids of proteobacteria are hotspots for new unique functions

Unusual living conditions are often concomitated with significant changes of proteomic repertoires (Ren and Paulsen, 2005). The overall genomic repertoire of *C. metallidurans* CH34 has been compared to those of some other closely related proteobacteria to shed light into these changes. This

includes the facultative hydrogen-oxidizing strain *C. eutrophus* H16, the xenobiotic degrader *C. eutrophus* JMP134, the plant-pathogenic strain *R. solanacearum* GMI1000*, B. xenovorans* LB400 and the pathogen *B. cepacia* AMMD.

The genome of *B. xenovorans* LB400 is one of the largest bacterial genomes being sequenced so far and this organism exhibits a broad metabolic versatility. It can degrade isoflavonoids, diterpenoids, sulfonates and biphenyl (PCB) compounds (Denef *et al.,* 2004; Denef *et al.,* 2006; Chain *et al.,* 2006; Bains and Boulanger**,** 2007).

B. cepacia AMMD is an opportunistic human pathogen. The *Burkholderia cepacia* complex (Bcc) is comprised of at least ten closely related species which are implicated in the occurrence of cystic fibrosis (Holmes *et al.,* 1998; Davies and Rubin, 2007; Thomas, 2007). Contamination with *B. cepacia* is also the major reason for recalls of pharmaceutical products (Jimenez, 2007).

A characteristic feature of the two strains of *C. eutrophus* is their ability to degrade recalcant aromatic compounds. *C. necator* JMP134 can grow on phenol as sole energy- and carbon source and it can produce polyhydroxyalkalonates as carbon and energy storage molecules (Pieper *et al.*, 1989; Uchino, *et al*. 2008; Steinbüchel, 1996).

C. metallidurans CH34, *C. eutrophus* JMP134 and *C. eutrophus* H16 each contain two circular chromosomes and *Burkholderia cepacia* exhibits even three chromosomes (Rodley *et al.,* 1995; Lessie *et al.* 1996; Parke and Gurian-Sherman 2001). Conserved housekeeping proteins (e.g. those for general metabolic pathways) are often encoded at the chromosome and they evolve usually rather slowly (Neidhardt *et al*., 1987; Koonin *et al.,* 1995). The additional 'lifestyle chromosomes' encode often largely proteins for a more specialized cellular function (Pohlmann *et al.,* 2006).

We have found an orthologous counterpart for the fast majority (76%) of all chromosomal encoded proteins of *C. metallidurans* CH34. Orthologs arise in general from a common ancestral gene and they usually keep at least a similar function (Tatusov *et al.;* 1997; Jothi *et al.*, 2006).

The smaller replicons encode fewer orthologs and the orthologs which are encoded at 'lifestyle chromosomes' were also often not as widely distributed among the other herein investigated *Burkholderiacea.* One quarter (26%) of all proteins in *C. metallidurans* CH34 fall into this group. Some 'rare orthologs' were only found in more distantly related strains. Not only strain specific gene duplications but also horizontal gene transfer and genome reduction have thus probably important for the evolution of these strains.

The plasmid-encoded proteins of the investigated proteobacteria exhibit an even higher genomic flexibility and specialisation. Plasmids have often a higher sequence evolution rates, they are prone to genomic rearrangements and they are thus often more ''plastic'' than larger replicons (Chain *et al*., 2006). The plasmids pMOL28 and pMOL30 of *C. metallidurans* CH34 encode many proteins for conjugative transfer and they exhibit several mobile genetic elements (Monchy *et al*., 2007). They have also a lower G+C contents (60.4 % for pMOL28; 60.1% for pMOL30) than the chromosome and

megaplasmid (on average 63.0%) and a common origin has been suggested for the plasmid pMOL28, plasmid pHG1 of *C. eutrophus* H16 and plasmid pSym of *R. taiwanensis* (Monchy *et al.*, 2007).

The comparison to other Burkholderiacea in this work has revealed, that nearly half (41%) of all pMOL30 encoded and also one quarter (23%) of all pMOL28 encoded proteins in *C. metallidurans* CH34 have recently been duplicated (paralogs). Not only in *C. metallidurans* CH34 does the ratio of plasmid encoded paralogs largely exceed the ratios of chromosomal encoded paralogs. This trend was also observed in related strains of the *Cupriavidus* cluster (*C. eutrophus* H16 and *C. eutrophus* JMP134). It has already been reported that plasmids are often responsible for the divergence of closely related bacterial strains (Frost *et al.*, 2005). The plasmids of these proteobacteria might probably be special hotspots for the development of specialized unique capabilities since such new functions are often provided by paralogs (Jothi *et al.*, 2006). Particularly those determinants that defend cells against viruses, toxins or varying environmental conditions evolve often more rapidly (Murphy 1993; Moxon and Thaler 1997; Jordan *et al.* 2002; Zheng *et al.*, 2004).

The most important unique feature of *C. metallidurans* CH34 is its extraordinary metal ion resistance and this special capability is predominantly achieved by transmembrane transport proteins (Nies, 1999; Nies, 2003). Further investigations were thus directed towards the genomic flexibility of these proteins.

4.5. The transport protein repertoires of proteobacteria

In eubacteria are usually approximately ten percent of all cellular proteins involved in the translocation of solutes across the membranes (Paulsen *et al.*, 1998). These transport proteins are not only responsible for nutrient uptake, but they can also facilitate the efflux of waste products, drugs and heavy metal ions (Ren and Paulsen, 2005). Some of these transport protein families arose gradually over evolutionary time, whereas others developed suddenly, due to bursts of intragenic or intergenic (or both) duplication events (Saier; 1996).

The transporter protein classification system (TC) has been adopted by IUBMB as the international standard for the classification of transport protein families (Zhou *et al.,* 2003*;* Das *et al.,* 2007*;* Saier *et al.,* 2006). The classification of transmembrane transport proteins is based on mode of transport, bioenergetics, topology, molecular phylogeny and substrate specifities (Ren and Paulsen, 2007). Conserved features of these and other transport protein families are described in the biological databases TRANSPORTDB (http://www.membranetransport.org/) and TCDB (at http://www.tcdb.org/) (Saier 2000).

The potential transport protein repertoires of *C. metallidurans* CH34 has been determined and compared with those of *C. eutrophus* H16, *C. eutrophus* JMP134, *R. solanacearum* GMI1000, *R. solanacearum* UW551, *B. xenovorans* LB400 and *B. cepacia* AMMD. Several transport protein families are almost equally distributed among all the investigated strains and about two third of the

transport proteins in the *Cupriavidus* cluster are orthologs. In *C. metallidurans* CH34 corresponds this to a housekeeping transporter repertoire of about 700 proteins. About 450 transporters have orthologs in all the investigated strains. Many uptake systems for nutrients and essential cations, plant pathogenesis factors and also certain heavy metal efflux systems fall into this group. The common extraordinary diversity of these uptake transporter families may supply the versatile nutritional pathways of these soil borne organisms (Salanoubat *et al.,* 2002; Genin and Boucher; 2006).

Many of these transporters can catalyze the thermodynamically unfavourable translocation of substrates against a concentration gradient through coupling to a second, energetically favourable process (Pimenta *et al*., 1996). The most commonly used energy sources to drive transport across the cytoplasmic membrane are chemical energy (ATP, PEP) and chemiosmotic energy (sodium or proton motive force) and a third group couples the transport to a covalent modification of their substrates (Maloney, 1990; Maloney, 1992; Paulsen *et al*., 2000). Accordingly differentiates the TC-system primary transporters e.g. ABC transporters (TC 3.A.1) (Saurin *et al*., 1999); secondary carriers (e.g. MFS (TC 2.A.1) - Pao *et al*., 1998) and group translocators (e.g. PTS (TC 4.A.1) – Reizer *at al*., 1991) (Ren *et al.,* 2004).

The distribution of these transporter types can be affected by the primary forms of energy generated, ecological niche and substrate availability and reflects thus the lifestyle and physiology of organisms (Paulsen *et al*., 1998; Ren and Paulsen*,* 2005). Many bacteria in nutrient rich environments have often an excess of pmf-driven permeases. (*E. coli* has 2.5-fold more pmf-driven permeases) but many free living bacteria (e.g. *Synechocystis*) in nutrient poor environments have often more primary transporters (Paulsen *et al*., 1998). More primary transporters than secondary carriers have also been found in *C. metallidurans* CH34 and in the other investigated proteobacteria.

Their low ratios of PTS systems (Lengeler 1990) and an unusual preponderance for di-and tricarboxylate transporters might suggests that all these strains have largely shifted their diet from sugars to amino acids. Many paralogous transporters in *C. metallidurans* CH34 are plasmid encoded and the plasmids of the other investigated strains exhibit also an unusual genomic flexibility. The plasmids of the investigated strains have however probably taken different evolutionary routes since they exhibit a very unequal ratio of certain transport protein families. The megaplasmid pHG1 (0.44 Mb) of *C. eutrophus* H16 contains no genetic determinants for heavy metal resistance (Mergeay *et al.,* 2003) and it has probably sacrificed genes for transition metal ion transport systems to make room for genes, which support an efficient chemolithoautotrophic lifestyle. In contrast to this has *C. metallidurans* CH34 an unusual variety of plasmid encoded transition metal ion transporters and this marks probably a specific adaptation to heavy metal burdened environments.

4.6. Mechanisms of metal ion homeostasis

The active accumulation of essential metal ions allows bacterial cells to potentially provide one third of all proteins with zinc or iron (Nies 2003; Finney and O'Halloran**,** *et al.,* 2003). In contrast to this are usually only a hundred thousand atoms of cobalt and nickel essential for the survival of bacteria and free cellular copper occurs probably only at zeptomolar levels (Changela *et al.,* 2003).
The cellular balance between metal ion overload and starvation in Gram-negative bacteria is achieved by the translocation of metal ions across the inner membrane, the outer membrane and the periplasmic space (Nies, 2004b). Low molecular weight compounds e.g. glutathione, organic acids, metallothioneins or polyphosphate can also facilitate the control of the cellular metal ion pool (Polatajko *et al.,* 2005; Oven *et al.*, 2002). Major strategies for heavy metal ion homeostasis are shown in figure 25.

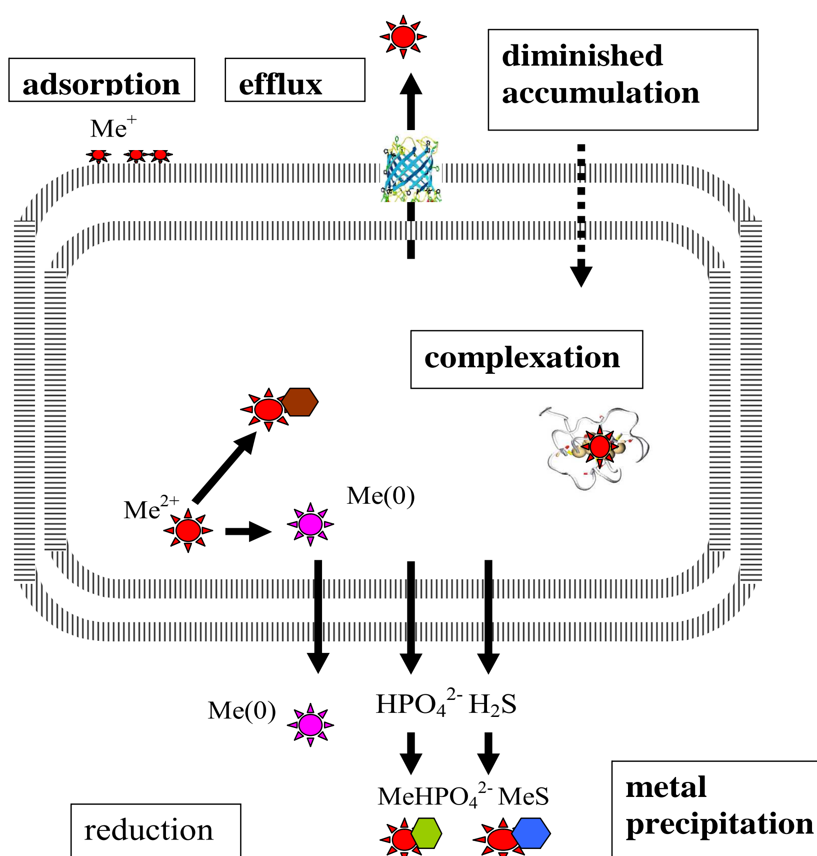


Fig. 25: Bacterial strategies for metal ion homeostasis (Valls and de Lorenzo, 2002 simplified)
Mechanisms of resistance to heavy metal ions include intra- or extracellular binding (immobilisation) of the metal ion with a cognate protein (frequently a metallothionein) or a matching anion. Another mechanism is biotransformation of the toxic ions into a less noxious or more volatile form and the dissimilatory reduction of the metal ions. The most important cytoplasmatic buffers for metal ions are low molecular weight components and cellular proteins. The entry of Pb^{2+}, Cd^{2+}, Zn^{2+} and Co^{2+} at increased pH can be reduced by the complexation with hydroxide or carbonates.

The most pathogenic bacteria need additional specialized mechanisms for the acquisition of iron because iron limitation is a common strategy by which host organisms suppress bacterial growth (Andrews, *et al*., 2003). This also regards the infection of human lungs as an iron-restricted

environment by the *Burkholderia cepacia* complex (Bcc) (Wang *et al*., 1996; Thomas 2007). Many bacteria synthesize siderophores as high-affinity iron sequestration molecules (Brown and Holden**,** 2002; Tauch *et al*., 2005). In *E. coli* is the oxidative release of iron from this siderophores in the periplasm probably carried out by the multicopper oxidase CueO (YacK) (Grass and Rensing, 2001) and a protein with similar function might be encoded by pMol30 of *C. metallidurans* CH34 (Roberts *et al.*, 2002; Monchy *et al*., 2006). The by products of this catechol oxidation (e.g. 2-carboxy-muconate) are probably involved in copper sequestration and CueO also oxidizes Cu (I) to the less toxic Cu (II) which confers copper tolerance at moderate copper concentrations (Grass, 2006; Singh *et al*., 2004; Outten *et al.,* 2001; Roberts *et al.*, 2002). The crystal structure of CueO (PDB-ID 1KV7) is available (Roberts *et al*., 2002) and has been used for initial structural investigations.

At excess concentrations are all divalent heavy metal cations potentially toxic (Nies 2004b). Glutathione is the major cellular thiol in cyanobacteria and proteobacteria, but it has only limited Cd^{2+},Ni^{2+},Pb^{2+} and Zn^{2+} binding capabilities and it is thus e.g. not involved in cadmium resistance of *C. metallidurans* CH34 (Fahey *et al*., 1978; Fahey, 2001; Satofuka *et al*. 2001; Helbig *et al*., 2008b;). Complexation is unfavourable from an energetical standpoint and binding of heavy metal ions to glutathione can cause oxidative stress (Nies 1999; Kachur *et al.,* 1998*;* Nies 1999). Several CPX-ATPases in *C. metallidurans* CH34 export glutathione-metal-complexes and they combine thus complexation and efflux (Anton 2001; Rensing *et al.,* 1999). Chromate is detoxified by a combination of reduction and efflux (Nies 2003) and all known bacterial mercury resistance mechanisms are also based on enzymatic reduction (Summers, A. O. 1988; Taghavi *et al.,* 1997b). Arsenate is reduced to arsenite to differentiate it from phosphate and to allow its selective transport out of the cell (Mukhopadhyay *et al*., 2002, Silver and Keach, 1982). A reduction is however only possible for certain heavy metal ions (Nies 1999).

The major mechanism for the homeostasis to transition metal ions in bacteria is transmembrane transport (Nies, 1999; Nies, 2003). Transport proteins for metal ions contain usually specific conserved charged (e.g. His, Glu, Asp) transmembrane residues (Argüello *et al.,* 2007). Pattern- , annotation-, and domain- based investigations (Andreini *et al.,* 2006) were carried out to identify potential metal binding sites in proteins of *C. metallidurans* CH34 to support further investigations of metal ion specific transporters with vaguely understood functions. The metal coordination by these charged groups can also reflect the substrate specifities and the function of metal binding proteins (Shi *et al*., 2005; Andreini *et al*., 2007b). Variations of these and other signature sequences in metal ion specific transport protein families have thus been used to support the assignment of these proteins to subgroups of metal ion specific transporter families (e.g. HME-RND transporters).

4.7. Transporter families for metal ion homeostasis

ATP-dependent nutrient transporters have often higher affinities for their substrates than pmf-driven systems, but the tight binding costs both time and energy (Paulsen *et al.*, 1998; Nies, 1999). For the transport across the cytoplasmic membrane have most bacteria thus constitutively expressed fast transporters with low substrate specifities and these are supplemented by the expression of high affinity transporters under ion deficiency (Grotz *et al.*, 1998; Guerinot, 2000; Hantke 2005). Figure 26 shows important uptake and efflux transporters of Gram-negative bacteria

The efflux of metal ions to the periplasmic space as the first step for the extrusion in Gram-negative cells is usually achieved by CPX P-type ATPases, Cation Diffusion Facilitators (CDF –TC 2.A.4) (Nies *et al.*, 1995) and by transporters of the Major Facilitator Superfamily (MFS) (Nies 2003). HME (RND) (TC 2.A.6.1) proteins and other components of CBA-type efflux systems (Nies 2003) are probably responsible for the efflux from the periplasmic space to the outside (Grass 2006).

The MerTP (TC 9.A.2), MFP (TC 8.A.1), MIT (TC 1.A.35), NiCoT (TC 2.A.52), OMF (1.B.17), OMR (1.B.14), CHR (2.A.51), and ZIP (2.A.5) transporter families are also important for metal ion homeostasis (Locher *et al.,* 1998; Dong and Mergeay, 1994,; Smith and Maguire, 1998; Saier, 2000; Nies *et al.*, 1995; Grotz *et al.*, 1998; Dinh *et al.*, 1994; Nies and Brown, 1998; Eitinger and Friedrich, 1991; Saurin *et al.*, 1999; Solioz and Vulpe, 1996; Dinh *et al.*, 1994; Qian *et al.*, 1998; Kammler *et al.*, 1993; Cao, *et al.*, 2007; Guerinot, 2000).

Important P-type-ATPases (TC 3.A.3) (Solioz and Vulpe, 1996) are MgtA and MgtB (Snavely *et al.,* 1989). Other important transporters are the NiCoT (2.A.52) (Eitinger and Friedrich, 1991) transporter HoxN (Eitinger *et al.,* 1997), the ZIP transporter ZupT (Grass and Rensing, 2001; Grass *et al.*, 2005), the NRAMP (2.A.55) (Cellier *et al.,* 1995) transporter MntH and probably the inorganic phosphate transporter family PiT (2.A.20) (Harris *et al.*, 2001) (Van Veen *et al.,* 1994, Beard *et al.*, 2000).

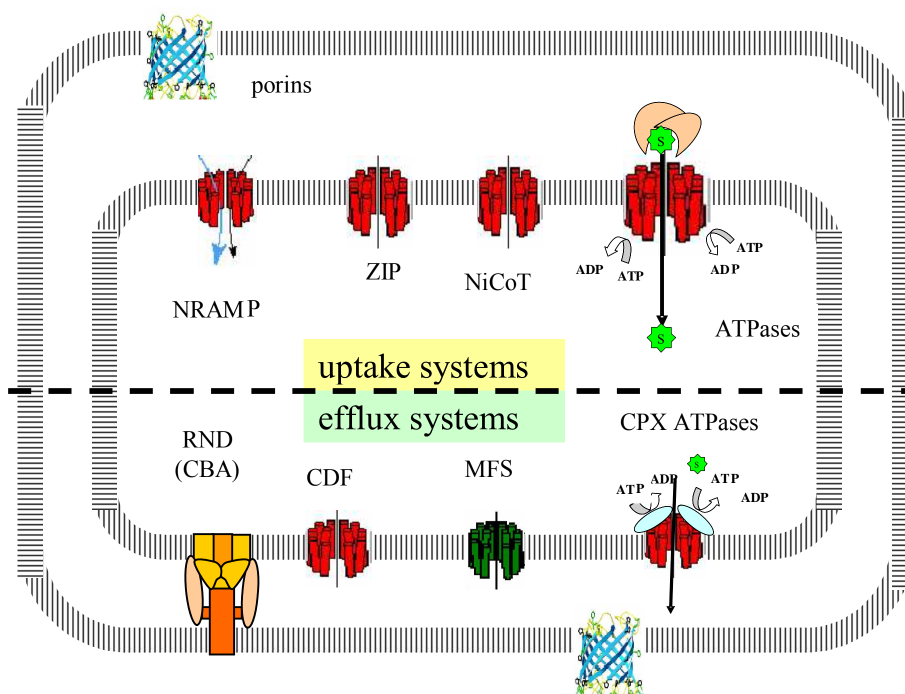


Fig. 26: Uptake and efflux transporters of Gram-negative bacteria
The figure shows major bacterial transport protein families for the translocation of metal ions (CPM cytoplasmic membrane, ABC ATP-binding cassette, RND resistance, nodulation, cell division, CHR chromate transport, MIT metal inorganic transport, CDF cation-diffusion facilitators).

Due to their common physicochemical properties are divalent cations difficult to be distinguished by metal ion uptake transporters and these transporters are often an open gate for many potentially toxic metal ions (Smith and Maguire, 1998, Nies 1999; Wackett *et al.*, 2004). The primary Mg^{2+} uptake transporter CorA (TC 1.A.35) (Lunin *et al.*, 2006) thus allows the entry of harmful $Ni^{2+,}$ and Co^{2+} ions (Smith and Maguire, 1998; Nies 1991; Nies 1999; Snavely *et al.*, 1989). Efflux transporters are probably the most important defence against excess concentrations of these and other heavy metal ions (Grass 2006).

4.8. Metal ion transport proteins in *C. metallidurans* CH34 and related bacteria

The potential transition metal ion transport protein repertoire of *C. metallidurans* CH34 has been compared to those of *C. eutrophus* H16, *C. eutrophus* JMP134, *R. solanacearum* GMI1000, *R. solanacearum* UW551, *B. xenovorans* LB400 and *B. cepacia* AMMD.
In *C. metallidurans* CH34 are about two thirds of its plasmid encoded transport proteins of pMOL30 and half of these of pMOL28 probably responsible for metal ion homeostasis. This corresponds to about three quarters of all its ion specific transporters and marks a strong prevalence for transition metal ion transporters. The plasmids of *C. eutrophus* H16 and *C. eutrophus* JMP134 encode much

fewer transition metal specific transporters and may largely be directed to the development of metabolic capabilities.

The distribution of the most metal ion transport protein families is similar in the investigated proteobacteria, but certain metal ion transporter families are unique to only some of these bacteria.

The NiCoT-transporter (2.A.52) (Eitinger and Friedrich, 1991) of *C. eutrophus* H16 facilitates a specific Ni^{2+} (and Co^{2+}) uptake that supports e.g. the biosynthesis of nickel containing hydrogenases (Hebbeln and Eitinger, 2004; Eitinger *et al*., 2005). *C. metallidurans* CH34 contains no NiCoT and no NRAMP transporter for uptake of divalent transition metal cations.

CPX-type ATPases are the basic defence against heavy metal ions and this subgroup of P-Type ATPases is ubiquously distributed among bacteria (Solioz and Vulpe, 1996; Nies, 2003). The copper- and silver specific P-type ATPases (Rensing *et al*., 1999; Gatti *et al*., 2000) in *C. metallidurans* CH34, as well as those, which are specific for potassium, are widely distributed among the herein investigated proteobacteria. Three of the four probably zinc/cadmium/lead specific P-type ATPases (CadA, PbrA CzcP) of *C. metallidurans* CH34 (Anton, 2001; Borremans *et al.,* 2001; Legatzki, 2003) are however probably the product of a recent multiplication of a common ancestor. The calcium specific P-type ATPases of *C. metallidurans* CH34 have also recently been duplicated (recent paralogs). CDF transporters may also be involved in cation uptake (Bloss *et al*., 2002, Nies, 2007). The development of CDF-proteins was probably an early evolutionary trait (Nies, 2003) and for the most CDF transporters of *C. metallidurans* CH34 has thus an orthologous counterpart been found among the investigated strains. The most proteobacteria contain two CDF proteins – one being specific for zinc, cobalt and cadmium and one for other cations such as iron (Nies; 2003). In *C. metallidurans* CH34 have three representatives of the CDF transport protein family been found. FieF (Grass *et al*., 2005b) and DmeF (Munkelt *et al*., 2004) are probably involved in iron, zinc and cobalt homeostasis (Munkelt *et al*., 2004, Grass, 2006) and the CDF-transporter CzcD (Nies and Silver, 1995; Anton *et al.*, 1999) is involved in zinc, cobalt and cadmium resistance (Nies *et al.,* 1987). CzcD probably also triggers the transcription of the CBA-type efflux complex CzcCBA (Nies 1992; Anton 2001, van der Lelie *et al*., 1997). All these proteins are old paralogs and the common ancestor of the *Burkholderia* / *R. solanacearum* cluster has thus probably already possessed a complete set of CDF transporters for $Fe^{2+}/Co^{2+}/Ni^{2+}/Zn^{2+}/Cd^{2+}$ homeostasis. CnrT is part of the *cnr* cluster and may have a similar significance for nickel homeostasis as does the CDF protein CzcD for cobalt, zinc and cadmium (Nies, 2003). CnrT-like nickel transport systems usually occur in only a few species since nickel is used seldom as a cofactor (Nies, 1999). And these transporters have only been found in *C. metallidurans* CH34 and two other bacteria.

The *chr* clusters of *C. metallidurans* CH34 probably confer resistance to chromate at different toxicity levels (Juhnke *et al*., 2002; Mergeay *et al.,* 2003). The four CHR transporters of *C. metallidurans* CH34 have all an orthologous counterpart in the other investigated proteobacteria and they are old paralogs. Thus was chromate efflux via these systems probably already present in the common

ancestor of these bacteria. Nearly the half of the MerT (9.A.2) and MerP proteins in *C. metallidurans* CH34, are paralogs. The MerT and MerP proteins are involved in the uptake of Hg^{2+} into bacterial cells in preparation for its reduction by the MerA mercuric reductase. The Hg^{o} produced by MerA is volatile and passively diffuses out of the cell (Brown *et al*., 2002b). Two MerT proteins have probably been duplicated into four transporters which yielded three active (Dressler *et al.,* 1991) systems for mercury homeostasis in *C. metallidurans* CH34. Many potassium uptake permeases of the KUP family (2.A.72) have also recently been duplicated (paralogs).

4.9. The HME RND subfamily is overrepresented in *C. metallidurans* CH34

The most RND transporters in proteobacteria belong usually to the HAE-subfamilies (HAE1 to HAE3) (TC 2.A.6.2) of RND transporters (Nies, 2003). The HAE1 subfamily (2.A.6.2) can translocate multiple drugs (e.g. tetracycline, chloramphenicol, fluoroquinolones, β-lactams, etc.). The HME subfamilies of RND (2.A.6.1) transporters are implicated in heavy metal (Co^{2+}, Zn^{2+}, Cd^{2+}, Ni^{2+}, Cu^{+} and Ag^{+}) homeostasis (Tseng *et al.*, 1999).

The most bacteria possess none or only one or two HME-RND-proteins and those are mostly putative copper/silver exporter of the HME4 group (Nies; 2003). *C. metallidurans* CH34 is however over equipped with HME-RND transporters. This not only regards the HME4 group, but also the other groups of heavy metal specific HME-RND transporters. Extensive gene duplications have occurred after the divergence of the *Cupriavidus* cluster from other *Burkholderiacea* and the HME4 and HME1 proteins were doubled and the HME3 proteins were triplicated (Nies; 2003). Even in the *Cupriavidus* cluster have only half of *C. metallidurans* CH34 encoded HME-proteins an orthologous counterpart. The other half has thus been developed during the specialization of *C. metallidurans* CH34. However is only one of these HME RND transport proteins constitutively expressed, another HME protein is induced by heavy metal ions while the remaining systems are silent (Nies, 2003; Nies *et al.*, 2006). They probably serve as surplus material for evolutionary genomic rearrangements or are backup systems (Nies *et al.,* 2006).

4.10. The composition of HME-CBA-type exporter clusters

Gene proximity is usually the result of selective pressure to associate genes that are co-regulated and possibly interacting (Dandekar *et al*., 1998; Tamames *et al.,* 1997; Enright *et al.,* 1999; Castillo-Davis and Hartl**,** 2003; Overbeek *et al***.,** 1999). Many metal resistance gene clusters at the plasmids pMOL28 and pMOL30 of *C. metallidurans* CH34 contain equivalents of chromosomal operons, but they are often flanked by additional genes that don't have equivalents in other genomes (e.g. copK, cnrY) (Mergeay *et al*., 2003). In the *czc* cluster (Nies *et al*., 1987) on plasmid pMOL30 and the *cnr*

(Liesegang *et al.*, 1993) operon on pMOL28 of *C. metallidurans* CH34 are genes of HME RND transporters combined with those of other transporters and regulators.

The *czc* and *cnr* clusters encode the transperiplasmatic HME-CBA-type efflux complex CzcCBA. CzcCBA is the first line of defence against excess concentrations of zinc, cobalt and cadmium and it is implicated in transenvelope-transport across both cytoplasmic and periplasmic membranes (Saier *et al.*, 1994; Johnson and Church, 1999; Pimenta *et al.*, 1996). The CzcCBA efflux complex consists of three major structural components. The most important component is the trimeric HME-RND transporter (TC 2.A.6.1) (Saier *et al.*, 1994, Dinh *et al.*, 1994) CzcA in the cytoplasmic membrane (Rensing *et al.* 1997, Anton *et al.,* 1999; Nies; 2003). Its function is supported by the membrane fusion protein (MFP- TC 8.A.1) (Saier *et al.*, 1994) CzcB and the outer membrane facillitator (OMF- TC 1.B.17) (Dong and Mergeay, 1994) CzcC. OMFs might actively accumulate substrates from the RND protein and they thus probably allow a rapid unloading to the outside by diffusion (Koronakis *et al.*, 2000; Rensing *et al.*, 1997). The structural assembly of this complex corresponds probably to the model for the AcrABTolC efflux complex in Fig. 27. The genes for these structural transporter components are surrounded by determinants for regulatory proteins. The *czc* operon and the related *ncc, cnr* and *sil* gene clusters encode also auxiliary transporters with similar substrate specifities. These are the CDF transporter CzcD, the MFS transporter NreB (Grass *et al.*, 2001b), the DMT transporter CnrT (Grass, 2000) and the P-Type ATPases SilP (Gupta *et al*, 1999). Proteins that function together as a pathway or protein complex exhibit often a convergent evolution and they are often inherited or eliminated in a co-dependent manner (Ren and Paulsen, 2005; Poteete *et al.* 1992; Silver and Ji, 1994; Ranea 2007).

All structural components of the HME1-CBA-type efflux complex CzcCBA (Nies, 1992) are old paralogs. Orthologs have been found in the Cupriavidus cluster and among the other investigated proteobacteria. The P-type ATPase CzcP, as well as the periplasmic metal binding protein CzcE (Hoffmann, 2001) are however paralogs. The assembly of the genes for individual transport protein components in the *czc* cluster is unique to this bacterium. The functional components of the pMOL28 encoded HME2-CBA-type system CnrCBA are however paralogs and their corresponding regulatory sigma factor CnrH (Lonetto, *et al.*, 1992) is unique to this organism (Nies, 2004). The components of the CnrCBA efflux pump are thus the product of a recent gene duplication event and they must have been formed after *C. metallidurans* CH34 and *C. eutrophus* JMP134 have diverged in evolution.

All taken together is a complex set of transporters for metal ion homeostasis a common trait in *C. eutrophus* JMP134, *C. eutrophus* H16 and *C. metallidurans* CH34. Several of these heavy metal specific transport families in *C. metallidurans* CH34 exhibit however an unusual genomic flexibility. Not only gene duplication, but also horizontally acquired genes, functional differentiation and the combination of genes into operons have shaped the outstanding heavy metal resistance of this organism.

4.11. The assembly of the CBA-type efflux complex

The protein structure model for a transporter of the HME-subfamily of RND transporters (e.g. CzcA or CnrA) has not yet been determined, but the crystal structures of the HAE-RND drug exporter, AcrB (TC #2.A.6.2.2) for hydrophobes and amphiphiles export in *E. coli* is available at 3.5 Å resolution (Murakami *et al*., 2002). The protein structure models for HAE RND transporters can facilitates the development of general models for the function RND-driven efflux complexes (Nies, 2003) and they have already been used for homology modelling of the homologous HAE-RND transporters EmhB (Pseudomonas *fluorescens* cLP6) and MexB of *Pseudomonas aeruginosa* (Murakami *et al.,* 2002; Hearn *et al*., 2006; Nehme and Poole, 2007). These RND proteins and other multidug exporters are the predominant choice for drug efflux in Gram-negative bacteria (Li *et al*., 1995*;* Okusu *et al*., 1996; Seeger *et al*., 2006; Paulsen *et al*., 1996). They spread rapidly through cell populations and new multidrug transporters can evolve relatively easily by mutations of their drug-binding and extensive gene duplication events were reported for this family (Higgins, 2007; Saier *et al*., 1998).
AcrB functions in conjunction with the OMF (1.B.17) transporter TolC and the MFP (8.A.1) protein AcrA as the transperiplasmic CBA-type efflux complex AcrABTolC- (figure 27) (Ma *et al*., 1993). A flexible coil-coil structure in the MFP (8.A.1.) protein AcrA (TC 8.A.1.6.1) may facilitate a flexible contact between the RND-pump AcrB and the OMF tube TolC (Touze *et al.,* 2004; Higgins *et al*., 2004; Zgurskaya and Nikaido, 1999; Lobedanz *et al*., 2007). A coiled coils region in the outer membrane transporter TolC seals off the channel to the periplasmic space and it is probably opened by the interaction with the RND or MFP components (Sharff *et al*., 2001; Andersen *et al*., 2000, Andersen *et al*., 2001; Andersen *et al*., 2002). Several OMFs (CnrC, CzcC and NccC) are functionally interchangeable (Grass, 2000) and basic structural features are thus probably common to these proteins. An alternate access pumping mechanism with large scale structural changes (outlined in figure 28) was suggested for AcrB and other RND transporters (Murakami *et al*., 2006).

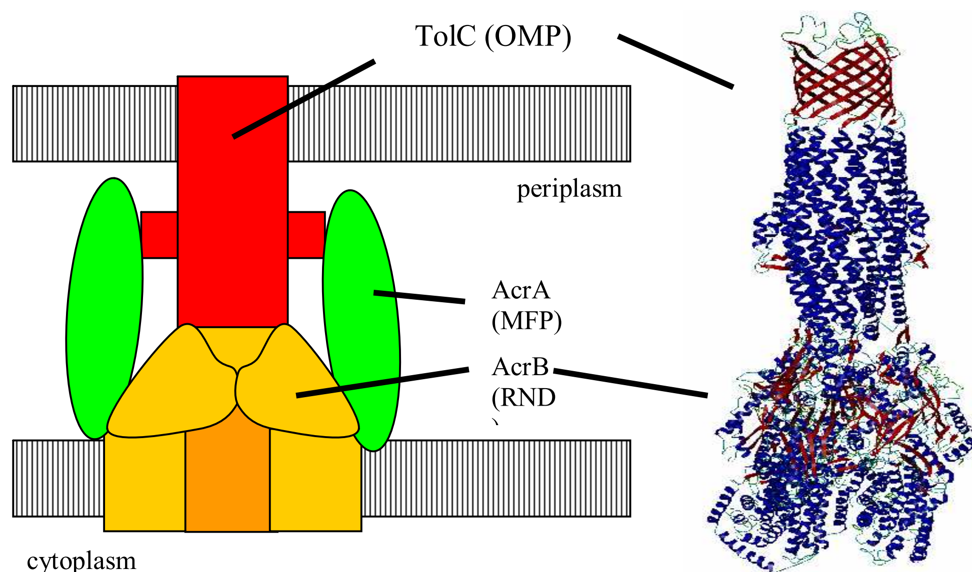


Fig. 27: Model for the AcrABTolC efflux complex (Murakami *et al*. 2002 simplified)
The AcrABTolC complex is composed of the RND transporter AcrB, the OMF transporter TolC and the MFP adapter protein AcrA. The OMF channel spans the periplasm and the outer membrane and its large central tube is composed of a 12 stranded β-barrel structure. AcrB is a homotrimer and the subunits are connected by intertwining loops in the periplasmic domains. The periplasmic domains also interact with the MFP component and the OMF transporter ("periplasmic kissing") and they are probably also involved in substrate recognition. As a potential adaptor protein brings AcrA probably the inner and outer membranes in close proximity and it also stabilizes the weak contact between TolC and AcrB. Substrates enter the CBA-type efflux complex from the periplasm or from the cytoplasmic space.

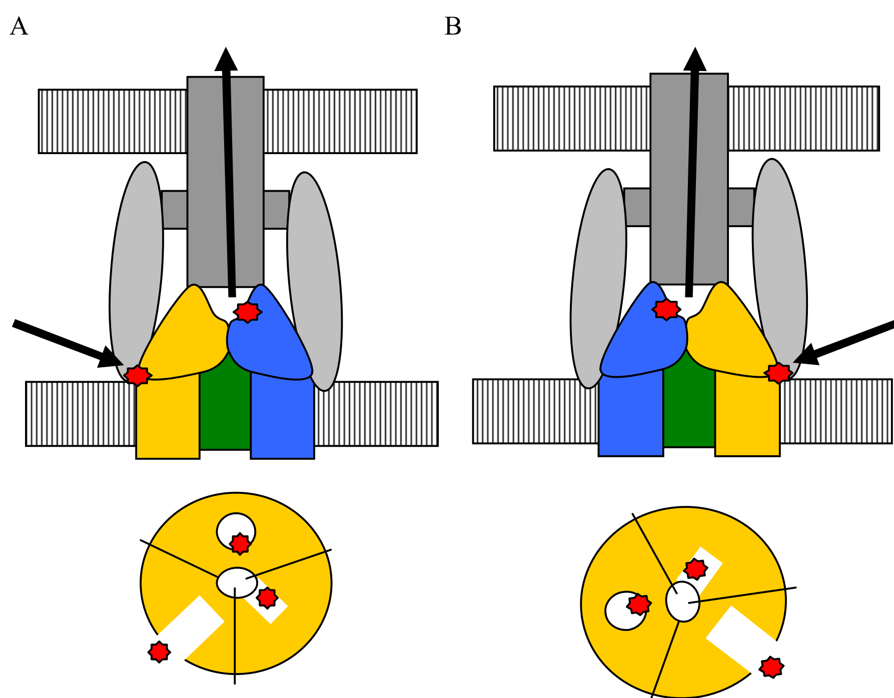


Fig. 28: Alternating site functional rotation transport mechanism of AcrABTolC (Stroebel *et al.*, 2007)
An alternating site functional rotation transport mechanism was suggested for the transport of substrates (red star) by the AcrABTolC CBA-type efflux complex and each subunit of the AcrB trimer may adopt a different conformation during this process. At each state of transport is one site occupied by substrate (green), a second site is only open to TolC (blue) and the third site is only open to the periplasm (yellow) to accept substrates. The conformational changes are probably induced by protons passing across the membrane down their electrochemical gradient.

Two separate transmembrane channels and several charged amino acids (residues D402, D408 and E415) are probably involved in a proton relay pathway that is essential for substrate transport (Goldberg *et al*., 1999; Mao *et al.,* 2002). The Position, geometry and conservation of these residues was explored to support functional investigations.

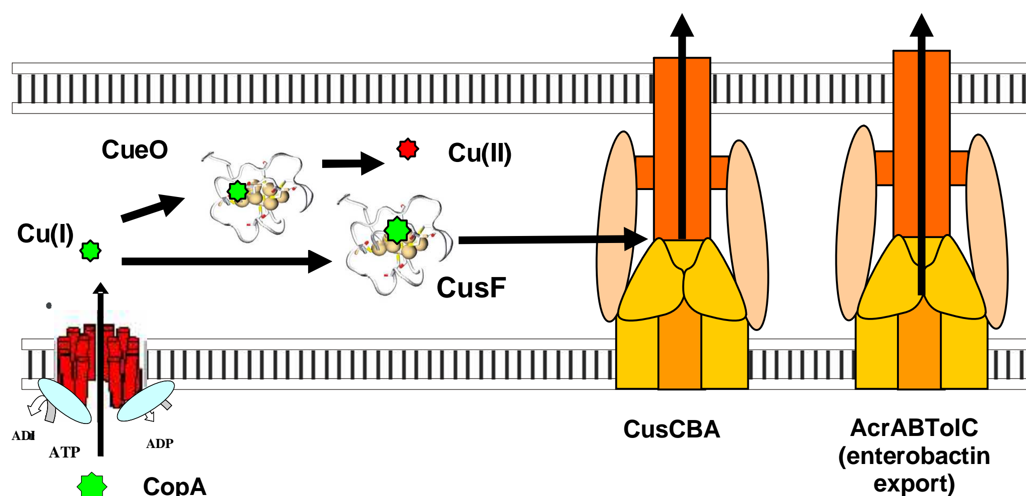


Fig. 29: Copper homeostasis of *E. coli* (Grass *et al.*, 2004 simplified)
The P-type ATPase CopA extrudes Cu (I) from the cytoplasm to the periplasmic space. The periplasmic multicopper oxidase CueO oxidizes Cu (I) to the less toxic Cu (II). As a second defence system binds the copper chaperon CusF periplasmic Cu (I) ions and mediates the extrusion of these ions to the outside via the CBA-type CusCBAF efflux system. CueO is also responsible for the release iron from enterobactin in the periplasm and enterobactin is probably also extruded by a CBA-type efflux system (AcrABTolC).

For the CBA-type efflux complex was initially a transenvelope transport across the cytoplasmic membrane, the periplasmic space and the outer membrane suggested (Saier *et al.* 1994). The presence of a periplasmic hydrophobic pocket in AcrB and in homologous HAE-RND-efflux transporters have however implied, that they probably take up their substrates from the periplasmic space or the outer leaflet of the cytoplasmic membrane (Murakami *et al.,* 2006; Hearn *et al.,* 2006; Lomovskaya and Totrov, 2005; Seeger *et al*., 2006).

The group of HME RND transporters might also take up substrates from the periplasmic space (Nikaido *et al*., 1998; Murakami *et al*., 2002; Legatzki *et al*., 2003). Thus is the periplasmic copper-binding protein CusF (Kittleson *et al*., 2006) involved in the copper and silver export via the CBA-type efflux system CusCFBA (Outten, *et al.,* 2001b) in *E. coli* (outlined in figure 29) (Grass and Rensing, 2001; Franke *et al*., 2003). Additionally can the CzcCBA complex probably not release metal ions from cytoplasmic metalloglutathionato complexes and this CBA-type efflux system also needs auxiliary transporters (CadA; ZntA, DmeF and CzcD) in the cytoplasmic membrane (Munkelt *et al*., 2004; Bagai *et al*., 2007; Grass, 2006). A periplasmic metal binding site in CzcA could however not be revealed, but a histidine rich cytoplasmic metal binding site is located between its TMH IV and V

of CzcA. This site has however probably a regulatory function and it might turn off the RND transporter if its activity has resulted in a low cytoplasmic concentration of trace element (Nies, personal communication).

4.12. The conservation of regulatory proteins

The extrusion of metal ions is energetically expensive and many transporters (e.g. CzcCBA) extrude not only toxic, but also essential cations (Nies and Silver, 1995). The metal ion resistance determinants have thus to be closely regulated (van der Lelie *et al.*, 1997). Various regulatory pathways are probably involved in the control of metal ion responsive loci in *C. metallidurans* CH34 (*cnr, cop*, *mer*, and *pbr*) and many regulators can respond to several distinct metal ions (multiple-metal-response) (Monchy *et al.,* 2007). Thus is the periplasmic metal binding protein CzcE (a paralog of CopH) not only a indirect regulator of the *czc* operon, but it can also bind copper and it might thus work as regulatory link between the copper homeostasis and cobalt/zinc/cadmium resistance in *C. metallidurans* CH34 (Grosse *et al.,* 2004a; Zoropogui et al. 2008).
Sigma factors and anti-sigma factors are also often implicated in the regulation of transport protein determinants (Kill *et al.,* 2005). The group of ECF (*extracytoplasmatic function*) -sigma factors responds to environmental stressors e.g. heat-, osmotic-, oxidative stress, cold shock, high pressure, acids and antibiotics (Lonetto *et al.*, 1994; Martinez-Salazar *et al.*, 1996). In *C. metallidurans* CH34 are ECF-sigma factors (RpoI, RpoJ and RpoK) also implicated in the regulation of transition metal homeostasis (Grosse *et al.*, 2007). Many of these and other transcription factors in prokaryotes have been subject to lineage-specific duplications and horizontal gene transfer (Babu *et al.,* 2004). The ECF sigma factor CnrH of *C. metallidurans* CH34 is thus unique to this organism and RpoK, RpoJ are paralogs. CnrH and the periplasmatic anti-sigma factor complex CnrXY are responsible for the regulation of the CBA-type efflux system CnrCBA (Grass *et al.*, 2000; Mergeay *et al.,* 2003). The related CBA-type efflux system CzcCBA is (as shown in figure 30) additionally controlled by the two-component-system CzcRS (Grosse *et al.*, 1999; van der Lelie *et al.*, 1997). Two-component-regulatory systems are also involved in the transcriptional control of related heavy metal homoeostasis systems in *E. coli* (PcoRS, CusRS) and in *P. syringae* (CopRS) (Mills *et al.*, 1994; Nies *et al.*, 1998; Munson *et al.,* 2000). The sensor kinase CzcS and the response regulator CzcR in *C. metallidurans* CH34 may have been acquired by horizontal gene transfer since neither orthologous nor paralogous counterparts have been found for these proteins.
The total repertoire of other putative two-component-regulatory systems in *C. metallidurans* CH34 was however not significantly different in respect to those in other investigated *Burkholderiacea* (*C. eutrophus* H16 and JMP134, *R. solanacearum* GMI1000, *B. xenovorans* LB400 and *B. cepacia* AMMD). One third of all two-component-systems in this group belong to a common orthologous core set and in the *Cupriavidus* cluster have even two third of these proteins an orthologous counterpart. Only very few (10%) paralogs were found for putative two-component-systems.

recent paralogs	middle aged paralogs	Old paralogs	
Ca^{2+} ATPases CzcP (pMOL30) PbrA (pMOL30) CadA (chrom1)		Cu/Ag^{2+} -type ZntA (chrom2) K^{+} -type	**P-type ATPases**
	ChrA2 (chrom2)	ChrA1 (pMOL28)	**CHR Proteins**
MerT1 (pMOL28) MerT4 (pMOL30)	MerT3 (chrom) MerT2 (pMOL30)		**MerT Proteins**
CnrA (pMOL28) NccA (pMOL30) CnrB (pMOL28)	SilA (pMOL30) CusA (chrom2) CusB (pMOL30) CusB (chrom2)	CzcD (pMOL30) DmeF (chrom1) FieF (chrom1) CzcA (pMOL30) CzcB (pMOL30)	**RND/MFP /CDF-Proteins**

Fig 30: Paralogs of putative heavy metal-specific proteins in C. metallidurans strain CH34
The distribution of recent, middle aged and old paralogous metal ion transporters after stepwise removal of organisms from consideration according to the phylogenetic distance of the proteobacteria: is shown (chrom., chromosome)

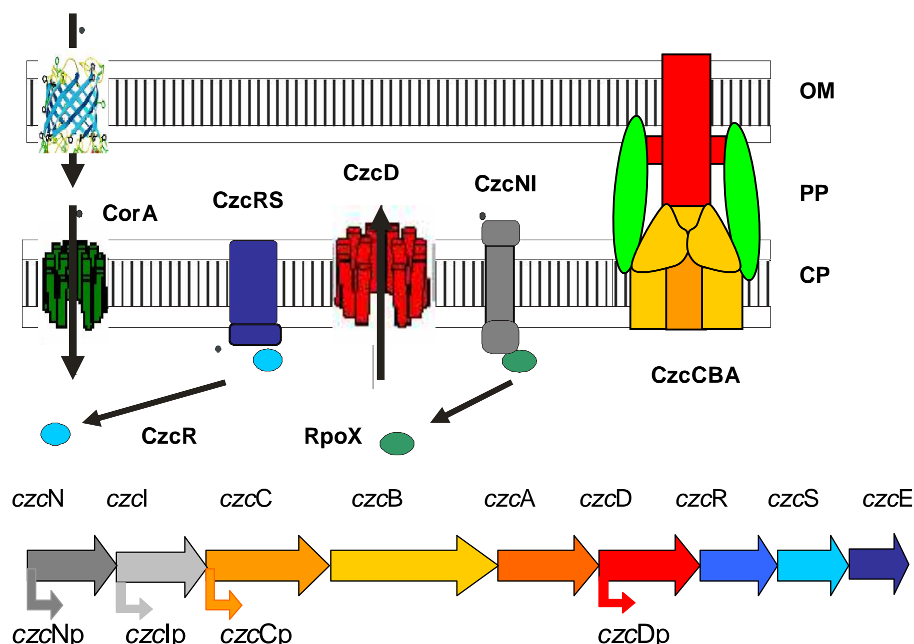


Fig. 31: The *czc* operon as an example for the role of ECF sigma factors and two-component-systems for the regulation of CBA-type efflux systems in of *C. metallidurans* CH34 **(Nies 2000 simplified)**
The cytoplasmic membrane (CP) and the outer membrane (OM) the periplasmic space (PP). Metal ions are taken up by unspecific uptake systems (e.g. CorA). The components of the CBA-type efflux pump CzcCBA span the entire periplasm. The *czc*CBA operon is probably regulated at the transcriptional level by the two-component-system CzcRS and the putative regulatory proteins CzcN and CzcI. These regulators respond probably to the cellular levels of zinc, cobalt or cadmium. An ECF sigmafactor of unknown identity (RpoX) is probably also involved, but no ECF sigma factor is encoded at the *czc* operon.

4.13. YedZ contains fused domains for transport and electron transfer

Iron homeostasis is not only of special importance for pathogens but an extensive accumulation of iron mineral magnetite (Fe_3O_4) is a special feature in magnetotactic bacteria (Bazylinski, 1995). The occurrence of these crystals in the in the martian meteorite ALH84001 (Thomas-Keprta, *et al.,* 2000) was also taken as evidence for pre-existing microbial life on Mars (McKay, et al 1996). In magnetotactic bacteria is magnetite stored in magnetosomes and these enable these bacteria to orient along magnetic field lines (Schübbe *et al.* 2003; Balkwill *et al.*, 1980; Bazylinski, 1995). This magnetotaxis is useful to bacteria to find and maintain a position in vertical oxygen concentration gradients in aquatic environments (Frankel *et al.,* 1997). Several transport proteins in the magnetosome membranes are part of a large genomic supercluster termed as the Magnetosome Island that are associated with magnetosome functions and iron (Fe^{2+} or Fe^{3+}) transport (Grünberg *et al.*, 2001; Tamegai and Fukumori, 1994; Fukumori *et al*., 1997; Schüler, D. 2004). A striking feature of the Magnetosome Island is a mosaic-like structure, a high genetic plasticity and a large number of mobile genetic elements and many magnetosome genes have been shuffled or exchanged between different sites (Schübbe *et al.*, 2003). Some unusual MFS transporters (2.A.1.43) (Msp2 and Mma2) are encoded at magnetosome islands (Bazylinsky, 1995) and are novel in that they contain an additional C-terminal YedZ-like domain. The YedZ-like domain may have a regulatory function it may mediate the response to oxygen or the redox state of the cell.

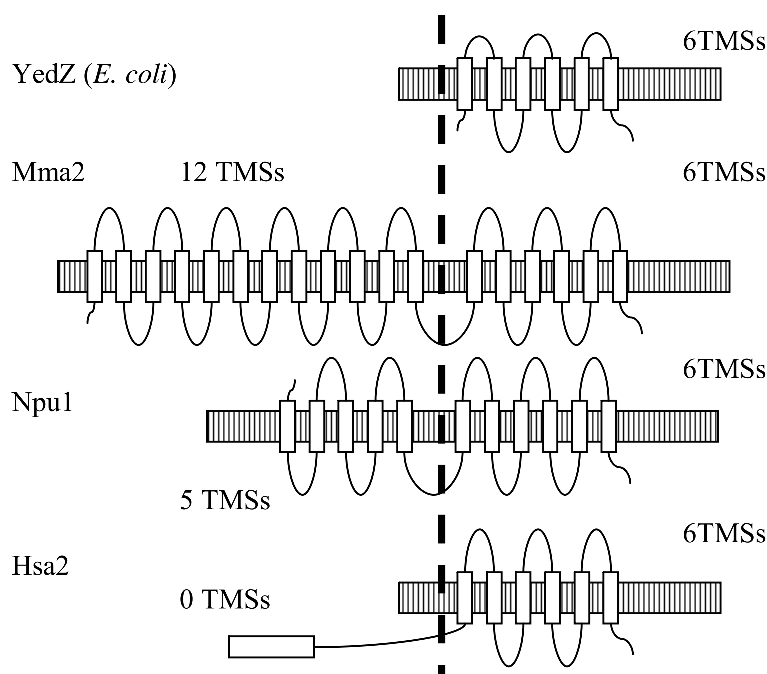


Fig. 32: overview of YedZ homologues
The figure shows a schematic representation of the membrane topologies of YedZ homologs. YedZ of *E.coli* and the most of its homologs (not shown) contain six transmembrane segments. The magnetosomal protein Mma2 is a novel MFS transporter and contains a C-terminal YedZ-like domain. The cyanobacterial Npu1 is distantly related to the putative transmembrane electron flow carriers of the DsbD family. It contains a C-terminal YedZ-like domain. The cyanobacterial heme exporter Hsa2 contains a fusion of an oxidoreductase to a YedZ-like domain.

These and other homologs of the YedZ (TC 9.B.43) transporter family are shown in figure 32.
The heme-containing membrane protein YedZ is part of the *yed*YZ operon that encodes a putative two subunit oxidoreductase in the inner membrane of *E. coli* (Drew *et al.,* 2002; Brokx *et al.*, 2005). YedZ anchors the catalytic soluble molybdenum cofactor-containing subunit YedY to the inner membrane (Loschi *et al*., 2004; Brokx *et al.*, 2005).

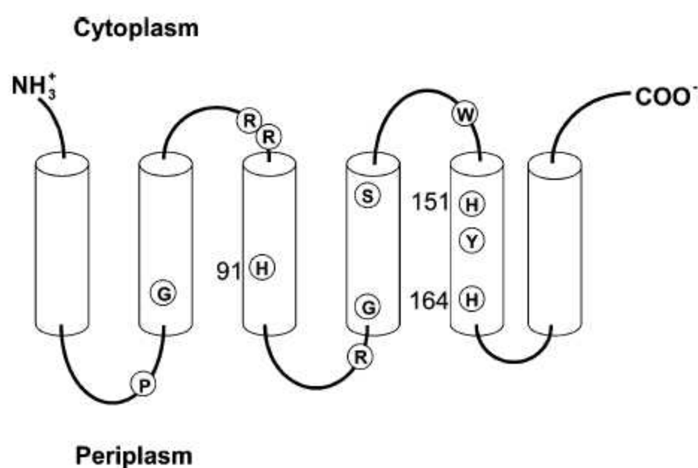


FIG. 33. Topology and distribution of conserved residues in the YedZ family (Workun *et al,* 2008).
The positions of the absolutely conserved residues within the bacterial YedZ family are shown. The tubes represent putative transmembrane helices. Circled letters indicate amino acid residues which are conserved in alignments of YedZ and conserved histidine residues, which may coordinate a heme, are numbered.

In this work have phylogenetic, structural and motif analyses been carried out to characterize the YedZ transport protein family. The most bacterial YedZ homologues contain six integral transmembrane spanning segments (TMS). Investigations of sequence similarities between these segments revealed that they have probably been formed by an intragenic triplication of a 2 TMS-encoding element. Several conserved histidyl residues (figure 33) within their transmembrane segments may have a heme binding function. Thus has a high significant sequence similarity to the putative heme export systems HEP (TC 9.B.14) and HemeE (TC 3.A.1.117) been observed. YedZ-like domains can also be found at the C-terminal end of several putative transmembrane electron carriers in cyanobacteria. These proteins are distantly related to the DsbD family (5.A.1) (Kimball *et al.,* 2003).

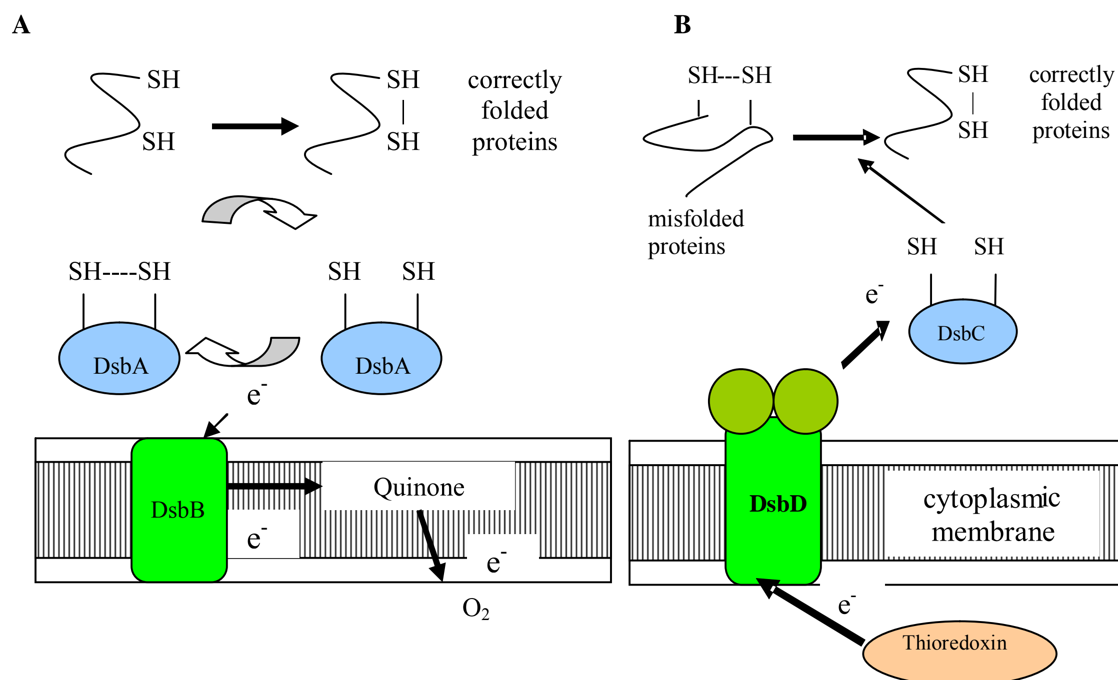


Fig. 34: DsbD and DsbB catalyze transmembrane electron flow in opposite directions (Rietsch *et al*., 1996; Nakamoto and Bardwell, 2004).
Disulfide bond formation in the *E. coli* periplasm is a catalyzed by the highly oxidizing DsbA/DsbB catalytic machinery (left side). The oxidizing power utilized by this system comes from a membrane-embedded electron transport system, which utilizes molecular oxygen as a final oxidant. DsbA is a very strong oxidant and also introduces non-native disulfide bonds into proteins. Proofreading of disulfide bond formation is performed by the DsbC/DsbD system (right side) that has chaperone-like activity and rearranges non-native disulfides. The reducing power for this process is provided by the cytoplasmic thioredoxin system, utilizing NADPH as the ultimate electron source.

As pointed out in figure 34, catalyze Dsb proteins the folding and oxidation of polypeptides in the periplasm. They also reduce wrongly paired disulfides and DsbD recycles the electrons for this process (Katzen and Beckwith, 2000; Krupp *et al*., 2001; Collet and Bardwell, 2002). The YedZ-like domains in the homologs of the bacterial DsbD family (TC #5.A.1) might either play a direct role in electron transfer or it regulates electron flow of this process.

Distant homologues of YedZ can also be found among coenzyme F420 -dependent oxidoreductases (Warkentin *et al*., 2001) in animals. This includes an epithelial plasma membrane antigen of prostate cancer cells and these STEAP proteins (9.B.43.2.1) (Knutson, 2007) may have a function in tight junctions, in gap junctions, or in cell adhesion. These proteins may be implicated in the proliferation and invasiveness of prostate cancer cells and are thus potential targets for prostate cancer therapy and diagnostic imaging (Hubert *et al*., 1999; Lalani *et al*., 1997). Taken together suggest the results of sequence analysis that YedZ homologues might facilitate or regulate oxidoreduction, transmembrane electron flow and transport.

4.14. SbtA combines similarity to secondary carriers and ABC transporters

A high genomic flexibility has also facilitated the occupation of local niches by the marine lineages of cyanobacteria e.g. *Synechococcus* and *Prochlorococcus* (Dufresne *et al.*, 2008). Freshwater cyanobacteria are subjected to large seasonal fluctuations especially in regard to the availability of nutrients and inorganic carbon (Ci) (Woodger *et al.*, 2005; Woodger *et al.*, 2007; Badger *et al.*, 2006). The primary carboxylating enzymes, the ribulose bisphosphate carboxylase-oxygenases (RuBisCO) of cyanobacteria have higher turnover rates, but much lower affinities for both CO_2 and O_2 than algal or higher plant counterparts (Badger and Andrews, 1987; Price *et al.*, 2008).
These constrain, and the restricted diffusion of Ci (CO_2 and HCO_3) species in water has triggered the development of special CO_2 concentrating mechanisms (CCMs) (outlined in figure 35) in these bacteria (Badger *et al.*, 2006; Woodger *et al.*, 2003). This includes e.g. the partitioning of RuBisCO into cellular micro-compartments (the carboxysomes) and the hydration of CO_2 inside the cells (Badger and Price, 2003; Price *et al.*, 2007; Price *et al.*, 1998).

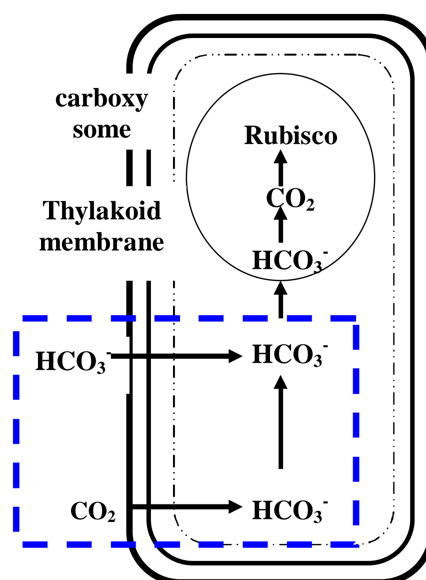


Fig 35: Roles of cyanobacterial CO_2 and HCO uptake systems (McGinn *et al.,* 2003).
Cyanobacterial cells have a highly differentiated membrane system consisting of the outer membrane (OM), plasma membrane and the thylakoid membrane. Carboxysomes are protein bodies which contain all the theCO_2-fixing RuBisCO proteins of these cells. The need to facilitate sufficient supply of this enzyme with inorganic carbon (Ci) has triggered the development of special CO_2 concentrating mechanisms (CCMs) in cyanobacteria. This includes at least five Ci transporters for CO_2 and HCO_3^-, the partitioning of RuBisCO into carboxysomes and the accumulation of HCO_3. Inside of the carboxysomes is HCO_3^- converted back to CO_2 to elevate the CO_2 level around the RuBisCO enzymes

The accumulated HCO_3 is much less permeable than the uncharged CO_2 molecule and only slowly leaks from the cell. The hydration of CO_2 also maintains an inwardly directed CO_2 concentration gradient such that external CO_2 continuously diffuses into cells (Badger and Price, 2003; Tchernov *et*

al., 2001). Inside of the carboxysomes is HCO_3^- converted back to CO_2 to elevate the CO_2 level around the RuBisCO enzymes (Amoroso *et al*., 2003; Ohkawa *et al*., 2000; Ogawa and Kaplan, 2003).

At least five Ci transporters are responsible for CO_2 and HCO_3^- homeostasis in these cyanobacteria and this includes constitutive and low- CO_2 inducible CO_2 uptake transporters (Badger *et al.,* 2006; Shibata *et al*. 2001; Sültemeyer *et al.* 1995).

At high pH levels occurs Ci mainly as bicarbonate ions and medium affinity transporter with high flux rate (BicA in *Synechocystis* PCC6803) and sodium dependent HCO_3^- transporters (e.g. IctB in *Synechocystis* PCC6803) facilitate HCO_3 uptake under these conditions (Zhang *et al*., 2004; Woodger *et al*., 2007; Badger *et al*., 2006; Bonfil *et al.,* 1998).

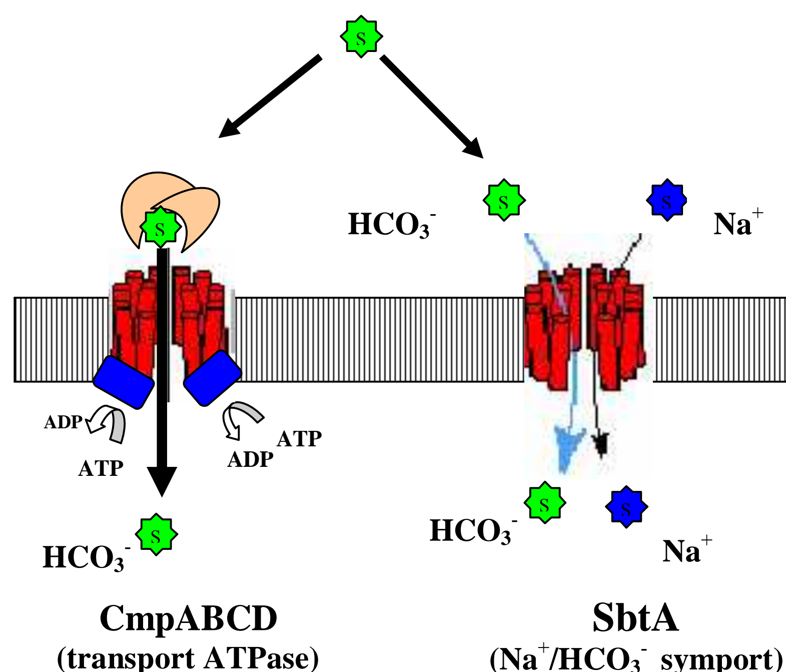


Fig. 36: Role of SbtA for HCO_3^- uptake in *Synechococcus* sp. (Woodger *et al*., 2005)
BCT1 (CmpABCD) is an inducible high affinity HCO_3^- transporter ATPase and SbtA is an inducible, sodium dependent HCO_3^- transporter. The sodium motif force (smf) is generated by the Na^+/H^+ antiporter NtpJ (not shown)

The most important HCO_3 transporters in the freshwater strain *Synechococcus* sp. strain PCC7942 are shown in figure 36 (Badger *et al*., 2006). The ABC transporter BCT1 (CmpABCD) is responsible for inducible high affinity HCO_3^- uptake at low CO_2 concentrations (Omata *et al*., 1999; Omata and Ogawa, 1986). The bicarbonate transporter SbtA (TC 2.A.83) of *Synechococcus* PCC6803 is probably responsible for a sodium dependent HCO^{3-} uptake (Shibata *et al.,* 2002; Zhang *et al*., 2004). These transporters account for most, if not all, of the HCO_3^- uptake at pH 9.3 (Price *et al*., 2004). The SbtA transporters have a variable representation among other β cyanobacteria (Badger *et al.,* 2002). Homologues of the SbtA transporter widely distributed among all species in freshwater environments but there may be a reduced requirement for these energy expensive transport systems in the HCO_3 rich

marine environments (Badger *et al.*, 2006). At sudden and severe CO_2 limitations can SbtA probably also post translationally be activated and a response to illumination was also suggested (Price *et al.,* 2007; Amoroso *et al.,* 2003; Zhang *et al.*, 2004).
Protein sequence alignments with the IC and GAP programs have revealed that the first halves of SBT family members are homologous to the second halves. The ten transmembrane segments (TMSs) of SbtA (39 kDa) are thus probably the product of an intragenic duplication event. The two halves have opposite orientation in the membrane. The first half of SbtA (TMSs 1–4) is also homolog to the first half (TMSs 2–5) of the 10 TMS containing ABC transporter AraH (3.A.1.2.2). As a sodium motif force (smf) driven transporter is SbtA however classified as a secondary carrier (TC 2.A.) (Shibata *et al.*, 2002). This unusual connection between ABC transporters and secondary carriers has only been found in the SbtA family.

4.15. Outlook

A high surface-to-volume ratio and a large contact area facilitate an efficient interaction between microorganisms and their surrounding environment. Proteobacteria have thus received much attention as potential tools for bioremediation purposes (Saier, 2005; Mergeay et al., 2003).
The β-proteobacterium C*upriavidus eutrophus* JMP134 is able to degrade major environmental pollutants and it might be of special value for the biodegradation of recalcant aromatic compounds (Johnson and Stanier*,* 1971). This strain can also build up polyhydroxyalkalonates (PHAs), which may serve as precursor for new biodegradable thermoplasts e.g. the novel polythioester BIOPOL® (Lee 2006*;* Pohlmann *et al.*, 2006*;* Schwartz *et al.*, 2003).
The use of proteobacteria for bioremediation may also include the clean up of metal-polluted environments since the adsorption, precipitation and active transport of heavy metal ions has been reported for many of these microorganisms (Ledin, 2000). A nickel resistance system (*ncc-nre* of *R. metallidurans* 31A) has already successfully been expressed in *B. cepacia* L.S.2.4 and the plant *Lupinus luteus* was subsequently inoculated with this strain. The heterologous expression of ncc-nre encoded nickel resistance was accompanied by nickel removal from the culture medium and nickel accumulation in roots (Lodewyckx, *et al.,* 2001; Mergeay *et al.,* 2003). The strain *C. metallidurans* CH34 has gained increasing interest for the development of biosensors and the in situ immobilisation of heavy metals on its surface have also been reported (Nies, 2000; Mergeay *et al.,* 2003; Saier 2005; Diels *et al.,* 1999; Valls and de Lorenzo, 2002). The removal of the bacteria from the polluted sites by flocculation might allow a clean up of mixed pollution (radionuclides, recalcitrant organics and heavy metals) (Valls and de Lorenzo, 2002). The development of strains which have an increased glutathione expression but missing efflux systems might also be useful for bioremediation purposes (Nies personal communication). The computational investigations of the genomic repertoire of heavy metal homeostasis determinants and their regulatory mechanisms may support these developments.

Special emphasis should e.g. be given to functional investigations of the chromate resistance, CDF proteins and the CzcCBA transporter complex as well as the regulatory networks controlling heavy metal ion homeostasis. CDFs are not only found in bacterial cell membranes but also in the vacuolar membrane of both plants and yeast and they are thus potentially useful in the engineering of hyperaccumulative phytoremediation systems (Haney *et al*., 2005).

Further structural investigations should also be carried out to reveal, how the closely related RND transporter families can transport such a variety of organic and mostly hydrophobic substrates and heavy metal cations. RND transporters can not only confer metal ion resistance, but they are also implicated in the problematic re-emergence of antibiotic-resistant infections (e.g. tuberculosis) and the resistance of human tumours to chemotherapeutic drugs (Frost *et al*., 2005; Higgins, 2007). Tularemia can not be treated with b-lactam antibiotics, since those are extruded by the widely distributed AcrAB RND efflux pumps of these strains (Cross and Jacobs, 1993; Bina *et al*., 2006; Bina *et al*., 2008; Sjostedt, 2007). Thus have aerosols of *F. tularensis* effectively been exploited by several countries for the development of biological weapons.

Some plant species such as maize and sugar cane possess biochemically and anatomically complex active uptake systems for the acquisition of CO_2 or HCO_3 (CCMs) known as the C4 cycle and those significantly reduce their water and nitrogen consumption (von Caemmerer and Furbank, 2003).

C3 crop plants lack these specialized systems and there is a significant scope to improve their water and nitrogen use efficiency through the establishment of a basal transgenic CCM (Price *et al*., 2008).

A promising approach could be to express a cyanobacterial HCO_3^- transporter in the C3 chloroplast and single-subunit HCO_3^- transporters such as BicA and SbtA are good first candidates (Price *et al*., 2008). These Na^+- coupled transporters SbtA and BicA may be supported by the potential sodium specific transporters of the *Arabidopsis* chloroplast envelope (Rolland *et al.,* 2003; Karley *et al.,* 2000; Price *et al*., 2008). The engineering of an active cyanobacterial HCO_3^-pump within the chloroplast would allow it to operate at a higher CO_2 level, allowing the plant the potential to use less water and nitrogen for the same crop yield (Price *et al*., 2008).

5. REFERENCES

Alberts, I. L., Nadassy, K., Wodak, S. J. 1998 Analysis of zinc binding sites in protein crystal structures. *Protein Sci.* 7:1700-1716.

Alexeyenko, A., Lindberg, J., Perez-Bercoff, A., Sonnhammer, E. L. L. 2006. Overview and comparison of ortholog databases. *Genome Technol*. 3:137-143.

Alexeyenko, A., Tamas, I., Liu, G., Sonnhammer, E. L. L. 2006b Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics*. 22:e9-e15.

Alisi, C, Lasinio, G. J., Dalmastri, C. *et al.* 2005. Metabolic profiling of *Burkholderia cenocepacia*, *Burkholderia ambifaria*, and *Burkholderia pyrrocinia* isolates from maize rhizosphere. *Microb Ecol.* 50:385-395.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403-10.

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.* 25:3389-402.

Amoroso, G.,Seimetz, N., Sültemeyer, D. 2003. The dc13 gene upstream of ictB is involved in rapid induction of the high affinity Na^+ dependent HCO_3^- transporter in cyanobacteria. *Photosyn Res.* 77:127-138.

Andersen, C., Hughes, C., Koronakis, V. 2000. Chunnel vision. Export and efflux through bacterial channel-tunnels. *EMBO Rep.* 1:313-318.

Andersen, C., Hughes, C., Koronakis, V. 2001. Protein export and drug efflux through bacterial channel-tunnels. *Curr Opin Cell Biol.* 13:412-416.

Andersen, C., Koronakis, E., Hughes, C., Koronakis, V. 2002. An aspartate ring at the TolC tunnel entrance determines ion selectivity and presents a target for blocking by large cations. *Mol Microbiol*. 44:1131-1139.

Andreini, C., Bestini, I., Rosato, A. 2004. A hint to search for metalloproteins in gene banks. *Bioinformatics*. 20:1373-1380.

Andreini, C., Banci, L., Bertini, I., Rosato, A. 2006. Counting the zinc-proteins encoded in the human genome. *J Proteome Res.*5:196-201.

Andreini, C., Banci, L., Bertini, I., Elmi, S., Rosato, A. 2007. Non-heme iron through the three domains of life. *Proteins*. 67:317-324.

Andreini, C., Banci, L., Bertini, I., Rosato, A. 2007b. Occurrence of Copper Proteins through the Three Domains of Life: A Bioinformatic Approach. *Proteome Res. 7*:209–216.

Andres, Y., Thouand, G., Boualam, M., Mergeay, M. 2000. Factors influencing the biosorption of gadolinium by micro-organisms and its mobilisation from sand. *Appl Microbiol Biotech*. 54:262-267.

Andrews, S. C., Robinson, A. K., Rodriguez-Quinones, F. 2003. Bacterial iron homeostasis. *FEMS Microbiol. Rev*. 27:215–237.

Anton, A. 2001. Genetische und biochemische Charakterisierung von CzcD und anderen Regulatoren der *czc*-vermittelten Schwermetallresistenz in *Ralstonia metallidurans*. *Dissertation. Martin-Luther-Universität Halle-Wittenberg, Halle.*

Anton, A., Grosse, C., Reissmann, J., Pribyl, T, Nies, D. H. 1999. CzcD is a heavy metal ion transporter involved in regulation of heavy metal resistance in *Ralstonia* sp. strain CH34. *J Bacteriol*. 181:6876-81.

Apatoff, A., Kim, E., Kliger, Y. 2008 Towards Alignment Independent Quantitative Assessment of Homology Detection. *PLoS ONE*. 1:e113

Argüello, J. M., Eren, E., Gonzalez-Guerrero, M. 2007. The structure and function of heavy metal transport P1B-ATPases. *Biometals*. 20:233-248.

Ashburner, M, Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., Sherlock, G. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet*. 25:25–29.

Babor, M., Greenblatt, H. M., Edelman, M., Sobolev, V. 2005. Flexibility of Metal Binding Sites in Proteins on a Database Scale. *Proteins*. 59:221–230.

Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M., Teichmann, S. A. 2004 Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol*. 14:283-291.

Badger, M. R., Andrews, T. J. 1987. Co-evolution of Rubisco and CO_2 concentrating mechanisms. *In: Biggins J, ed. Progress in photosynthesis research. Proceedings of the VIIth International Congress on Photosynthesis Providence. Rhode Island. 10–15 Dordrecht: Martinus Nijhoff Publishers.* 601–609.

Badger, M. R., Hanson, D. T., Price, G. D. 2002. Evolution and diversity of CO_2 concentrating mechanisms in cyanobacteria. *Functional. Plant Biol*. 29:407-416.

Badger, M. R., Price, G. D. 2003. CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J Exp Bot*. 54:609–622.

Badger, M. R., Price, G. D., Long, B. M., Woodger, F. J. 2006. The environmental plasticity and ecological genomics of the cyanobacterial CO_2 concentrating mechanism. *J Exp Bot.* 57:249–265.

Bagai, I., Liu, W., Rensing, C., Blackburn, N. J., McEvoy, M. M. 2007. Substrate-linked Conformational Change in the Periplasmic Component of a Cu(I)/Ag(I) Efflux System. *J Biol Chem*. 282:35695-35702.

Bains, J., Boulanger, M. J. 2007. Biochemical and structural characterization of the paralogous benzoate CoA ligases from *Burkholderia xenovorans* LB400: defining the entry point into the novel benzoate oxidation (box) pathway. *J Mol Biol.* 373:965-77.

Bairoch, A., Bucher, P., Hofmann, K. 1997. The PROSITE database, its status in 1997. *Nucleic Acids Res.* 25:217–221.

Balkwill, D., Maratea, D., Blakemore, R. 1980. Ultrastructure of a magnetotactic spirillum. *J Bacteriol.* 141:1399–1408.

Battistuzzi, G., Borsari, M., Menabue, L., Saladini, M., Sola, M. 1996. Amide group coordination of the Pb2. ion. *Inorg Chem*. 35:4239-4247

Bazylinski, D. 1995. Structure and function of the bacterial magnetosome. *ASM News*. 61:337–343.

Beard, S. J., Hashim, R., Wu, G., Binet, M. R., Hughes M. N., Poole, R. K. 2000. Evidence for the transport of zinc(II) ions via the pit inorganic phosphate transport system in *Escherichia coli. FEMS Microbiol Lett*. 184:231-5.

Bendtsen, J. D., Nielsen, H., von Heijne, G, Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*. 340:783-795.

Bennett, P. M. 2004. Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. *Methods Mol Biol.* 266:71–113.

Berg, G., Eberl, L., Hartmann, A. 2005. The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environ Microbiol.* 7:1673-1685

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, Jr., E. F., Brice,M. D., Rodgers, J. R., Kennard, 0., Shimanouchi, T., Tasumi, M. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol.* 112:535-542.

Bertini, I., Rosato, A. 2007. From genes to metalloproteins: a bioinformatic approach. *European Journal of Inorganic Chemistry*. 18:2546-2555

Bina, J. E., Provengano, D., Wang, C., Bina, X. R., Mekalanos, J. J. 2006. Characterization of the Vibrio cholerae vexAB and vexCD efflux systems. *Arch Microbiol*. 186:171-181.

Bina, X. R., Lavine, C. L., Miller, M. A., Bina, J. E. 2008. The AcrAB RND efflux system from the live vaccine strain of *Francisella tularensis* is a multiple drug efflux system that is required for virulence in mice. *FEMS Microbiol Lett*. 279:226-233.

Bloss, T., Clemens, S., Nies, D. H., 2002. *Planta*. 214:783.

Bock, C. W., Katz, A. K., Glusker, J. P. 1995. Hydration of zinc ions: A comparison with magnesium and beryllium ions. *J Am Chem Suc*. 117:3754-3765.

Bonfield, J. K., Smith, K., Staden, R. 1995. A new DNA sequence assembly program. *Nucleic Acids Res*. 23:4992-9.

Bonfil, D. J., Ronen-Tarazi, M., Sültemeyer, D., Lieman-Hurwitz, J., Schatz, D., Kaplan, A. 1998. A putative HCO_3^- transporter in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *FEBS Lett*. 430:236-240.

Borremans, B., Hobman, J. L., Provoost, A., Brown, N. L., van Der Lelie, D. 2001. Cloning and functional analysis of the pbr lead resistance determinant of *Ralstonia metallidurans* CH34. *J Bacteriol*. 183:5651-5658.

Boucher, C., Genin, S., Arlat, M. 2001. Current concepts of pathogenicity in plant pathogenic bacteria. *Life Sciences*. 324:915-922.

Braun, V., Braun, M. 2002. Iron transport and signaling in *Escherichia coli. FEBS Lett.* 529:78-85

Brenner, S. E., Chothia, C., Hubbard, T. J. P. 1998. Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships *Proc Natl Acad Sci*. 95:6073–6078.

Bröker, D., Arenskötter, M., Legatzki, A., Nies, D. H., Steinbüchel, A. 2004. Characterization of the 101-Kilobase-Pair megaplasmid pKB1, isolated from the rubber-degrading bacterium *Gordonia westfalica* Kb1. *Microbiology*. 186:212-225.

Brokx, S. J., Rothery, R. A., Zhang, G., Ng, D. P., Weiner, J. H. 2005. Characterization of an *Escherichia coli* sulfite oxidase homologue reveals the role of a conserved active site cysteine in assembly and function. *Biochemistry*. 44:10339–10348.

Brown, J. S., Holden, D. W. 2002. Iron acquisition by gram-positive bacterial pathogens. *Microbes Infect.* 4:1149–1156.

Brown, N., Shih, Y., Leang, C., Glendinning, K., Hobman, J., Wilson, J., 2002b. Mercury transport and resistance. *Biochem Soc Trans*. 30:715–718.

Bundesumweltministerium. 1997. UBA-Bericht:"Daten zur Umwelt - Der Zustand der Umwelt in Deutschland" *BMU-Pressedienst* 038/97.

Busch, W., Saier, M. H. 2004. The IUBMB-endorsed transporter classification system. *Mol Biotechnol*. 27:253-62.

Büttner, D., Bonas, U. 2002**.** Getting across—bacterial type III effector proteins on their way to the plant cell. *EMBO J*. 21:5313-5322.

Caballero-Mellado, J., Onofre-Lemus, J., Estrada-de los Santos, P., Martinez-Aguilar, L. 2007. The Tomato Rhizosphere, an Environment Rich in Nitrogen-Fixing *Burkholderia* Species with Capabilities of Interest for Agriculture and Bioremediation. *Appl Environ Microbiol*. 73:5308-5319.

Cao, Z., Klebba, P. E. 2002. Mechanisms of colicin binding and transport through outer membrane porins. *Biochimie*. 84:399-412.

Cao, J., Woodhall, M. R. Alvarez, J., Cartron, M. L., Andrews, S. C. 2007. EfeUOB (YcdNOB) is a tripartite, acid-induced and CpxAR-regulated, low-pH Fe transporter that is cryptic in *Escherichia coli* K-12 but functional in *E. coli* O157:H7. *Mol Microbiol*. 65:857-875.

Castillo-Davis, C. I., Hartl, D. L. 2003. Conservation, relocation and duplication in genome evolution. *Trends Genet*. 19:593-597.

Cellier, M., Prive, G., Belouchi, A., Kwan, T., Rodrigues, V., Chia, W., Gros, P. 1995. Nramp defines a family of membrane proteins. *Proc Natl Acad Sci*. 92:10089-10093.

Chain, P. S., Denef, V. J., Konstantinidis, K. T., Vergez, L. M., Agullo, L., Reyes, V. L., Hauser, L., Cordova, M., Gomez, L., Gonzalez, M., Land, M., Lao, V., Larimer, F., LiPuma, J. J., Mahenthiralingam, E., Malfatti, S. A., Marx, C. J., Parnell, J. J., Ramette, A., Richardson, P., Seeger, M., Smith, D., Spilker, T., Sul, W. J., Tsoi, T. V., Ulrich, L. E., Zhulin, I. B., Tiedje, J. M. 2006 *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc Natl Acad Sci.* 103:15280-7.

Chance, M. R., Fiser, A., Sali, A., Pieper, U., Eswar, N., Xu, G., Fajardo, J. E., Radhakannan, T., Marinkovic, N. 2004. High-throughput computational and experimental techniques in structural genomics. *Genome Res*. 14:2145-2153.

Changela, A., Chen, K., Xue, Y., Holschen, J., Outten, C. E., O'Halloran, T. V., Mondragon, A., 2003. Molecular basis of metal-ion selectivity and zeptomolar sensitivity by CueR. *Science*. 301:1383-1387.

Christianson, D. W. 1991. Structural biology of zinc. *Adv Protein Chem*. 42:281-355.

Coenye, T., Vandamme, P. 2003. Simple sequence repeats and compositional bias in the bipartite *Ralstonia solanacearum* GMI1000 genome. *BMC Genomics*. 4:10.

Coenye, T., Falsen, E., Vancanneyt, M., Hoste, B., Govan, J. R., Kersters, K., Vandamme, P. 1999. Classification of *Alcaligenes faecalis*-like isolates from the environment and human clinical samples as *Ralstonia gilardii* sp. nov. *Int J Syst Bacteriol*. 49:405-413.

Collet, J. F., Bardwell, J. C. A. 2002 Oxidative protein folding in bacteria *Mol Microbiol.* 44:1–8.

Corbisier, P., van der Lelie, D., Borremans, B., Provoost, A., de Lorenzo, V., Brown, N. L., Lloyd, J. R., Hobman, J. L., Csöregi, E., Johansson, G., Mattiasson, B. 1999. Whole cell- and protein-based biosensors for the detection of bio available heavy metals in environmental samples. *Anal Chim Acta*. 387:235-244.

Cornelis, G. R., Van Gijsegem, F. 2000**.** Assembly and function of type III secretory systems. *Annu Rev Microbiol.* 54:735-774.

Coulson R. M, Enright A. J, Ouzounis C. A. 2001 Transcription-associated protein families are primarily taxon-specific. *Bioinformatics.* 17:95-97.

Cross J. T., Jacobs R. F. 1993 Tularemia: treatment failures with outpatient use of ceftriaxone. *Clin Infect Dis*. 17:976-980.

Cunnac, S., Occhialini, A., Barberis, P., Boucher, C., Genin, S. 2004. Inventory and functional analysis of the large Hrp regulon in *Ralstonia solanacearum*: identification of novel effector proteins translocated to plant host cells through the type III secretion system. *Mol Microbiol.* 53:115-128.

Frausto da Silva, J. J. R., Williams, R. J. P. 2001. The Biological Chemistry of the Elements: the Inorganic Chemistry of Life. Oxford University Press, New York.

Dandekar, T., Snel, B., Huynen, M., Bork, P. 1998. Conservation of gene order: a finger-print of proteins that physically interact. *Trends Biochem Sci.* 23:324-328.

Dartigalongue, C., Missiakas, D., Raina, S. 2001. Characterization of the *Escherichia coli* sigma E regulon. *J Biol Chem*. 276:20866-75.

Das, S.,Saier, M. H. Jr., Elkan, C. 2007. A Pipeline to Automate the Updating of a Specialized Protein Database. *Bioinformatics*

Davies J. C., Rubin B. K. 2007. Emerging and unusual gram-negative infections in cystic fibrosis. *Semin Respir Crit Care Med*.28:312-21.

Degtyarenko,K. N., North,A. C. T., Findlay,J. B. C. 1999. PROMISE: a database of bioinorganic motifs. *Nucleic Acids Res*. 27:233–236.

Dehal, P. S., Boore, J. L. 2006. A phylogenomic gene cluster resource: The Phylogenetically Inferred Groups (PhIGs) database. *BMC Bioinformatics. 7*:201.

Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, 0., Salzberg, S. L. 1999. Alignment of whole genomes. *Nucleic Acids Res*. 27:2369-2376.

Denef, V. J., Park, J., Tsoi, T. V., Rouillard, J. M., Zhang, H., Wibbenmeyer, J. A., Verstraete, W., Gulari, E., Hashsham, S. A., Tiedje, J. M. 2004. Biphenyl and benzoate metabolism in a genomic context:outlining genome-wide metabolic networks in *Burkholderia xenovorans* LB400. *Appl Environ Microbiol*. 70:4961-70.

Denef, V. J., Konstantinidis, K. T., Vergez, L. M., Agullo, L., Reyes, V. L., Hauser, L., Cordova, M., Gomez, L., Gonzalez, M., Land, M., Lao, V., Larimer, F., LiPuma, J. J., Mahenthiralingam, E., Malfatti, S. A.,

Marx, C. J., Parnell, J. J., Ramette, A., Richardson, P., Seeger, M., Smith, D., Spilker, T., Sul, W. J., Tsoi, T. V., Ulrich, L. E., Zhulin, I. B., Tiedje, J. M. 2006. *Burkholderia xenovorans* LB400 harbors a multi-replicon., 9.73-Mbp genome shaped for versatility. *Proc Natl Acad Sci.* 103:15280-7.

Devereux, J., Haeberli, P., Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387-95.

Diels, L., Faelen, M., Mergeay, M., Nies, D. H. 1985. Mercury transposons from plasmids governing multiple resistance to heavy metals in *Alcaligenes eutrophus* CH34. *Arch Int Physiol Biochem.* 93:B27-B28.

Diels, L., Dong, Q., van der Lelie, D., Baeyens, W., Mergeay, M. 1995. The *czc* operon of *Alcaligenes eutrophus* CH34: from resistence mechanism to the removal of heavy metals. *J Ind Microbiol.* 14:142-53.

Diels, L., De Smet, M., Hooyberghs, L., Corbisier, P. 1999. Heavy metals bioremediation of soil. *Mol Biotechnol.* 12:149-158.

Dinh, D., Paulsen, I. T., Saier, M. H. Jr. 1994. A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of gram-negative bacteria. *J Bacteriol.* 176:3825-3831.

Dong, Q., Mergeay, M. 1994. Czc/cnr efflux: a three-component chemiosmotic antiport pathway with a 12-transmembrane-helix protein. *Mol Microbiol.* 14:185-187.

Dräger, A. 2005. Automatische und vergleichende Analyse bakterieller Genome mit Schwerpunkt auf *Ralstonia/Cupriavidus* Arten sowie verwandten Proteobakterien. *Diplomarbeit. Martin-Luther-Universität Halle-Wittenberg.*

Dressler, C., Kues, U., Nies, D. H., Friedrich, B. 1991. Determinants encoding resistance to several heavy metals in newly isolated copper-resistant bacteria. *Appl Environ Microbiol.* 57:3079-3085.

Drew, D., Sjostrand, D., Nilsson, J., Urbig, T., Chin, C. N., de Gier, J. W., von Heijne, G. 2002. Rapid topology mapping of *Escherichia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis. *Proc Natl Acad Sci.* 99:2690-2695.

Dröge, M., Pühler, A., Selbitschka, W. 1998. Horizontal gene transfer as a biosafety issue: a natural phenomenon of public concern. *J Biotechnol.* 64:75–90

Dröge, M., Pühler, A., Selbitschka, W. 1999. Horizontal gene transfer among bacteria in terrestrial and aquatic habitats as assessed by microcosm and field studies. *Biol Fertil Soils*. 29:221–245.

Dufresne, A., Ostrowski, M., Scanlan, D. J., Garczarek, L., Mazard, S., Palenik, B. P., Paulsen, I. T., de Marsac, N. T., Wincker, P., Dossat, C., Ferriera, S., Johnson, J., Post, A. F., Hess, W. R., Partensky, F. 2008. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol.* 9:R90.

Eitinger, T., Friedrich, B. 1991. Cloning, nucleotide sequence, and heterologous expression of a high-affinity nickel transport gene from *Alcaligenes eutrophus. J Biol Chem*. 266:3222-3227.

Eitinger, T., Wolfram, L., Degen, O., Anthon, **C.** 1997. A Ni^{2+} binding motif is the basis of high affinity transport of the *Alcaligenes eutrophus* nickel permease. *J Biol Chem*. 272:17139-17144.

Eitinger T., Suhr J., Moore L., Smith J. A. C. 2005 Secondary transporters for nickel and cobalt ions: Theme and variations. *Biometals*. 18:399-405.

Eng, B. H., Guerinot, M. L., Eide, D., Saier, M. H. Jr. 1998. Sequence analyses and phylogenetic characterization of the ZIP family of metal ion transport proteins. *J Membr Biol*. 166:1-7.

Enright, A. J., Iliopoulos, I., Kyrpides, N. C., Ouzounis, C. A. 1999. Protein interaction maps for complete genomes based on gene fusion events. *Nature*. 402:86-90.

Fahey, R. C. 2001. Novel thiols of prokaryotes. *Annu Rev Microbiol.* 55:333-356.

Fahey, R. C., Brown, W. C., Adams, W. B., Worsham, M. B. 1978. Occurrence of glutathione in bacteria. *J Bacteriol.* 133:1126–29.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J. A., Hofmann, K., Bairoch, A. 2002. The PROSITE database, its status in 2002. *Nucleic Acids Res*. 30:235–238.

Feng, D. F., Doolittle, R. F. 1990. Progressive alignment and phylogenetic tree construction of protein sequences. *Methods Enzymol*. 183:375-87.

Fenton, H. J. H. 1876. On a new reaction of tartaric acid. *Chem News*. 33:190- 190.

Fichant, G. A., Quentin, Y. 1995. A frameshift error detection algorithm for DNA sequencing projects. *Nucleic Acids Res*. 23:2900-8.

Finney, L. A., O'Halloran, T. V. 2003. Transition metal speciation in the cell:insights from the chemistry of metal ion receptors. *Science*. 300:931-6.

Fitch, W. M. 1970**.** Distinguishing homologous from analogous proteins. *Syst Zool*. 19:99-106.

Fitch, W. M., Ayala F. J. 1994 Tempo and mode in evolution *Proc Natl Acad Sci.* 91:6717-6720.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A. Merrick, J. M. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*. 269:496–512.

Franke, S., Grass, G., Rensing, C., Nies, D. H. 2003. Molecular analysis of the copper-transporting CusCFBA efflux system from *Escherichia coli*. *J Bacteriol*. 185:3804-3812.

Frankel, R. B., Papaefthymiou, G. C., Blakemore, R. P., O'Brien, W. 1983. Fe_3O_4 precipitation in magnetotactic bacteria. *Biochim Biophys Acta*. 763:147–159.

Frankel, R. B., Bazylinski, D. A., Johnson, M. S., Taylor, B. L. 1997. Magneto-aerotaxis in marine coccoid bacteria. *Biophys J.* 73:994–1000.

Fraser, C. M., Eisen, J. A., Nelson, K. E., Paulson, I. T., Salzberg, S. L. 2002. The value of complete microbial genome sequencing (you get what you pay for). *J Bacteriol*. 184:6403-6405.

Frost, L. S., Leplae, R., Summers, A. O. Toussaint, A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol*. 3:722–732.

Fukumori, Y., Oynagi, H., Yoshimatsu, K., Noguchi, Y., Fujiwara, T. 1997. Enzymatic iron oxidation and reduction in magnetite synthesizing *Magnetospirillum magnetotacticum. J Phys* IV. 7:659–662.

Fulton, D. L., Li, Y. Y., Laird1, M. R., Horsman, B. G. S., Roche F. M., Brinkman F. S. L., 2006 Improving the specificity of high-throughput ortholog prediction *Bioinformatics*. 7:270.

Gabriel, D. W., Allen, C., Schell, M., Denny T. P., Greenberg J. T., Duan Y. P., Flores-Cruz Z., Huang Q., Clifford J. M., Presting G., Gonzalez E. T., Reddy J., Elphinstone J., Swanson J., Yao J., Mulholland V., Liu L., Farmerie W., Patnaikuni M., Balogh B., Norman D., Alvarez A., Castillo J. A., Jones J., Saddler G., Walunas T., Zhukov A., Mikhailova N. 2006. Identification of open reading frames unique to a select agent: *Ralstonia solanacearum* race 3 biovar 2. *Mol Plant Microbe Interact*. 19:69-79.

Gadd, G. M. 1992. Metals and microorganisms:a problem of definition. *FEMS Microbiol Lett*. 79:197-203.

Galán, J. E., Collmer, A. 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science*. 284:1322-1328.

Gantt, E. 1994 In: Bryant DA (ed) The molecular biology of cyanobacteria. *Kluwer Dordrecht.* 119-138

Gatti, D., Mitra, B., Rosen, B. P. 2000. *Escherichia coli* soft metal ion-translocating ATPases. *J Biol Chem*. 275:34009-12.

Gauthier, A., Thomas, N. A., Finlay, B. B. 2003. Bacterial injection machines. *J Biol Chem*. 278:25273-25276.

Gehring, W. J., Affolter, M., Burglin, T.. 1994. Homeodomain proteins. *Annu Rev Biochem.* 63:487–526.

Gene Ontology Consortium. 2004, The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 32:D258-D261.

Genin, S., Boucher, C. 2002. *Ralstonia solanacearum*: secrets of a major pathogen unveiled by analysis of its genome. *Mol Plant Pathol*. 3:111-118.

Genin, S., Boucher, **C.** 2004. Lessons learned from the genome analysis of *Ralstonia solanacearum*. *Annu Rev Phytopathol.* 42:107-134.

Goesmann, A., Linke, B., Rupp, O., Krause, L., Bartels, D., Dondrup, M., McHardy, A.C., Wilke, A., Pühler, A., Meyer, F., 2003. Building a BRIDGE for the integration of heterogeneous data from functional genomics into a platform for systems biology. *J Biotechnol*. 106:157–167.

Gogarten, J. P., Olendzenski, L. 1999. Orthologs, paralogs and genome comparisons. *Curr Opin Genet Dev*. 9:630-636.

Goldberg, M., Pribyl, T., Juhnke, S., Nies, D. H. 1999 Energetics and topology of CzcA, a cation/proton antiporter of the resistance-nodulation-cell division protein family. *J Biol Chem*. 274:26065-26070.

Gordon, D., Abajian, C., Green, P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res*. 11:614–625.

Goris, J., De Vos, P., Coenye, T., Hoste, B., Janssens, D., Brim, H., Diels, L., Mergeay, M., Kersters, K., Vandamme, P. 2001 Classification of metal-resistant bacteria from industrial biotopes as *Ralstonia campinensis* sp. nov., *Ralstonia metallidurans* sp. nov. and *Ralstonia basilensis* Steinle *et al*. 1998 emend. *Int J Syst Evol Microbio*. l51:1773-1782.

Grass, G. 2000 Molekulargenetische und biochemische Charakterisierung der cnr Cobalt/Nickel-Resistenz-Determinante aus *Ralstonia metallidurans* CH34. *Dissertation. Martin-Luther-Universität Halle-Wittenberg. Halle.*

Grass, G., Rensing, C. 2001. Genes involved in copper homeostasis in *Escherichia coli. J Bacteriol*. 183:2145-2147.

Grass, G., Rensing, C., 2001b. CueO is a multi-copper oxidase that confers copper tolerance in *Escherichia coli*. *Biochem Biophys Res Commun.* 286:902-8.

Grass, G., Fan, B., Rosen, B. P., Lemke, K., Schlegel, H. G., Rensing. **C.** 2001. NreB *from Achromobacter xylosoxidans* 31A is a nickel-induced transporter conferring nickel resistance. *J Bacteriol*. 183:2803–2807

Grass, G., Thakali, K., Klebba, P. E., Thieme, D., Müller, A., Wildner, G. F., Rensing, C. 2004. Linkage between catecholate siderophores and the multicopper oxidase CueO in *Escherichia coli*. *J Bacteriol*. 186:5826-5833.

Grass, G., Franke, S., Taudte, N., Nies, D. H., Kucharski, L. M. Maguire, M. E., Rensing, C. 2005. The metal permease ZupT from *Escherichia coli* is a transporter with a broad substrate spectrum. *J Bacteriol*. 187:1604-1611.

Grass, G., Haney, C., Otto, M., Rensing, C., Nies, D. H., Munkelt, D. 2005b. FieF (YiiP) from *E. coli* mediates decreased cellular accumulation of iron and relieves iron stress. *Arch Microbiol.* 183:9-18.

Grass, G., 2006 Biometall-Homöostase in *Escherichia coli Habilitationsschrift Martin-Luther-Universität Halle-Wittenberg. Halle.*

Grass, G. 2006b. Iron transport in *Escherichia coli*: All has not been said and done. *BioMetals*. 19:159–172.

Gribskov, M., McLachlan, A. D., Eisenberg, D. 1987. Profile Analysis: Detection of distantly related Proteins. *Proc Natl Acad Sci.* 84:4355-4358.
Gribskov, M., Veretnik, S. 1996. Identification of sequence pattern with profile analysis. *Methods Enzymol.* 266:198-212.
Grosse, C., Grass, G., Anton, A., Franke, S., Santos, A. N., Lawley, B., Brown, N. L., Nies, D. H. 1999. Transcriptional organization of the *czc* heavy-metal homeostasis determinant from *Alcaligenes eutrophus*. *J Bacteriol*. 181:2385-93.
Grosse, C., Anton, A., Hoffmann, T., Franke, S., Schleuder, G., Nies, D. H. 2004 Identification of a relulatory pathway that controls the heavy-metal resistance Czc via promoter *czcNp* in *Ralstonia metallidurans*. *Arch Microbiol.* 182:109-118.
Grosse, C., Friedrich, S., Nies, D. H., 2007 Contribution of extracytoplasmic function sigma factors to transition metal homeostasis in *Cupriavidus metallidurans* strain CH34. *J Mol Microbiol Biotechnol. 1*2:227-240.
Grotz, N., Fox, T., Connolly, E., Park, W., Guerinot, M. L., Eide, D. 1998. Identification of a family of zinc transporter genes from *Arabidopsis* that respond to zinc deficiency. *Proc Natl Acad Sci..* 95:7220-7224.
Grover, A., Azmi, W., Gadewar, A. V., Pattanayak, D., Naik, P. S., Shekhawat, G. S., Chakrabarti, S. K. 2006. Genotypic diversity in a localized population of *Ralstonia solanacearum* as revealed by random amplified polymorphic DNA markers *J Appl Microbiol*. 101:798-806.
Grünberg, K., Wawer, C., Tebo, B. M., Schüler, D. 2001. A large gene cluster encoding several magnetosome proteins is conserved in different species of magnetotactic bacteria. *Appl Environ Microbiol*. 67:4573–4582.
Guan, L., Nakae, T. 2001. Identification of essential charged residues in transmembrane segments of the multidrug transporter MexB of *Pseudomonas aeruginosa*. *J Bacteriol*. 183:1734-9.
Guerinot, M. L. 2000. The ZIP family of metal transporters. *Biochim Biophys Acta*. 1465:190-8.
Guex, N., Peitsch, M. C. 1997. SWISS-MODEL and the Swiss- PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 18:2714–2723.
Guidot, A., Prior, P., Schoenfeld, J., Carrere, S., Genin, S., Boucher, C. 2007. Genomic structure and phylogeny of the plant pathogen *Ralstonia solanacearum* inferred from gene distribution analysis. *J Bacteriol. 1*89:377-387.
Gupta, A., Matsui, K., Lo, J. F., Silver, S. 1999. Molecular basis for resistance to silver cations in *Salmonella*. *Nat Med*. 5:183-8.
Hall, R. M. 1997. Mobile gene cassettes and integrons: moving antibiotic resistance genes in Gram-negative bacteria. *Ciba Found Symp*. 207:192–202; discussion 202–205.
Haney, C. J., G. Grass, S. Franke, Rensing, C. 2005. New developments in the understanding of the cation diffusion facilitator family. *J Ind Microbiol Biotechnol*. 32:215- 226.
Hantke, K. 2005. Bacterial zinc uptake and regulators. *Curr Opin Microbiol*. 8:196–202.
Harris, R. M., Webb, D. C., Howitt, S. M., Cox, G.B. 2001. Characterization of PitA and PitB from *Escherichia coli*. *J Bacteriol.* 183:5008-5014.
Hauser, L., Larimer, F., Land, M., Shah, M., Uberbacher, E. 2004. Analysis and annotation of microbial genome sequences. *Genet Eng*. 26:225-238.
Hausner, M., Wuertz, S. 1999. High rates of conjugation in bacterial biofilms as determined by quantitative in situ analysis. *Appl Environ Microbiol*. 65:3710-3713.
Hearn, E. M., Gray, M. R., Foght, J. M. 2006. Mutations in the central cavity and periplasmic domain affect efflux activity of the resistance-nodulationdivision pump EmhB from *Pseudomonas fluorescens* cLP6a. *J Bacteriol*. 188:115-123.
Hebbeln, P., Eitinger, T. 2004. Heterologous production and characterization of bacterial nickel/cobalt permeases. *FEMS Microbiol Lett.* 230:129-135.
Helbig, K. 2003. Aufklärung der Funktion des Schwermetall-Transporters CzcA aus *Ralstonia metallidurans* CH34 durch gerichtete Mutagenese. *Diplomarbeit. Martin-Luther-Universität Halle-Wittenberg, Halle.*
Helbig, K., Bleuel, C., Krauss, G. J., Nies D. H. 2008. Glutathione and transition metal homeostasis in *Escherichia coli*. *J Bacteriol.* 190:5431-8.
Helbig, K., Grosse, C., Nies, D. H. 2008. Cadmium toxicity in glutathione mutants of *Escherichia coli*. *J Bacteriol.* 190:5439-54.
Henikoff, S, Henikoff, J. G., Pietrokovski, S. 1999. Blocks: A nonredundant database of protein alignment blocks derived from multiple compilations. *Bioinformatics*. 15:471–479.
Henikoff, S., Henikoff, J. G. 2000 Amino acid substitution matrices. *Adv Protein Chem.,* 54:73-97.
Higgins, M. K., Eswaran, J., Edwards, P., Schertler, G. F., Hughes, C., Koronakis, V. 2004. Structure of the ligand-blocked periplasmic entrance of the bacterial multidrug efflux protein TolC. *J Mol Biol.* 342:697–702
Higgins, C. F. 2007. Multiple molecular mechanisms for multidrug resistance transporters. *Nature*. 446:749-757.
Hoffmann, T. 2001. Untersuchung der Regulation der Schwermetallresistenz in *R. metallidurans* CH34 durch Zweidimensionale Gelelektrophorese und Reportergenfusion. *Diplomarbeit. Martin-Luther-Universität Halle-Wittenberg, Halle.*
Holmes, A., Govan, J., Goldstein, R. 1998. Agricultural use of *Burkholderia* (*Pseudomonas*) *cepacia* :a threat to human health?. *Emerg Infect Dis*. 4:221-7.
Huang, X., Madan, A. 1999 CAP3: a DNA sequence assembly program. *Genome Res.* 9:868–877.

Hubert, R. S., Vivanco, I., Chen, E., Rastegar, S., Leong, K., Mitchell, S. C., Madraswala, R., Zhou, Y., Kuo, J., Raitano, A. B., Jakobovits, A., Saffran, D. C., Afar, D. E. 1999. STEAP: A prostate-specifi c cell surface antigen highly expressed in human prostate tumors. *Proc Natl Acad Sci*. 96:14523–14528.

Jimenez, L. 2007. Microbial diversity in pharmaceutical product recalls and environments. *PDA J Pharm Sci Technol*. 61:383-99.

Johnson, B. F., Stanier, R. Y. 1971. Dissimilation of aromatic compounds by *Alcaligenes eutrophus. J Bacteriol*. 107:468-75.

Johnson, C. M. and R. F. Schleif. 1995. In vivo induction kinetics of the arabinose promoters in Escherichia coli. *J Bacteriol*. 177:3438-3442.

Johnson, J. M., Church, G. M. 1999. Alignment and structure prediction of divergent protein families: periplasmic and outer membrane proteins of bacterial efflux pumps. *J Mol Biol*. 287:695-715.

Jordan, I. K., Rogozin, I. B., Wolf, Y. I., Koonin, E. V. 2002. Microevolutionary genomics of bacteria. *Theor Popul Biol*. 61:435–447.

Jothi, R., Zotenko, E., Tasneem, A., Przytycka1, T. M. 2006. COCO-CL: hierarchical clustering of homology relations based on evolutionary correlations. *Bioinformatics*. 22:779–788.

Juhnke, S., Peitzsch, N., Hubener, N., Grosse, C., Nies, D. H. 2002. New genes involved in chromate resistance in *Ralstonia metallidurans* strain CH34. *Arch Microbiol*. 179:15-25.

Kachur, A. V., Koch, C. J., Biaglow, J. E. 1998. Mechanism of copper-catalyzed oxidation of glutathione. *Free Radic Res*. 28:259-69.

Kammler, M., Schön, C., Hantke, K. 1993. Characterization of the ferrous iron uptake system of *Escherichia coli. J Bacteriol*. 175:6212-6219.

Kanehisa, M., Goto, S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 28:27–30.

Karley, A. J., Leigh, R. A., Sanders, D. 2000. Where do all the ions go? The cellular basis of differential ion accumulation in leaf cells (Review). *Trends Plant Sci*. 5:465–470.

Karwath, A., King, R. D. 2002. Homology Induction: the use of machine learning to improve sequence similarity searches. *BMC Bioinformatics*. 3:11.

Kataoka, K., Komori, H., Ueki, Y., Konno, Y., Kamitaka, Y., Kurose, S., Tsujimura, S., Higuchi, Y., Kano, K., Seo, D., Sakurai, T. 2007. Structure and Function of the Engineered Multicopper Oxidase CueO from *Escherichia coli* Deletion of the Methionine-Rich Helical Region Covering the Substrate-Binding Site *J Mol Biol*. 373:141-152.

Katayama, A., Tsujii, A., Wada, A., Nishino, T., Ishihama, A. 2002. Systematic search for zinc-binding proteins in *Escherichia coli*. *Eur J Biochem*. 269:2403-2413.

Katz, A. K, Glusker, J. P, Beebe, S. A, Bock, C. W. 1996. Calcium ion coordination: A comparison with that of beryllium, magnesium, and zinc. *J Am Chem Soc*. 118:5752-5763.

Katz, L. A. 1996. Transkingdom transfer of the phosphoglucose isomerase gene. *J Mol Evol*. 43:453-459.

Katzen, F., Beckwith, J. 2000. Transmembrane electron transfer by the membrane protein DsbD occurs via a disulfide bond cascade. *Cell*. 103:769–779.

Thomas-Keprta, K. L., D. A. Bazylinski, J. L. Kirchvink, S. J. Clemett, D. S. McKay, S. J. Wentworth, H. Vali, E. K. Gibson, Jr., Romanek, C. S. 2000. Elongated prismatic magnetite crystals in ALH84001 carbonate globules: potential Martian magnetofossils. *Geochim Cosmochim Acta* 64:4049–4081**.**

Kershaw, C. J., Brown, N. L., Constantinidou, C., Patel, M. D., Hobman, J. L. 2005**.** The expression profile of *Escherichia coli* K-12 in response to minimal, optimal and excess copper concentrations. *Microbiology*. 151:1187– 1198.

Kill, K., Binnewies, T. T., Sicheritz-Ponten, T., Willenbrock, H., Hallin, P. F., Wassenaar, T. M., Ussery, D. W. 2005. Genome update: sigma factors in 240 bacterial genomes. *Microbiology*. 151:3147-50.

Kim, C., Lorenz, W. W., Hoopes, J. T., Dean, J. F. 2001. Oxidation of phenolate siderophores by the multicopper oxidase encoded by the *Escherichia coli* yacK gene. *J Bacteriol*. 183:4866–4875.

Kimball, R. A., Martin, L., Saier, M. H. 2003. Reversing transmembrane electron flow :the DsbD and DsbB protein families. *J Mol Microbiol Biotechnol*. 5:133-49.

Kittleson, J. T., Loftin, I. R., Hausrath, A. C., Engelhardt, K. P., Rensing, C., McEvoy, M. M. 2006. Periplasmic metal-resistance protein CusF exhibits high affinity and specificity for both Cu-I and Ag-I. *Biochemistry*. 45:11096-11102.

Knutson, M. D. 2007. Steap proteins: implications for iron and copper metabolism. *Nutr Rev*. 65:335-40.

Koonin, E. V. 2005. Orthology, paralogy, and evolutionary genomics. *Annu Rev Genet*. 39:309-38.

Koonin, E. V., Tatusov, R. L., Rudd, K. E. 1995 Sequence similarity analysis of *Escherichia coli* proteins: Functional and evolutionary implications. *Proc Natl Acad Sci*. 92:11921–11925.

Koonin, E. V., Tatusov, R. L., Rudd, K. E. 1996. Protein sequence comparison at genome scale. *Methods Enzymol*. 266:295-322.

Koppenol, W. H. 2001. The Haber-Weiss cycle--70 years later. *Redox Rep*. 6:229-234.

Kormutakova, R., Klucar, L., Turna, J. 2000. DNA sequence analysis of the tellurite-resistance determinant from clinical strain of *Escherichia coli* and identification of essential genes. *Biometals*. 13:135-139.

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., Hughes, C. 2000. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature*. 405:914-9.

Krauss, G., Bärlocher, F., Schreck, P., Wennrich, R., Glässer, W., Krauss, G. J. 2001. Aquatic hyphomycetes occur in hyperpolluted waters in Central Germany. *Nova Hedwigia*. 72:419–428.

Krieger, E., Koraimann, G., Vriend, G. 2002. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins*. 47:393–402.

Krupp, R., Chan, C., Missiakas, D. 2001. DsbD-catalyzed transport of electrons across the membrane of *Escherichia coli. J Biol Chem*. 276:3696–3701.

Kunin, V., Ouzounis, C. A. 2003. The balance of driving forces during genome evolution in prokaryotes. *Genome Res*. 13:1589–1594.

Labedan, B, Riley, M., 1999. Genetic inventory: *Escherichia coli* as a window on ancestral proteins. In Organization of the Prokaryotic Genome. Edited by Charlebois R. Washington DC: *ASM Press*:311-329.

Lalani, E. N., Laniado, M. E., Abel, P. D. 1997. *Cancer Metastasis Rev*. 16:29–66.

Larbig, K. D., Christmann, A., Johann, A., Klockgether, J., Hartsch, T., Merkl, R., Wiehlmann, L., Fritz, H. J., Tuemmler, B. 2002 Gene islands integrated into tRNAGly genes confer genome diversity on a *Pseudomonas aeruginosa* clone. *J Bacteriol*. 184:6665-6680.

Ledin, M. 2000. Accumulation of metals by microorganisms - processes and importance for soil systems. *Earth Sci Rev*. 51:1-31.

Ledrich, M. L., Stemmler, S., Laval-Gilly, P., Foucaud, L., Falla, J. 2005. Precipitation of silver-thiosulfate complex and immobilization of silver by *Cupriavidus metallidurans* CH34. *BioMetals*. 18:643-650.

Lee, S. Y. 2006. Deciphering bioplastic production. *Nat Biotechnol*. 24:1227-9.

Legatzki, A. 2003 Charakterisierung zweier CPx-Typ ATPasen und deren Zusammenwirken mit anderen Metal-Effluxsystemen in *Ralstonia metallidurans*. *Dissertation. Martin-Luther-Universität Halle-Wittenberg. Halle.*

Legatzki, A., Grass, G., Anton, A., Rensing, C., Nies, D. H. 2003. Interplay of the Czc system and two P-type ATPases in conferring metal resistance to *Ralstonia metallidurans. J Bacteriol*. 185:4354-4361.

Lejeune, P., Mergeay, M., Van Gijsegem, F., Faelen, M., Gerits, J., Toussaint, A. 1983. Chromosome transfer and R-prime plasmid formation mediated by plasmid pULB113 (RP4: :mini-Mu) in *Alcaligenes eutrophus* CH34 and *Pseudomonas fluorescens*. 6.2. *J Bacteriol.* 155:1015-1026.

Lengeler, J. W. 1990. Molecular analysis of the enzyme II complexes of the bacterial phosphotransferase system (PTS) as carbohydrate transport system. *Biochim Biophys Acta*. 1018:155-159.

Lessie, T. G., Hendrickson, W., Manning, B. D., Devereux, R. 1996. Genomic complexity and plasticity of *Burkholderia cepacia*. *FEMS Microbiol Lett*. 144:117-128.

Li, X. Z., Nikaido, H., Poole, K. 1995. Role of MexA-MexB-OprM in antibiotic efflux in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother*. 39:1948-1953.

Liesegang, H., Lemke, K., Siddiqui, R. A., Schlegel, H. G. 1993. Characterization of the inducible nickel and cobalt resistance determinant *cnr* from pMOL28 of *Alcaligenes eutrophus* CH34. *J Bacteriol.* 175:767-778.

Liochev, S. I., Fridovich, I. 2002. The Haber-Weiss cycle—70 years later: an alternative view. *Redox Rep*. 7:55-57, 59-60.

Lobedanz, S., Bokma, E., Symmons, M. F., Koronakis, E., Hughes, C. Koronakis, V. 2007. A periplasmic coiled-coil interface underlying TolC recruitment and the assembly of bacterial drug efflux pumps. *Proc Natl Acad Sci*.. 104:4612-4617**.**

Locher, K. P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J. P., Moras, D. 1998. Transmembrane signaling across the ligand-gated FhuA receptor: crystal structures of free and ferrichrome-bound states reveal allosteric changes. *Cell*. 95:771-778.

Lodewyckx, C., Taghavi, S., Mergeay, M., Vangronsveld, J., Clijsters, H., van der Lelie, D. 2001. The effect of recombinant heavy metal resistant endophytic bacteria in heavy metal uptake by their host plant. *Int J Phytoremediat*. 3:173-187.

Lomovskaya, O., Totrov, M. 2005. Vacuuming the periplasm. *J Bacteriol*. 187:1879-1883.

Lonetto, M., Gribskov, M., Gross, C. A. 1992. The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol*. 174:3843–3849

Lonetto, M. A., Brown, K. L., Rudd, K. E., Buttner, M. J. 1994. Analysis of the *Streptomyces coelicolor* sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci*. 91:7573-7.

Loschi, L., Brokx, S. J., Hills, T. L., Zhang, G., Bertero, M. G., Lovering, A. L., Weiner, J. H., Strynadka, N. C. 2004. Structural and biochemical identification of a novel bacterial oxidoreductase. *J Biol Chem*. 279:50991-50400.

Lovely, D. R. 2000. Environmental Microbe-Metal Interactions. *ASM Press, Washington, D.C.*

Ludwig, W., Klenk, H. P. 2000. Overview:A Phylogenetic Backbone and Taxonomic Framework for Procaryotic Systematics. 49-65. *In D. R. Boone, R. W. Castenholz, and G. M. Garrity (ed.), Bergey's manual of systematic bacteriology, 2nd ed., vol. 1. Springer-Verlag, New York, N.Y*

Lunin, V., Dobrovetsky, E., Khutoreskaya, G., Zhang, R., Joachimiak, A., Doyle, D. A., Bochkarev, A., Maguire, M. E., Edwards, A. M., Koth, C. M. 2006 Crystal structure of the CorA Mg^{2+} transporter. *Nature*. 440:833-837

Ma, D., Cook, D. N., Alberti, M., Pon, N. G., Nikaido, H., Hearst, J. E. 1993. Molecular cloning and characterization of acrA and acrE genes of *Escherichia coli. J Bacteriol.* 175:6299-313.

Maidak., B. L., Cole., J. R., Lilburn., T. G., Parker, C. T., Saxman., P. R., Farris., R. J., Garrity., G. M., Olsen., G. J., Schmidt., T. M., Tiedje.,J. M. 2001. The RDP-II Ribosomal Database Project. *Nucleic acids Res.* 29:73- 74.

Maloney, P. C. 1990**.** Microbes and membrane biology. *FEMS Microbiol Rev*. 87:91-102.

Maloney, P. C. 1992**.** The molecular and cell biology of anion transport by bacteria. *BioEssays*. 14:757-762.

Mao, W. *et al*. 2002. On the mechanism of substrate specificity by resistance nodulation division (RND)-type multidrug resistance pumps: The large periplasmic loops of MexD from *Pseudomonas aeruginosa* are involved in substrate recognition. *Mol Microbiol*. 46:889-901.

Maret, W. 2001. Crosstalk of the group IIa and IIb metals calcium and zinc in cellular signaling. *PNAS*. 98:12325–12327

Maret, W. 2004**.** Zinc and Sulfur: A Critical Biological Partnership. *Biochemistry*. 43:3302-3309.

Markowitz, V. M, Korzeniewski, F., Palaniappan, K., *et al.* 2006. The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* 34:D344-D348.

Martinez-Salazar, J. M., Moreno, S., Najera, R., Boucher, J. C., Espin, G., Soberon-Chavez, G., Deretic, V. 1996. Characterization of the genes coding for the putative sigma factor AlgU and its regulators MucA., MucB., MucC and MucD in *Azotobacter vinelandii* and evaluation of their roles in alginate biosynthesis. *J Bacteriol.* 178:1800-8.

McCutcheon, J. P., Moran, N. A. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc Natl Acad Sci*. 104:19392-7.

McKay, D. S., E. K. Gibson, Jr., K. L. ThomasKeprta, H. Vali, C. S. Romanek, S. J. Clemett, X. D. F. Chillier, C. R. Maechling, Zare, R. N. 1996. Search for past life on Mars: possible relic biogenic activity in Martian meteorite ALH84001. *Science* 273:924–930.

McGinn, P. J., Price, G. D., Maleszka, M., Badger, M. R. 2003. Inorganic Carbon Limitation and Light Control the Expression of Transcripts Related to the CO_2-Concentrating Mechanism in the Cyanobacterium *Synechocystis* sp. Strain PCC68031. *Plant Physiol.. 1*32:218-229.

Mergeay, M, Houba, C, Gerits, J. 1978. Extrachromosomal inheritance controlling resistance to cadmium, cobalt and zinc ions: evidence for curing a *Pseudomonas*. *Arch Int Physiol Biochim*. 86:440–441.

Mergeay, M., Nies, D. H., Schlegel, H. G., Gerits, J., Charles, P., van Gijsegen, F. 1985. *Alcaligenes eutrophus* CH34 is a fakultative chemolithotroph with plasmid bound resistance to heavy metals. *J Bacteriol.* 162:328-334.

Mergeay, M. 2000. Bacteria adapted to industrial biotopes:Metal-resistant *Ralstonia metallidurans Bacterial Stress Responses ASM Press Washington. 2*6:403-414.

Mergeay, M., Monchy, S., Vallaeys, T., Auquier, V., Benotmane, A., Bertin, P., Taghavi, S., Dunn, J., van der Lelie, D., Wattiez, R. 2003. *Ralstonia metallidurans*., a bacterium specifically adapted to toxic metals:towards a catalogue of metal-responsive genes. *FEMS Microbiol Rev*. 27:385-410.

Metzgar, D., Wills, C. 2000. Evolutionary changes in mutation rates and spectra and their influence on the adaptation of pathogens. *Microbes Infect*. 2:1513- 22.

Mills, S. D., Lim, C. K., Cooksey, D. A. 1994. Purification and characterization of CopR., a transcriptional activator protein that binds to a conserved domain (*cop* box) in copper-inducible promoters of *Pseudomonas syringae*. *Mol Gen Genet.* 244:341-51.

Missiakas, D., Mayer, M. P., Lemaire, M., Georgopoulos, C., Raina, S. 1997. Modulation of the *Escherichia coli* sigmaE (RpoE) heat-shock transcription-factor activity by the RseA., RseB and RseC proteins. *Mol Microbiol.* 24:355-71.

Missiakas, D., Raina, S. 1998. The extracytoplasmic function sigma factors : role and regulation. *Mol Microbiol.* *2*8:1059-66.

Monchy, S., Benotmane, M. A., Wattiez, R., van Aelst, S., Auquier, V., Borremans, B. *et al.* 2006. Transcriptomic and proteomic analyses of the pMOL30encoded copper resistance in *Cupriavidus metallidurans* strain CH34. *Microbiology*. 152:1765–1776.

Monchy, S., Benotmane, M. A., Janssen, P., Vallaeys, T., Taghavi, S., van der Lelie, D., Mergeay, M. 2007. Plasmids pMOL28 and pMOL30 of *Cupriavidus metallidurans* are specialized in the maximal viable response to heavy metals. *J Bacteriol*. 189:7417–7425.

Moran, N. A. 2007. Symbiosis as an adaptive process and source of phenotypic complexity. *PNAS*. 104:8627– 863

Moxon, E. R., Rainey, P. B., Nowak, M. A., Lenski, R. E. 1994 Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol*. 4:24-33.

Moxon, E. R., Thaler, D. S. 1997. Microbial genetics. The tinkerer's evolving toolbox. *Nature*. 387:659:661–652.

Mukhopadhyay, R., Rosen, B. P., Phung, L. T., Silver, S. 2002. Microbial arsenic: from geocycles to genes and enzymes. *FEMS Microbiol Rev*. 26:311–325.

Munkelt, D., Grass, G., Nies, D. H., 2004. The chromosomally encoded cation diffusion facilitator proteins DmeF and FieF from *Wautersia metallidurans* CH34 are transporters of broad metal specificity. *J Bacteriol*. 186:8036-8043.
Munson, G. P., Lam, D. L., Outten, F. W., O'Halloran, T. V. 2000. Identification of a copper-responsive two-component system on the chromosome of *Escherichia coli* K-12. *J Bacteriol*. 182:5864-71.
Murakami, S., Nakashima, R., Yamashita, E., Yamaguchi, A. 2002. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature. 4*19:587-593.
Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T., Yamaguchi, A. 2006. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature*. 443:173-9.
Murphy, P. M. 1993. Molecular mimicry and the generation of host defense protein diversity. *Cell. 7*2:823–826.
Mushegian, A. R., Garey, J. R., Martin, J., Liu, L. X. 1998. Large-scale taxonomic profiling of eukaryotic model organisms : a comparison of orthologous proteins encoded by the human., fly., nematode., and yeast genomes. *Genome Res*. 8:590-8.
Nakai, K., Kanehisa, M. 1991. Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins*. 11:95–110.
Nakamoto, H., Bardwell, J. C. A. 2004. Catalysis of disulfide bond formation and isomerization in *the Escherichia coli* periplasm. *Biochim Biophys Acta*. 1694:111 –119.
Navarro, C., Wu, L. F., Mandrand-Berthelot, M. A. 1993 The nik operon of *Escherichia coli* encodes a periplasmic binding-proteindependent transport system for nickel. *Mol Microbiol*. 9:1181-1191.
Needleman, S. B., Wunsch, C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. 48:443-53.
Nehme, D., Poole, K. 2007. Assembly of the MexAB-OprM multidrug pump of *Pseudomonas aeruginosa*: component interactions defined by the study of pump mutant suppressors. *J Bacteriol*. 189:6118-6127.
Neidhardt, F.,Ingraham, J. L., Low, K. B., Magasanik,B., Schaechter, U., Umbarger, H. E., *et al.*, 1987. *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. *Am Soc Microbiol*. 7–20.
Nelson, D., Kennedy, E. 1971. Magnesium transport in *Escherichia coli*. Inhibition by cobaltous ion. *J Biol Chem.* 246:3042-3049.
Nicholls, A., Sharp, K. A., Honig, B. 1991. Protein folding and association : insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*. 11:281-96.
Nies, D. H. 1991. Effluxsysteme als Grundlage bakterieller Metallresistenzen. *Bioforum. 7*-8:251-256.
Nies, D. H. 1992. CzcR and CzcD., gene products affecting regulation of resistance to cobalt., zinc., and cadmium (*czc* system) in *Alcaligenes eutrophus*. *J Bacteriol*. 174:8102-10.
Nies, D. H. 1995. The cobalt, zinc, and cadmium efflux system CzcABC from *Alcaligenes eutrophus* functions as a cation-proton antiporter in *E. coli*. *J Bacteriol*. 177:2707-2712.
Nies, D. H. 1999. Microbial heavy-metal resistance. *Appl Microbiol Biotechnol*. 51:730-50.
Nies, D. H. 2000. Heavy metal-resistant bacteria as extremophiles : molecular physiology and biotechnological use of *Ralstonia* sp. CH34. *Extremophiles*. 4:77-82.
Nies, D. H. 2003. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol Rev*. 27:313-339.
Nies, D. H. 2004. Incidence and function of sigma factors in *Ralstonia metallidurans* and other bacteria. *Arch Microbiol*. 181:255-68.
Nies, D. H. 2004b. Essential and Toxic Effects of Elements on Microorganisms. Metals and their compounds in the environment. *K Anke M Ihnat and M Stoeppler Weinheim Wiley-VCH*:*Part II.1.*
Nies, D. H. 2007. How cells control zinc homeostasis. *Science. 3*17:1695-1696.
Nies, D. H., Mergeay, M., Friedrich, B., Schlegel, H. G. 1987. Cloning of plasmid genes encoding resistance to cadmium., zinc., and cobalt in *Alcaligenes eutrophus* CH34. *J Bacteriol*. 169:4865-8.
Nies, D. H., Nies, A., Chu, L., Silver, S. 1989. Expression and nucleotide sequence of a plasmid-determined divalent cation efflux system from *Alcaligenes eutrophus*. *Proc Natl Acad Sci.* 86:7351-7355.
Nies, D. H., Silver, S. 1989b. Plasmid-determined inducible efflux is responsible for resistance to cadmium., zinc and cobalt in Alcaligenes eutrophus. *J Bacteriol*. 171:896-900.
Nies, A., Nies, D. H., Silver, S. 1990 Nucleotide sequence and expression of a plasmid-encoded chromate resistance determinant from Alcaligenes eutrophus. *J Biol Chem*. 265:5648-5653.
Nies, D. H., Silver, S. 1995 Ion efflux systems involved in bacterial metal resistances. *J Ind Microbiol.* 14:186-199
Nies, D. H., Brown, N. L. 1998. Two-component systems in regulation of heavy metal resistance. Metal ions in gene regulation. *Chapman & Hal, London England*. 77-103.
Nies, D. H., Koch, S., Wachi, S., Peitzsch, N., Saier, M. H. Jr. 1998 CHR, a novel family of prokaryotic proton motive force-driven transporters probably containing chromate/sulfate transporters. *J Bacteriol.* 180:5799-5802.
Nies, D. H., Rehbein, G., Hoffmann, T., Baumann, C., Grosse, C. 2006. Paralogs of genes encoding metal resistance proteins in *Cupriavidus metallidurans* strain CH34. *J Mol Microbiol Biotechnol*. 11:82-93.
Nikaido, H. 1996. Multidrug efflux pumps of Gram-negative bacteria. *J Bacteriol. 1*78:5853-9.
Nikaido, H. 1998. Multiple antibiotic resistance and efflux. *Curr Opin Microbiol*. 1:516-523.

Nikaido, H., Basina, M., Nguyen, V., Rosenberg, E. Y. 1998. Multidrug efflux pump AcrAB of *Salmonella typhimurium* excretes only those b-lactam antibiotics containing lipophilic side chains. *J Bacteriol*. 180:4686-4692.

Novikova, N., De Boever, P., Poddubko, S., Deshevaya, E., Polikarpov, N., Rakova, N., Coninx, I., Mergeay, M. 2006 Survey of environmental biocontamination on board the International Space Station. *Res Microbiol*. 157:5-12.

O'Brien, I. G., Gibson, F. 1970 The structure of enterochelin and related 2,3-dihydroxy-N-benzoylserine conjugates from *Escherichia coli*. *Biochim Biophys Acta*. 215:393–402.

Ochman, H., Lawrence, J. G., Groisman, E. A. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 405:299–304.

Ochman, H., Moran, N. A. 2001. Genes lost and found: evolution of bacterial pathogenesis and symbiosis. *Science*. 292:1096-1099.

Ogawa, T., Kaplan, A. 2003. Inorganic carbon acquisition systems in cyanobacteria. *Photosyn Res*. 77:105-115.

Ohkawa, H., Pakrasi, H. B., Ogawa, T. 2000. Two types of functionally distinct NAD(P)H dehydrogenases in *Synechocystis* sp. strain PCC6803. *J Biol Chem*. 275:31630-31634.

Okusu, H., Ma, D., Nikaido, H. 1996. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants. *J Bacteriol*. 178:306-308.

Omata, T., Ogawa, T. 1986. Biosynthesis of a 42KD polypeptide in the cytoplasmic membrane of the cyanobacterium *Anacystis nidulans* strain R2 during adaptation to low CO_2 concentration. *Plant Physiol.* 80:525-530.

Omata, T., Price, G. D., Badger, M. R., Okamura, M., Gotha, S., Ogawa, T. 1999 Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *PNAS*. 96:13571-13576.

Opella, S. J., DeSilva, T., Veglia, G. 2002 Structural biology of metal-binding sequences. *Curr Opin Chem Biol*. 6:217-223.

Outten, C. E., O'Halloran, T. V.. 2001. Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. *Science*. 292:2488-2492.

Outten, F. W., Huffman, D. L., Hale, J. A., O'Halloran, T. V. 2001b. The Independent cue and cus Systems Confer Copper Tolerance during Aerobic and Anaerobic Growth in *Escherichia coli. J Biol Chem*. 276:30670-30677.

Oven, M., Grill, E., Golan-Goldhirsh, A., Kutchan, T. M., Zenk, M. H. 2002 Increase of free cysteine and citric acid in plant cells exposed to cobalt ions. *Phytochemistry*. 60:467-474.

Overbeek, R., Fonstein, M., D'Souza, M., Pusch,G. D., Maltsev,N. 1999. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci*. 96:2896–2901.

Overbeek, R. 2000. Genomics: what is realistically achievable. *Genome Biol., 1, Comment.* 2002.

Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E. Jr., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I. *et al*. 2003. The ERGO genome analysis and discovery system. *Nucleic Acids Res.* 31:1-8.

Pao, S.S., Paulsen, I. T., Saier, M. H. Jr. 1998. The major facilitator superfamily. *Microbiol Mol Biol Rev*. 62:1–32.

Park, M. H., Wong, B. B., Lusk, J. E. 1976. Mutants in three genes affecting transport of magnesium in *Escherichia coli*: genetics and physiology. *J Bacteriol.* 126:1096-1103.

Park, J., Karplus, K., Barrett, C. Hughey, R., Haussler, D., Hubbard, T. & Chothia, C. 1998. Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods *J Mol Biol*. 284:1201–1210.

Parke, J. L., Gurian-Sherman, D., 2001. Diversity of the *Burkholderia cepacia* complex and implications for risk assessment of biological control strains. *Annu Rev Phytopathol*. 39:225-258.

Passerini, A., Andreini, C., Menchetti, S., Rosato, A., Frasconi, P. 2007 Predicting zinc binding at the proteome level, *BMC bioinformatics,* 5:8-39.

Patel, K., Kumar, A., Durani, S., 2007. Analysis of the structural consensus of the zinc coordination centers of metalloprotein structures. *Biochim Biophys Acta*. 1774:1247-1253.

Patzer, S. I., Hantke, K. 1998. The ZnuABC high-affinity zinc uptake system and its regulator Zur in *Escherichia coli*. *Mol Microbiol*. 28:1199-210.

Paulsen, I. T., Brown, M. H., Skurray R. A. 1996 Proton-Dependent Multidrug Efflux Systems. *Microbiol Rev*. 60:575-608.

Paulsen, I. T., Park, J. H., Choi, P. S., Saier, M. H. Jr. 1997. A family of Gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from Gram-negative bacteria. *FEMS Microbiol Lett*. 156:1-8.

Paulsen, I. T., A. M. Beness, Saier, M. H. Jr. 1997b. Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria. *Microbiology*. 143:2685–2699.

Paulsen, I. T., Saier, M. H. Jr. 1997 A novel family of ubiquitous heavy metal ion transport proteins. *J Membr Biol*. 156:99-103.

Paulsen, I. T., Sliwinski, M. K., Saier, M. H. Jr. 1998. Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. *J Mol Biol*. 277:573-592.

Paulsen, I. T., Nguyen, L., Sliwinski, M. K., Rabus, R., Saier, M. H. Jr. 2000. Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J Mol Biol.* 301:75-100.

Payne, G. W., Ramette, A., Rose, H. L., *et al.,* 2006. Application of a recA gene-based identification approach to the maize rhizosphere reveals novel diversity in *Burkholderia* species. *FEMS Microbiol Lett*. 259:126-132.

Peterson, J. D., Umayam, L. A., Dickinson, T., Hickey, E. K., White, O. 2001. The Comprehensive Microbial Resource. *Nucleic Acids Res*. 29:123-5.

Pieper, H. D., Engesser, K. H., Knackmuss, H. J. 1989. Regulation of catabolic pathways of phenoxyacetic acids and phenols in *Alcaligenes eutrophus* JMP134. *Arch Microbiol*. 151:356-371.

Pierson, D. L. 2001 Microbial contamination of spacecraft. *Gravit Space Biol Bull.* 14:1-6.

Pimenta, A., Blight, M., Clarke, D., Holland, I. B. 1996. The Gram-negative cell envelope “springs” to life: coiled-coil trans-envelope proteins. *Mol Microbiol*. 19:643-645.

Pohlmann, A., Fricke, W. F., Reinecke, F., Kusian, B., Liesegang, H., Cramm, R., Eitinger, T., Ewering, C., Potter, M., Schwartz, E., Strittmatter, A., Voss, I., Gottschalk, G., Steinbuchel, A., Friedrich, B., Bowien, B. 2006. Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. *Nat Biotechnol.* 24:1257-62.

Polatajko, A., Banas, B., Encinar, J. R., Szpunar, J. 2005. Advances in analytical methodology for bioinorganic speciation analysis : metallomics., metalloproteomics and heteroatom-tagged proteomics and metabolomics. *Analyst*. 130:442-65.

Pollack, J. R., Neilands, J. B. 1970 Enterobactin, an iron transport compound from *Salmonella typhimurium. Biochem Biophys Res Commun*. 38:989–992.

Poole, K, Heinrichs, D. E., Neshat, S. 1993 Cloning and sequence analysis of an EnvCD homologue in *Pseudomonas aeruginosa*: regulation by iron and possible involvement in the secretion of the siderophore pyoverdine. *Mol Microbiol*. 10:529–544.

Poteete, A. R., Renndell, D., Bouvier, S. E. 1992 Functional significance of conserved amino acid residues. *Proteins*. 13:38-40.

Price, G. D., Sültemeyer, D., Klughammer, B., Ludwig M., Badger, M. R. 1998 The functioning of the CO_2 concentrating mechanism in several cyanobacterial strains: A review of general physiological characteristics, genes, proteins and recent advances. *Can J Bot*. 76:973-1002.

Price, G. D., Woodger, F. J., Badger, M. R., Howitt, S. M., Tucker, L. 2004. Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. *Proc Natl Acad Sci*.101:18228-18233.

Price, G. D., Badger, M. R., Woodger, F. J., Long, B. M. 2008. Advances in understanding the cyanobacterial CO_2 concentrating- mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J Exp Bot.* 59:1441–1461.

Pruitt, K. D., Tatusova, T., Maglott, D. R. 2005. NCBI reference sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 33:D501-504.

Qian, H., Sahlman, L., Eriksson, P. O., Hambraeus, C., Edlund, U., Sethson, I. 1998. NMR solution structure of the oxidized form of MerP, a mercuric ion binding protein involved in bacterial mercuric ion resistance. *Biochemistry*. 37:9316-9322.

Quentin, Y., Fichant, G. 2000. ABCdb: an ABC transporter database. *J Mol Microbiol Biotechnol*. 2:501-504.

Ragan, M. A. 2001. Detection of lateral gene transfer among microbial genomes. *Curr Opin Genet Dev*. 11:620–626.

Raivio, T. L., Silhavy, T. J. 2001. Periplasmic stress and ECF sigma factors. *Annu Rev Microbiol*. 55:591–624.

Ranea, J. A., 2007. Predicting protein function with hierarchical phylogenetic profiles: the Gene3D Phylo-Tuner method applied to eukaryotic genomes. *PLoS Comput Biol.* 3:e237.

Raymond, K. N., Dertz, E. A.,Kim, S. S. 2003. Enterobactin : An archetype for microbial iron transport *PNAS*. 100:3584-3588.

Reith, F., Lengke, M. F., Falconer, D., Craw, D., Southam, G. 2007. The geomicrobiology of gold. *ISME J.* 1:567-84.

Reizer, A., Pao, G. M., Saier, M. H. Jr. 1991. Evolutionary relationships among the permease proteins of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. Construction of phylogenetic trees and possible relatedness to proteins of eukaryotic mitochondria. *J Mol Evol.* 33:179-193.

Remm, M., Storm, C. E., Sonnhammer, E. L. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol*. 314:1041-1052.

Ren, Q., Kang, K. H., Paulsen, I. T. 2004. TransportDB: a relational database of cellular membrane transport systems. *Nucleic Acids Res*. 32 Database issue:D284-D288.

Ren, Q., Paulsen, I. T. 2005. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. *PLoS Comput Biol*. 1:e27.

Ren, Q., Paulsen, I. T. 2007. Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes. *J Mol Microbiol Biotechnol*. 12:165-79.

Rensing, C., Pribyl, T., Nies, D. H. 1997. New functions for the three subunits of the CzcCBA cation-proton antiporter. *J Bacteriol.* 179:6871-9.

Rensing, C., Ghosh, M., Rosen, B. P. 1999 Families of soft metal ATPases. *J Bacteriol*. 181:5891-5897.

Rietsch, A., Belin, D., Martin, N., Beckwith, J. 1996 An in vivo pathway for disulfide bond isomerization in *Escherichia coli*. *Proc Natl Acad Sci*. 93:13048– 13053.

Riley, M. L., Schmidt, T., Wagner, C., Mewes, H. W., Frishman,D. 2005. The PEDANT genome database in 2005. *Nucleic Acids Res*. 33:D308–D310.

Roberts, S. A., Weichsel, A., Grass, G., Thakali, K., Hazzard, J. T., Tollin, G., Rensing, C., Montfort, W. R. 2002. Crystal structure and electron transfer kinetics of CueO., a multicopper oxidase required for copper homeostasis in *Escherichia coli*. *Proc Natl Acad Sci.* 99:2766-71.

Roberts, S. A., Wildner, G. F., Grass, G., Weichsel, A., Ambrus, A., Rensing, C., Montfort, W. R. 2003. A labile regulatory copper ion lies near the T1 copper site in the multicopper oxidase CueO. *J Biol Chem*. 278:31958–. 31963.

Rodley, P. D., Romling, U., Tummler, B. 1995. A physical genome map of the *Burkholderia cepacia* type strain. *Mol Microbiol*. 17:57-67.

Rolland, N., Ferro, M., Seigneurin-Berny, D., Garin, J., Douce, R., Joyard, J. 2003. Proteomics of chloroplast envelope membranes. *Photosyn Res*. 78:205–230.

Rosen, B. P., 2002. Transport and detoxification systems for transition metals, heavy metals and metalloids in eukaryotic and prokaryotic microbes. *Comp Biochem Physiol* 133:689–693

Roux, M., Sarret, G., Pignot-Paintrand, I., Coves J. 2001 Mobilization of selenite by *R. metallidurans* CH34. *Appl Environ Microbiol*. 67:769-773.

Sacha, A. F. T., van Hijum, Zomer, A. L., Kuipers, O. P., Kok, J. 2003. Projector : automatic contig mapping for gap closure purposes. *Nucleic Acids Res.* 31:1093.

Sacha, A. F. T., van Hijum, Zomer, A. L., Kuipers, O. P., Kok, J. 2005. Projector2: contig mapping for efficient gap-closure of prokaryotic genome sequence assemblies. *Nucleic Acids Res.* 2005

Saier, M. H. Jr. 1994. Computer-aided analyses of transport protein sequences : gleaning 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 superfamilies. *Microb Comp Genom.* 1:129-150.

Saier, M. H. Jr. 2000. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol R*. 64:354-411.

Saier, M. H. Jr. 2000b. Families of transmembrane sugar transport proteins. *Mol Microbiol*. 35:699-710.

Saier, M. H. Jr. 2000c. Families of proteins forming transmembrane channels. *J Membr Biol*. 175:165-180.

Saier, M.H., Jr. 2000d. Families of transmembrane transporters selective for amino acids and their derivatives. *Microbiology* 146:1775-1795.

Saier, M. H. Jr. 2005. Beneficial bacteria and bioremediation. *J Mol Microbiol Biotechnol*. 9:63-4.

Saier, M. H. Jr., Tam, R., Reizer, A., Reizer, J. 1994. Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. *Mol Microbiol*. 11:841-847.

Saier, M. H., Jr., Paulsen, I. T., Sliwinski, M. K., Pao, S. S., Skurray, R. A., Nikaido, H. 1998. Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. *FASEB J*. 12:265–274.

Saier, M. H. Jr., Tran, C. V., Barabote, R. D. 2006. TCDB : the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res*. 34:D181-6.

Sakurai, T., Kataoka, K. 2007. Basic and applied features of multicopper oxidases, CueO, bilirubin oxidase, and laccase. *Chem Rec*. 7:220-229.

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangeno, T. S., Arlat, M., Billault, A., Brottier, P., Camus, C. J., Cattolico, L., Chandler, M., Choisne, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., Boucher, B. C. A. 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature*. 415:497-502.

Sanner, M. F., Olson, A. J., Spehner, J. C. 1996. Reduced surface : an efficient way to compute molecular surfaces. *Biopolymers*. 38:305-20.

Santo, C. E., Taudte, N., Nies, D. N., Grass, G. 2008. Contribution of Copper Ion Resistance to Survival of *Escherichia coli* on Metallic Copper Surfaces. *Appl Environ Microbiol*., 72:977-986

Satofuka, H., Fukui, T., Takagi, M., Atomi, H., Imanaka, T. 2001 Metal-binding properties of phytochelatin-related peptides. *J Inorg Biochem*. 86:595-602.

Saurin, W., Hofnung, M., Dassa, E. 1999. Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J Mol Evol*. 48:22-41.

Schreck, P. 1997 Environmentally sustainable disposal of radioactive sludge rich in heavy metals from copper shale smelting in Central Germany. *In: Engineering geology and the environment*. 2:2145-2149.

Schmidt, T., Schlegel, H. G., 1994. Combined nickel-cobalt-cadmium resistance encoded by the *ncc* locus of *Alcaligenes xylosoxidans* 31A. *J Bacteriol.* 176:7045-7054

Schübbe, S., Kube, M., Scheffel, A., Wawer, C., Heyen, U., Meyerdierks, A., Madkour, M., Mayer, F., Reinhardt, R., Schüler, D. 2003. Characterization of a spontaneous nonmagnetic mutant of *Magnetospirillum gryphiswaldense* reveals a large deletion comprising a putative magnetosome island. *J Bact*. 185:5779–5790.

Schüler, D, Baeuerlein. E 1996. Iron-limited growth and kinetics of iron uptake in *Magnetospirillum gryphiswaldense*. *Arch Microbiol*. 166:301–307.

Schüler, D. 2004. Molecular analysis of a subcellular compartment: the magnetosome membrane in *Magnetospirillum gryphiswaldense. Arch Microbiol*. 181:1–7.

Schwartz, E., Friedrich, B. 2001 A physical map of the megaplasmid pHG1, one of three genomic replicons in *Ralstonia eutropha* H16. *FEMS Microbiol Lett*. 201:213-219.

Schwartz, E., Henne, A., Cramm, R., Eitinger, T., Friedrich, B., Gottschalk, G. 2003. Complete nucleotide sequence of pHG1 : a *Ralstonia eutropha* H16 megaplasmid encoding key enzymes of H(2)-based lithoautotrophy and anaerobiosis. *J Mol Biol*. 332:369-83.

Seeger, M. A., Schiefner, A., Eicher, T., Verrey, F., Diederichs, K., Pos, K. M. 2006. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science*. 313:1295-8.

Sharff, A., Fanutti, C., Shi, J., Calladine, C., Luisi, B., 2001. The role of the TolC family in protein transport and multidrug efflux. From stereochemical certainty to mechanistic hypothesis. *Eur J Biochem*. 268:5011-26.

Shi, W., Zhan, C., Ignatov, A., Manjasetty, B. A., Marinkovic, N., Sullivan, M., Huang, R., Chance, M. R. 2005. Metalloproteomics: high-throughput structural and functional annotation of proteins in structural genomics**.** *Structure*. 13:1473-1486.

Shibata, M., Ohkawa, H., Kaneko, T., Fukuzawa, H., Tabata, S., Kaplan, A., Ogawa, T. 2001 Distinct constitutive and low-CO_2-induced CO_2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. *PNAS*. 98:11789-11794.

Shibata, M., Katoh, H., Sonoda, M., Ohkawa, H., Shimoyama, M., Fukuzawa, H., Kaplan, A., Ogawa, T., 2002. Genes Essential to Sodium-dependent Bicarbonate Transport in Cyanobacteria. *J Biol Chem*. 277:18658-18664.

Siddiqui, R. A., Schlegel, H. G., Meyer, M. 1988. Inducible and constitutive expression of pMOL28-encoded nickel resistance in *Alcaligenes eutrophus* N9A. *J Bacteriol*. 170:4188-93.

Sidman, K. E., George, D. G., Barker, W. C., Hunt, L. T. 1988. The protein identification resource (PIR). *Nucleic Acids Res*. 16:1869– 1871.

Silver, S., Keach, D. 1982. Energy-dependent arsenate efflux: the mechanism of plasmid-mediated resistance. *Proc. Natl Acad Sci*. 79:6114– 6118.

Silver, S., Ji, G. 1994. Newer systems for bacterial resistances to toxic heavy metals. *Environ Health Perspect.* 102:107-113.

Singh, S. K., Grass, G., Rensing, C., Montfort, W. R. 2004. Cuprous oxidase activity of CueO from *Escherichia coli. J Bacteriol*. 186**:**7815-7817.

Sjostedt, A. 2007. Tularemia: history, epidemiology, pathogen physiology, and clinical manifestations. *Ann NY Acad Sci*. 1105:1-29.

Smith, T. F., Waterman, M. S. 1981. Identification of common molecular subsequences. *J Mol Biol.* 147:195-7.

Smith, R. L., Maguire, M. E. 1998. Microbial magnesium transport : unusual transporters searching for identity. *Mol Microbiol*. 28:217-26.

Snavely, M. D., Florer, J. B., Miller, C. G., Maguire, M. E. 1989. Magnesium transport in *Salmonella typhimurium* : Mg^{2+} transport by the CorA, MgtA, and MgtB systems. *J Bacteriol*. 171:4761-6.

Solioz, M., Vulpe, C. 1996. CPx-type ATPases: a class of P-type ATPases that pump heavy metals. *Trends Biochem Sci*. 21:237-41.

Sonnhammer, E. L., Eddy, S. R., Birney, E., Bateman, A., Durbin, R. 1998. Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res*. 26:320– 322.

Sonnhammer, E. L. L., von, Heijne, G., Krogh, A. 1998b. A hidden Markov model for predicting transmembrane helices in protein sequences. *In Proc of Sixth Int Conf on Intelligent Systems for Molecular Biology*. 175-182.

Sonnhammer, E. L. L., Koonin, E. V. 2002. Orthology, paralogy and pro-posed classification for paralog subtypes. *Trends Genet. 1*8:619-620.

Staden, R., Beal, K. F., Bonfield, J. K. 2000. The Staden package, 1998. *Methods Mol Biol.* 132:115–130.

Steinbüchel, A., 1996. PHB and other Polyhydroxyalkanoic Acids. *In: Rehm, H.-J., Reed, G., Roehr, M. (Sds.) Biotechnology. VCH Weinheim.* Volume 6/ Chapter 13, 404-464.

Stewart, J.B. Hermodson, M.A. 2003. Topology of RbsC, the membrane component of the *Escherichia coli* ribose transporter. *J Bacteriol*. 185: 5234-5239.

Storm, C. E., Sonnhammer, E. L. 2002 Automated ortholog inference from phylogenetic trees and calculation of orthology reliability. *Bioinformatics.* 18:92-99.

Stroebel, D., Sendra, V., Cannella, D., Helbig, K., Nies, D. H., Covès, J. 2007. Oligomeric behavior of the RND transporters CusA and AcrB in micellar solution of detergent. *Biochim Biophys Acta.*1768:1567-73.

Suhre, K., Claverie, J. M., 2004. FusionDB : a database for in-depth analysis of prokaryotic gene fusion events. *Nucleic Acids Res*. 32:D273-6.

Sültemeyer, D., Price, G. D., Yu, J. W., Badger, M. R. 1995. Characterisation of carbon dioxide and bicarbonate transport during steady-state photosynthesis in the marine cyanobacterium *Synechococcus* strain PCC 7002. *Planta*. 197:597-607.

Summers, A. O. 1988. Biotransformations of mercury compounds. *Basic Life Sci.* 45:105–109.

Sutton, G., White, O., Adams, M., Kerlavage, A. R. 1995. TIGR Assembler: a new tool for assembling large shotgun sequencing projects. *Genome Sci Technol.* 1:9-19.

Szpirer, C., Top, E. M., Couturier, M., Mergeay, M. 1999. Retrotransfer or gene capture: a feature of conjugative plasmids, with ecological and evolutionary significance. *Microbiology*. 145:3321-3329.

Szurek, B., O. Rossier, G. Hause, Bonas, U. 2002**.** Type III-dependent translocation of the *Xanthomonas* AvrBs3 protein into the plant cell. *Mol Microbiol*. 46:13-23.

Tabacchioni, S., Bevivino, A., Dalmastri, C., Chiarini, L. 2002. *Burkholderia cepacia* complex in the rhizosphere: a minireview. *Ann Microbiol*. 52:103–117.

Taghavi, S., Mergeay, M., van der Lelie, D. 1997. Genetic and physical maps of the *Alcaligenes eutrophus* CH34 megaplasmid pMOL28 and its derivative pMOL50 obtained after temperature-induced mutagenesis and mortality. *Plasmid*. 37:22-34.

Taghavi, S., Mergeay, M., Nies, D., van der Lelie, D. 1997b *Alcaligenes eutrophus* as a model system for bacterial interactions with heavy metals in the environment. *Res Microbiol. 1*48:536-51.

Tamames, J., Casari, G., Ouzounis, C., Valencia, A. 1997. Conserved clusters of functionally related genes in two bacterial genomes. *J Mol Evol*. 44:66-73.

Tamegai, H., Fukumori, Y. 1994. Purification and some molecular and enzymatic features of a novel *ccb* -type cytochrome c oxidase from a microaerobic denitrifier, *Magnetospirillum magnetotacticum. FEBS Lett.* 347:22–26.

Tampakaki, A. P., Fadouloglou, V. E., Gazi, A. D., Panopoulos, N. J., Kokkinidis, M. 2004**.** Conserved features of type III secretion. *Cell Microbiol. 6*:805-816.

Tatusov, R.L., Altschul, S. F., Koonin, E. V. 1994. Detection of conserved segments in proteins: Iterative scanning od sequence databases with alignment blocks. *Proc Nat Acad Sci*. 91:12091- 12095.

Tatusov, R. L., Koonin, E. V., Lipman, D. J. 1997. A genomic perspective on protein families. *Science*. 278:631-7

Tauch, A., Kaiser, O., Hain, T. *et al.* 2005. Complete genome sequence and analysis of the multiresistant nosocomial pathogen *Corynebacterium jeikeium* K411, a lipid-requiring bacterium of the human skin flora. *J Bacteriol*. 187:4671–4682.

Tchernov, D., Helman, Y., Keren, N., Luz, B., Ohad, I., Reinhold, L., Ogawa, T., Kaplan, A. 2001. Passive entry of CO_2 and its intracellular conversion to HCO_3^- in cyanobacteria are driven by a photosystem I-generated DmHþ. *J Biol Chem.* 276:23450–23455.

Teichmann, S. A., Chothia, C., Gerstein, M., 1999 Advances in structural genomics. *Curr Opin Struct Biol.* 9:390-399.

Thomas, C. M., Nielsen, K. M. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol.* 3:711–721.

Thomas, M. S. 2007 Iron acquisition mechanisms of the *Burkholderia cepacia* complex. *Biometals*. 2007 20:431-452.

Thompson, J. D., Higgins, D. G., Gibson, T. J. 1994. CLUSTAL W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-80.

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*. 25:4876-82.

Tindall, B. J. 2008. Rule 15 of the international code of nomenclature of bacteria: a current source of confusion. *Int J Syst Evol Microbiol.* 58:1775-1778

Tong, L., Nakashima, S., Shibasaka, M., Katsuhara, M. Kasamo, K. 2002. A novel histidine-rich CPx-ATPase from the filamentous cyanobacterium *Oscillatoria brevis* related to multiple-heavy-metal cotolerance. *J Bacteriol.* 184:5027-5035.

Top, E. M., Springael, D. 2003 The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr Opin Biotechnol*. 14:262–269.

Toussaint, A., Merlin, C., Monchy, S., Benotmane, M. A., Leplae, R., *et al.* 2003. The biphenyl-and 4-chlorobiphenyl-catabolic transposon Tn4371,a member of a new family of genomic islands related to IncP and Ti plasmids. *Appl Environ Microbiol.* 69:4837-45.

Touze, T., Eswaran, J., Bokma, E., Koronakis, E., Hughes, C., Koronakis, V. 2004 Interactions underlying assembly of the *Escherichia coli* AcrAB-TolC multidrug efflux system. *Mol Microbiol*. 53:697-706.

Tseng, T. T., Gratwick, K. S., Kollman, J., Park, D., Nies, D. H., Goffeau, A., Saier, M. H. Jr. 1999. The RND Permease Superfamily: An ancient, ubiquitous and diverse family that includes human disease and development proteins. *J Mol Microbiol Biotechnol.* 1:107-125.

Tusnady, G. E., Simon, I. 1998. Principles governing amino acid composition of integral membrane proteins : application to topology prediction. *J Mol Biol*. 283:489-506.

Uchino, K., Saito, T., Jendrossek, D. 2008. Poly(3-hydroxybutyrate) (PHB) depolymerase PhaZa1 is involved in mobilization of accumulated PHB in *Ralstonia eutropha* H16. *Appl Environ Microbiol.* 74:1058-63.

Vallee, B. L., Auld, D. S. 1990. Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry*. 29:5647-5659.

Valls, M., de Lorenzo, V. 2002. Exploiting the genetic and biochemical capacities of bacteria for the remediation of heavy metal pollution. *FEMS Microbiol Rev*. 26:327-38.

Valls, M., Genin, S., Boucher, C. 2006 Integrated regulation of the Type III secretion system and other virulence determinants in *Ralstonia solanacearum*. *PLoS Pathog*. 2:e82.

van der Lelie, D., Schwuchow, T., Schwidetzky, U., Wuertz, S., Baeyens, W., Mergeay, M., Nies, D. H. 1997. Two-component regulatory system involved in transcriptional control of heavy-metal homoeostasis in *Alcaligenes eutrophus*. *Mol Microbiol*. 23:493-503.

van der Lelie, D., Barac, T., Taghavi, S., Vangronsveld, J. 2005. Response to Newman: New uses of endophytic bacteria to improve phytoremediation. *Trends Biotechnol.* 23:8-9.

van Elsas, J. D. 1992. Antibiotic resistance gene transfer in the environment: an overview. In: Wellington EMH, van Elsas JD (eds) Genetic interactions among microorganisms in the natural environment. *Pergamon Oxford.* 17–39.

van Veen, H. W., Abee, T., Kortstee, G. J., Konings, W. N., Zehnder, A. J. 1994. Translocation of metal phosphate via the phosphate inorganic transport system of *Escherichia coli*. *Biochemistry*. 33:1766-1770.

Vandamme, P., Coenye, T. 2004 Taxonomy of the genus *Cupriavidus*: a tale of lost and found. *Int J Syst Evol Microbiol*. 54:2285-2289.

Vaneechoutte, M., Kämpfer, P., De Baere, T., Falsen, E., Verschraegen, G. 2004. *Wautersia* gen. nov., a novel genus accommodating the phylogenetic lineage including *Ralstonia eutropha* and related species, and proposal of *Ralstonia* (*Pseudomonas*) *syzygii* comb. nov. *Int J Syst Evol Microbiol.* 54:317-327.

von Caemmerer, S., Furbank, R. T. 2003. The C-4 pathway: an efficient CO_2 pump. *Photosyn Res.* 77:191–207.

von Rozycki, T., Nies, D. H., Saier, M. H. Jr. 2005. Genomic analyses of transport proteins in *Ralstonia metallidurans*. *Comp Func Genom*. 6:17-56.

Wackett, L. P., Dodge, A. G., Ellis, L. B. M. 2004. Microbial Genomics and the Periodic Table. *Appl Environ Microbiol*. 70:647–655.

Wang, J., Lory, S., Ramphal, R., Jin, S. 1996. Isolation and characterisation of *Pseudomonas aeruginosa* genes inducible by respiratory mucus derived from cystic fibrosis patients. *Mol Microbiol*. 22:1005-1012.

Wang, J., Wong, G. K., Ni, P., Han, Y., Huang, X., Zhang, J., Ye, C., Zhang, Y., Hu, X., Zhang, K., Xu, X., Cong, L., Lu, H., Ren, X., Ren, X., He, J., Tao, L., Passey, D. A., Wang, J., Yang, H., Yu, J., Li, S. 2002 RePS:A sequence assembler that masks exact repeats identified from the shotgun data. *Genome Res.* 12:824-831.

Wapinski, I., Pfeffer, A., Friedman, N., Regev, A. 2007. Automatic genome-wide reconstruction of phylogenetic gene trees. *Bioinformatics*. 23:549–558.

Warkentin, E., Mamat, B., Sordel-Klippert, M., Wicke, M., Thauer, R. K., Iwata, M., Iwata, S., Ermler, U., Shima, S. 2001. Structures of F 420 H 2 :NADPH $^+$ oxidoreductase with and without its substrates bound. *EMBO J*. 20:6561–6569.

Weltrowski, A. 2005. Die Struktur und Funktion des CDF-Proteins CzcD aus *Cupriavidus metallidurans Dissertation Martin-Luther-Universität Halle-Wittenberg. Halle*.

Woodger, F. J., Badger, M. R., Price, G. D. 2003. Inorganic carbon limitation induces transcripts encoding components of the CO_2-concentrating mechanism in *Synechococcus* sp. PCC7942 through a redox-independent pathway. *Plant Physiol*. 133:2069–2080.

Woodger, F. J., Badger, M. R., Price, G. D. 2005. Sensing of Inorganic Carbon Limitation in *Synechococcus* PCC7942 Is Correlated with the Size of the Internal Inorganic Carbon Pool and Involves Oxygen. *Plant Physiol.* 139:1959-1969.

Woodger, F. J., Bryant, D. A., Price, G. D. 2007. Transcriptional Regulation of the CO_2-Concentrating Mechanism in a Euryhaline, Coastal Marine Cyanobacterium, *Synechococcus* sp. Strain PCC 7002: Role of NdhR/CcmR. *J Bacteriol. 1*89:3335–3347.

Workun, G. J., Moquin, K., Rothery, R. A., Weiner, J. H. 2008. Evolutionary Persistence of the Molybdopyranopterin-Containing Sulfite Oxidase Protein Fold. *Microbiol Mol Biol Rev*. 72:228–248.

Zgurskaya, H. I., Nikaido, H. 1999. AcrA is a highly asymmetric protein capable of spanning the periplasm. *J Mol Biol*. 285:409-420.

Zgurskaya, H. I., Nikaido, H. 1999b. Bypassing the periplasm: reconstitution of the AcrAB multidrug efflux pump of *Escherichia coli*. *Proc Natl Acad Sci*. 96:7190–7195.

Zhai, Y., Saier, M. H., 2001. A web-based program (WHAT) for the simultaneous prediction of hydropathy., amphipathicity., secondary structure and transmembrane topology for a single protein sequence. *J Mol Microbiol Biotechnol.* 3:501-2.

Zhai, Y., Saier, M. H., 2002. A simple sensitive program for detecting internal repeats in sets of multiply aligned homologous proteins. *J Mol Microbiol Biotechnol*. 4:375-7.

Zhang, Z., Schaffer, A. A., Miller, W., Madden, T. L., Lipman, D. J., Koonin, E. V., Altschul, S. F. 1998. Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res.* 26:3986–3990

Zhang, P., Battchikova, N., Jansen, T., Appel, J., Ogawa, T., Aro, E. M. 2004. Expression and functional roles of the two distinct NDH-1 complexes and the carbon acquisition complex NdhD3/ NdhF3/CupA/Sll1735 in *Synechocystis* sp. PCC 6803. *Plant Cell*. 12:3326-3340.

Zheng, Y., Roberts, R. J., Kasif, S. 2004. Segmentally Variable Genes: A New Perspective on Adaptation. *PLoS Biol*. 2:0452- 0464.

Zhou, X., Yang, N. M., Tran, C. V., Hvorup, R. N., Saier, M. H. 2003. Web-based programs for the display and analysis of transmembrane alpha-helices in aligned protein sequences. *J Mol Microbiol Biotechnol*. 5:1-6.

Zoropogui, A., Gambarelli, S., Coves, J. 2008. CzcE from *Cupriavidus metallidurans* CH34 is a copper-binding protein. *Biochim Biophys Acta*. 365:735-739.

6. OVERVIEW OF PUBLICATIONS

1:

von Rozycki, T., Yen, M.R., Lende, E.E., Saier, M. H. Jr. 2004

The YedZ family: possible heme binding proteins that can be fused to transporters and electron carriers.

J Mol Microbiol Biotechnol 8:129-40.

2:

von Rozycki, T., Schultzel, M.A., Saier, M. H. Jr. 2004

Sequence analyses of cyanobacterial bicarbonate transporters and their homologues.

J Mol Microbiol Biotechnol 7:102-8.

3:

von Rozycki, T., Nies, D.H., Saier, M. H. Jr. 2005

Genomic analyses of transport proteins in *Ralstonia metallidurans*.

Comp Func Genom. 6:17-56.

4:

von Rozycki, T., Nies, D.H. 2008

Cupriavidus metallidurans: evolution of a metal-resistant bacterium

(accepted in 2008 *Antonie Van Leeuwenhoek*)

Danksagung

Mein besonderer Dank gilt Herrn Prof. Dr. D. H. Nies für das in mich gesetzte Vertrauen, die konstruktive Kritik und die Möglichkeit diese Arbeit in seiner Gruppe durchzuführen.

Professor Nies verdanke ich auch den Anstoss für die ertragreiche Arbeit in der Arbeitsgruppe von Herrn Prof. Dr. M. H. Saier an der UCSD.

Die Veröffentlichung meiner Arbeiten wurden jedoch erst durch die bereitwillige und weitreichende Unterstützung von Prof. Saier ermöglicht. Auch haben Fr Dr Cornelia Grosse und Dr Gregor Grass immer ein offenes Ohr für mich gehabt.

Allen ehemaligen und derzeitigen Mitgliedern des Instituts für Mikrobiologie in der Arbeitsgruppe von Prof. Dr. D. H. Nies – besonders Antje, Monique, Nadine, Doreen, Grit, Kerstin, Sylvia und Susann und Mitgliedern benachbarter Arbeitsgruppen und Institute verdanke ich eine stets freundschaftliche Atmosphäre und breite Unterstützung.

Ein ganz besonderer Dank gebührt meinen Eltern und meinem Bruder Christian die stets für mich da waren.

Eidesstattliche Erklärung

Hiermit erkläre ich, daß ich die vorliegende Dissertation selbständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe. Aus anderen Werken entnommene Stellen wurden als solche kenntlich gemacht.

Halle (Saale), den
Torsten von Rozycki

Lebenslauf

Torsten von Rozycki

Persönliche Angaben

Familienstand: ledig

Nationalität: deutsch

Staatsangehörigkeit: deutsch

Geburtsdatum: 4.7.1976

Geburtsort: Halle / Saale

Ausbildung

Polytechnische Oberschule

1983-91 Pädagogischen Oberschule „Kurt Wabbel“ ***Halle /Saale***

Gymnasium

1991-95 *Thomas Müntzer Gymnasium Halle / Saale*

1995 Abschluß des Abiturs

Studium

1995-2001 Biochemie - Martin- Luther- Universität Halle Wittenberg

2001 Forschungsgruppenpraktikum - Thema : ‚Klonierung der Gene von P-typ ATPasen aus *Pseudomonas aeruginosa*’ am Institut für Mikrobiologie (AG Nies)

2001-2002 Diplomarbeit zum Thema: „Der ECF (*extracytoplasmic function*)-Sigma factor RpoE aus *Ralstonia metallidurans* CH34 “

2002 Abschluß des Diploms

Promotion

seit 2002 wissenschaftlicher Mitarbeiter in der Arbeitsgruppe von Prof. D. H. Nies

2003 wissenschaftlicher Mitarbeiter an der Universität von Kalifornien San Diego in der Arbeitsgruppe von Prof. M. H. Saier

2008 Abschuss Promotionsarbeit

Short Communication

Journal of Molecular Microbiology and Biotechnology

J Mol Microbiol Biotechnol 2004;7:102–108
DOI: 10.1159/000078653

Sequence Analyses of Cyanobacterial Bicarbonate Transporters and Their Homologues

Torsten von Rozycki[1] Mark A. Schultzel[1] Milton H. Saier Jr.

Division of Biological Sciences, University of California at San Diego, La Jolla, Calif., USA

Key Words
Bicarbonate · Transport · Photosynthesis · Cyanobacteria · Evolution

Abstract
The primary HCO_3^- uptake system in the cyanobacterium *Synecocystis* is the Na^+-dependent transporter SbtA. SbtA and its homologues were identified and shown to display a common topology of ten transmembrane segments (TMSs). These proved to have arisen by an intragenic duplication event from an ancestral gene encoding a five TMS protein product. A region of SbtA shows sufficient similarity to 10 TMS ABC-type integral membrane transport proteins to suggest a common origin. Phylogenetic analyses of the SbtA family revealed two clusters of cyanobacterial homologues with all non-cyanobacterial family members outside of these two clusters. The tree topology suggests that SbtA family members display multiple transport functions.

Copyright © 2004 S. Karger AG, Basel

Introduction

Cyanobacteria possess multiple CO_2 and HCO_3^- concentrating systems allowing rapid CO_2 fixation by ribulose-1,5-bisphosphate carboxylase/oxygenase during the photosynthetic reaction cycle [Shibata et al., 2002]. Under neutral conditions, the predominant source of carbon for this reaction is available as CO_2 [Ritchie et al., 1996]. This molecule seems to be imported via a mechanism that depends on at least two sets of distinct NADPH dehydrogenase-like subunits (NdhD3/NdhF3 and NdhD4/NdhF4) induced by low concentrations of CO_2 (NdhD3/NdhF3) or constitutively expressed (NdhD4/NdhF4). They work in cooperation with soluble but membrane associated proteins, CupA and CupB, respectively, that are found in association with the NDH complexes noted above [Ohkawa et al., 2000a, b; Omata et al., 1999; Shibata et al., 2001, 2002]. The CupA/B proteins seem to be necessary for the conversion of CO_2 to HCO_3^- [Ohkawa et al., 2000b], whereas carbonic anhydrase (EcaA) facilitates conversion back to CO_2 in close proximity to ribulose-1,5-bisphosphate carboxylase/oxygenase in the inner thylakoid membrane [McGinn et al., 2003; Shibata et al., 2001; Soltes-Rak et al., 1997].

Under alkaline conditions, characteristic of the growth of the cyanobacterium *Synechocystis* [Fogg, 1956], most CO_2 is present in its hydrated form, HCO_3^- [Shibata et al., 2002]. The transport of HCO_3^- involves at least two transport systems. CmpABCD is a high-affinity ATP-dependent system, whereas SbtA is a low-affinity Na^+-dependent system that requires the presence of NtpJ, encoding a putative Na^+/K^+ transporter, for normal func-

[1] These two authors contributed equally to the work reported.

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2004 S. Karger AG, Basel
1464–1801/04/0073–0102$21.00/0

Accessible online at:
www.karger.com/mmb

Milton H. Saier, Jr.
Division of Biological Sciences
University of California at San Diego
La Jolla, CA 92093-0116 (USA)
Tel. +1 858 534 4084, Fax +1 858 534 7108, E-Mail msaier@ucsd.edu

Table 1. Homologues of SbtA

Species name	Abbre-viation	Gi number	Number of TMSs	Length (# aa's)
Cyanobacteria				
Nostoc sp. PCC 7120	Nsp1	17,229,626	10	370
Nostoc punctiforme	Npu	23,126,744	10	324
Nostoc sp. PCC 7120	Nsp2	17,228,481	10	322
Prochlorococcus marinus subsp. *marinus* str. CCMP1375	Pma1	33,239,693	10	331
Prochlorococcus marinus str. MIT 9313	Pma2	33,863,482	10	339
Prochlorococcus marinus subsp. *pastoris* str. CCMP1378	Pma3	33,860,773	10	332
Synechocystis sp. PCC 6803	Ssp	16,330,612	10	374
Gram-negative bacteria				
Caulobacter crescentus CB15	Ccr	16,127,739	10	365
Chloroflexus aurantiacus (fragment)[1]	Cau	22,972,265	9	294
Cytophaga hutchinsonii	Chu	23,137,488	10	327
Leptospira interrogans serovar lai str. 56601[2]	Lin	24,216,969	10	322
Shewanella oneidensis MR-1	Son	24,375,252	10	319
Gram-positive bacteria				
Bacillus halodurans	Bha	15,616,431	10	337
Mycobacterium tuberculosis H37Rv	Mtu	15,609,524	10	417

[1] The DNA fragment sequenced lacks the portion coding for the C-terminal part of this protein.
[2] The sequence provided in the NCBI database lacks the N-terminal 22 amino acyl residues because of an incorrect initiation codon choice.

tion [Bonfil et al., 1998; Shibata et al., 2002]. A Na^+ symport mechanism has been proposed [Shibata et al., 2002] in spite of earlier opposing considerations [Ritchie et al., 1996]. SbtA is not demonstrably homologous to mammalian HCO_3^-:Na^+ symporters [Sciortino and Romero, 1999; unpublished observation].

Computer Methods

PSI BLAST searches (e value $\leq 10^{-4}$) of the NCBI protein database were carried out with the protein sequence of SbtA as the query sequence [Altschul et al., 1997]. Redundant sequences were eliminated using an unpublished program [S. Singhi and M.H. Saier, Jr., unpublished]. The Clustal X program [Thompson et al., 1997] and the TREE program [Feng and Doolittle, 1990] were used for multiple alignment of homologous sequences and construction of a phylogenetic tree with the aid of the BLOSUM30 scoring matrix and the TREEVIEW drawing program. Family assignments were based upon phylogenetic results and the statistical analyses obtained with the GAP program [Devereux et al., 1984]. The standard for establishing homology between two proteins is nine standard deviations for regions of at least sixty residues that are compared with the GAP program, using 500 random shuffles with a gap opening penalty of eight and a gap extension penalty of two [Saier, 1994]. Sequence comparisons between multiple homologues were conducted using the IC program [Zhai and Saier, 2002], and individual pairs of protein sequences were compared using the GAP program [Devereux et al., 1984]. The TMS SPLIT program [Zhou et al., 2003] was used to generate fragmented protein sequences used for detection of internal repeats using the IC program [Zhai and Saier, 2002], the GAP program [Devereux et al., 1984], and the TMS-ALIGN program [Zhou et al., 2003]. The TMHMM [Krogh et al., 2001], HMMTOP [Tusnady and Simon, 1998], and WHAT [Zhai and Saier, 2001a] programs were used to estimate the topology of individual membrane proteins. The AveHAS program [Zhai and Saier, 2001b] was used for plotting the average hydropathy, similarity and amphipathicity as a function of alignment position for each family after aligning the sequences with the CLUSTAL X program [Thompson et al., 1997].

Results

Table 1 lists the homologues of SbtA identified by conducting PSI BLAST searches of the NCBI protein database. Homologues were derived from a variety of non-photosynthetic Gram-negative and Gram-positive bacteria in addition to several species of photosynthetic cyano-

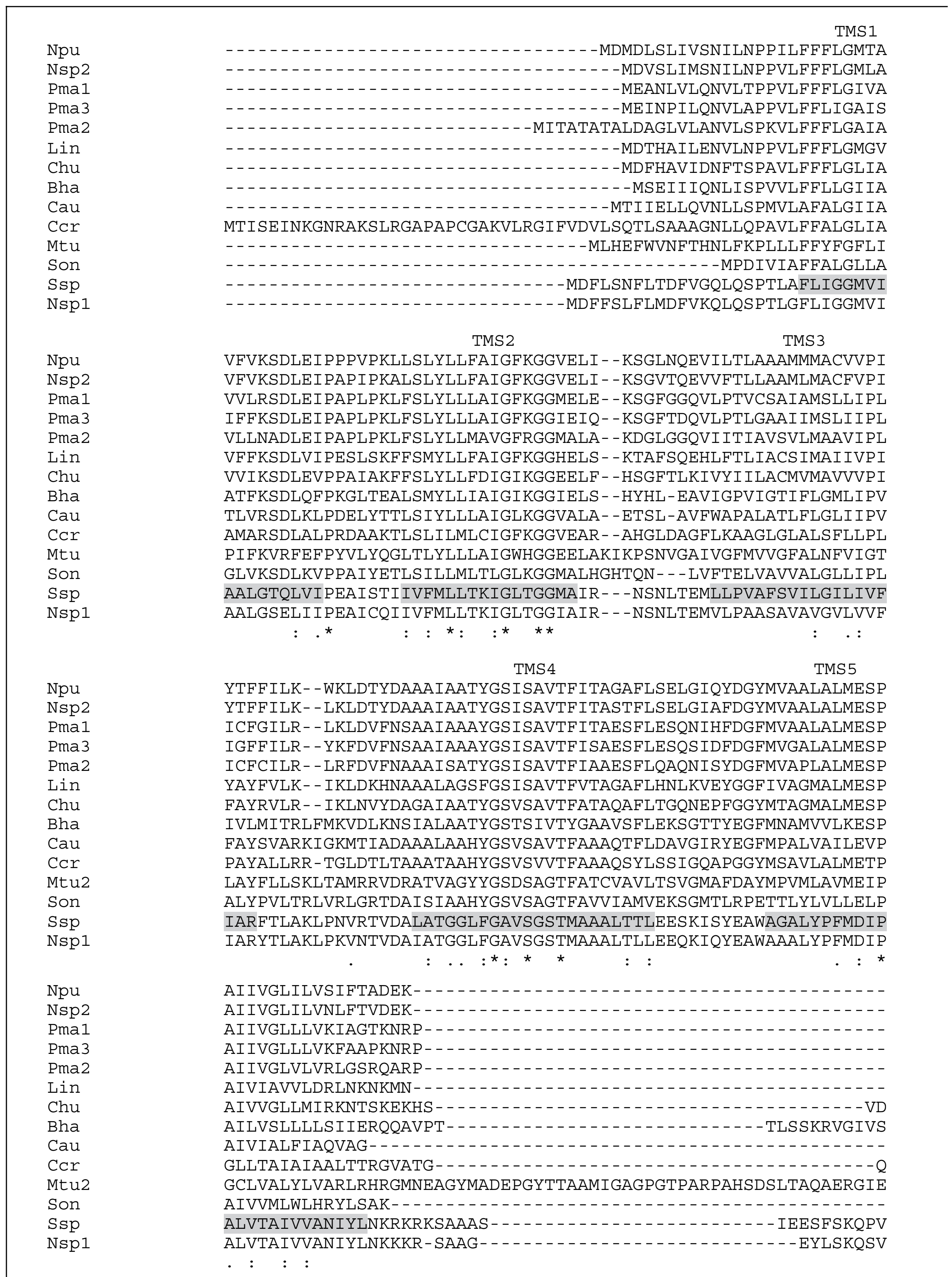


Fig. 1. Multiple sequence alignment of SbtA and its homologues. The Clustal X program was used to derive the alignment. * = Fully conserved residues; **:** = close conservation at a single position; **.** = distant conservation. The positions of the TMSs in SbtA (Ssp) are shown.

```
                                                                   TMS6
Npu            --REFAWS------------------------EVLQEAFLNSSVFLLVGSLLIGVLTGER
Nsp2           --REFAWS------------------------EVLQEAFLNSSVFLLVGSLFIGFLTGEH
Pma1           DSREMKWS------------------------TIIRESLLNGSVYLLLGSLLIGFLTAAH
Pma3           NSRKMHLS------------------------SILHESLLNGSVYLLLSSLIVGFLTAFS
Pma2           GSDGMNWR------------------------KVLHESMLNGYVVYLIAGSLVIGFIASIY
Lin            GGGAINWK------------------------ALLHEALFGSSIYLLVGALIVGYITGDS
Chu            AAHGSKMG------------------------EVLREAFFNGSIVLLVGSMIIGYLGGEH
Bha            GHQGLIDK------------------------EVLKESFFGKSIILLLGSLIIGLVVGER
Cau            GPQGGDWR------------------------EALRELITSKSILLLVGGMIIGWLTGPR
Ccr            DADRASAG------------------------KLAHEVLLNAASVVLIGGFLIGLITGEA
Mtu            EELELSLEKREHPNWDEDGVKDSGTNASIFSRELLQEVFLNPGLVLLFGGIVIGLISGLQ
Son            QPLQATVPNTQQS-------------------SILHEALTSRGVVLLVGGVVIGWLYGPT
Ssp            AAGDYGDQTDYPRTRQEYLSQQEPEDNRVKIWPIIEESLQGPALSAMLLGLALGIFTKPE
Nsp1           AAGEYPDQQDYPSSRQEYLRKQQSADNRVKIWPIVKESLQGPALSAMLLGIALGLFTQPE
                                                   .* : .     :  .. :* .

                                     TMS7                           TMS8
Npu            ---GWHVLEPFAQGLFYGILTFFLLDMGLVAARRIKDLQKTGVFLILFAILIPILNAGIG
Nsp2           ---GWQVLEPFTQGLFYGALTFFLLDMGLVAARRIKDLQKTGFFLILFAILIPILNAGIG
Pma1           NPIGVEKMQPFTGKLFYGAECFFLLDMGIVAAQRLPGLKKAGSFLIFFAVLIPLFNAFLG
Pma3           NPAAISKMEPFTGQLFYGAECFFLLDMGIVAAQRLPRLKNAGSYLIGFAIFMPLFNAFIG
Pma2           SPAGVEKMEPFVYKFFYGVLCFFLLDMGIVAAQRFKDLKKAGAFLIFFAILMPMFNALIG
Lin            ---GWKAEKVFADDLFKGILTFFLLDMGISAARRFKELANVGLFLIIAAIALMIVNATLG
Chu            ---GEVDLKPFVGGIFKGMLCLYLLDMGIVAGSRLSALKQSGVFLVAFGILTPIINATLG
Bha            ---AIPMVQPLFIDLYQSVLILFLLYMGLTVGERLPEVKKHGIKLILFGVLTPILLGALG
Cau            ---GGKEVAPLFVDLFKGALTFFLLELGMVAARRFRDLPSAGLFLLGFGIIMPILNGLLG
Ccr            ---GGERLKTFTGPVFQGVLCVFLLDLGVRAGRQLAAARGMNLGVLALGIVLPILGGVVA
Mtu            GQKVLHDDDNFFVAAFQGVLCLFLLEMGMTASRKLKDLASAGSGFVFFGLLAPNLFATLG
Son            GLAAISP---VLLGGFKTLLALFLLEMGLVTAKVCLPLPLQQWRLLVFAAVTPFALAWCG
Ssp            -----SVYEGFYDPLFRGLLSILMLIMGMEAWSRIGELRKVAQWYVVYSLIAPIVHGFIA
Nsp1           -----SVYKSFYDPLFRGLLSILMLVMGMEAWSRIGELRKVAQWYVVYSVVAPLVHGFIA
                          .    :     . :* :*: .              :  .       .  .

                                         TMS9                           TMS10
Npu            LAIA------KFIGMPPGNSLLFAVLCASASYIAVPAAMRMTVPEANPSLYVSTALAVTF
Nsp2           LLIA------KFIGMPAGDSLLFAVLSASASYIAVPAAMRLTVPEANPSLYVSTALAVTF
Pma1           VFVA------KALMLGPGNALLFAVLCASASYLAVPAAMRMTVPEAKASYYISTTLGLTF
Pma3           VFVA------RFLSLGPGNALLFVVLCASASYLAVPAAMRMTVPEARSSYYISTTLGLTF
Pma2           GLVA------RALGLGYGNALLFIILCSSASYIAVPTAMRMTVPEANPRYYISSALGLTF
Lin            LILT------KVIEMPAGDALMFVVLCASASYIAVPAAMKDMIPEANPSIYLTVALSIVF
Chu            IIIS------YLLGLNHGDALLFTILCASASYIAVPAAMRMAVPQANMSLLLPMSLGVTF
Bha            VLVG------TLAGLSVGGATLMGILAGSASYIAAPAALRTSVPEANPSIYLGLSLGVTF
Cau            VWFG------SLAGLSVGGSTILGVLAASASYIAAPAAVRIALPQANPEFYA--------
Ccr            LTLG------WMAGLPAGDLAALAVLAASASYIAAPAAMSMALPKADAGVYLTLSLGVTF
Mtu            IIVAHGYAYVTNNDFAPGTYVLFAVLCGAASYIAVPAVQRLAIPEASPTLPLAASLGLTF
Son            IGVG------LWLELPPGSILVLAGLSASASYIAAPAAIRAAIPEANIGLAMLASLGITF
Ssp            FGLG--MIAHYATGFSLGGVVVLAVIAASSSDISGPPTLRAGIPSANPSAYIGSSTAIGT
Nsp1           FGLG--MIAHYATGFSLGGVVILAVIAASSSDISGPPTLRAGIPSANPSAYIGASTAIGT
                  .           :  *     :  :..::* :: *..    :*.*


Npu            PFNIIVGIPLYLYGINLFWR----------
Nsp2           PFNIIVGIPLYQYGINLFWR----------
Pma1           PFNIVIGIPLYMGLVNNIIPSAG-------
Pma3           PFNIVLGIPIYMSLVNKIIPLSPL------
Pma2           PFNHTIGIPLYMGLVYKLIPASI-------
Lin            PINIVAGIPLYYYLVTLT------------
Chu            TFNIVVGIPVYYGIITSLI-----------
Bha            PFNLIFGIPLYFEFAKLLH-----------
Cau            ------------------------------
Ccr            PFNLTIGIPLLAIAAARLAGG---------
Mtu            SYNVTIGIPLYIEIARIVGQWFPATGASIG
Son            PVNVLIGLPLYQHWVMQITG----------
Ssp            PIAIGVCIPLFIGLAQTLGAG---------
Nsp1           PIAIGLAIPLFLGLAQAIGGR---------
```

bacteria. These proteins varied in length between 319 and 430 amino acyl residues (aa's) and exhibited 10 putative transmembrane segments (TMSs). Two shorter proteins (Lin and Cau) appeared to exhibit 9 TMSs, but one (Lin) proved to be truncated at the N-terminus because of an incorrect initial codon assignment (this error has been corrected in figure 1 and table 1 – see footnote 2 to table 1) while the other (Cau) is truncated at the C-terminus because DNA sequencing was incomplete (see fig. 1 and footnote 1 to table 1).

The multiple alignment was examined for well-conserved residues (fig. 1). Nine fully conserved residues occur in the first half of the alignment, and nine more occur in the second half. These residues in the first halves were (1) a P at the end of putative TMS 1, (2) an LX_4GX_2GG motif in TMS 2, (3) a GX_2SX_2T motif in TMS 4, and (4) a P in the middle of TMS 5. In the second half, fully conserved residues were: (1) an E in front of, and a G within TMS 6, (2) an LX_2G motif in TMS 7, and (3) a $GX_{12}SX_4PX_7PXA$ motif within and following TMS 9. It is apparent that motifs conserved in the first halves of these proteins are not the same as those conserved in the second halves. Although these two halves derive from a common ancestral sequence (see below), we conclude that they have diverged to serve dissimilar functions.

From the alignment of the 14 members of the SbtA family shown in figure 1, average hydropathy and average similarity plots were derived (fig. 2). As for the hydropathic profile of SbtA, the average hydropathy plot revealed 5 well-conserved peaks in the first half of the alignment (1–5) separated from 5 additional well-conserved peaks (6–10). A poorly conserved region separates the two halves of these proteins. In the two halves, similar hydropathy profiles were observed: Peaks 1, 3 and 5 in the first half of the alignment as well as peaks 6, 8 and 10 in the second half are more hydrophobic than the intervening peaks (2 and 4 in the first half; 7 and 9 in the second half). Because the average amphipathicity plot did not display noteworthy characteristics, it is not shown.

Using the IC and GAP programs, the first halves of SbtA family members could be shown to be homologous to the second halves. Thus, when a 75-residue segment from the first half of Mtu was compared with a 77-residue segment of the second half of Pma3, a comparison score of 9.0 standard deviations (SD) was obtained, establishing a common origin (fig. 3). These results suggest that SbtA homologues arose by an internal gene duplication event.

Using the TC-BLAST search tool, followed by statistical analysis of the aligned sequences using the GAP program, we could show that residues 3–106 of SbtA align

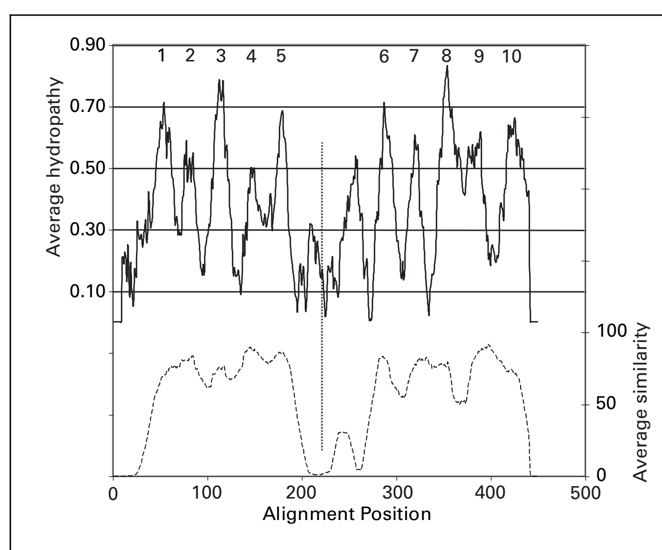


Fig. 2. Average hydropathy (top) and average similarity (bottom) plots for the 14 SbtA family members. The plot was generated using the AveHas program. The central vertical line indicates where the protein sequences were cut in order to compare the two halves of the proteins.

with residues 42–135 of the integral membrane arabinose transporter protein, AraH, of the ABC superfamily (TC 3.A.1.2.2). The alignment shown in figure 4 exhibited 35% identity and gave a comparison score of 9.8 SD, sufficient to strongly suggest homology. Surprisingly, TMSs 1–4 in SbtA align with TMSs 2–5 in AraH. As the AraH protein, like SbtA, appears to have 10 TMSs, as do most integral membrane constituents of the CUT2 family within the ABC superfamily, we suggest that SbtA and AraH arose from a common ancestral polypeptide chain but that intragenic rearrangement of one relative to the other occurred during their evolution.

The phylogenetic tree for the SbtA family is shown in figure 5. Surprisingly, the cyanobacterial proteins fall into two distant clusters. All other protein homologues are distantly related to these proteins as well as to each other. The phylogenies of the proteins do not follow the phylogenies of the source organisms. It is therefore clear that these proteins do not exhibit orthologous relationships.

Discussion

SbtA is likely to be a Na^+-dependent secondary bicarbonate uptake carrier. The available evidence suggests that it functions to accumulate intracellular bicarbonate

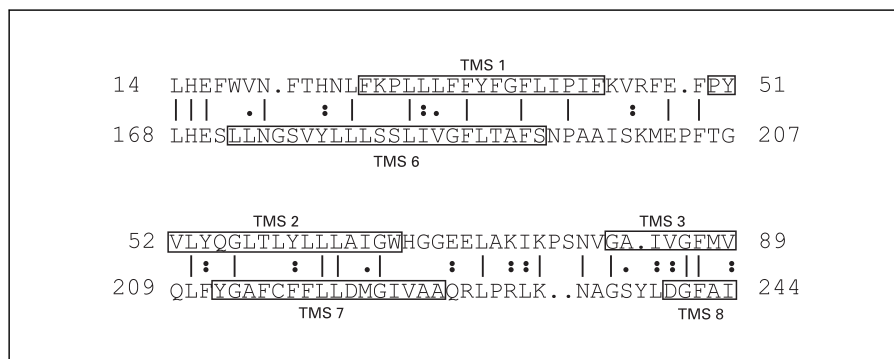


Fig. 3. Alignment of a 75 aa segment of the first half of the *Mycobacterium tuberculosis* homologue of SbtA (Mtu; top) with a 72 aa segment of the second half of a *Prochlorococcus marinus* homologue (Pma3; bottom). The two sequences show 30% identity and gave a comparison score of 9.0 SD. Residue numbers are provided at the beginning and end of each line. The positions of three putative TMSs in both proteins are shown.

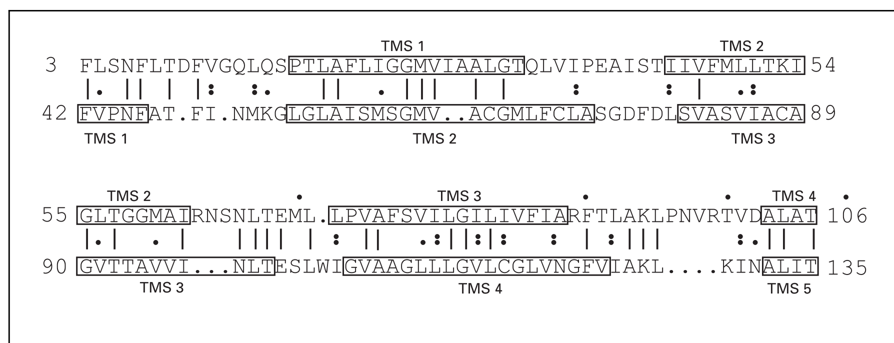


Fig. 4. Alignment of the N-terminal region of SbtA (top) with a portion of the N-terminal sequence of the integral membrane protein of the ABC-type L-arabinose transport protein, AraH (bottom). The alignment gave 35% identity and a comparison score of 9.8 SD. Note that TMSs 1–4 in SbtA align with TMSs 2–5 in AraH.

using a Na^+:HCO_3^- symport mechanism [Shibata et al., 2002]. As for a few other families of secondary carriers (e.g., DMT; TC # 2.A.8), we could show that the SbtA family arose by intragenic duplication of a 5 TMS repeat-encoding element to give a protein product of 10 TMSs. Assuming an odd number of TMSs per repeat unit, the two halves would be expected to have opposite orientation in the membrane [Saier, 2003]. Such an arrangement has been established for several families of transporters with odd numbers of TMSs in the repeat unit [Saier, 2003].

TMSs 1–4 in the 5 TMS integral membrane repeat element of SbtA homologues appeared to be homologous to TMSs 2–5 in the 10 TMS ABC-type transporter AraH. To the best of our knowledge, this is the first bioinformatic evidence that an ABC transporter shows a common origin with a secondary carrier. However, an independent report

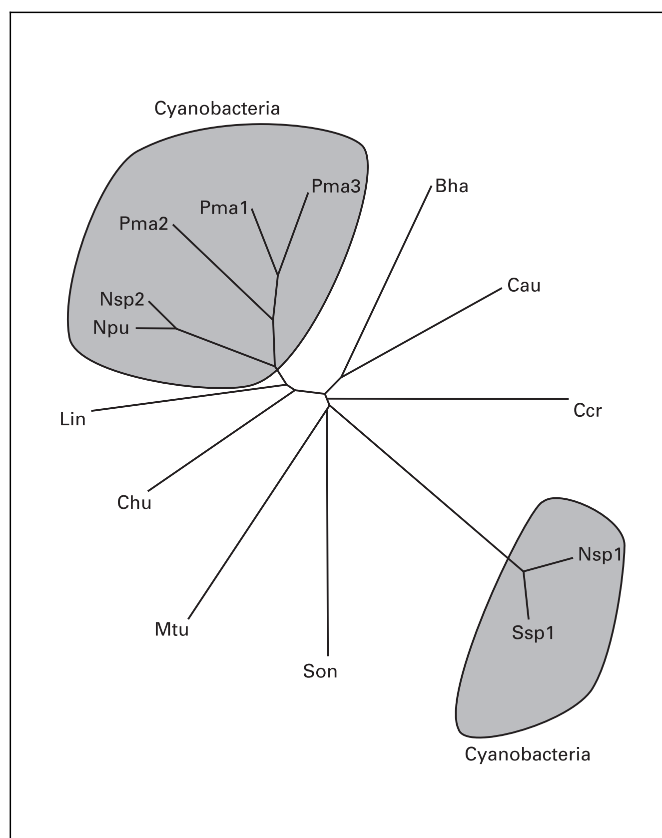


Fig. 5. Phylogenetic tree of the SbtA family. The two shaded regions show two separate clusters of the cyanobacterial proteins. Proteins from non-photosynthetic bacteria are outside of these clusters and branch from points near the center of the unrooted tree. The tree was based on the alignment shown in figure 1. Protein abbreviations are provided in table 1.

has recently shown that an ABC porter, LmrA of *Lactococcus lactis*, when severed from its ATP-hydrolyzing domain, can catalyze secondary active transport [Venter et al., 2003].

It is interesting to note that 5 TMS integral membrane constituents of ABC transporters are normally present as homo- or heterodimers, depending on the system, and that duplication of such a 5 TMS element to give a 10 TMS polypeptide chain should generate a full-length integral membrane transporter. The rarity of such an event within the ABC superfamily may be due to the fact that duplication of an odd number of TMSs requires that the two domains assume opposite orientation in the membrane, even though the primordial homodimeric transporter presumably has both subunits with the same orientation. The ease with which integral membrane protein domains can undergo topological inversion [Bogdanov and Dowhan, 1995, 1998; Bogdanov et al., 2002; Wang et al., 2002], attests to the plausibility of such an evolutionary pathway. Such topological inversions may involve simple transmembrane hairpin structures instead of an entire protein domain [Zhang et al., 2003]. Further experiments will be required to establish the functional consequences of the structural findings reported here.

References

Altschul SF, Madden TL, Schaeffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.

Bogdanov M, Dowhan, W: Phosphatidylethanolamine is required for in vivo function of the membrane-associated lactose permease of *Escherichia coli*. J Biol Chem 1995;270:732–739.

Bogdanov M, Dowhan W: Phospholipid-assisted protein folding: Phosphatidylethanolamine is required at a late step of the conformational maturation of the polytopic membrane protein lactose permease. EMBO J 1998;17:5255–5264.

Bogdanov M, Heacock PN, Dowhan W: A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition. EMBO J 2002;21:2107–2116.

Bonfil D, Ronen-Tarazi M, Sultemeyer D, Liemn-Hurwitz J, Schatz D, Kaplan A: A putative HCO_3^- transporter in the cyanobacterium Synechococcus sp. strain PCC 7942. FEBS 1998; 430:236–240.

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

Feng DF, Doolittle RF: Progressive alignment and phylogenetic tree construction of protein sequences. Methods Enzymol 1990;183:375–387.

Fogg GE: The comparative physiology and biochemistry of the blue-green algae. Bacteriol Rev 1956;20:148–165.

Krogh A, Larsson B, von Heijne G, Sonnhammer E: Predicting transmembrane protein topology with a hidden Markov model. Application to complete genomes. J Mol Biol 2001;305:567–580.

McGinn P, Price G, Maleszka R, Badger M: Inorganic carbon limitation and light control the expression of transcripts related to the CO_2-concentrating mechanism in the cyanobacterium Synechocystis sp. strain PCC6803. Plant Physiol 2003;132:218–229.

Ohkawa H, Pakrasi HB, Ogawa: Two types of functionally distinct NAD(P)H dehydrogenases in Synechocystis sp. strain PCC6803. J Biol Chem 2000a;275:31630–31634.

Ohkawa H, Price G, Badger M, Ogawa T: Mutation of *ndh* genes leads to inhibition of CO_2 uptake rather than HCO_3^- uptake in Synechocystis sp. strain PCC6803. J Bacteriol 2000b;182:2591–2596.

Omata T, Price G, Badger M, Okamura M, Gohta S, Ogawa T: Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium Synechococcus sp. strain PCC7942. Proc Natl Acad Sci USA 1999;96:13571–13576.

Ritchie R, Nadolny C, Larkun AWD: Driving forces for bicarbonate transport in the cyanobacterium Synechococcus R-2 (PCC7942). Plant Physiol 1996;112:1573–1584.

Saier MH Jr: Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution. Microbiol Rev 1994;58:71–93.

Saier MH Jr: Tracing pathways of transport protein evolution. Mol Microbiol 2003;48:1145–1156.

Sciortino CM, Romero MF: Cation and voltage dependence of rat kidney electrogenic Na^+-HCO_3^- cotransporter, rkNBC, expressed in oocytes. Am J Physiol 1999;277:611–623.

Shibata M, Katoh H, Sonoda M, Ohkawa H, Shimoyama M, Fukuzawa H, Kaplan A, Ogawa T: Genes essential to sodium-dependent bicarbonate transport in cyanobacteria: function and phylogenetic analysis. J Biol Chem 2002; 277:18658–18664.

Shibata M, Ohkawa H, Kaneko T, Fukuzawa H, Tabata S, Kaplan A, Ogawa T: Distinct constitutive and low-CO_2-induced CO_2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. Proc Natl Acad Sci USA 2001;98:11789–11794.

Soltes-Rak E, Mulligan M, Coleman J: Identification and characterization of a gene encoding a vertebrate-type carbonic anhydrase in cyanobacteria. J Bacteriol 1997;179:769–774.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876–4882.

Tusnady GE, Simon I: Principles governing amino acid composition of integral membrane proteins: application to topology prediction. J Mol Biol 1998;283:489–506.

Venter H, Shilling RA, Velamakanni S, Balakrishnan L, van Veen HW: An ABC transporter with a secondary-active multidrug translocator domain. Nature 2003;426:866–870.

Wang X, Bogdanov M, Dowhan W: Topology of polytopic membrane protein subdomains is dictated by membrane phospholipid composition. EMBO J 2002;21:5673–5681.

Zhai Y, Saier MH Jr: A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence. J Mol Microbiol Biotechnol 2001a;3:501–502.

Zhai Y, Saier MH Jr: A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins. J Mol Microbiol Biotechnol 2001b;3:285–286.

Zhai Y, Saier MH Jr: A simple sensitive program for detecting internal repeats in sets of multiply aligned homologous proteins. J Mol Microbiol Biotechnol 2002;4:29–31.

Zhang W, Bogdanov M, Pittard AJ, Dowhan W: Reversible topological organization within a polytopic membrane protein is governed by a change in membrane phospholipid composition. J Biol Chem 2003;278:50128–50135.

Zhou X, Yang NM, Tran CV, Hvorup RN, Saier MH Jr: Web-based programs for the display and analysis of transmembrane α-helices in aligned protein sequences. J Mol Microbiol Biotechnol 2003;5:1–6.

Research Article

Journal of Molecular Microbiology and Biotechnology

J Mol Microbiol Biotechnol 2004;8:129–140
DOI: 10.1159/000085786

The YedZ Family: Possible Heme Binding Proteins That Can Be Fused to Transporters and Electron Carriers

Torsten von Rozycki[1] Ming-Ren Yen[1] Erik E. Lende Milton H. Saier, Jr.

Division of Biological Sciences, University of California at San Diego, La Jolla, Calif., USA

Key Words
Bioinformatics · Membrane proteins · Phylogeny · Heme · Transport · Electron transfer

Abstract
YedZ of *Escherichia coli* is an integral 6 transmembrane spanning (TMS) protein of unknown function. We have identified homologues of YedZ in bacteria and animals but could not find homologues in Archaea or the other eukaryotic kingdoms. YedZ homologues exhibit conserved histidyl residues in their transmembrane domains that may function in heme binding. Some of the homologues encoded in the genomes of magnetotactic bacteria and cyanobacteria have YedZ domains fused to transport and electron transfer proteins, respectively. One of the animal homologues is the 6 TMS epithelial plasma membrane antigen of the prostate (STAMP1) that is overexpressed in prostate cancer. Animal homologues have YedZ domains fused C-terminal to homologues of coenzyme F_{420}-dependent NADP oxidoreductases. YedZ homologues are shown to have arisen by intragenic triplication of a 2 TMS-encoding element. They exhibit slight but statistically significant sequence similarity to two families of putative heme export systems and one family of cytochrome-containing electron carriers. We propose that YedZ homologues function as heme-binding proteins that can facilitate or regulate oxidoreduction, transmembrane electron flow and transport.

Copyright © 2004 S. Karger AG, Basel

Introduction

The transporter classification (TC) system [Busch and Saier, 2003; Saier, 2000] includes over 60 families of putative transport proteins (TC subclass 9.B). Reliable functional data are not available for any of the members of these families. A major task of the molecular biologist is to provide functional information about these and other uncharacterized proteins revealed by genome sequencing projects. This immense task will require the concerted efforts of molecular geneticists, biochemists, physiologists and bioinformaticists.

One of the families in TC subclass 9.B is the YedZ family (TC #9.B.43). YedZ of *Escherichia coli* has been examined topologically and has 6 established TMSs with

[1] These two authors contributed equally to the work reported.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2004 S. Karger AG, Basel

Accessible online at:
www.karger.com/mmb

Milton H. Saier, Jr.
Division of Biological Sciences
University of California at San Diego
La Jolla, CA 92093-0116 (USA)
Tel. +1 858 534 4084, Fax +1 858 534 7108, E-Mail msaier@ucsd.edu

both the N- and C-termini in the cytoplasm [Drew et al., 2002]. It belongs to the uncharacterized SwissProt UPF0191 family but has not been the subject of computational or molecular biological analyses. Nothing is known about the functions of it or any of its homologues.

In this paper, we examine the YedZ family using bioinformatic tools. We show that homologues of YedZ are found in a wide variety of bacteria including magnetotactic bacteria [Frankel et al., 1997, 1998; Schubbe et al., 2003; Schüler, 1999, 2002] where C-terminal YedZ domains are fused to 12 TMS putative permeases of the major facilitator superfamily [MFS; TC #2.A.1; Pao et al., 1998] and cyanobacteria where C-terminal YedZ domains are fused to 5 TMS putative transmembrane electron carriers distantly related to the DsbD family [TC #5.A.1; Kimball et al., 2003]. Additionally, we show that YedZ is related to a family of animal proteins of unknown function including the 6 TMS epithelial plasma membrane antigen of the prostate (STAMP1; gi15418732) that is overexpressed in prostate cancer [Hubert et al., 1999; Korkmaz et al., 2002; Yang et al., 2001]. In this protein and the other animal homologues, the YedZ domain is fused C-terminally to a 200-residue N-terminal domain that resembles coenzyme F_{420}-dependent $NADP^+$ oxidoreductases [Warkentin et al., 2001]. Finally, we note distant relationships between YedZ family members and various heme-containing and heme-transporting proteins including those found in the putative heme exporter (HemeE) family of ABC transporters (TC #3.A.1.107), the putative heme exporter protein (HEP) family (TC #9.B.14), and the heme-containing Cytb family (TC #5.B.2) of transmembrane electron carriers. These relationships and the presence of conserved histidyl residues in YedZ homologues indicate that the members of this family are heme-binding proteins. We present phylogenetic, structural and motif analyses that serve to characterize the YedZ family. The work presented provides a guide for researchers conducting functional studies concerned with members of the YedZ family and its distant relatives.

Results

Prokaryotic YedZ Homologues

Table 1 presents the prokaryotic YedZ homologues retrieved from the NCBI protein database. These proteins are derived exclusively from bacteria. Several different gram-negative bacterial kingdoms are represented. These kingdoms include the α-, β-, γ-, and ε-proteobacteria, cyanobacteria, *Chloroflexus* and *Deinococcus*. Only one gram-positive bacterium, *Clostridium acetobutylicum*, is represented. One homologue of YedZ (gi17988649) from *Brucella melitensis* (see table 1) is listed in the database as a 'bicyclomycin resistance protein', but no published paper describes the characterization of this protein. Another homologue (gi31194633) is listed in the database as being from an insect (*Anopheles gambiae*; table 1), but since this protein is 47% identical to the *E. coli* YedZ protein, this assignment is likely to be in error, being due to the sequencing of contaminating bacterial DNA. Alternatively, it could be the result of a relatively recent horizontal transfer event.

The majority of bacterial YedZ homologues are about 200 residues long with 6 putative TMSs (table 1). However, two proteins from magnetotactic bacteria of over 600 residues have 18 putative TMSs while four proteins from cyanobacteria of about 400 residues have 10 or 11 putative TMSs. As will be shown below, the two 18 TMS proteins have YedZ-like domains fused to MFS-type transport proteins with 12 TMSs, while the 10 or 11 TMS homologues have the YedZ domains fused to DsbD-like domains, probably with 5 TMSs. In these fusion proteins, the YedZ-like domains are C-terminal to the other domains.

Using the 73 sequences tabulated in table 1, a multiple alignment was generated (see our website: http://biology.ucsd.edu/~msaier/supmat/YedZ). Few gaps appeared in the multiple alignment. Preceding putative TMS 3 is a well- but not fully conserved motif

$$RRHyHyGLX_8\overset{*}{H}Hy$$

(Hy = a hydrophobic residue, X = any residue) where the histidyl residue indicated by the asterisk is fully conserved. Encompassing all of TMS 4 is a well-conserved motif

$$RP(F/Y)I(Hy)_3\overset{*}{G}(Hy)_2A(Hy)_9(S/T)\overset{*}{S}$$

with two fully conserved residues, a G and an S, both indicated by asterisks. Finally, TMS 5 includes two nearly, but not fully conserved histidyl residue [within the well-conserved motif, $HX(Hy)_2Y(Hy)_5(GA)(Hy)_2H(Hy)_4$]. Both histidines in this motif are conserved in all but the most distant homologues. The presence of multiple well or fully conserved histidines as well as a fully conserved serine suggests a common function, which might involve binding of a prosthetic group (such as heme) or formation of a transmembrane hydrophilic channel (such as for proton conduction).

Table 1. Prokaryotic members of the YedZ family

Species name	Proteobacterial subgroup	Protein abbreviation	Protein length (amino acids)	Number of TMSs[1]	gi No.[2]
Proteobacteria					
Actinobacillus pleuropneumoniae	γ	Apl	197	6	46143483
Agrobacterium tumefaciens str. C58	α	Atu	215	6	17935810
Azotobacter vinelandii	γ	Avi	200	6	23104538
Bordetella bronchiseptica RB50	β	Bbr	220	6	33602740
Bordetella parapertussis	β	Bpa	206	6	47117485
Bordetella pertussis Tohama I	β	Bpe	206	6	33591691
Bradyrhizobium japonicum USDA 110	α	Bja	222	6	27381604
Brucella melitensis 16M	α	Bme	220	6	17988649
Brucella suis 1330	β	Bsu	220	6	23500715
Burkholderia cepacia R1808	β	Bce	230	6	46319598
Burkholderia fungorum LB400	β	Bfu	196	6	48781581
Burkholderia mallei ATCC 23344	β	Bma	237	6	53723889
Burkholderia pseudomallei K96243	β	Bps	237	6	53720785
Campylobacter jejuni	ε	Cje	179	6	15791745
Caulobacter crescentus	α	Ccr	210	6	20178257
Chromobacterium violaceum ATCC 12472	β	Cvi	205	6	34102099
Dechloromonas aromatica RCB	β	Dar	221	6	41722900
Erwinia carotovora subsp. atroseptica	γ	Eca	199	6	50119223
Escherichia coli K12	γ	Eco	211	6	16129918
Haemophilus ducreyi 35000HP	γ	Hdu	197	6	33152896
Haemophilus somnus 129PT	γ	Hso	199	6	23467055
Magnetococcus sp. MC-1	α	Msp1	194	5	48833906
Magnetococcus sp. MC-1	α	Msp2	642	18	48832757
Magnetospirillum magnetotacticum MS-1	α	Mma1	218	6	46201352
Magnetospirillum magnetotacticum MS-1	α	Mma2	644	18	46201608
Methylococcus capsulatus str. Bath	γ	Mca1	229	6	53805011
Methylococcus capsulatus str. Bath	γ	Mca2	222	6	53803306
Microbulbifer degradans 2–40	γ	Mde	197	6	48864589
Novosphingobium aromaticivorans	α	Nar	194	6	48848208
Pasteurella multocida subsp. multocida	γ	Pmu	206	6	15602403
Photobacterium profundum	γ	Ppr	213	6	54302417
Polaromonas sp. JS666	β	Psp	216	6	54032238
Pseudomonas aeruginosa PAO1	γ	Pae1	202	6	15599885
Pseudomonas aeruginosa PAO1	γ	Pae2	209	6	15599298
Pseudomonas putida KT2440	γ	Ppu	197	6	26991359
Pseudomonas syringae pv. tomato	γ	Psy	210	6	28851443
Ralstonia eutropha JMP134	β	Reu	230	6	53760677
Ralstonia metallidurans CH34	β	Rme	224	6	48767809
Ralstonia solanacearum	β	Rso	217	6	17547699
Rhodobacter sphaeroides 2.4.1	α	Rsp	201	6	22957957
Rhodospirillum rubrum	α	Rru	208	6	48763376
Rubrivivax gelatinosus PM1	β	Rge	217	6	47574440
Salmonella enterica subsp. enterica	γ	Sen	199	6	29143626
Salmonella typhimurium LT2	γ	Sty	199	6	16766673
Shewanella oneidensis MR-1	γ	Son	215	6	24373603
Shigella flexneri 2a str. 2457T	γ	Sfl	211	6	30063409
Silicibacter sp. TM1040	α	Ssp	199	6	52009857
Sinorhizobium meliloti 1021	α	Sme	216	6	15965135
Wolinella succinogenes DSM 1740	ε	Wsu	189	6	34558056
Xanthomonas axonopodis pv. citri	γ	Xax	218	6	21242396
Xanthomonas campestris pv. campestris	γ	Xca	218	6	21231041
Yersinia pestis	γ	Ype	206	6	16123804

Table 1 (continued)

Species name	Proteobacterial subgroup	Protein abbreviation	Protein length (amino acids)	Number of TMSs[1]	gi No.[2]
Cyanobacteria					
Anabaena variabilis		Ava	402	10	45506513
Gloeobacter violaceus PCC 7421		Gvi1	221	6	37523126
Gloeobacter violaceus PCC 7421		Gvi2	202	6	37519978
Nostoc punctiforme PCC		Npu	394	10	53688732
Nostoc sp. PCC 7120		Nsp	423	11	17232203
Prochlorococcus marinus		Pma	204	6	33864529
Synechococcus elongatus		Sel1	393	11	45512098
Other gram-negative bacteria					
Chloroflexus aurantiacus		Cau	141	4	53795874
Deinococcus radiodurans R1		Dra	202	6	15807521
Gram-positive bacteria					
Clostridium acetobutylicum		Cac	235	7	15893845
Eukaryote					
Anopheles gambiae		Aga	201	6	31194633

All homologues are derived from gram-negative bacteria except for Cac from a gram-positive bacterium and Aga purportedly from a eukaryote. The latter protein may be a bacterial contaminant (see text).

[1] Number of putative transmembrane α-helical spanners as estimated using the WHAT program [Zhai and Saier, 2001b].

[2] GenBank index number.

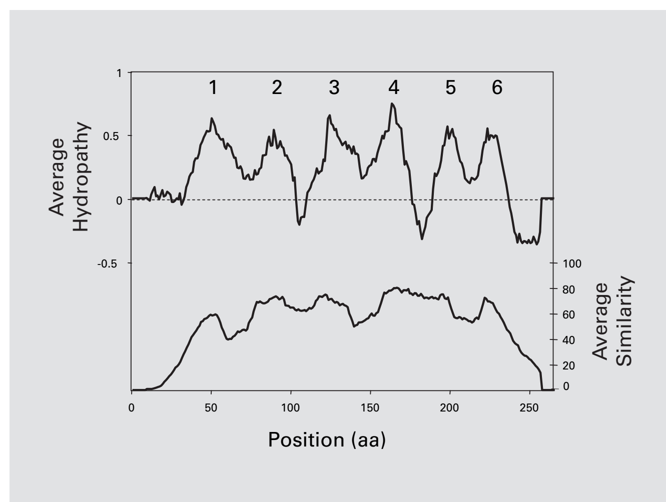


Fig. 1. Average hydropathy and similarity plots for the prokaryotic 6 TMS YedZ domains. The proteins and their abbreviations are listed in table 1. These plots were generated with the AveHAS program [Zhai and Saier, 2001b] using the multiple alignment generated with the Clustal X program. The N-terminal transport and oxidoreductase domains in the fused homologues from magnetobacteria and cyanobacteria, respectively, were removed before the plots were generated. The multiple alignment upon which these plots were based can be viewed on our website in figure S1 (http://biology.ucsd.edu/~msaier/supmat/YedZ).

Average hydropathy and similarity plots for the 6 TMS prokaryotic YedZ domains (fig. 1) revealed that TMSs 1 and 6 are less well conserved than TMSs 2–5. This observation correlates with the fact that the three well-conserved motifs noted above encompass TMSs 3–5.

The phylogenetic tree for the prokaryotic YedZ family is shown in figure 2. The proteins fall into clusters that correlate partially with their organismal origins. Thus, most of the α-proteobacterial homologues comprise three clusters, one of which includes the MFS-fused homologues in the magnetotactic bacteria; most of the β-proteobacterial homologues comprise another cluster, and most of the γ-proteobacterial homologues fall into three distinct clusters. One of these γ-proteobacterial clusters includes the putative mosquito protein, Aga, suggesting that this protein either is a contaminant derived from a γ-proteobacterium (most likely) or resulted from a recent horizontal transfer event (less likely). The remaining proteins are from different organismal types: (1) the gram-positive bacterium *C. acetobutylicum*, (2) the ε-proteobacterium *Campylobacter jejuni*, (3) two gram-negative bacteria that fall within their own kingdoms, *Chloroflexus aurantiacus* and *Deinococcus radiodurans*, and (4) cyanobacteria. All of the cyanobacterial proteins cluster together in figure 2.

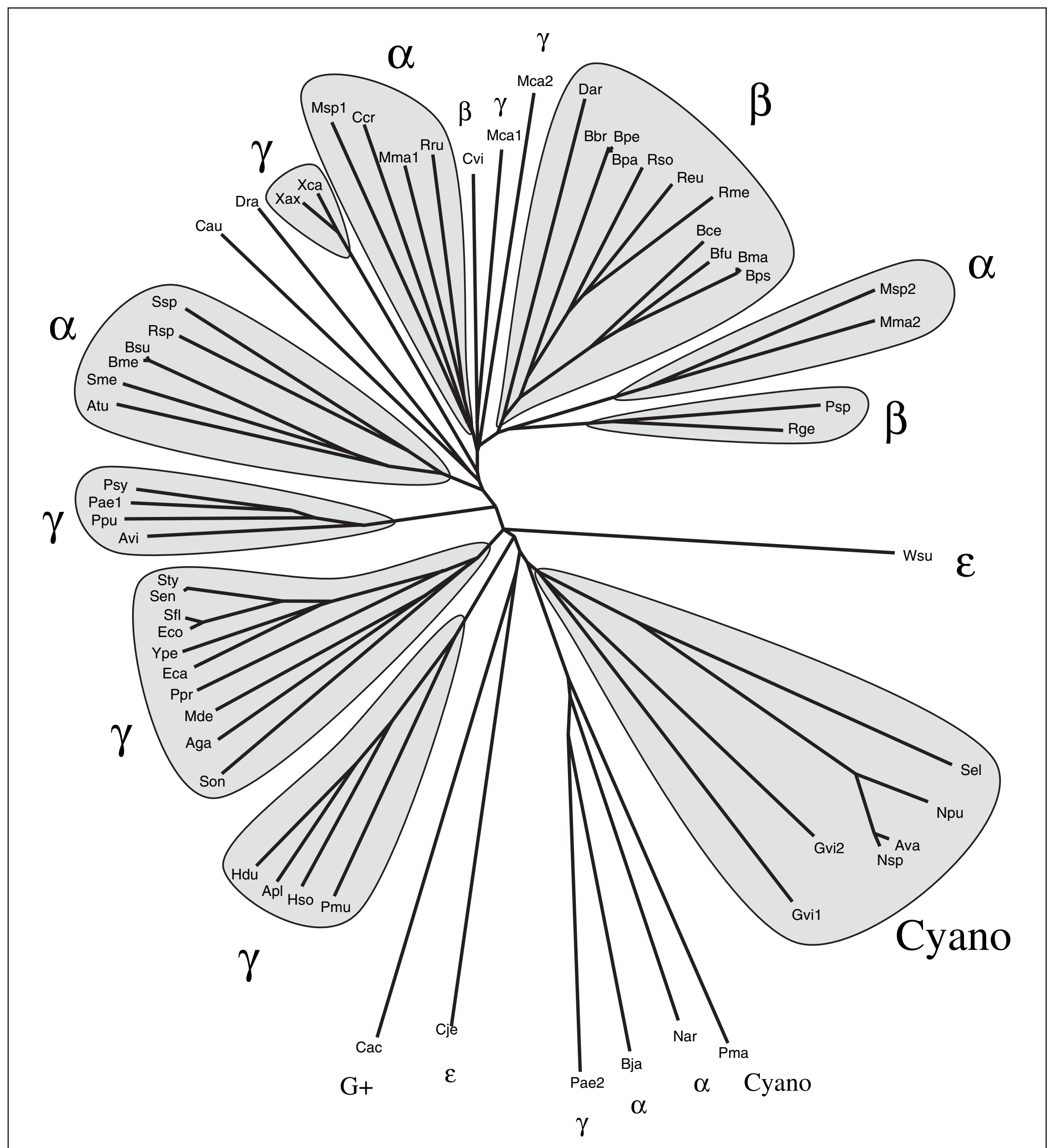


Fig. 2. Phylogenetic tree including the members of the bacterial YedZ family listed in table 1. The Clustal X program was used to generate the multiple alignment (see fig. S1 on our website) upon which the tree was based. The tree was drawn using the TREEVIEW program [Zhai et al., 2002]. Greek letters refer to the proteobacterial subgroups from which the protein members of a cluster were derived. Cyano = Cyanobacterial proteins; G+ = a protein from the gram-positive bacterium, *C. acetobutylicum*.

Eukaryotic YedZ Homologues

Among the mammalian paralogues were proteins annotated variously as 6 TMS epithelial and endothelial antigens of the prostate, tumor suppressors and regulators of the cell cycle and apoptosis [Hubert et al., 1999; Yang et al., 2001] (table 2). These and other animal homologues ranged in size from 339 to 754 aa, showing that none of them consists merely of a YedZ-like domain. The two large proteins, Gga3 and Gga4, from the chicken have an insertion and a C-terminal extension, respectively, that

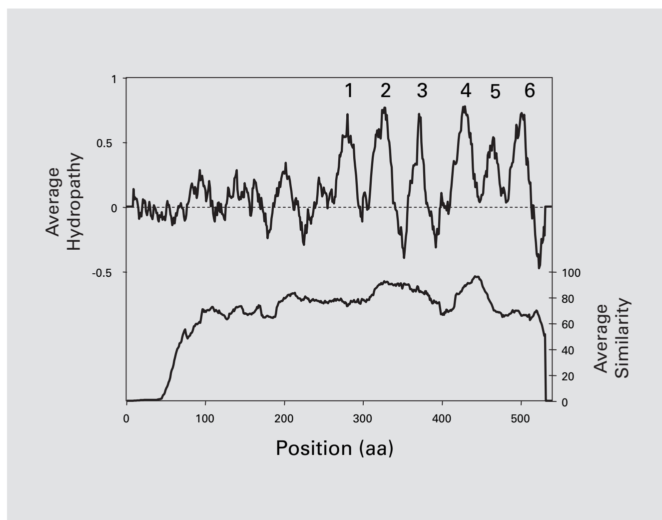


Fig. 3. Average hydropathy and similarity plots for the eukaryotic YedZ homologues. The proteins and their abbreviations are listed in table 2. The AveHAS program was used to generate the plot using the full-length proteins. The nonhomologous regions of the two large chicken homologues, Gga3 and Gga4, were removed before generation of the multiple alignment (fig. S2 on our website) upon which the plot was based.

were not homologous to anything else in the NCBI database. Those regions were removed for the analyses presented. In all of the remaining proteins, the YedZ-like domains proved to be C-terminal. They are distantly related to the bacterial YedZ proteins listed in table 1 as demonstrated below.

Table 2 shows that the three mammals represented (man, mouse and rat) as well as the chicken *(Gallus gallus)* each have 4 YedZ paralogues. The puffer fish, *Tetraodon nigroviridis*, has three while the frog, *Xenopus laevis*, has two. These 21 homologues were multiply aligned after removal of the nonhomologous regions of Gga3 and Gga4 (table 2; see fig. S2 on our website). Among the best-conserved regions in the eukaryotic homologues were segments that overlapped the three well-conserved motifs in the prokaryotic sequences. Thus, the three best-conserved regions overlapped TMSs 3, 4 and 5. These motifs are presented and compared with the corresponding prokaryotic motifs in table 3. Similarities, presented in bold print, are easily recognized.

Average hydropathy and average similarity plots were derived (fig. 3) based on the multiple alignment of the

Table 2. Eukaryotic members of the YedZ family

Species name	Abbreviation	Length (amino acids)	Number of TMSs	gi No.
Gallus gallus	Gga1	475	6	50732451
Gallus gallus	Gga2	489	6	50732453
Gallus gallus	Gga3	754	6	50750740
Gallus gallus	Gga4	715	4	50732447
Homo sapiens	Hsa1	486	6	51094921
Homo sapiens	Hsa2	490	6	15418732
Homo sapiens	Hsa3	488	6	27769110
Homo sapiens	Hsa4	459	6	33312462
Mus musculus	Mmu1	339	6	38174593
Mus musculus	Mmu2	489	6	28501136
Mus musculus	Mmu3	526	6	28372380
Mus musculus	Mmu4	470	6	14349146
Rattus norvegicus	Rno1	473	6	34854290
Rattus norvegicus	Rno2	529	6	34854293
Rattus norvegicus	Rno3	488	6	21717655
Rattus norvegicus	Rno4	464	4	33086486
Tetraodon nigroviridis	Tni1	379	4	47225469
Tetraodon nigroviridis	Tni2	463	6	47225470
Tetraodon nigroviridis	Tni3	491	6	47208402
Xenopus laevis	Xla1	474	6	38197323
Xenopus laevis	Xla2	458	6	37747633

The conventions of presentation are as described in the footnote to table 1.

animal homologues. The first half of the alignment proved to be hydrophilic while the second half includes the six peaks of hydrophobicity typical of YedZ-like domains. As for the bacterial homologues, the first and last putative TMSs showed the least sequence similarity (fig. 3).

The phylogenetic tree for the eukaryotic YedZ homologues is shown in figure 4. Four major branches stem from the center of the unrooted tree. All of these clusters contain the three mammalian orthologues plus the chicken orthologue. Each set of these orthologues shows the same branching arrangements but with differing degrees of sequence divergence in the order $2 < 3 < 1 < 4$. Thus, cluster 2 shows the least sequence divergence while cluster 4 shows the most. One of the *Tetraodon nigroviridis* paralogues is found in cluster 3 while the remaining two as well as the two *X. laevis* paralogues occur in cluster 4.

Table 3. Comparison of conserved motifs in the bacterial versus the animal YedZ homologues

Organismal group	Motif and location
(1) Bacteria	TMS #3 * **RR**OO**GL**OOOOOOOO**H**OOXOOOOXXX
Animals	* * * * * * * * * **RK**QO**GL**HSOOOAOO**H**OOYSOOOPOR A T
(2) Bacteria	TMS #4 * * O**K**RPFIOOO**G**OO**A**OOOOOOOOO**TS**XXX Y
Animals	* * ** * ** W**R**XDXYOSO**G**OO**G**OOOOOLLAO**TS**LPS **L** E A **A** G I
(3) Bacteria	TMS #5 LHXOOYOOOOOAOO**H**OOOO G
Animals	** * * * * QSXOLGOALOOXTO**H**OOOO T

The first (1), second (2) and third (3) motifs presented overlap TMSs 3, 4 and 5, respectively, in all YedZ domains. Fully conserved residues in each group are indicated with asterisks. Bold residues indicate well-conserved residues shared by most of the prokaryotic and eukaryotic proteins. O = Any hydrophobic residue; X = any residue. The letters below the primary amino acid sequences indicate the second most prevalent residue at a particular position.

Based on the tree configuration, it can be concluded that the duplication events that gave rise to the four clusters of paralogues occurred relatively early, before the different animal types shown diverged from each other.

YedZ Fusion Proteins in Bacteria

As noted above, two magnetotactic bacteria have YedZ-like domains fused C-terminally to 12 TMS MFS-type permeases (Msp2 and Mma2 in table 1; see fig. 5). The encoding genes were present in 'magnetosome islands' associated with magnetosome functions [Bazylinsky, 1995]. When the N-terminal MFS domains were TC blasted, the nearest functionally characterized homologue in the TCDB proved to be the *Staphylococcus aureus* tetracycline resistance protein (TC #2.A.1.3.6; PO2983) with a 146 residue segment that showed 20.5% identity, 30.0% similarity and a comparison score of 9 SD. Both magnetotactic bacterial orthologues showed extensive sequence similarity with a 12 TMS (428 aa) protein MamH of *Magnetospirillum gryphiswaldense* (Acc. #CAE12030; 39% identity; 59% similarity) [Schubbe et al., 2003] as well as an MFS homologue from *Chlorobium tepidum* (gi|21674295, 25% identity; 44% similarity). MamH is 29.6% identical and 36.4% similar throughout its length with the *E. coli* Tet^R protein (P02982; TC #2.A.1.2.4; gap score of 15.6 SD). It can therefore be concluded that the MFS homologues fused to YedZ domains in magnetotactic bacteria comprise a novel family within the MFS (TC #2.A.1.43), which, however, may be more closely related to drug efflux pumps of MFS families 2.A.1.2 and 2.A.1.3 than to other currently recognized families within the MFS [Saier, 2000].

The four fused cyanobacterial proteins showed N-terminal regions that exhibited five putative TMSs (fig. 6), all nearly equally well conserved. Npu exhibits sequence similarity to an 80-residue segment of the DsbD protein of *E. coli* (TC #5.A.1.1.1; 29% identity; 48% similarity) as well as to a 180-residue segment in the DsbD protein of *Pasteurella multocida* (Q9CP40). This last comparison included putative TMSs 2–5 in both proteins and gave a comparison score of 11.4 SD (32% identity; 42% similarity). In DsbD, TMSs 2–4 and TMSs 5–7 are homologous and arose by an intragenic duplication event [Kimball et al., 2003]. We therefore attempted to detect an internal repeat in the cyanobacterial proteins, but these attempts were unsuccessful.

A tentative suggestion from these observations is that the four cyanobacterial homologues listed in table 1 may be involved in transmembrane electron transport, as are all DsbD homologues. It should be noted that DsbD ho-

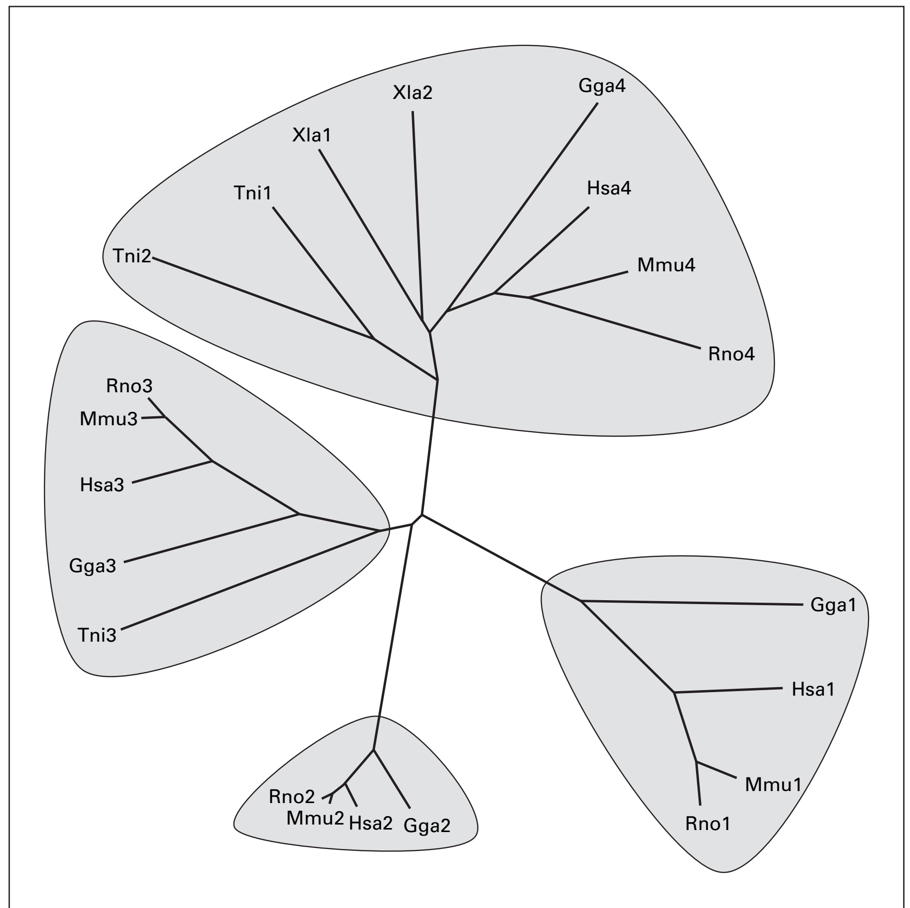


Fig. 4. Phylogenetic tree for the eukaryotic members of the YedZ family listed in table 2. The tree, based on the multiple alignment shown in figure S2 on our website, was prepared as described in the legend to figure 2.

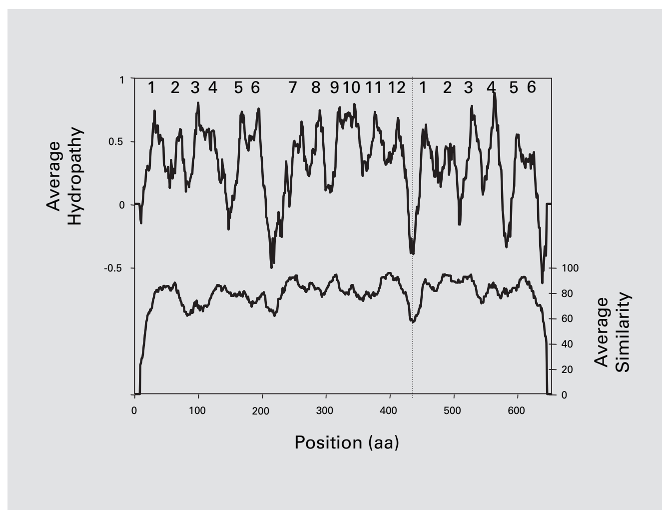


Fig. 5. Average hydropathy and similarity plots for the two YedZ fusion proteins found in magnetotactic bacteria. The vertical dotted line indicates the boundary between the MFS permease domain and the fused YedZ domain.

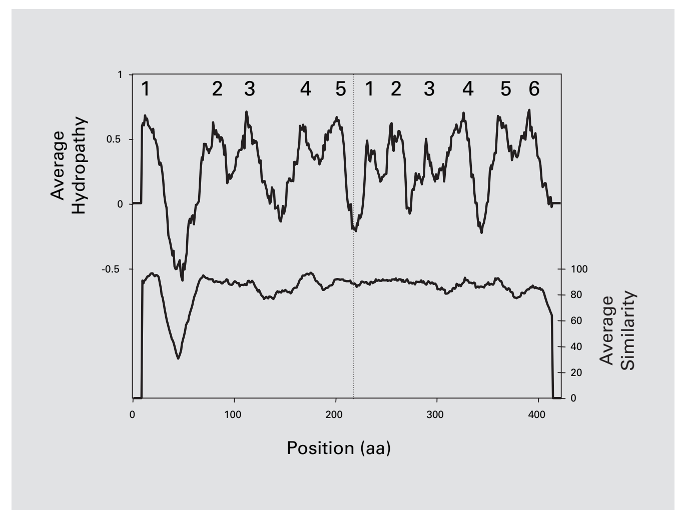


Fig. 6. Average hydropathy and similarity plots for the four cyanobacterial YedZ fusion proteins. The vertical dotted line indicates the boundary between the DsbD-like domain and the C-terminal YedZ domain.

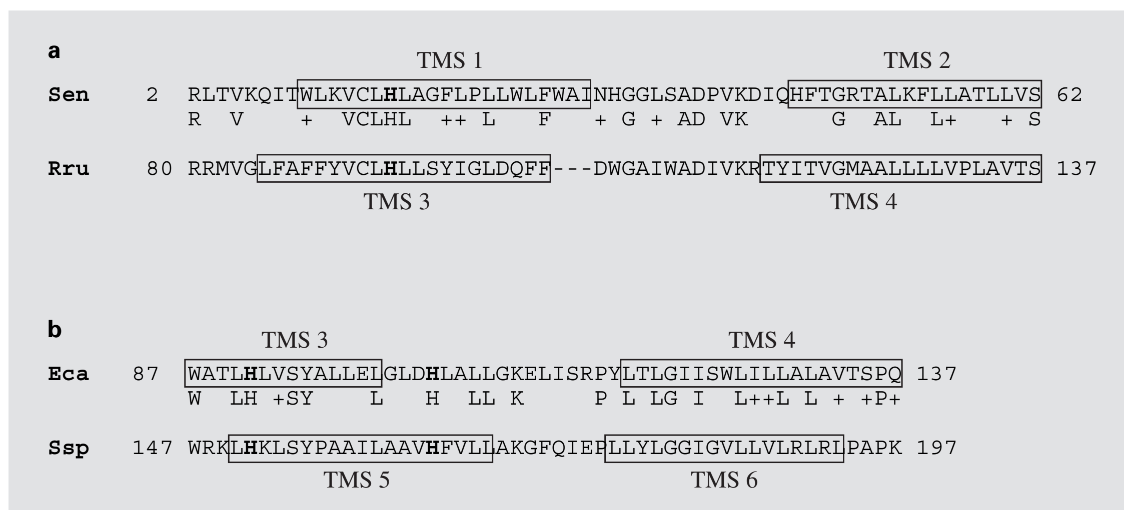


Fig. 7. Alignments revealing the intragenic triplication event that probably gave rise to members of the YedZ family. **a** TMSs 1–2 of Sen aligned with TMSs 3–4 of Rru (table 1). **b** TMSs 3–4 of Eca aligned with TMSs 5–6 of Ssp. The boxes show the positions of the putative TMSs, and the bold numbers above them indicate the numbers of these TMSs in the proteins. Residue numbers within the proteins are presented at the beginning and end of each line. Conserved residues between the two sequences are indicated between these sequences. + indicates aligned residues of a similar character. The Hs in bold indicate fully conserved histidines. Only the first histidine in TMS 3 and the second histidine in TMS 5 are fully conserved in all of the eukaryotic homologues.

mologues serve a variety of functions including (1) cytochrome c biogenesis, (2) extracytoplasmic dithiol:disulfide exchange, (3) Cu^{2+} and Hg^{2+} resistance, and (4) methylamine utilization [Kimball et al., 2003]. Since the cyanobacterial homologues listed in table 1 are only distantly related to these proteins and exhibit 5 rather than 6–9 putative TMSs, we cannot assign specific functions to them.

Evidence that YedZ Arose by Intragenic Triplication

The IC and GAP programs were used to attempt to find internal repeats in the YedZ homologues. The results clearly suggested that the 6 TMS topology arose by triplication of a primordial 2 TMS unit. For example, as shown in figure 7a, TMSs 1 and 2 of Sen align with TMSs 3 and 4 of Rru, with a sequence identity of 31% and a similarity of 42% (GAP comparison score of 8.1 SD) while TMSs 3 and 4 of Eca align with TMSs 5 and 6 of Ssp with a sequence identity of 37% and a similarity of 48% (GAP comparison score of 10.7 SD). These values, particularly the latter one, are strongly indicative of homology. In both alignments, the conserved histidines are in the odd numbered TMSs. In figure 7b, there are two well-conserved histidines, and these are separated by 12 residues. When drawn in a helical wheel, almost all of the residues conserved between the two sequences localized to one side of the helix (data not shown). The only other family of 6 TMS proteins to have been shown to have arisen by intragenic triplication is the eukaryotic mitochondrial carrier family (TC #2.A) [Kuan and Saier, 1993].

Distant YedZ Homologues Involved in Heme Recognition

YedZ homologues proved to exhibit weak similarity with cytochrome c biogenesis proteins/heme exporters of the putative HEP family (TC #9.B.14) and the HemeE family (TC #3.A.1.107) of the ABC superfamily. Members of these families have topologies that vary between 6 and 15 TMSs. They correspond to two distantly related families in the SwissProt database, the CcmF/CycK (HEP) and CcmC/CycZ (HemeE) families. Some of these proteins are known to exhibit heme-binding sites. When the YedZ homologue, Pmu (table 1), was compared with the 6 TMS cytochrome c biogenesis protein of *Bradyrhizobium japanicum* (CycK; TC #9.B.14.1.2), an overlapping region of 230 residues gave 26% identity, 37% similarity and a comparison score of 8.5 SD. Two of the well-

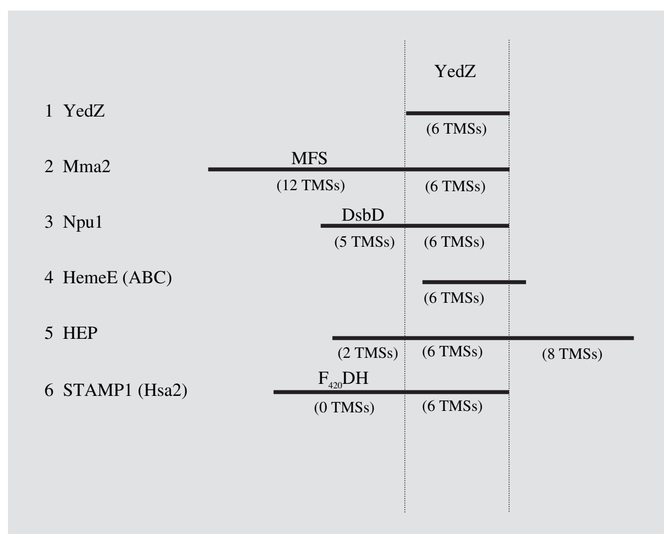


Fig. 8. Schematic depiction of YedZ homologues: (1) YedZ of *E. coli*; (2) Mma2, a fusion protein between a putative magnetosome MFS permease and a YedZ domain; (3) Npu, a cyanobacterial fusion protein between a putative transmembrane electron carrier domain, distantly related to the DsbD family, and a YedZ domain; (4 and 5) homology between YedZ and members of the two heme exporter families, HEP and HemeE, respectively, and (6) STAMP1 (Hsa2), a fusion protein between a YedZ domain and a coenzyme F_{420}-dependent $NADP^+$ oxidoreductase homologue.

conserved histidyl residues in the YedZ family were conserved in this alignment. When a 74-residue segment of Pmu (130–203) was compared with the corresponding region of CycK (residues 226–303) within the 230 region cited above, a comparison score of 10.6 SD (30.1% identity and 38.4% similarity) was obtained. These results suggest an evolutionary interrelationship between (1) the putative heme exporter (HemeE) family of the ABC superfamily (TC #3.A.1.107), (2) the putative heme exporter (HEP) family (TC #9.B.14), and (3) the YedZ family (TC #9.B.43). It is therefore possible that the YedZ, HEP and HemeE families are all related and are concerned with heme biosynthesis, reception and/or transport.

Discussion

In this paper we have used bioinformatic approaches to characterize a family of 6 TMS integral membrane proteins which we have called the YedZ family. In bacteria, these proteins proved to exist (1) as single 6 TMS proteins of about 200 aa, (2) as C-terminal domains fused to proteins that comprise a novel family of the MFS (TC #2.A.1.43), probably found in the magnetosome membranes of magnetotactic bacteria [Tamegai and Fukumori, 1994], and (3) as C-terminal domains fused to a novel family of cyanobacterial putative transmembrane electron flow carriers that resemble cytochrome biogenesis/disulfide:dithiol exchange proteins of the DsbD family (TC #5.A.1). Moreover, in animals, the YedZ domain is fused C-terminally to homologues of ubiquitous coenzyme F_{420}-dependent $NADP^+$ oxidoreductases [Warkentin et al., 2001]. All of these protein types are schematized in figure 8.

A surprising observation concerns the distribution of YedZ homologues in the living world. Thus, gram-negative bacteria, especially proteobacteria and cyanobacteria, have many such homologues, but only one gram-positive bacterium, *C. acetobutylicum* proved to have such a homologue. Because the completely sequenced genomes of many gram-positive bacteria are available for analysis, this fact is surprising. Evidently, the prokaryotic YedZ family is largely restricted to gram-negative bacteria. Perhaps *C. acetobutylicum* acquired its YedZ homologues by horizontal transfer from a gram-negative bacterium. Moreover, no homologues were identified in Archaea, and in eukaryotes, they could be found only in animals. In spite of several searches conducted using a variety of approaches, we did not detect YedZ homologues in plants, fungi or protozoans. Perhaps an ancient but rare horizontal transfer event, involving gene transfer from a gram-negative bacterium to a primordial animal occurred, giving rise to the surprising distribution of these proteins.

Further examination of the proteins of the YedZ family revealed that these proteins are distantly related to two families of integral membrane transporters, both putatively concerned with heme export (HEP, TC #9.B.14 and HemeE, TC #3.A.1.117; see fig. 8). The latter family is a constituent family within the ABC superfamily. We therefore propose that YedZ homologues are integral membrane heme binding proteins. When they occur fused to novel MFS transporters in magnetotactic bacteria, they may regulate or function in conjunction with the transporters. The substrates of the fused permeases might, for example, be Fe^{2+} or Fe^{3+} [Frankel et al., 1983; Fukumori et al., 1997], and the YedZ domains might allow sensitivity of the transporter to oxygen or the redox state of the cell. When these domains are fused to distant members of the bacterial DsbD family (TC #5.A.1), or to members of the ubiquitous coenzyme F_{420}-dependent oxidoreductases, the fused YedZ domain might either play a direct role in electron transfer or regulate electron flow catalyzed

by the N-terminal domain. It should be noted that these suggestions are speculative, and that another function of the YedZ domains might, for example, be heme transport instead of heme binding and regulation of the N-terminal domain. Further experimentation will be required to establish the significance of the findings reported here.

Computer Methods

With the protein sequence of YedZ from *E. coli* as query, the PSI-Blast search tool was used to identify proteins of similar sequences [Altschul et al., 1997]. These sequences were retrieved from the NCBI database (e value $\leq 10^{-4}$). Redundant sequences were eliminated using an unpublished program [Singhi and Saier, unpubl.]. The Clustal X program [Thompson et al., 1997] and the TREE program [Feng and Doolittle, 1990] were used for multiple alignment of homologous sequences and construction of a phylogenetic tree with the aid of the BLOSUM30 scoring matrix and the TREEVIEW drawing program. Family assignments were based upon the phylogenetic results, and statistical analyses were performed with the GAP program [Devereux et al., 1984]. The standard for establishing homology between two proteins is 9 SDs for regions of at least 60 residues that are compared with the GAP program, using 500 random shuffles with a gap opening penalty of 8 and a gap extension penalty of 2 [Saier, 1994]. Sequence comparisons between multiple homologues were performed using the IC program [Zhai and Saier, 2002], and individual comparisons were made using the GAP program [Devereux et al., 1984]. The TMS SPLIT program [Zhou et al., 2003] was used to generate fragmented protein sequences used for analysis of internal duplications using the IC program [Zhai and Saier, 2002], the GAP program [Devereux et al., 1984] and the TMS-ALIGN program [Zhou et al., 2003]. The TMHMM [Krogh et al., 2001], HMMTOP [Tusnady and Simon, 1998], and WHAT [Zhai and Saier, 2001b] programs were used to estimate the topologies of individual membrane proteins [von Heijne, 1986, 1991]. The AveHAS program [Zhai and Saier, 2001a] was used for plotting the average hydropathy, similarity and amphipathicity as a function of alignment position for each family after aligning the sequences with the Clustal X program [Thompson et al., 1997].

The FusionDB database (http://igs-server.cnrs-mrs.fr/FusionDB/) was used to verify gene fusion events. FusionDB is a database of bacterial and archaeal gene fusions [Suhre and Claverie, 2004]. Annotated genes were checked for fusions by screening 89 fully sequenced bacterial and archaeal genomes. A fusion event between two genes from a given reference genome in a given target genome [Yanai et al., 2003] was defined as follows: (1) Each of the two reference genes must match the same open reading frame (ORF) in the target genome as their highest scoring BLAST hit. (2) Both reference genes must not be homologous to each other. (3) The overlap between the BLAST hits of both genes must be less than 10% of the size of the smaller one of the two target genes. (4) When split in the middle between the two BLAST hits, the two halves of the target ORF must match back to the original two reference genes as their best BLAST hit. All triplets of genes (the two genes from the reference genome and the ORF from the target genome) that fulfill the above criteria are called putative fusion events. Each putative fusion event is subjected to a scoring scheme based on different evaluations of its pair-wise and multiple alignments.

Acknowledgments

This work was supported by NIH grant GM64368. We thank Mary Beth Hiller for her assistance in the preparation of the manuscript.

References

Altschul, S.F., Madden, T.L., Schaeffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Bazylinski, D.A. 1995. Structure and function of the bacterial magnetosome. ASM News 61: 337–343.

Busch, W., and Saier, M.H., Jr. 2003. The IUBMB-endorsed transporter classification system. Methods Mol. Biol. 227:21–36.

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

Drew, D., Sjöstrand, D., Nilsson, J., Urbig, T., Chin, C., de Gier, J.-W., and von Heijne, G. 2002. Rapid topology mapping of *Escherichia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis. Proc. Natl. Acad. Sci. USA 99:2690–2695.

Feng D.-F., and Doolittle, R.F. 1990. Progressive alignment and phylogenetic tree construction of protein sequences. Methods Enzymol. 183: 375–387.

Frankel, R.B., Bazylinski, D.A., Johnson, M.S., and Taylor, B.L. 1997. Magneto-aerotaxis in marine coccoid bacteria. Biophys. J. 73:994–1000.

Frankel, R.B., Bazylinski, D.A., and Schüler, D. 1998. Biomineralization of magnetic iron minerals in magnetotactic bacteria. Supramol. Sci. 5:383–390.

Frankel, R.B., Papaefthymiou, G.C., Blakemore, R.P., and O'Brien, W. 1983. Fe_3O_4 precipitation in magnetotactic bacteria. Biochim. Biophys. Acta 763:147–159.

Fukumori, Y., Oynagi, H., Yoshimatsu, K., Noguchi, Y., and Fujiwara, T. 1997. Enzymatic iron oxidation and reduction in magnetite synthesizing *Magnetospirillum magnetotacticum*. J. Phys. IV 7:659–662.

Hubert, R.S., Vivanco, I., Chen, E., Rastegar, S., Leong, K., Mitchell, S.C., Madraswala, R., Zhou, Y., Kuo, J., Raitano, A.B., Jakobovits, A., Saffran, D.C., and Afar, D.E. 1999. STEAP: A prostate-specific cell surface antigen highly expressed in human prostate tumors. Proc. Natl. Acad. Sci. USA 96:14523–14528.

Kimball, R.A., Martin, L., and Saier, M.H., Jr. 2003. Reversing transmembrane electron flow: The DsbD and DsbB protein families. J. Mol. Microbiol. Biotechnol. 5:133–149.

Korkmaz, K.S., Elbi, C.C., Korkmaz, C.G., Loda, M., Hager, G.L., and Saatcioglu, F. 2002. Molecular cloning and characterization of STAMP1, a highly prostate specific six-transmembrane protein that is overexpressed in prostate cancer. J. Biol. Chem. 277:36689–36696.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. 2001. Predicting transmembrane protein topology with a hidden Markov model. Application to complete genomes. J. Mol. Biol. 305:567–580.

Kuan, J., and Saier, M.H., Jr. 1993. The mitochondrial carrier family of transport proteins: Structural, functional and evolutionary relationships. Crit. Rev. Biochem. Mol. Biol. 28: 209–233.

Nilsson, I., Saaf, A., Whitley, P., Gafvelin, G., Waller, C., and von Heijne, G. 1998. Proline-induced disruption of a transmembrane α-helix in its natural environment. J. Mol. Biol. 284:1165–1175.

Pao, S.S., Paulsen, I.T., and Saier, M.H., Jr. 1998. The major facilitator superfamily. Microbiol. Mol. Biol. Rev. 62:1–32.

Saier, M.H., Jr. 1994. Computer-aided analyses of transport proteins sequences: Gleaning evidence concerning function, structure, biogenesis, and evolution. Microbiol. Rev. 58:71–93.

Saier, M.H., Jr. 2000. A functional-phylogenetic classification system for transmembrane solute transporters. Microbiol. Mol. Biol. Rev. 64: 354–411.

Schubbe, S., Kube, M., Scheffel, A., Wawer, C., Heyen, U., Meyerdierks, A., Madkour, M.H., Mayer, F., Reinhardt, R., and Schüler D. 2003. Characterization of a spontaneous nonmagnetic mutant of *Magnetospirillum gryphiswaldense* reveals a large deletion comprising a putative magnetosome island. J. Bacteriol. 185:5779–5790.

Schüler, D. 1999. Formation of magnetosomes in magnetotactic bacteria. J. Mol. Microbiol. Biotechnol. 1:79–86.

Schüler, D. 2002. The biomineralization of magnetosomes in *Magnetospirillum gryphiswaldense*. Int. Microbiol. 5:209–214.

Suhre, K., and Claverie, J.M. 2004. Fusion DB: A database for in-depth analysis of prokaryotic gene fusion events. Nucleic Acids Res. 32: D273–D276.

Tamegai, H., and Fukumori, Y. 1994. Purification and some molecular and enzymatic features of a novel *ccb*-type cytochrome c oxidase from a microaerobic denitrifier, *Magnetospirillum magnetotacticum*. FEBS Lett. 347:22–26.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and 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.

Tusnady, G.E., and Simon, I. 1998. Principles governing amino acid composition of integral membrane proteins: Application to topology prediction. J. Mol. Biol. 283:489–506.

von Heijne, G. 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. EMBO J. 5:3021–3027.

von Heijne, G. 1991. Proline kinks in transmembrane alpha-helices. J. Mol. Biol. 218:499–503.

Warkentin, E., Mamat, B., Sordel-Klippert, M., Wicke, M., Thauer, R.K., Iwata, M., Iwata, S., Ermler, U., and Shima, S. 2001. Structures of $F_{420}H_2$:NADPH$^+$ oxidoreductase with and without its substrates bound. EMBO J. 20: 6561–6569.

Yanai, I., Wolf, Y.I., and Koonin, E.V. 2003. Evolution of gene fusions: Horizontal transfer versus independent events. Genome Biol. 4:301.

Yang, D., Holt, G.E., Velders, M.P., Kwon, E.D., and Kast, W.M. 2001. Murine six-transmembrane epithelial antigen of the prostate, prostate stem cell antigen, and prostate-specific membrane antigen: Prostate-specific cell-surface antigens highly expressed in prostate cancer of transgenic adenocarcinoma mouse prostate mice. Cancer Res. 61:5857–5860.

Zhai, Y., and Saier, M.H., Jr. 2001a. A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins. J. Mol. Microbiol. Biotechnol. 3:285–286.

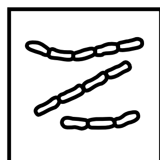
Zhai Y., and Saier, M.H., Jr. 2001b. A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence. J. Mol. Microbiol. Biotechnol. 3:501–502.

Zhai, Y., and Saier, M.H., Jr. 2002. A simple sensitive program for detecting internal repeats in sets of multiply aligned homologous proteins. J. Mol. Microbiol. Biotechnol. 4:29–31.

Zhai, Y., Tchieu, J., and Saier, M.H., Jr. 2002. A web-based Tree View (TV) program for the visualization of phylogenetic trees. J. Mol. Microbiol. Biotechnol. 4:69–70.

Zhou, X., Yang, N.M., Tran, C.V., Hvorup, R.N., and Saier, M.H., Jr. 2003. Web-based programs for the display and analysis of transmembrane α-helices in aligned protein sequences. J. Mol. Microbiol. Biotechnol. 5: 1–6.

Comparative and Functional Genomics
Comp Funct Genom 2005; **6**: 17–56.
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/cfg.454



Research Paper

Genomic analyses of transport proteins in *Ralstonia metallidurans*

Torsten von Rozycki[1], Dietrich H. Nies[2] and Milton H. Saier Jr[1]*
[1] *Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093-0116, USA*
[2] *Molekulare Mikrobiologie, Institut für Mikrobiologie, Martin-Luther-Universität Halle-Wittenberg, 06099 Halle, Germany*

**Correspondence to: Milton H. Saier Jr, Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093-0116, USA. E-mail: msaier@ucsd.edu*

Received: 14 May 2004
Revised: 14 December 2004
Accepted: 15 December 2004

Abstract

Ralstonia (Wautersia, Cupriavidus) metallidurans **(Rme) is better able to withstand high concentrations of heavy metals than any other well-studied organism. This fact renders it a potential agent of bioremediation as well as an ideal model organism for understanding metal resistance phenotypes. We have analysed the genome of Rme for genes encoding homologues of established and putative transport proteins; 13% of all genes in Rme encode such homologues. Nearly one-third of the transporters identified (32%) appear to function in inorganic ion transport with three-quarters of these acting on cations. Transporters specific for amino acids outnumber sugar transporters nearly 3 : 1, and this fact plus the large number of uptake systems for organic acids indicates the heterotrophic preferences of these bacteria. Putative drug efflux pumps comprise 10% of the encoded transporters, but numerous efflux pumps for heavy metals, metabolites and macromolecules were also identified. The results presented should facilitate genetic manipulation and mechanistic studies of transport in this remarkable bacterium.** **Copyright © 2005 John Wiley & Sons, Ltd.**

Keywords: **bioinformatics; transport proteins; comparative genomics**

Introduction

Ralstonia metallidurans (Rme; previously *Alcaligenes eutrophus,* renamed in 2004 *Wautersia metallidurans* and then *Cupriavidus metallidurans*; Goris *et al*., 2001; Vandamme and Coenye, 2004; Vaneechoutte *et al*., 2004), is a Gram-negative facultative chemolithoautotrophic β-proteobacterium. It was first identified in 1976, when it was isolated from industrial sediments, soils and wastes that were polluted with high concentrations of various heavy metals, such as cobalt, zinc, nickel and cadmium (Mergeay *et al*., 1985). The concentrations of these metals that can exist in the habitats of Rme greatly exceed the values that are lethal to almost any other living organisms. Rme is related to the important plant pathogen *Ralstonia solanacearum* (Boucher *et al*., 2001), which is resistant to a wide variety of drugs and toxic compounds. The complete genome sequence of the latter organism is available (Salanoubat *et al*., 2002).

The properties of Rme render it potentially important for purposes of bioremediation, such as for the degradation of aromatic compounds and xenobiotics, even in the presence of heavy metals as additional pollutants. Rme is also able to synthesize polyhydroxyalkalonates (PHAs), which accumulate as carbon and energy sources and might be useful for the development of biodegradable plastics. The extraordinary heavy metal resistance of Rme and its ability to accumulate these metals on its surface make it a candidate for a variety of clean-up purposes (Legatzki *et al*., 2003a; Mergeay *et al*., 2003; Nies, 2003).

Two low copy number plasmids, pMOL30 (238 kb; Mergeay *et al*., 1985) and pMOL28 (180 kb; Taghavi *et al*., 1997), that are stably carried by Rme strain CH34, are primary determinants of the remarkable heavy metal resistance characteristic of Rme (Legatzki *et al*., 2003a,b). Both are self-transferable at low frequencies, potentially offering a new approach for inserting resistance genes into other organisms. Rme lacks the RecBCD

Copyright © 2005 John Wiley & Sons, Ltd.

pathway for DNA degradation — a property that allows it to serve as an acceptor for foreign resistance genes. The fact that specific transport systems responsible for the uptake and export of various metabolites and heavy metals (Andres *et al.*, 2000; Borremans *et al.*, 2001; Goris *et al.*, 2001; Juhnke *et al.*, 2002; Mergeay *et al.*, 2003; Nies, 2003; Roux *et al.*, 2001) have been better characterized in Rme than in any other bacterium (Nies, 2003), renders Rme a model organism for basic research on metal resistance and homeostasis.

It has been suggested that the resistance of Rme to heavy metals and toxic compounds results from multiple layers of efflux pumps with overlapping substrate specificities (Juhnke *et al.*, 2002; Nies, 2003; Silver, 2003). However, comprehensive genome analyses of the transporters in Rme are still lacking. In this paper we correct this deficiency, reporting bioinformatic studies of all recognizable transporters encoded within the genome of Rme.

Computer methods

The protein sequences of Rme were extracted from the JGI database and downloaded for all of the analyses reported here. The sequencing work done at JGI (**http://genome.jgi-psf.org/draft_microbes/ralme/ralme.home.html**) and the annotation project performed by the CH34 annotation consortium (**http://genome.ornl.gov/microbial/rmet/**) formed the basis of this work and are acknowledged at this point. Since the names of the CH34 genes have changed many times in the past, as has the name of the organism, cross-reference tables are supplied as supplementary material (**http://bionomie.mikrobiologie.uni-halle.de/SupMat/SupplMat.htm**). Computer-aided searches were conducted to retrieve all proteins encoded within the genome that are recognizably homologous to transport system constituents included in the Transporter Classification Database (TCDB; Busch and Saier 2002; Tran *et al.*, 2003). Briefly, all proteins encoded within the genome were blasted in an automated manner (using BLASTP) against TCDB. Additional databases used for protein functional analysis were the non-redundant SWISSPROT and TrEMBL protein sequence databases. Several protein pattern databases (conserved domain databases at NCBI and Pfam) were also used. Charge bias analyses of membrane protein topology were performed using the TMHMM (Krogh *et al.*, 2001) and WHAT (Zhai and Saier, 2001) programs.

Results and discussion

Topological predictions for membrane transporter homologues

The proteome of Rme was analysed for topological predictions; 59% (4072) of the 6985 proteins identified have no predicted TMSs, while 21% (1434) have only one putative TMS. While most of the former proteins are likely to be cytoplasmic, many of the latter will undoubtedly prove to be periplasmic and outer membrane proteins; 8% (580) have two or three TMSs, 5% (320) have four to six TMSs, and 3% each (196 and 223) have seven to 10 and $>$10 TMSs, respectively. Relative to most other prokaryotes analysed, Rme has increased proportions of integral membrane proteins of all topological types (Paulsen *et al.*, 2000).

All putative transport protein constituents recognized in the proteome of Rme were similarly analysed for topology; 932 putative transporter proteins (13%) were recognized in the proteome of Rme. This percentage is higher than observed for most other organisms with fully sequenced genomes (Paulsen *et al.*, 2000). About 24% (227) of these proteins may be cytoplasmic, as they exhibit no putative TMSs. All others are potential integral membrane constituents. Of these, 21% (196) are predicted to have one TMS, 9% (88) have two or three TMSs, 16% (146) have four to six TMSs, 10% (94) have seven to 10 TMSs, and 19% (179) have $\geqslant$11 TMSs. Many of the one-TMS proteins displayed typical leader sequences at their respective amino-termini and may be secreted via the Sec and Tat export systems (see below). They may be receptors for ABC-, TRAP-T- and TTT-type transport systems (see below). Since transporter families include proteins that are almost always concerned exclusively with transport (Saier, 2003), it is probable that nearly all of these proteins function in transmembrane transport.

Copyright © 2005 John Wiley & Sons, Ltd.

Classes of transporters found in *R. metallidurans*

According to the transporter classification (TC) system, transporters are classified into five well-defined categories (classes 1–5) and two poorly defined categories (classes 8 and 9). The well-defined categories are; (a) channels; (b) secondary carriers; (c) primary active carriers; (d) group translocators; and (e) transmembrane electron flow carriers (Busch and Saier, 2002; Saier, 2000). The less well-defined proteins include (8) auxiliary transport proteins and (9) transporters or putative transporters of unknown mechanism of action or function (Saier, 2000).

Table 1 summarizes the distribution of the 932 transporter protein constituents from Rme in each of the major TC categories and also provides a breakdown of these proteins found in the various TC subclasses; 123 channel proteins, most of them outer membrane porins, were identified. However, the majority of defined transport proteins found are secondary carriers (304) and constituents of primary active transporters (343).

Only one phosphoenolpyruvate-dependent, sugar transporting phosphotransferase system (PTS) permease, which catalyses group translocation of hexoses, was found. Further, only 10 transmembrane electron flow system constituents were identified. This latter fact may in part reflect the limited representation of transmembrane electron flow carriers in the Transporter Classification Database (TCDB). Thirty-one auxiliary proteins of TC class 8 and 65 putative transporters of TC class 9 were identified (Table 1). The probable functional identities of the individual proteins will be discussed below.

Classes of substrates transported

Table 2 summarizes the numbers of transporter proteins concerned with the transport of various types of substrates; 300 proteins are putative transport protein homologues concerned with the uptake or efflux of inorganic ions, and nearly three-quarters of them are concerned with inorganic cation transport. This observation undoubtedly relates to the remarkable heavy metal resistance of Rme.

Forty-one systems specific for sugars and their derivatives and 110 systems specific for amino acids and their derivatives were identified. These facts suggest that amino acid metabolism may be more important to Rme than sugar metabolism for heterotrophic growth. This substrate preference of Rme has been observed before (Mergeay *et al*., 1985). Rme has 142 transport protein homologues putatively concerned with carboxylate transport, which also agrees with the substrate spectrum of this bacterium (Mergeay *et al*., 1985). This fact, together with the greater number of secondary carriers relative to primary active transporters, points to a strong metabolic dependency on respiration rather than fermentation. Ninety-one

Table 1. Categories of recognized transport proteins found in *Ralstonia metallidurans*

TC class	No. of transporters (%)	TC subclass	No. of transporters (%)
1 Channels	123 (13)	1.A. α-Type channel-forming proteins and peptides	27 (3)
		1.B. Outer membrane porins (ß-structure)	94 (10)
		1.C. Pore-forming toxins (proteins and peptides)	1 (0.1)
		1.E. Holins	1 (0.1)
2 Secondary carriers	304 (33)	2.A. Carrier-type facilitators	299 (32)
		2.C. Ion-gradient-driven energizers	5 (1)
3 Primary transporters	343 (37)	3.A. P-P bond hydrolysis-driven transporters	290 (31)
		3.B. Decarboxylation-driven active transporters	2 (0.2)
		3.D. Oxidoreduction-driven active transporters	51 (5)
4 Group translocators (PTS)	2 (0.2)	4.A. Phosphotransferase systems	2 (0.2)
5 Transmembrane electron carriers	10 (1)	5.A. Transmembrane electron transfer carriers	10 (1)
8 Auxiliary transport proteins	31 (3)	8.A. Auxiliary transport proteins	31 (3)
9 Poorly-defined systems	65 (7)	9.A. Transporters of unknown classification	10 (1)
		9.B. Putative uncharacterized transporters	55 (6)
Unclassified	54 (6)	Unclassified	54 (6)
Total number	932 (100)		932 (100)

Copyright © 2005 John Wiley & Sons, Ltd.

Table 2. Breakdown of transport proteins according to predicted substrate types in *Ralstonia metallidurans*

Substrate class	No. of transporters (%)	Substrate subclass	No. of transporters (%)
1 Inorganic compounds	300 (32)	Cations	221 (24)
		Anions	78 (8)
		H_2O	1 (0.1)
2 Organic compounds	400 (43)	Sugars/sugar metabolites	41 (4)
		Amino acids/polyamines	110 (12)
		Mono-, di-, tricarboxylates	
		Fatty acids	142 (15)
		Drugs/toxic compounds	91 (10)
		Nucleotides/nucleosides	4 (0.4)
		Aromatics	13 (1)
3 Macromolecules	102 (11)	Lipoproteins/proteins	75 (8)
		Lipopolysaccharides/polysaccharides	20 (2)
		DNA	5 (0.5)
		Lipids	1 (0.1)
4 Miscellaneous/unknown	130 (14)	Miscellaneous	15 (2)
		Unknown	115 (12)
Total	932 (100)		932 (100)

proteins are predicted to be concerned with transport of drugs and hydrophobic substances, while 130 proteins fall into the miscellaneous/unknown category.

Global analysis of transporters in Rme and their family associations

Table 3 summarizes the results of our detailed analyses of transporters found in Rme. On the left, the family TC number, the name of the family and its standard abbreviation can be found (columns 1–3). Column 4 presents the types of substrates known to be transported by members of the respective family. Column 5 gives the number of family members identified in Rme, while column 6 presents the gene designation used in the draft version (02jul03) of the Rme genome analysed here. A full version of this table that contains all of the various names of the CH34 genes is provided as supplementary material (**http://bionomie.mikrobiologie.uni-halle.de/SupMat/Roz_05/Table 3.htm**). Column 7 gives the protein size in number of amino acyl residues, and column 8 provides an estimate of the number of putative transmembrane spanning regions (TMSs) for each protein. The TC number of the protein in TCDB that shows greatest similarity to the Rme ORF under consideration is presented in column 9. Finally, column 10 presents the level of confidence for the functional assignment (1 = sure, 2 = probable, 3 = uncertain or unknown).

Channels

In category 1A (α-type channels), Rme possesses two members of the VIC family (1.A.1), both probably K^+ channels. Two members of the MIP family of aqua/glycerol porins are also present. Four putative chloride channels (ClC family) were found, as well as one CytB homologue. This last system may function primarily in transmembrane electron flow, but no bacterial member of this family has been characterized (Kimball and Saier, 2002).

MscL (1.A.22), MscS (1.A.23) and MIT (1.A.35) families are all well represented with one, nine and four members, respectively. All four MIT family members are probably divalent cation transporters, while the MscL and MscS proteins are most likely non-specific channels for protection against osmotic stress (Busch and Saier, 2002; Nottebrock *et al*., 2003; Pivetti *et al*., 2003). Rme exhibits two paralogues within the hsp70 family of chaperone proteins, some of which have been shown to be capable of forming transmembrane channels (Arispe and De Maio, 2000). No other channel-type proteins of TC class 1.A could be recognized.

A tremendous number of putative outer membrane porins were identified. For example, just within the general bacterial porin (GBP) family (1.B.1), 29 paralogues were found. Most of these proteins are of 300–400 amino acids in length and probably consist largely of β-structure. A trimeric

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Putative transport proteins identified in *Ralstonia metallidurans*[a]

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
I.A. α-Type channel-forming proteins and peptides									
I.A.1	Voltage-gated ion channel	VIC	Na^+, K^+, Ca^{2+}, multiple cations		Contig372gene5732	307	5	I.A.1.2.3(1)	3
				2	Contig373gene6187	229	7	I.A.1.13.1(1)	3
I.A.8	Major intrinsic protein	MIP	H_2O, glycerol, urea, polyols, NH_3, CO_2		Contig375gene7720	250	6	I.A.8.13.1(1)	3
				2	Contig375gene8643	234	6	I.A.8.3.1(1)	2
I.A.11	Chloride channel	ClC	Cl^-, anions		Contig365gene3384	376	4	I.A.11.6.1(1)	3
					Contig365gene3245	657	12	I.A.11.6.1(1)	3
					Contig352gene1115	560	8	I.A.11.6.1(1)	3
				4	Contig367gene3837	522	10	I.A.11.6.1(1)	3
I.A.20	$gp91_{phox}$ Phagocyte NADPH oxidase-associated cytochrome b558	CytB	H^+	1	Contig363gene2857	447	6	I.A.20.6.1(1)	2
I.A.22	Large conductance mechanosensitive ion channel	MscL	Proteins, ions (slightly cation selective)	1	Contig367gene3927	144	2	I.A.22.1.3(1)	2
I.A.23	Small conductance mechanosensitive ion channel	MscS	Ions (slight anion selectivity)		Contig350gene938	456	6	I.A.23.1.1(1)	2
					Contig373gene6759	357	4	I.A.23.1.1(1)	3
					Contig375gene8191	275	3	I.A.23.1.1(1)	3
					Contig375gene8051	832	9	I.A.23.1.1(1)	2
					Contig375gene7869	288	4	I.A.23.2.1(1)	2
					Contig372gene5633	284	4	I.A.23.2.1(1)	2
					Contig371gene5100	771	13	I.A.23.3.1(1)	3
					Contig358gene1757	570	5	I.A.23.4.1	3
				9	Contig361gene2514	447	5	I.A.23.4.1(1)	3
I.A.30	H^+- or Na^+-translocating bacterial flagellar motor 1ExbBD outer membrane transport energizer	Mot/Exb-Mot	H^+, Na^+		Contig371gene5340	299	4	I.A.30.1.1(2)	2
				2	Contig371gene5341	325	1	I.A.30.1.1(2)	2
I.A.33	Cation channel-forming heat shock protein-70	Hsp70	Ions, polypeptides		Contig370gene5054	621	0	I.A.33.1.2(1)	2
				2	Contig372gene5852	648	0	I.A.33.1.2(1)	2
I.A.35	CorA metal ion transporter	MIT	Heavy-metal ions, Mg^{2+}, Mn^{2+}, Co^{2+}, Ni^{2+}, Fe^{2+}, Al^{3+}, Mn^{2+}		Contig363gene2888	320	2	I.A.35.1.2(1)	2

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
					Contig374gene7317	383	3	1.A.35.3.1(1)	3
					Contig365gene3224	362	3	1.A.35.3.1(1)	3
				4	Contig367gene3951	393	2	1.A.35.3.1(1)	3
1.B. Outer membrane porins (β-structure)									
1.B.1	General bacterial porin	GBP	Ions, small (M_r of <1000 Da) molecules		Contig358gene1907	367	1	1.B.1.4.1(1)	2
					Contig340gene366	432	3	1.B.1.4.1(1)	3
					.00Dec2000-Contig485gene1110			1.B.1.4.1(1)	3
					Contig373gene6049	381	1	1.B.1.4.1(1)	3
					Contig373gene6801	393	6	1.B.1.4.1(1)	3
					Contig374gene7003	111	1	1.B.1.4.1(1)	2
					Contig338gene292	382	1	1.B.1.4.1(1)	3
					Contig375gene8510	352	0	1.B.1.4.1(1)	2
					Contig375gene8258	379	1	1.B.1.4.1(1)	2
					Contig375gene8489	355	3	1.B.1.4.1(1)	2
					Contig375gene9145	353	1	1.B.1.4.1(1)	2
					Contig375gene9285	341	0	1.B.1.4.1(1)	3
					Contig354gene1343	377	1	1.B.1.4.1(1)	2
					Contig364gene3106	371	1	1.B.1.4.1(1)	2
					Contig375gene8681	352	1	1.B.1.4.1(1)	2
					Contig373gene6282	386	1	1.B.1.4.1(1)	2
					Contig371gene5423	391	1	1.B.1.4.1(1)	2
					Contig372gene5670	387	2	1.B.1.4.1(1)	2
					Contig372gene5912	358	0	1.B.1.4.1(1)	2
					Contig372gene5687	362	0	1.B.1.4.1(1)	2
					Contig370gene4663	361	1	1.B.1.6.1(1)	3
					Contig373gene6196	382	1	1.B.1.6.1(1)	2
					Contig357gene1691	374	1	1.B.1.6.1(1)	2
					Contig358gene1746	354	1	1.B.1.6.1(1)	2
					Contig361gene2442	371	1	1.B.1.6.1(1)	2
					Contig373gene6139	355	3	1.B.1.6.1(1)	2
					Contig373gene6638	371	1	1.B.1.6.1(1)	3
					Contig371gene5227	363	1	1.B.1.6.1(1)	2
				29	Contig369gene4254	355	3	1.B.1.6.1(1)	2
1.B.6	OmpA-OmpF porin	OOP	Ions, small molecules		Contig370gene4785	217	0	1.B.6.1.1(1)	2
					Contig365gene3355	643	2	1.B.6.1.2(1)	3
				3	Contig373gene6115	217	1	1.B.6.1.3(1)	3
1.B.9	FadL outer membrane protein	FadL	Fatty acid, toluene, *m*-xylene and benzyl alcohol	1	Contig343gene479	464	1	1.B.9.2.1(1)	2

Copyright © 2005 John Wiley & Sons, Ltd.

I.B.11	Outer membrane fimbrial usher porin	FUP	Protein folding and subunit assembly		Contig358gene1879	761	0	I.B.11.3.1(I)	2
					Contig365gene3393	854	I	I.B.11.3.1(I)	2
				3	Contig354gene1279	850	I	I.B.11.3.1(I)	2
I.B.12	Autotransporter	AT	N-terminal protein domains	I	Contig365gene3360	1741	I	I.B.12.1.3(I)	2
I.B.14	Outer membrane receptor	OMR	Iron-siderophore complexes, vitamin B_{12}, Cu^{2+}, colicin, DNA of various phages		Contig374gene7240	733	I	I.B.14.1.2(I)	3
					Contig372gene5928	731	I	I.B.14.1.2(I)	2
					Contig372gene5930	717	0	I.B.14.1.2(I)	2
					Contig370gene4894	764	I	I.B.14.1.4(I)	2
					Contig361gene2288	742	I	I.B.14.1.4(I)	2
					Contig374gene7149	744	I	I.B.14.1.4(I)	2
					Contig374gene7151	804	0	I.B.14.1.4(I)	3
					Contig363gene2909	753	I	I.B.14.1.4(I)	2
					Contig369gene4344	728	I	I.B.14.1.4(I)	2
					Contig369gene4384	761	2	I.B.14.1.4(I)	3
					Contig375gene8531	741	0	I.B.14.1.4(I)	2
					Contig374gene6949	719	0	I.B.14.1.6(I)	3
					Contig373gene6356	661	0	I.B.14.1.6(I)	2
					Contig375gene8072	821	I	I.B.14.1.8(I)	3
					Contig366gene3585	698	I	I.B.14.3.1(I)	2
					Contig375gene8595	724	0	I.B.14.4.1(I)	2
				17	Contig369gene4334	815	I	I.B.14.9.1(I)	2
I.B.17	Outer membrane factor	OMF	Heavy metal cations, drugs, oligosaccharides, proteins, etc.		Contig366gene3474	493	0	I.B.17.1.1(I)	2
					Contig369gene4234	448	I	I.B.17.2.1(I)	I
					Contig368gene3997	418	0	I.B.17.2.1(I)	I
					Contig357gene1641	460	0	I.B.17.2.2(I)	I
					Contig371gene5461	445	0	I.B.17.2.2(I)	3
					Contig374gene7266	431	I	I.B.17.2.2(I)	I
					Contig375gene8615	418	0	I.B.17.2.2(I)	I
					Contig374gene7202	520	2	I.B.17.2.3(I)	3
					Contig373gene6079	433	0	I.B.17.2.3(I)	3
					Contig375gene9177	496	0	I.B.17.3.1(I)	2
					Contig375gene8190	485	I	I.B.17.3.1(I)	2
					Contig353gene1195	455	0	I.B.17.3.2(I)	3
					Contig360gene2100	519	0	I.B.17.3.2(I)	2
					Contig354gene1322	418	I	I.B.17.3.2(I)	2
					Contig364gene3065	504	I	I.B.17.3.2(I)	2
					Contig358gene1809	486	2	I.B.17.3.3(I)	2
					Contig375gene7574	497	I	I.B.17.3.3(I)	2
					Contig359gene2067	488	I	I.B.17.3.3(I)	2

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
					Contig353gene1181	488	1	1.B.17.3.3(1)	2
					Contig373gene6314	518	1	1.B.17.3.3(1)	2
					Contig373gene6558	511	3	1.B.17.3.3(1)	2
					Contig375gene8564	495	0	1.B.17.3.3(1)	2
					Contig358gene1815	519	0	1.B.17.3.4(1)	2
					Contig373gene6386	589	0	1.B.17.3.4(1)	3
					Contig362gene2648	512	0	1.B.17.3.4(1)	2
					Contig353gene1190	497	1	1.B.17.3.5(1)	2
					Contig375gene8587	476	0	1.B.17.3.5(1)	2
				28	Contig375gene7766	491	2	1.B.17.3.5(1)	2
1.B.18	Outer membrane auxiliary protein	OMA	Exo- or capsular polysaccharide		Contig375gene8672	606	0	1.B.18.1.2(1)	2
				2	Contig372gene5594	362	0	1.B.18.3.1(1)	2
1.B.19	Glucose-selective OprB porin	OprB	Ions, small molecules	1	Contig359gene1948	492	1	1.B.19.1.1(1)	3
1.B.20	Two-partner secretion	TPS	Proteins		Contig373gene6550	588	0	1.B.20.1.1(1)	2
				2	Contig371gene5256	558	1	1.B.20.3.1(1)	3
1.B.22	Outer bacterial membrane secretin	Secretin	Proteins		Contig371gene5305	473	0	1.B.22.1.1(1)	2
					Contig365gene3336	620	1	1.B.22.1.2(1)	3
					Contig375gene7610	783	1	1.B.22.1.2(1)	2
					Contig367gene3787	710	1	1.B.22.2.1(1)	2
					Contig375gene9238	734	1	1.B.22.4.1(1)	2
				6	Contig368gene4122	600	0	1.B.22.7.1(1)	3
1.B.39	Bacterial porin, OmpW	OmpW	Methyl viologen and benzyl viologen		Contig375gene9331	286	1	1.B.39.1.1(1)	3
				2	Contig372gene5565	245	0	1.B.39.1.1(1)	3
1.C. Pore-forming toxins (proteins and peptides)									
1.C.1	Channel-forming colicin	Colicin	Ions, small molecules	1	Contig353gene1187	443	1	1.C.1.3.1(1)	3
1.E. Holins									
1.E.14	LrgA holin	LrgA Holin	Zn^{2+}, Fe^{2+}	1	Contig372gene5735	128	3	1.E.14.1.1(1)	3
2.A. Carrier type facilitators									
2.A.1	Major facilitator superfamily	MFS	Various small molecules						
		- SP (1)	Sugars	1	Contig373gene6468	484	12	2.A.1.1.15(1)	3
		- DHA1 (12 spanner) (2)	Drugs		Contig374gene7546	418	12	2.A.1.2.4(1)	3
					Contig358gene1790	411	12	2.A.1.2.4(1)	3
					Contig370gene4856	418	12	2.A.1.2.7(1)	2
					Contig375gene9203	426	11	2.A.1.2.7(1)	2
					Contig369gene4402	634	12	2.A.1.2.9(1)	3
					Contig375gene8690	408	12	2.A.1.2.8(1)	3
					Contig371gene5419	415	12	2.A.1.2.8(1)	3

Copyright © 2005 John Wiley & Sons, Ltd.

			Contig373gene6337	404	12	2.A.1.2.10(1)	3
			Contig373gene6438	275	6	2.A.1.2.10(1)	3
			Contig375gene8780	408	12	2.A.1.2.10(1)	3
			Contig369gene4487	426	11	2.A.1.2.10(1)	3
			Contig373gene6352	388	12	2.A.1.2.14(1)	2
			Contig375gene9023	421	12	2.A.1.2.14(1)	2
			Contig375gene7618	400	12	2.A.1.2.14(1)	3
			Contig375gene7724	398	12	2.A.1.2.18(1)	3
		16	Contig369gene4338	395	12	2.A.1.2.20(1)	3
- DHA2 (14 spanner) (3)	Drugs		Contig360gene2102	537	14	2.A.1.3.2(1)	2
			Contig358gene1921	532	14	2.A.1.3.2(1)	2
			Contig364gene3067	519	14	2.A.1.3.2(1)	2
			Contig366gene3427	396	12	2.A.1.3.5(1)	3
			Contig375gene7796	528	14	2.A.1.3.5(1)	2
			Contig339gene328	480	14	2.A.1.3.11(1)	2
			Contig335gene226	548	14	2.A.1.3.12(1)	2
			Contig373gene6800	504	14	2.A.1.3.17(1)	3
			Contig374gene7112	490	14	2.A.1.3.17(1)	2
			Contig375gene8085	473	14	2.A.1.3.17(1)	2
			Contig373gene6605	513	14	2.A.1.3.17(1)	2
			Contig351gene1053	508	14	2.A.1.3.18(1)	2
			Contig375gene8189	524	13	2.A.1.3.18(1)	2
			Contig375gene8551	545	14	2.A.1.3.18(1)	2
		15	Contig371gene5277	525	14	2.A.1.3.18(1)	3
- MHS (6)	Dicarboxylates, tricarboxylates		Contig373gene6395	434	12	2.A.1.6.1(1)	2
			Contig361gene2390	429	12	2.A.1.6.1(1)	2
			Contig374gene7196	476	12	2.A.1.6.1(1)	2
			Contig375gene8112	436	12	2.A.1.6.1(1)	2
			Contig373gene6395	434	0	2.A.1.6.1(1)	3
			Contig354gene1283	262	7	2.A.1.6.1(1)	2
			Contig372gene5669	443	12	2.A.1.6.1(1)	2
			Contig375gene7621	461	12	2.A.1.6.1(1)	2
			Contig372gene5548	437	12	2.A.1.6.3(1)	2
			Contig352gene1110	565	0	2.A.1.6.4(1)	3
			Contig358gene1834	255	6	2.A.1.6.5(1)	2
			Contig358gene1835	300	6	2.A.1.6.5(1)	2
			Contig352gene1110	565	12	2.A.1.6.5(1)	2
			Contig373gene6129	459	12	2.A.1.6.6(1)	2
		15	Contig375gene8177	439	12	2.A.1.6.6(1)	2
- NNP (8)	Nitrate, nitrite		Contig375gene7921	437	12	2.A.1.8.4(1)	2
			Contig360gene2126	427	12	2.A.1.8.11(1)	2
		3	Contig360gene2127	460	12	2.A.1.8.11(1)	2
- OFA (11)	Oxalate, formate		Contig364gene3127	441	12	2.A.1.11.1(1)	2

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
					Contig355gene1391	1472	12	2.A.1.11.1(1)	3
				3	Contig357gene1707	475	12	2.A.1.11.1(1)	3
		- SHS (12)	Sialate, lactate, pyruvate	1	Contig359gene2081	397	12	2.A.1.12.1(1)	3
		- ACS (14)	Organic acids		Contig353gene1192	443	12	2.A.1.14.1(1)	3
					Contig375gene9216	453	12	2.A.1.14.1(1)	2
					Contig371gene5115	433	12	2.A.1.14.1(1)	2
					Contig365gene3217	444	12	2.A.1.14.1(1)	2
					Contig365gene3305	418	12	2.A.1.14.2(1)	3
					Contig346gene689	437	12	2.A.1.14.3(1)	2
					Contig361gene2423	441	12	2.A.1.14.3(1)	2
					Contig359gene1940	453	12	2.A.1.14.8(1)	2
				9	Contig375gene8175	432	12	2.A.1.14.8(1)	3
		- AAHS(15)	Aromatic acids		Contig364gene3071	413	12	2.A.1.15.1(1)	3
					Contig375gene8062	459	12	2.A.1.15.1(1)	2
					Contig371gene5360	441	12	2.A.1.15.1(1)	2
					Contig373gene6741	395	12	2.A.1.15.3(1)	3
				5	Contig375gene9515	441	12	2.A.1.15.4(1)	2
		- CP (17)	Cyanate	1	Contig370gene4875	423	12	2.A.1.17.1(1)	3
		- OCT (19)	Organic cations	1	Contig373gene6048	526	12	2.A.1.19.4(1)	2
		- SET (20)	Sugars	1	Contig373gene6742	434	0	2.A.1.20.2(1)	3
		- DHA3 (12 spanner) (21)	Drugs	1	Contig351gene976	493	12	2.A.1.21.3(1)	3
		- VNT (22)	Neurotransmitter	1	Contig375gene7913	514	12	2.A.1.22.1(1)	3
		- BST (23)	Unknown	1	Contig375gene8194	436	12	2.A.1.23.1(1)	3
		- PAT (25)	Peptides, AcCoA	1	Contig374gene6932	466	12	2.A.1.25.2(1)	2
		- UMC-terminal fragment (26)	Unknown	1	Contig368gene4202	413	12	2.A.1.26.1(1)	3
		- PPP (27)	Phenylpropionate	1	Contig369gene4405	365	11	2.A.1.27.1(1)	2
		- ADT (30)	Abietane diterpenoid	1	Contig372gene5630	468	12	2.A.1.30.1(1)	3
		- Nre (31)	Ni^{2+}	1	Contig369gene4238	408	12	2.A.1.31.1(1)	2
		- Fsr (35)	Fosmidomycin	1	Contig375gene7801	409	12	2.A.1.35.1(1)	2
		- AtoE (37)	Short chain fatty		Contig364gene2968	467	14	2.A.1.37.1(1)	2
	total 83			2	Contig375gene9151	484	14	2.A.1.37.1(1)	2
2.A.3	Amino acid-polyamine-organocation	APC	Amino acids, polyamines, choline						
		- AAA (1)	Amino acids		Contig361gene2312	493	12	2.A.3.1.2(1)	2
					Contig375gene8013	510	12	2.A.3.1.2(1)	2
					Contig375gene8010	462	12	2.A.3.1.3(1)	2

Copyright © 2005 John Wiley & Sons, Ltd.

					Contig373gene6637	475	12	2.A.3.1.9(1)	2
				5	Contig373gene6757	474	12	2.A.3.1.9(1)	2
	total 6	- CAT (3)	Cationic amino acids	1	Contig374gene7465	469	14	2.A.3.3.1(1)	2
2.A.4	Cation diffusion facilitator	CDF	Cd^{2+}, Co^{2+}, Zn^{2+}		Contig375gene8618	316	5	2.A.4.1.1(1)	1
					Contig375gene9479	337	6	2.A.4.1.1(1)	3
				3	Contig374gene6900	436	6	2.A.4.1.2(1)	2
2.A.5	Zinc (Zn^{2+})—iron (Fe^{2+}) permease	ZIP		1	Contig356gene1473	291	6	2.A.5.4.1(1)	3
2.A.6	Resistance-nodulation-cell division	RND							
		- HME (1)	Heavy metal ions, multiple drugs, oligosaccharides, organic solvents, fatty acids, phospholipids, cholesterol		Contig369gene4237	1076	11	2.A.6.1.1(1)	1
					Contig368gene3999	1076	12	2.A.6.1.1(1)	1
					Contig331gene151	1043	12	2.A.6.1.2(1)	1
					Contig371gene5463	840	0	2.A.6.1.X	3
					Contig371gene5464	184	5	2.A.6.1.X	1
					Contig373gene6557	1039	12	2.A.6.1.2(1)	1
					Contig373gene6563	1045	12	2.A.6.1.2(1)	1
					Contig375gene8617	1063	12	2.A.6.1.2(1)	1
					Contig361gene2416	1036	12	2.A.6.1.2(1)	1
					Contig375gene8282	1023	11	2.A.6.1.2(1)	1
					Contig363gene2863	1009	7	2.A.6.1.2(1)	1
					Contig375gene8486	691	6	2.A.6.1.2(1)	1
					Contig375gene8119	365	5	2.A.6.1.2(1)	1
					Contig373gene6081	1055	14	2.A.6.1.3(1)	2
					Contig357gene1642	384	0	2.A.6.1.4(4)	3
					Contig369gene4482	521	1	2.A.6.1.4(4)	2
				17	Contig369gene4483	1056	14	2.A.6.1.4(1)	2
		- HAE1 (2)	Hydrophobe/amphiphile substrates		Contig375gene7765	1050	12	2.A.6.2.2(1)	2
					Contig369gene4331	1044	12	2.A.6.2.7(1)	2
					Contig358gene1808	1063	12	2.A.6.2.9(1)	2
					Contig375gene7572	1051	12	2.A.6.2.12(1)	2
					Contig375gene7573	1100	12	2.A.6.2.12(1)	2
					Contig365gene3373	1066	0	2.A.6.2.12(1)	2
					Contig365gene3373	1066	12	2.A.6.2.12(1)	2
					Contig353gene1179	1065	12	2.A.6.2.12(1)	2
				9	Contig375gene7759	1069	12	2.A.6.2.12(1)	2
		- SecDF (4)	Sec secretory accessory proteins		Contig372gene5750	636	5	2.A.6.4.1(2)	2
				2	Contig372gene5751	324	6	2.A.6.4.1(2)	2
		- HAE2 (5)	Hydrophobe/amphiphile substrates	1	Contig364gene3205	858	9	2.A.6.5.1(1)	3

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
	total 30	- ORF4 (8)	Hydrophobe/amphiphile substrates	1	Contig364gene3146	786	11	2.A.6.8.1(1)	2
2.A.7	Drug/metabolite transporter	DMT	Multiple drugs and dyes (mostly cationic)						
		- SMR (1)	Drugs		Contig338gene297	109	4	2.A.7.1.3(1)	3
				2	Contig374gene7258	123	5	2.A.7.1.3(1)	3
		- BAT (2)	Unknown		Contig356gene1583	362	11	2.A.7.2.1(1)	3
				2	Contig375gene9035	143	5	2.A.7.2.1(1)	2
		- DME (3)	Drugs, metabolites		Contig375gene9463	297	10	2.A.7.3.2(1)	2
					Contig364gene3008	306	10	2.A.7.3.2(1)	3
					Contig374gene7235	312	10	2.A.7.3.2(1)	3
					Contig352gene1130	301	10	2.A.7.3.2(1)	3
					Contig375gene7949	297	10	2.A.7.3.2(1)	3
					Contig363gene2886	347	11	2.A.7.3.2(1)	3
					Contig375gene8660	337	11	2.A.7.3.3(1)	3
					Contig370gene4864	532	10	2.A.7.3.3(1)	3
					Contig361gene2346	300	10	2.A.7.3.4(1)	3
					Contig375gene8332	345	10	2.A.7.3.4(1)	3
					Contig348gene822	319	10	2.A.7.3.6(1)	3
				12	Contig375gene7919	288	10	2.A.7.3.6(1)	2
		- RarD (7)	Chloramphenicol		Contig375gene7722	342	10	2.A.7.7.1(1)	2
	total 18			2	Contig375gene8931	311	10	2.A.7.17.1(1)	2
2.A.9	Cytochrome oxidase bio-genesis	Oxa1	Proteins	1	Contig375gene9312	555	4	2.A.9.3.1(1)	2
2.A.10	2-Keto-3-deoxygluconate transporter	KDGT	2-Keto-3-deoxygluconate	1	Contig356gene1558	327	10	2.A.10.1.1(1)	2
2.A.11	Citrate-Mg^{2+} : H^+ (CitM) Citrate-Ca^{2+} : H^+ (CitH) Symporter	CitMHS	Citrate	1	Contig375gene7824	485	11	2.A.11.1.1(1)	3
2.A.12	ATP : ADP antiporter	AAA	ATP, ADP	1	Contig375gene7893	453	10	2.A.12.3.1(1)	3
2.A.14	Lactate permease	LctP	Lactate	1	Contig372gene5556	566	16	2.A.14.1.2(1)	2
2.A.19	Ca^{2+}:cation antiporter	CaCA	Ca^{2+}	1	Contig375gene7970	360	11	2.A.19.1.1(1)	2
2.A.20	Inorganic phosphate transporter	PiT	Inorganic phosphate	1	Contig344gene557	336	9	2.A.20.2.4(1)	2
2.A.21	Solute : sodium symporter	SSS	Sugars, amino acids, vitamins, nucleosides, inositols, iodide, urea		Contig375gene7737	461	13	2.A.21.4.1(1)	2
					Contig375gene8758	683	14	2.A.21.7.1(1)	2
					Contig372gene5917	553	14	2.A.21.7.1(1)	2
					Contig359gene1964	478	13	2.A.21.8.1(1)	2
				5	Contig374gene7322	967	4	2.A.21.9.1(1)	3
2.A.23	Dicarboxylate/amino acid : cation (Na^+ or H^+) symporter	DAACS	C4-dicarboxylates, acidic and neutral amino acids	5	Contig365gene3380	435	9	2.A.23.1.2(1)	2
					Contig369gene4393	430	8	2.A.23.1.2(1)	2

Copyright © 2005 John Wiley & Sons, Ltd.

					Contig375gene7654	467	10	2.A.23.1.3(1)	2
					Contig369gene4353	452	8	2.A.23.1.3(1)	2
					Contig374gene7480				
2.A.24	Citrate : cation symporter	CCS	Mono-, di-, and tricarboxylates	1	Contig373gene6794	448	13	2.A.24.2.5(1)	2
2.A.36	Monovalent cation : proton antiporter-1	CPA1	Na^+/H^+, Na^+ or K^+/H^+	1	Contig373gene6805	435	12	2.A.36.6.1(1)	3
2.A.37	Monovalent cation : proton antiporter-2	CPA2	Na^+/H^+ or K^+/H^+		Contig358gene1749	404	12	2.A.37.1.1(2)	2
					Contig375gene9498	219	0	2.A.37.1.1(2)	2
					Contig375gene9499	604	13	2.A.37.1.1(2)	2
					Contig374gene7542	674	13	2.A.37.1.2(2)	2
					Contig375gene8748	408	13	2.A.37.1.2(2)	2
				6	Contig375gene9414	406	12	2.A.37.1.2(2)	3
2.A.40	Nucleobase : cation symporter-2	NCS2	Nucleobases, urate		Contig370gene4820	482	13	2.A.40.1.1(1)	2
					Contig372gene5621	453	13	2.A.40.1.1(1)	3
				3	Contig371gene5375	445	14	2.A.40.3.1(1)	2
2.A.45	Arsenite–antimonite	ArsB	Arsenite, antimonite	1	Contig375gene8078	419	12	2.A.45.1.1(1)	3
2.A.46	Benzoate : H^+ symporter	BenE	Benzoate	1	Contig338gene295	395	12	2.A.46.1.1(1)	2
2.A.47	Divalent anion : Na^+ symporter	DASS	Dicarboxylates, phosphate, sulphate		Contig369gene4367	181	5	2.A.47.3.1(1)	2
					Contig369gene4366	532	7	2.A.47.3.1(1)	2
					Contig365gene3403	507	15	2.A.47.3.3(1)	2
				4	Contig373gene6264	530	15	2.A.47.3.3(1)	3
2.A.49	Ammonium transporter	Amt	Ammonium		Contig375gene7868	400	0	2.A.49.X	2
				2	Contig374gene7490	510	13	2.A.49.1.1(1)	2
2.A.51	Chromate ion transporter	CHR	Chromate, sulphate (uptake or efflux)		Contig355gene1410	193	5	2.A.51.1.1(1)	3
					Contig371gene5134	390	12	2.A.51.1.1(1)	1
					Contig368gene4196	401	12	2.A.51.1.(0)	1
				4	Contig375gene7933	408	11	2.A.51.1.2(1)	3
2.A.52	Ni^{2+}-Co^{2+} transporter	NiCoT	Ni^{2+}, Co^{2+}	1	Contig334gene209	278	7	2.A.52.1.2(1)	3
2.A.53	Sulphate permease	SulP	Sulphate		Contig325gene78	603	11	2.A.53.1.4(1)	2
					Contig375gene8514	492	13	2.A.53.3.1(1)	2
					Contig365gene3367	599	11	2.A.53.4.1(1)	2
					Contig375gene8575	578	11	2.A.53.4.1(1)	2
				5	Contig371gene5371	586	13	2.A.53.4.1(1)	2
2.A.56	Tripartite ATP-independent periplasmic transporter	TRAP-T	C4-dicarboxylates, acidic amino acids, sugars?		Contig369gene4416	434	13	2.A.56.1.1(3)	3
					Contig361gene2330	327	1	2.A.56.1.1(3)	3
					Contig361gene2332	436	11	2.A.56.1.1(0)	2
					Contig369gene4417	343	1	2.A.56.1.1(3)	2
					Contig366gene3497	180	4	2.A.56.1.2(0)	2
				6	Contig366gene3498	574	13	2.A.56.1.2(0)	2

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
2.A.58	Phosphate : Na^+ symporter	PNaS	Inorganic phosphate		Contig353gene1224	632	9	2.A.58.2.1(1)	2
				2	Contig374gene7545	558	9	2.A.58.2.1(1)	2
2.A.59	Arsenical resistance-3	ACR3	Arsenite	1	Contig359gene2078	354	10	2.A.59.1.1(1)	3
2.A.64	Twin arginine targeting	Tat	Redox proteins		Contig367gene3817	77	1	2.A.64.1.1(4)	3
					Contig367gene3818	168	1	2.A.64.1.1(4)	3
					Contig367gene3819	260	5	2.A.64.1.1(4)	2
				4	Contig367gene3820	401	1	2.A.64.1.1(4)	2
2.A.66	Multidrug/oligosaccharidyl-lipid/polysaccharide	MOP	Drugs, lipid-linked oligosaccharide precursors						
		- MATE (1)	Drugs		Contig362gene2523	449	12	2.A.66.1.1(1)	2
					Contig367gene3916	455	12	2.A.66.1.1(1)	3
				3	Contig375gene8978	492	12	2.A.66.1.3(1)	3
		- PST (2)	Polysaccharides	1	Contig366gene3637	419	12	2.A.66.2.4(1)	2
		- MVF (4)	Unknown	1	Contig372gene5715	534	14	2.A.66.4.1(1)	2
2.A.67	Oligopeptide transporter	OPT	Peptides		Contig372gene5640	668	17	2.A.67.3.1(1)	3
				2	Contig363gene2777	676	18	2.A.67.4.1(1)	2
2.A.69	Auxin efflux carrier	AEC	Auxin (efflux)		Contig356gene1506	293	10	2.A.69.1.1(1)	3
				2	Contig375gene9534	351	10	2.A.69.2.1(1)	3
2.A.72	K^+ uptake permease	KUP	K^+ (uptake)	1	Contig349gene850	656	11	2.A.72.1.1(1)	2
2.A.75	L-Lysine exporter	LysE	Basic amino acids	1	Contig371gene5136	216	6	2.A.75.1.1(1)	2
2.A.76	Resistance to homoserine/threonine	RhtB	Neutral amino acids and their derivatives		Contig355gene1462	205	5	2.A.76.1.1(1)	3
					Contig362gene2710	223	6	2.A.76.1.1(1)	3
					Contig355gene1382	209	6	2.A.76.1.1(1)	3
					Contig363gene2791	212	6	2.A.76.1.1(1)	3
					Contig371gene5511	208	6	2.A.76.1.1(1)	3
					Contig374gene7486	214	6	2.A.76.1.1(1)	2
					Contig375gene7623	212	6	2.A.76.1.1(1)	3
					Contig372gene5890	212	6	2.A.76.1.1(1)	3
					Contig373gene6271	205	5	2.A.76.1.2(1)	3
					Contig373gene6137	203	6	2.A.76.1.2(1)	3
				11	Contig372gene5541	204	6	2.A.76.1.2(1)	3
2.A.78	Branched chain amino acid exporter	LIV-E	Carboxylates, amino acids, amines (efflux)	1	Contig352gene1134	265	4	2.A.78.1.1(2)	3
2.A.80	Tricarboxylate transporter	TTT	Tricarboxylate		Contig373gene6710	326	1	2.A.80.1.1(3)	3
					Contig364gene3003	322	1	2.A.80.1.1(3)	3
					Contig364gene2995	320	1	2.A.80.1.1(3)	3
					Contig358gene1831	337	0	2.A.80.1.1(3)	3
					Contig370gene4730	554	1	2.A.80.1.1(3)	3

Copyright © 2005 John Wiley & Sons, Ltd.

Contig345gene580	327	4	2.A.80.1.1(3)	3
Contig360gene2191	326	1	2.A.80.1.1(3)	3
Contig373gene6749	322	3	2.A.80.1.1(3)	3
Contig373gene6096	328	3	2.A.80.1.1(3)	3
Contig373gene6578	330	0	2.A.80.1.1(3)	3
Contig371gene5514	328	4	2.A.80.1.1(3)	3
Contig371gene5517	325	2	2.A.80.1.1(3)	3
Contig373gene6354	322	0	2.A.80.1.1(3)	3
Contig370gene4900	323	4	2.A.80.1.1(3)	3
Contig357gene1683	334	4	2.A.80.1.1(3)	3
Contig375gene9115	328	2	2.A.80.1.1(3)	3
Contig358gene1825	327	1	2.A.80.1.1(3)	3
Contig373gene6531	321	0	2.A.80.1.1(3)	3
Contig357gene1608	332	1	2.A.80.1.1(3)	3
Contig372gene5872	334	0	2.A.80.1.1(3)	3
Contig335gene234	311	0	2.A.80.1.1(3)	2
Contig361gene2500	327	1	2.A.80.1.1(3)	3
Contig341gene384	336	1	2.A.80.1.1(3)	3
Contig353gene1176	332	4	2.A.80.1.1(3)	3
Contig374gene7144	366	0	2.A.80.1.1(3)	3
Contig374gene7146	327	0	2.A.80.1.1(3)	3
Contig373gene6763	348	0	2.A.80.1.1(3)	3
Contig345gene622	323	1	2.A.80.1.1(3)	3
Contig345gene628	363	1	2.A.80.1.1(3)	3
Contig357gene1594	329	2	2.A.80.1.1(3)	3
Contig366gene3580	336	3	2.A.80.1.1(3)	3
Contig354gene1306	353	2	2.A.80.1.1(3)	3
Contig371gene5098	332	2	2.A.80.1.1(3)	3
Contig371gene5502	341	0	2.A.80.1.1(3)	3
Contig370gene4704	500	0	2.A.80.1.1(3)	2
Contig370gene4705	325	0	2.A.80.1.1(3)	3
Contig375gene8996	328	1	2.A.80.1.1(3)	3
Contig358gene1912	333	0	2.A.80.1.1(3)	3
Contig375gene7952	334	0	2.A.80.1.1(3)	3
Contig370gene4597	333	0	2.A.80.1.1(3)	3
Contig375gene8980	500	12	2.A.80.1.1(3)	2
Contig375gene8884	330	2	2.A.80.1.1(3)	3
Contig356gene1500	337	2	2.A.80.1.1(3)	3
Contig373gene6518	331	0	2.A.80.1.1(3)	3
Contig354gene1258	345	2	2.A.80.1.1(3)	3
Contig375gene8567	348	0	2.A.80.1.1(3)	3
Contig375gene8579	353	0	2.A.80.1.1(3)	3
Contig371gene5324	346	4	2.A.80.1.1(3)	3

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
					Contig367gene3835	323	1	2.A.80.1.1(3)	3
					Contig364gene3052	320	0	2.A.80.1.1(3)	3
					Contig364gene3101	382	2	2.A.80.1.1(3)	3
					Contig373gene6242	333	0	2.A.80.1.1(3)	3
					Contig373gene6267	322	1	2.A.80.1.1(3)	3
					Contig373gene6280	330	1	2.A.80.1.1(3)	3
					Contig373gene6586	331	4	2.A.80.1.1(3)	3
					Contig375gene7775	323	1	2.A.80.1.1(3)	3
					Contig371gene5154	322	1	2.A.80.1.1(3)	3
					Contig371gene5178	326	2	2.A.80.1.1(3)	3
					Contig371gene5421	318	3	2.A.80.1.1(3)	3
					Contig371gene5426	326	1	2.A.80.1.1(3)	3
					Contig371gene5429	320	1	2.A.80.1.1(3)	3
					Contig371gene5443	325	2	2.A.80.1.1(3)	3
					Contig373gene6160	341	1	2.A.80.1.1(3)	3
					Contig373gene6671	329	1	2.A.80.1.1(3)	3
					Contig373gene6675	320	0	2.A.80.1.1(3)	3
					Contig372gene6001	328	1	2.A.80.1.1(3)	3
					Contig375gene9477	330	3	2.A.80.1.1(3)	3
					Contig375gene9392	385	1	2.A.80.1.1(3)	3
					Contig375gene7729	322	1	2.A.80.1.1(3)	2
					Contig375gene7731	504	13	2.A.80.1.1(3)	2
					Contig375gene8159	333	1	2.A.80.1.1(3)	2
					Contig375gene8161	551	12	2.A.80.1.1(3)	2
					Contig375gene8171	329	0	2.A.80.1.1(3)	3
				74	Contig375gene8944	513	11	2.A.80.1.1(3)	2
2.A.81	Aspartate : alanine exchanger	AAE	Aspartate, alanine		Contig344gene526	561	11	2.A.81.1.1(1)	2
				2	Contig375gene7947	567	11	2.A.81.1.1(1)	2
2.C. Ion-gradient-driven energizers									
2.C.1	TonB-ExbB-ExbD/TolA-TolQ-TolR family of auxiliary proteins for energization of outer membrane receptor (OMR)-mediated active transport	TonB	H^{+}?, drives solute uptake across outer bacterial membranes		Contig375gene8338	243	3	2.C.1.1.1(3)	3
					Contig366gene3670	227	3	2.C.1.2.1(6)	2
					Contig366gene3671	145	1	2.C.1.2.1(6)	3
				5	Contig366gene3673	446	1	2.C.1.2.1(6)	3

Copyright © 2005 John Wiley & Sons, Ltd.

3.A P-P bond hydrolysis-driven transporters									
3.A.1	ATP-binding cassette	ABC	All sorts of inorganic and organic molecules of small, intermediate, and large sizes, from simple ions to macromolecules						
					Contig375gene8339	137	1	2.C.1.1.1(3)	3
		- CUT1 (1)	Sugars, metabolites						
					Contig360gene2245	293	6	3.A.1.1.3(4)	2
					Contig360gene2246	282	6	3.A.1.1.3(4)	2
					Contig360gene2247	367	1	3.A.1.1.3(4)	2
					Contig362gene2677	395	1	3.A.1.1.X	2
					Contig362gene2678	366	0	3.A.1.1.X	2
					Contig362gene2679	294	6	3.A.1.1.X	2
					Contig362gene2680	276	6	3.A.1.1.X	2
					Contig362gene2682	580	1	3.A.1.1.X	3
					Contig375gene7943	352	1	3.A.1.1.12(4)	2
					Contig365gene3399	279	0	3.A.1.1.16(4)	2
					Contig375gene9297	464	0	3.A.1.1.X	3
					Contig375gene9298	371	1	3.A.1.1.X	2
					Contig375gene9299	310	6	3.A.1.1.X	2
				15	Contig375gene9300	295	6	3.A.1.1.X	2
		- CUT2 (2)	Sugars, metabolites		Contig349gene903	298	9	3.A.1.2.1(4)	3
					Contig370gene4724	537	0	3.A.1.2.X	2
					Contig370gene4725	364	10	3.A.1.2.X	3
				4	Contig370gene4726	306	9	3.A.1.X	3
		- PAAT (3)	Polar amino acids		Contig346gene661	302	1	3.A.1.3.4(4)	2
					Contig369gene4394	303	0	3.A.1.3.4(4)	3
					Contig359gene1941	302	1	3.A.1.3.4(4)	2
					Contig346gene653	282	1	3.A.1.3.4(4)	2
					Contig346gene654	231	5	3.A.1.3.4(4)	2
					Contig346gene655	447	6	3.A.1.3.4(4)	2
					Contig346gene656	249	0	3.A.1.3.4(4)	2
					Contig374gene7255	304	0	3.A.1.3.4(4)	2
					Contig338gene294	310	1	3.A.1.3.4(4)	2
					Contig359gene1987	299	0	3.A.1.3.4(4)	2
					Contig359gene1988	242	5	3.A.1.3.4(4)	2
					Contig359gene1989	227	5	3.A.1.3.4(4)	2
					Contig359gene1990	244	0	3.A.1.3.4(4)	2
				14	Contig371gene5478	274	0	3.A.1.3.10(3)	2
		- HAAT (4)	Hydrophobic amino acids		Contig350gene948	384	9	3.A.1.4.1(6)	2
					Contig350gene949	258	0	3.A.1.4.1(6)	2
					Contig350gene950	238	0	3.A.1.4.1(6)	2
					Contig374gene7171	361	0	3.A.1.4.X	3
					Contig374gene7172	285	7	3.A.1.4.X	2

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
					Contig374gene7174	255	0	3.A.1.4.1(6)	2
					Contig361gene2483	437	11	3.A.1.4.1(6)	3
					Contig361gene2484	259	0	3.A.1.4.1(6)	2
					Contig361gene2485	237	0	3.A.1.4.1(6)	2
					Contig355gene1439	479	1	3.A.1.4.1(6)	2
					Contig355gene1440	308	9	3.A.1.4.1(6)	2
					Contig370gene4963	425	1	3.A.1.4.1(6)	3
					Contig355gene1441	424	11	3.A.1.4.1(6)	2
					Contig355gene1442	255	0	3.A.1.4.1(6)	2
					Contig355gene1443	233	0	3.A.1.4.1(6)	2
					Contig340gene370	398	3	3.A.1.4.1(6)	3
					Contig375gene8086	287	7	3.A.1.4.X	2
					Contig375gene8087	342	10	3.A.1.4.X	2
					Contig375gene8088	254	1	3.A.1.4.1(6)	2
					Contig375gene8089	235	2	3.A.1.4.1(6)	2
					Contig375gene8090	390	1	3.A.1.4.X	3
					Contig349gene849	238	0	3.A.1.4.1(6)	2
					Contig372gene5996	379	2	3.A.1.4.1(6)	3
					Contig374gene7472	313	0	3.A.1.4.1(6)	2
					Contig374gene7473	304	8	3.A.1.4.1(6)	2
					Contig374gene7474	358	10	3.A.1.4.X	3
					Contig374gene7476	271	0	3.A.1.4.1(6)	2
					Contig375gene9380	257	0	3.A.1.4.X	2
					Contig375gene9381	241	0	3.A.1.4.1(6)	2
					Contig375gene9382	402	1	3.A.1.4.X	3
					Contig375gene9387	382	1	3.A.1.4.X	3
					Contig375gene9388	350	9	3.A.1.4.1(6)	3
					Contig375gene9389	617	10	3.A.1.4.1(6)	2
					Contig375gene9390	247	0	3.A.1.4.X	2
					Contig361gene2481	416	1	3.A.1.4.X	3
					Contig361gene2482	323	8	3.A.1.4.2(5)	3
					Contig375gene9184	288	7	3.A.1.4.2(5)	3
					Contig366gene3541	389	1	3.A.1.4.2(5)	3
					Contig374gene6823	263	0	3.A.1.4.2(5)	2
					Contig350gene946	401	1	3.A.1.4.3(4)	2
					Contig375gene9383	294	8	3.A.1.4.3(4)	2
					Contig375gene9384	344	9	3.A.1.4.3(4)	3
					Contig375gene9412	383	1	3.A.1.4.3(4)	2
					Contig358gene1906	412	1	3.A.1.4.4(5)	3
				45	Contig370gene5000	230	0	3.A.1.4.4(5)	3
		- PepT (5)	Peptide, opine, nickel		Contig373gene6123	348	0	3.A.1.5.X	2

Copyright © 2005 John Wiley & Sons, Ltd.

			Contig373gene6124	337	0	3.A.1.5.X	2
			Contig373gene6120	527	0	3.A.1.5.X	3
			Contig373gene6122	299	5	3.A.1.5.X	2
			Contig336gene262	306	0	3.A.1.5.2(5)	2
			Contig336gene263	300	6	3.A.1.5.X	2
			Contig336gene265	275	0	3.A.1.5.2(5)	2
			Contig362gene2744	661	1	3.A.1.5.2(5)	3
			Contig362gene2745	349	6	3.A.1.5.2(5)	2
			Contig362gene2746	376	6	3.A.1.5.2(5)	2
			Contig362gene2747	549	0	3.A.1.5.2(5)	2
			Contig370gene4823	344	0	3.A.1.5.2(5)	3
			Contig374gene7247	318	6	3.A.1.5.X	2
			Contig374gene7248	308	6	3.A.1.5.X	2
			Contig374gene7249	575	1	3.A.1.5.X	2
			Contig374gene7250	325	0	3.A.1.5.X	2
			Contig374gene7251	367	0	3.A.1.5.X	2
			Contig355gene1375	545	1	3.A.1.5.4(5)	2
			Contig357gene1662	259	1	3.A.1.5.3(5)	2
			Contig340gene353	535	1	3.A.1.5.X	2
			Contig340gene354	325	6	3.A.1.5.X	2
			Contig340gene355	309	5	3.A.1.5.X	2
			Contig340gene356	332	0	3.A.1.5.X	2
			Contig340gene357	354	0	3.A.1.5.X	2
			Contig358gene1735	289	1	3.A.1.5.X	3
			Contig358gene1736	347	6	3.A.1.5.X	2
			Contig358gene1737	279	6	3.A.1.5.X	2
			Contig358gene1738	547	0	3.A.1.5.X	2
			Contig361gene2374	586	0	3.A.1.5.X	2
			Contig361gene2375	316	6	3.A.1.5.X	2
			Contig361gene2376	295	5	3.A.1.5.X	2
			Contig361gene2377	359	0	3.A.1.5.X	2
		33	Contig361gene2378	337	0	3.A.1.5.X	2
- SulT (6)	Sulphate, tungstate		Contig351gene1029	335	1	3.A.1.6.1(5)	2
			Contig351gene1036	335	6	3.A.1.6.1(5)	2
			Contig351gene1037	305	6	3.A.1.6.1(5)	2
			Contig351gene1038	367	0	3.A.1.6.1(5)	2
			Contig374gene6978	279	0	3.A.1.6.1(5)	2
		6	Contig367gene3795	232	0	3.A.1.6.3(4)	2
- PhoT (7)	Phosphate		Contig362gene2718	343	1	3.A.1.7.1(4)	2
			Contig362gene2719	321	6	3.A.1.7.1(4)	2
			Contig362gene2720	300	6	3.A.1.7.1(4)	2
			Contig362gene2721	262	0	3.A.1.7.1(4)	2
			Contig372gene5607	333	1	3.A.1.7.1(4)	2
		6	Contig375gene8133	355	3	3.A.1.7.1(4)	2

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
		- MolT (8)	Molybdate		Contig372gene5726	232	5	3.A.1.8.1(3)	3
					Contig375gene7940	272	6	3.A.1.8.1(3)	2
					Contig365gene3396	258	1	3.A.1.8.1(3)	2
				4	Contig365gene3398	238	5	3.A.1.8.1(3)	2
		- PhnT (9)	Phosphonate		Contig349gene904	264	0	3.A.1.9.1(3)	2
					Contig370gene4744	326	0	3.A.1.9.1(3)	2
					Contig370gene4745	349	6	3.A.1.9.1(3)	2
					Contig372gene5787	279	0	3.A.1.9.1(3)	2
					Contig372gene5788	292	1	3.A.1.9.1(3)	3
				6	Contig372gene5789	266	5	3.A.1.9.1(3)	2
		- POPT(11)	Polyamine, opine, phosphonate		Contig364gene3011	364	0	3.A.1.11.1(4)	2
					Contig364gene3012	338	6	3.A.1.11.1(4)	2
					Contig364gene3013	260	6	3.A.1.11.1(4)	2
					Contig364gene3014	362	1	3.A.1.11.X	2
					Contig371gene5479	259	6	3.A.1.11.1(4)	3
					Contig375gene7939	340	1	3.A.1.11.X	3
					Contig375gene7942	291	6	3.A.1.11.X	3
				8	Contig372gene5727	229	0	3.A.1.11.2(4)	2
		- QAT (12)	Quaternary amine		Contig370gene4824	217	5	3.A.1.12.6(3)	3
					Contig357gene1649	516	6	3.A.1.12.3(4)	2
					Contig370gene4872	316	1	3.A.1.12.4(4)	2
					Contig370gene4873	216	5	3.A.1.12.4(4)	2
				5	Contig370gene4874	398	0	3.A.1.12.4(4)	2
		- VB12T(13)	Vitamin B_{12}	1	Contig366gene3515	300	0	3.A.1.13.1(3)	2
		- FeCT (14)	Iron chelate		Contig366gene3586	335	9	3.A.1.14.X	2
					Contig366gene3587	269	0	3.A.1.14.X	2
					Contig373gene6359	283	1	3.A.1.14.5(3)	2
					Contig373gene6360	333	9	3.A.1.14.5(3)	2
				5	Contig373gene6361	261	0	3.A.1.14.5(3)	2
		- MZT (15)	Manganese, zinc, iron chelate	1	Contig375gene9102	264	0	3.A.1.15.1(3)	3
		- NitT (16)	Nitrate, nitrite, cyanate		Contig374gene6840	434	3	3.A.1.16.X	2
					Contig374gene6841	303	6	3.A.1.16.1(4)	2
					Contig374gene6842	267	1	3.A.1.16.1(4)	2
					Contig360gene2192	347	2	3.A.1.16.2(3)	3
					Contig360gene2193	270	0	3.A.1.16.2(3)	2
					Contig362gene2587	341	1	3.A.1.16.2(3)	3
					Contig362gene2588	347	8	3.A.1.16.2(3)	2
					Contig362gene2589	262	0	3.A.1.16.X	2
					Contig350gene955	317	3	3.A.1.16.X	3

Copyright © 2005 John Wiley & Sons, Ltd.

				Contig350gene956	291	6	3.A.1.16.2(3)	2
				Contig350gene957	259	0	3.A.1.16.2(3)	2
				Contig364gene3098	256	1	3.A.1.16.2(3)	2
				Contig375gene8294	304	0	3.A.1.16.2(3)	2
			14	Contig367gene3893	581	12	3.A.1.16.3(4)	3
	- TauT (17)	Taurine		Contig360gene2194	263	5	3.A.1.17.1(3)	2
				Contig351gene1030	345	1	3.A.1.17.2(3)	2
				Contig351gene1032	268	6	3.A.1.17.2(3)	2
				Contig351gene1033	291	0	3.A.1.17.2(3)	2
				Contig350gene955	317	3	3.A.1.17.1(3)	3
				Contig359gene1981	352	1	3.A.1.17.1(3)	3
				Contig359gene1982	281	0	3.A.1.17.1(3)	2
				Contig359gene1983	259	6	3.A.1.17.X	2
				Contig367gene3894	448	0	3.A.1.17.1(3)	2
				Contig364gene3097	348	1	3.A.1.17.X	3
				Contig375gene8293	341	0	3.A.1.17.X	3
				Contig364gene3099	298	7	3.A.1.17.X	3
			13	Contig375gene8295	388	7	3.A.1.17.X	3
uptake—total 181	- BIT (20)	Fe^{3+}	1	Contig375gene7762	371	1	3.A.1.20.1(6)	3
	- CPSE (101)	Capsular polysaccharides		Contig375gene8669	262	6	3.A.1.101.1(2)	2
			2	Contig375gene8670	221	0	3.A.1.101.1(2)	2
	- LOSE (102)	Lipo-oligosaccharide		Contig358gene1866	919	6	3.A.1.102.1(2)	2
				Contig355gene1354	303	1	3.A.1.102.1(2)	2
				Contig355gene1355	285	6	3.A.1.102.1(2)	2
				Contig359gene2065	328	1	3.A.1.102.1(2)	2
				Contig367gene3801	316	0	3.A.1.102.1(2)	2
				Contig362gene2639	83	0	3.A.1.102.1(2)	3
			7	Contig362gene2647	372	7	3.A.1.102.1(2)	3
	- DrugE1 (105)	Drugs		Contig359gene2066	384	6	3.A.1.105.2(2)	3
			2	Contig375gene9090	316	1	3.A.1.105.2(2)	2
	- HemeE (107)	Heme		Contig375gene8794	207	0	3.A.1.107.1(3)	2
				Contig375gene8795	228	6	3.A.1.107.1(3)	2
				Contig375gene8796	245	6	3.A.1.107.1(3)	2
				Contig357gene1721	211	0	3.A.1.107.1(3)	2
				Contig357gene1722	222	6	3.A.1.107.1(3)	3
			6	Contig357gene1723	259	6	3.A.1.107.1(3)	2
	- Prot1E (109)	Proteins	1	Contig370gene4796	767	6	3.A.1.109.2(1)	2
	- Prot2E (110)	Proteins	1	Contig372gene5991	232	1	3.A.1.110.2(1)	3
	- Drug RA1 (120)	Drugs		Contig342gene443	540	0	3.A.1.120.1(1)	2
				Contig355gene1406	659	0	3.A.1.120.1(1)	2
				Contig373gene6751	536	0	3.A.1.120.2(1)	2

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
					Contig353gene1211	554	0	3.A.1.120.3(1)	3
				5	Contig367gene3831	670	0	3.A.1.120.4(1)	2
		- Drug RA2 (121)	Drugs	1	Contig336gene249	347	0	3.A.1.121.2(1)	2
		- MacB (122)	Macrolide	1	Contig353gene1239	234	0	3.A.1.122.1(1)	2
		- LPT (125)	Lipoproteins		Contig361gene2469	208	0	3.A.1.125.1(3)	2
					Contig370gene5060	416	5	3.A.1.125.1(3)	2
				3	Contig370gene5061	249	0	3.A.1.125.1(3)	2
	export -total 30	- HMT (210)	Heavy metals		Contig355gene1389	610	6	3.A.1.210.2(1)	2
	total 211			2	Contig359gene1980	630	7	3.A.1.210.3(1)	2
3.A.2	H^+- or Na^+-translocating F-type, V-type and A-type ATPase	F-ATPase	H^+, Na^+		Contig375gene9399	289	6	3.A.2.1.1(8)	2
					Contig375gene9400	88	0	3.A.2.1.1(8)	3
					Contig375gene9401	156	1	3.A.2.1.1(8)	3
					Contig375gene9402	180	0	3.A.2.1.1(8)	2
					Contig375gene9403	513	0	3.A.2.1.1(8)	2
					Contig375gene9404	291	0	3.A.2.1.1(8)	2
					Contig375gene9405	467	1	3.A.2.1.1(8)	2
				8	Contig375gene9406	138	0	3.A.2.1.1(8)	2
3.A.3	P-type ATPase	P-ATPase	Na^+, H^+, K^+, Ca^{2+}, Mg^{2+}, Cd^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, Co^{2+}, Ni^{2+}, Ag^+, phospholipids (flipping)		Contig373gene6510	920	10	3.A.3.2.4(1)	2
					Contig375gene9376	813	8	3.A.3.5.1(1)	2
					Contig369gene4263	805	8	3.A.3.5.5(1)	2
					Contig375gene7707	66	0	3.A.3.5.5(1)	3
					Contig375gene8429	752	0	3.A.3.5.7(1)	3
					Contig373gene6415	829	6	3.A.3.6.1(1)	2
					Contig374gene7074	794	6	3.A.3.6.4(1)	2
					Contig375gene8357	984	8	3.A.3.6.4(1)	2
					Contig373gene6441	799	8	3.A.3.6.4(1)	2
					Contig374gene7319	610	12	3.A.3.7.1(3)	2
					Contig374gene7320	743	7	3.A.3.7.1(3)	2
				12	Contig374gene7321	203	1	3.A.3.7.1(3)	2
3.A.5	General secretory pathway	IISP	Proteins		Contig363gene2920	463	0	3.A.5.1.1(11)	2
					Contig363gene2773	930	1	3.A.5.1.1(11)	2
					Contig374gene6838	948	0	3.A.5.1.1(11)	2
					Contig367gene3758	447	10	3.A.5.1.1(11)	2
				5	Contig372gene5749	108	1	3.A.5.1.1(11)	3
3.A.6	Type III (virulence-related) secretory pathway	IIISP	Proteins		Contig373gene6292	156	0	3.A.6.1.2(10)	3

Copyright © 2005 John Wiley & Sons, Ltd.

					Contig373gene6293	186	2	3.A.6.1.2(10)	3
					Contig373gene6294	264	4	3.A.6.1.2(10)	2
					Contig373gene6295	89	2	3.A.6.1.2(10)	3
					Contig373gene6296	253	6	3.A.6.1.2(10)	2
					Contig373gene6256	563	2	3.A.6.1.2(10)	2
					Contig373gene6258	278	0	3.A.6.1.2(10)	2
					Contig373gene6259	486	0	3.A.6.1.2(10)	2
					Contig371gene5351	380	4	3.A.6.1.2(10)	2
				10	Contig371gene5352	695	8	3.A.6.1.2(10)	2
3.A.7	Type IV (conjugal DNA-protein transfer or VirB) secretory pathway	IVSP	Proteins, protein–DNA complexes		Contig351gene1012	423	0	3.A.7.4.1(10)	2
					Contig342gene418	669	2	3.A.7.X	2
					Contig342gene420	358	0	3.A.7.4.1(10)	2
					Contig351gene1006	819	0	3.A.7.4.1(10)	2
					Contig351gene1007	245	1	3.A.7.4.1(10)	3
					Contig351gene1009	459	7	3.A.7.X	3
					Contig351gene1010	234	1	3.A.7.4.1(10)	2
					Contig351gene1011	330	0	3.A.7.4.1(10)	2
					Contig342gene423	809	0	3.A.7.4.1(10)	2
					Contig342gene424	241	1	3.A.7.4.1(10)	3
					Contig342gene427	234	0	3.A.7.4.1(10)	3
					Contig342gene428	333	0	3.A.7.4.1(10)	2
					Contig342gene429	422	1	3.A.7.4.1(10)	2
					Contig368gene4065	818	2	3.A.7.4.1(10)	2
					Contig368gene4066	252	1	3.A.7.4.1(10)	3
					Contig368gene4116	303	0	3.A.7.4.1(10)	3
					Contig368gene4117	414	2	3.A.7.4.1(10)	2
					Contig371gene5307	438	0	3.A.7.5.1(10)	2
					Contig365gene3342	456	0	3.A.7.5.1(10)	2
				20	Contig368gene4062	349	0	3.A.7.5.1(10)	2
3.A.11	Bacterial competence-related DNA transformation transporter	DNA-T	Single-stranded DNA	1	Contig370gene5063	851	9	3.A.11.1.1(3)	2
3.A.12	Septal DNA translocator	S-DNA-T	DNA, DNA–protein complexes		Contig357gene1647	1123	3	3.A.12.1.2(1)	2
				2	Contig346gene649	775	4	3.A.12.1.2(1)	2
3.A.13	Filamentous phage exporter	FPhE	DNA	1	Contig374gene7161	358	0	3.A.13.1.1(1)	2
3.A.15	Outer membrane protein secreting main terminal branch	MTB	Pilin/fimbrilin		Contig363gene2929	573	1	3.A.15.2.1(10)	3
					Contig363gene2930	421	4	3.A.15.2.1(10)	2
					Contig363gene2931	289	7	3.A.15.2.1(10)	2
					Contig346gene667	154	1	3.A.15.2.1(10)	3
					Contig372gene5842	347	0	3.A.15.2.1(10)	2
					Contig372gene5843	381	0	3.A.15.2.1(10)	2
					Contig375gene9231	202	1	3.A.15.X	2

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
					Contig375gene9232	442	3	3.A.15.X	2
					Contig375gene9233	568	0	3.A.15.X	2
					Contig352gene1091	182	1	3.A.15.2.1(10)	3
					Contig352gene1092	234	1	3.A.15.2.1(10)	3
					Contig374gene7449	147	2	3.A.15.2.1(10)	3
					Contig375gene8624	167	2	3.A.15.2.1(10)	3
					Contig368gene4125	635	0	3.A.15.2.1(10)	3
					Contig375gene7605	284	1	3.A.15.1.1(14)	3
					Contig375gene7606	327	1	3.A.15.1.1(14)	3
					Contig375gene7611	513	0	3.A.15.1.1(14)	2
				18	Contig375gene7612	405	4	3.A.15.1.1(14)	2
3.B. Decarboxylation-driven active transporters									
3.B.1	Na^+-transporting carboxylic acid decarboxylase	NaT-DC	Na^+		Contig358gene1826	539	4	3.B.1.1.2(5)	2
				2	Contig365gene3364	535	3	3.B.1.1.2(5)	2
3.D. Oxidoreduction-driven active transporters									
3.D.1	Proton-translocating NADH dehydrogenase	NDH	H^+ or Na^+ (efflux)		Contig356gene1471	467	3	3.D.1.1.1(14)	2
					Contig370gene4970	119	3	3.D.1.2.1(14)	2
					Contig370gene4971	160	1	3.D.1.2.1(14)	2
					Contig370gene4972	199	0	3.D.1.2.1(14)	2
					Contig370gene4973	417	0	3.D.1.2.1(14)	2
					Contig370gene4974	168	1	3.D.1.2.1(14)	3
					Contig370gene4975	431	1	3.D.1.2.1(14)	2
					Contig370gene4976	828	1	3.D.1.2.1(14)	2
					Contig370gene4977	354	8	3.D.1.2.1(14)	2
					Contig370gene4979	163	0	3.D.1.2.1(14)	2
					Contig370gene4980	225	5	3.D.1.2.1(14)	3
					Contig370gene4981	101	3	3.D.1.2.1(14)	2
					Contig370gene4982	692	17	3.D.1.2.1(14)	2
					Contig370gene4984	491	14	3.D.1.2.1(14)	2
					Contig365gene3228	518	2	3.D.1.X	2
					Contig365gene3229	957	0	3.D.1.X	2
					Contig369gene4352	414	2	3.D.1.1.1(14)	2
					Contig364gene3085	402	1	3.D.1.1.1(14)	3
				19	Contig370gene4983	488	14	3.D.1.3.1(14)	2
3.D.2	Proton-translocating transhydrogenase	PTH	H^+ (efflux)		Contig334gene207	101	3	3.D.2.2.1(3)	3
					Contig334gene208	457	10	3.D.2.2.1(3)	2
					Contig326gene94	257	1	3.D.2.2.1(3)	2
					Contig372gene5764	401	0	3.D.2.2.1(3)	2

Copyright © 2005 John Wiley & Sons, Ltd.

					Contig372gene5765	152	3	3.D.2.2.1(3)	2
				6	Contig372gene5766	490	10	3.D.2.2.1(3)	2
3.D.3	Proton-translocating quinol : cytochrome *c* reductase	QCR	H^+ (efflux)		Contig367gene3822	205	1	3.D.3.1.1(3)	2
					Contig367gene3823	467	13	3.D.3.1.1(3)	2
				3	Contig367gene3824	247	2	3.D.3.X	3
3.D.4	Proton-translocating cytochrome oxidase	COX	H^+ (efflux)		Contig375gene8425	518	13	3.D.4.2.1(1)	3
					Contig360gene2255	482	12	3.D.4.3.1(1)	3
					Contig364gene3091	529	13	3.D.4.3.1(1)	3
					Contig370gene4992	322	3	3.D.4.5.1(5)	2
					Contig370gene4993	657	14	3.D.4.5.1(5)	2
					Contig370gene4994	214	5	3.D.4.5.1(5)	2
					Contig370gene4995	116	3	3.D.4.5.1(5)	3
					Contig374gene7516	308	9	3.D.4.5.1(5)	2
					Contig374gene7508	536	12	3.D.4.6.1(2)	2
					Contig374gene7512	286	7	3.D.4.X	2
					Contig375gene8971	391	3	3.D.4.X	3
					Contig375gene8972	585	12	3.D.4.X	2
					Contig375gene8973	222	5	3.D.4.X	3
					Contig375gene8974	234	5	3.D.4.X	3
					Contig372gene5641	319	3	3.D.4.5.1(5)	2
					Contig372gene5642	658	15	3.D.4.5.1(5)	2
					Contig372gene5643	226	5	3.D.4.5.1(5)	2
					Contig372gene5644	121	3	3.D.4.5.1(5)	3
					Contig375gene8913	349	4	3.D.4.5.1(5)	2
					Contig375gene8914	667	15	3.D.4.5.1(5)	2
					Contig375gene8915	218	5	3.D.4.5.1(5)	2
					Contig375gene8916	142	0	3.D.4.5.1(5)	3
				23	Contig374gene7507	422	3	3.D.4.7.1(3)	3
4. Phosphotransferase systems									
4.A.6	PTS mannose–fructose–sorbose	Man	Glucose, mannose, fructose, sorbose, etc.		Contig362gene2525	316	3	4.A.6.1.1(3)	3
				2	Contig374gene7493	151	0	4.A.6.1.2(4)	3
5.A. Transmembrane electron transfer carriers									
5.A.1	Disulphide bond oxido-reductase D	DsbD	2 e^-		Contig375gene8804	278	4	5.A.1.1.1(1)	3
				2	Contig367gene3767	624	9	5.A.1.1.1(1)	2
5.A.2	Disulphide bond oxido-reductase B	DsbB	2 e^-	1	Contig336gene256	255	4	5.A.2.1.1(1)	3
5.A.3	Prokaryotic molybdopterin-containing oxidoreductase	PMO	Proton translocation		Contig340gene359	701	0	5.A.3.2.1(3)	3
					Contig366gene3609	1025	1	5.A.3.2.1(3)	2

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
					Contig366gene3610	226	0	5.A.3.2.1(3)	3
					Contig366gene3612	418	6	5.A.3.2.1(3)	2
					Contig360gene2128	1252	0	5.A.3.1.1(3)	2
					Contig360gene2129	517	0	5.A.3.1.1(3)	2
				7	Contig360gene2131	227	5	5.A.3.1.1(3)	2
8.A. Auxiliary transport proteins									
8.A.1	Membrane fusion protein	MFP	Proteins, peptides, lipopolysaccharides, drugs, dyes, signalling molecules, heavy metal ions, etc.		Contig358gene1816	378	1	8.A.1.1.1(1)	2
					Contig374gene7200	349	1	8.A.1.1.1(1)	3
					Contig360gene2101	413	2	8.A.1.1.1(1)	2
					Contig354gene1324	322	1	8.A.1.1.1(1)	3
					Contig375gene9176	328	1	8.A.1.1.1(1)	3
					Contig375gene8188	381	1	8.A.1.1.1(1)	2
					Contig375gene8550	380	2	8.A.1.1.1(1)	2
					Contig375gene8586	392	3	8.A.1.1.1(1)	3
					Contig364gene3066	423	2	8.A.1.1.1(1)	2
					Contig371gene5462	405	0	8.A.1.2.1(1)	3
					Contig373gene6080	505	1	8.A.1.2.1(1)	2
					Contig373gene6556	385	1	8.A.1.2.1(1)	3
					Contig373gene6562	404	1	8.A.1.2.1(1)	3
					Contig375gene8616	520	0	8.A.1.2.1(1)	1
					Contig361gene2415	523	0	8.A.1.2.1(1)	3
					Contig363gene2862	407	1	8.A.1.2.1(1)	3
					Contig369gene4235	93	1	8.A.1.2.1(1)	3
					Contig369gene4236	292	0	8.A.1.2.1(1)	1
					Contig368gene3998	395	1	8.A.1.2.1(1)	1
					Contig329gene132	387	1	8.A.1.6.1(1)	2
					Contig353gene1238	387	0	8.A.1.6.1(1)	3
					Contig358gene1807	412	0	8.A.1.6.1(1)	3
					Contig353gene1180	398	2	8.A.1.6.1(1)	2
					Contig375gene7758	407	3	8.A.1.6.1(1)	2
				25	Contig375gene7764	415	0	8.A.1.6.1(1)	2
8.A.3	Cytoplasmic membrane-periplasmic auxiliary-1 (MPA1) protein with cytoplasmic (C) domain	MPA1	Complex polysaccharides		Contig366gene3603	362	0	8.A.3.2.2(2)	3
					Contig372gene5596	748	1	8.A.3.3.1(1)	2
				3	Contig372gene5968	777	2	8.A.3.3.2(1)	2

Copyright © 2005 John Wiley & Sons, Ltd.

8.A.4	Cytoplasmic membrane-periplasmic auxiliary-2	MPA2	Complex polysaccharides	1	Contig375gene8671	368	2	8.A.4.1.1(1)	2
8.A.7	Phosphotransferase system enzyme I	EI	Sugars	1	Contig374gene7495	585	0	8.A.7.1.1(1)	2
8.A.8	Phosphotransferase system HPr	HPr	Sugars	1	Contig374gene7494	89	1	8.A.8.1.1(1)	3
9.A. Transporters of unknown classification									
9.A.2	MerTP mercuric ion (Hg^{2+}) permease	MerTP	Hg^{2+} (uptake)		Contig375gene8504	88	0	9.A.2.1.1(1)	3
					Contig375gene8370	95	0	9.A.2.1.1(1)	3
				3	Contig369gene4509	91	0	9.A.2.1.1(1)	2
9.A.8	Ferrous iron uptake	FeoB	Fe^{2+} (uptake)	1	Contig372gene5560	620	11	9.A.8.1.1(1)	2
9.A.10	Oxidase-dependent Fe^{2+} transporter	OFeT	Fe^{2+} (uptake)		Contig375gene9242	504	1	9.A.10.1.1(2)	3
					Contig369gene4470	614	1	9.A.10.1.1(2)	3
				3	Contig375gene8124	605	1	9.A.10.1.1(2)	3
9.A.17	Lead	PbrT	Lead resistance		Contig373gene6437	642	7	9.A.17.1.1(1)	2
				2	Contig369gene4270	254	1	9.A.17.1.1(1)	2
9.A.21	ComC DNA uptake competence	ComC	DNA, proteins	1	Contig375gene8629	1102	1	9.A.21.1.1(1)	3
9.B. Putative uncharacterized transporters									
9.B.3	Putative bacterial murein precursor exporter	MPE	Lipid-linked murein precursors such as NAG-NAM-pentapeptide pyrophosphoryl undecaprenol (lipid II)		Contig363gene2762	413	9	9.B.3.1.1(1)	2
				2	Contig374gene7332	380	9	9.B.3.1.2(1)	2
9.B.4	Putative efflux transporter	PET	Unknown		Contig367gene3857	790	12	9.B.4.1.1(1)	2
					Contig353gene1242	664	11	9.B.4.1.2(1)	3
					Contig354gene1321	728	12	9.B.4.1.2(1)	3
				4	Contig375gene9174	659	11	9.B.4.1.2(1)	2
9.B.10	6 TMS putative MarC transporter	MarC	Multiple antibiotic resistance	1	Contig367gene3809	207	6	9.B.10.1.1(1)	3
9.B.14	Putative heme exporter protein	HEP	Heme		Contig375gene8799	653	15	9.B.14.1.1(1)	2
				2	Contig357gene1726	680	15	9.B.14.1.1(1)	2
9.B.17	Putative fatty acid transporter	FAT	Fatty acyl CoA ligases (fatty acyl CoA synthases), carnitine CoA ligases, and putative transporters		Contig349gene843	549	4	9.B.17.1.4(1)	2
					Contig353gene1184	617	2	9.B.17.1.4(1)	2
					Contig340gene362	560	1	9.B.17.1.4(1)	2
					Contig362gene2701	629	2	9.B.17.1.4(1)	2
					Contig362gene2706	553	0	9.B.17.1.4(1)	2
					Contig373gene6071	631	0	9.B.17.1.4(1)	3
					Contig358gene1823	548	1	9.B.17.1.4(1)	2

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
					Contig355gene1404	516	1	9.B.17.1.4(1)	2
					Contig374gene7145	564	0	9.B.17.1.4(1)	2
					Contig345gene613	555	2	9.B.17.1.4(1)	2
					Contig366gene3582	517	0	9.B.17.1.4(1)	2
					Contig375gene8255	630	0	9.B.17.1.4(1)	2
					Contig366gene3706	566	0	9.B.17.1.4(1)	2
					Contig373gene6519	510	0	9.B.17.1.4(1)	2
					Contig373gene6406	515	0	9.B.17.1.4(1)	2
					Contig375gene8238	509	0	9.B.17.1.4(1)	2
					Contig367gene3834	500	0	9.B.17.1.4(1)	2
					Contig371gene5171	545	0	9.B.17.1.4(1)	2
					Contig371gene5425	523	0	9.B.17.1.4(1)	2
					Contig371gene5430	510	3	9.B.17.1.4(1)	2
					Contig371gene5447	501	0	9.B.17.1.4(1)	2
					Contig365gene3363	570	0	9.B.17.1.4(1)	2
					Contig373gene6687	517	0	9.B.17.1.4(1)	2
					Contig372gene5610	518	0	9.B.17.1.4(1)	2
					Contig375gene8757	660	0	9.B.17.1.6(1)	2
					Contig370gene4532	626	1	9.B.17.1.6(1)	2
					Contig373gene6475	545	1	9.B.17.1.6(1)	2
					Contig371gene5414	567	2	9.B.17.1.6(1)	3
					Contig373gene6157	527	1	9.B.17.1.6(1)	2
				30	Contig373gene6682	525	1	9.B.17.1.6(1)	2
9.B.20	Putative Mg^{2+} transporter-C	MgtC	Mg^{2+}		Contig375gene8497	152	4	9.B.20.2.1(1)	3
				2	Contig385gene222	240	4		
9.B.22	Putative permease	PerM	Unknown	1	Contig357gene1729	361	7	9.B.22.1.3(1)	3
9.B.24	Testis-enhanced gene transfer	TEGT	Glucose (and fructose?) uptake or metabolism, cell death	1	Contig360gene2157	235	7	9.B.24.2.1(1)	2
9.B.26	PC-terminal fragment 7	PC-terminal fragment 7	Unknown	1	Contig370gene4655	241	5	9.B.26.1.1(1)	3
9.B.30	Hly III	Hly III	Unknown	1	Contig375gene8924	205	7	9.B.30.1.1(1)	2
9.B.32	Putative vectorial glycosyl polymerization	VGP	Polysaccharides		Contig353gene1189	658	7	9.B.32.1.3(1)	3
				2	Contig369gene4253	367	2	9.B.32.1.3(1)	3
9.B.37	HlyC/CorC	HCC	Ions?		Contig353gene1170	530	7	9.B.37.1.2(1)	3
					Contig346gene665	438	5	9.B.37.2.1(1)	2
				3	Contig375gene7878	437	3	9.B.37.2.1(1)	2
9.B.40	DotA/TraY	DotA/TraY	Unknown	1	Contig366gene3548	751	7	9.B.40.1.2(1)	2
9.B.42	ExeAB	ExeAB	Secretin	1	Contig375gene8813	277	0	9.B.42.1.1(2)	3
9.B.43	YedZ	YedZ	Unknown	1	Contig367gene3778	224	6	9.B.43.1.1(1)	2

Copyright © 2005 John Wiley & Sons, Ltd.

9.B.45	YnfA	YnfA	Unknown	1	Contig375gene8050	105	4	9.B.45.1.1(1)	2
9.B.53	Unknown IT-6	UIT6	Unknown	1	Contig347gene696	476	12	9.B.53.1.1(0)	3
Unclassified									
	Unclassified	Unclassified	Unknown		Contig358gene1867	397	6	N/A(0)	3
					Contig353gene1240	384	4	N/A(0)	3
					Contig370gene4727	380	0	N/A(0)	3
					Contig370gene4852	195	0	N/A(0)	3
					Contig355gene1383	409	10	N/A(0)	3
					Contig366gene3534	372	6	N/A(0)	3
					Contig366gene3535	387	6	N/A(0)	3
					Contig375gene7560	363	9	N/A(0)	3
					Contig358gene1814	388	8	N/A(0)	3
					Contig375gene8797	63	1	N/A(0)	3
					Contig365gene3385	367	7	N/A(0)	3
					Contig342gene421	127	0	N/A(0)	3
					Contig342gene426	461	9	N/A(0)	3
					Contig342gene428	333	0	N/A(0)	3
					Contig373gene6385	3750	1	N/A(0)	3
					Contig367gene3796	258	6	N/A(0)	3
					Contig367gene3797	179	1	N/A(0)	3
					Contig375gene9280	83	0	N/A(0)	3
					Contig372gene5736	241	6	N/A(0)	3
					Contig362gene2646	389	6	N/A(0)	3
					Contig371gene5477	273	1	N/A(0)	3
					Contig371gene5133	335	1	N/A(0)	3
					Contig375gene9091	273	6	N/A(0)	3
					Contig375gene8503	115	3	N/A(0)	3
					Contig374gene6814	955	0	N/A(0)	3
					Contig374gene6977	376	6	N/A(0)	3
					Contig374gene7342	316	10	N/A(0)	3
					Contig365gene3237	268	1	N/A(0)	3
					Contig356gene1538	232	7	N/A(0)	3
					Contig352gene1090	231	7	N/A(0)	3
					Contig352gene1095	336	0	N/A(0)	3
					Contig375gene8369	116	3	N/A(0)	3
					Contig375gene8372	562	0		3
					Contig364gene3037	253	7	N/A(0)	3
					Contig372gene5975	467	12	N/A(0)	3
					Contig367gene3796	258	6	N/A(0)	3
					Contig375gene7603	174	1	N/A(0)	3
					Contig375gene7604	134	1	N/A(0)	3

Copyright © 2005 John Wiley & Sons, Ltd.

Table 3. Continued

Family (1)	Family (2)	Abbreviation (3)	Typical substrates (4)	Total # (5)	Gene (02jul03) (6)	Length (#aas) (7)	# TMSs (8)	Nearest homologue in TCDB (9)	Evidence (10)
					Contig375gene7607	509	0	N/A(0)	3
					Contig375gene7608	188	1	N/A(0)	3
					Contig375gene7609	268	1	N/A(0)	3
					Contig369gene4471	427	0	N/A(0)	3
					Contig369gene4472	132	1	N/A(0)	3
					Contig369gene4473	305	8	N/A(0)	3
					Contig369gene4474	158	0	N/A(0)	3
					Contig369gene4481	435	0	N/A(0)	3
					Contig375gene9429	402	12	N/A(0)	3
					Contig375gene8485	366	0	N/A(0)	3
					Contig375gene8120	419	1	N/A(0)	3
					Contig375gene8125	360	1	N/A(0)	3
					Contig375gene8126	128	0	N/A(0)	3
					Contig369gene4508	116	3	N/A(0)	3
					Contig368gene4000	351	10	N/A(0)	3
					Contig368gene4195	324	0	N/A(0)	3
				55	Contig368gene4197	197	0	N/A(0)	3
				932					

[a] A full version of the table containing all the various names of the CH34 genes is provided as on-line supplementary material at: **http://bionomie.mikrobiologie.uni-halle.de/SupMat/Roz_05/Table 3.htm**

Copyright © 2005 John Wiley & Sons, Ltd.

structure is established for several members of this family. Three members of the OmpA-OmpF porin (OOP) family and a single FadL homologue, presumably concerned with transport of fatty acids across the outer membrane, were identified.

The next two families listed in Table 3, the FUP and AT families, with three members and one member, respectively, are concerned with export of proteins across the outer membrane. The three FUP ushers probably export fimbrial subunits for the assembly of 3 structurally and functionally distinct fimbriae. AT family members export their own N-terminal domains, which in this case may be a large cell surface protein. However, no surface layer could be observed for Rme (D. Neumann and D. H. Nies, unpublished data).

Seventeen OMR family members were identified. Fifteen of these are probably concerned with uptake of iron siderophore complexes (subfamilies 1 and 9). One is probably the Rme vitamin B_{12} porin (subfamily 3). The single member of subfamily 4 may be concerned with copper acquisition.

Outer membrane factors (OMFs; TC #1.B.17) generally mediate efflux of heavy metals, drugs and macromolecules across the outer membrane in conjunction with an active efflux pump in the inner membrane. Twenty-eight homologues were identified. Of these, one is in subfamily 1 (a general OMF able to interact with multiple efflux pumps), eight are in subfamily 2 (concerned with heavy metal ion efflux), and 19 are in subfamily 3 (concerned with export of macromolecules, drugs and metals). Two members of this last subfamily resemble oligosaccharide exporters; four most resemble protein exporters; seven may be involved in export of drugs and other hydrophobic substances; and three may function in copper ion efflux.

Two members of the OMA family (1.B.18) are presumed to function in exopolysaccharide export, one member of the OprB family (1.B.19) probably allows facilitation of small molecules across the outer membrane, and the two members of the TPS family (1.B.20) most likely export proteins. Most of the six secretins (1.B.22) also probably function in protein export. Finally, the two OmpW family members (1.B.39) may export drugs and other hydrophobic molecules.

A channel-forming colicin-like protein (1.C.1), resembling colicin A of *Citrobacter freundii*, was found. A single holin (1.E.14), presumably involved in autolysin export for the purpose of promoting cell death, is also present.

Secondary carriers

By far the largest number secondary carriers encoded within the Rme genome are members of the major facilitator superfamily (MFS). Rme has 83 recognizable MFS carriers. As shown in Table 3 and summarized in Table 4, 32 of these MFS permeases are putative drug/amphiphile/hydrophobe transporters of MFS families DHA1 (16 members), DHA2 (15 members) and DHA3 (1 member) (Busch and Saier, 2002). Some of these are likely to serve as lipid exporters, but others undoubtedly play primary roles in defence, in toxic substance export or in metabolite export.

Just one sugar transporter (SP family), one organophosphate porter (OPA family), 15 metabolite transporters (MHS family), three nitrate/nitrite transporters (NNP family), and three oxalate : formate antiporters (OFA) of the MFS allow uptake of essential nutrients. Additionally, one SHS porter, nine ACS porters, five AAHS porters, and one CP porter all probably function to bring organoanions into the cell. The OCT porter may transport organocations. Other MFS paralogues represented, with usually a single protein member in any one family, undoubtedly transport a wide range of other substances (Table 3).

Six amino acid/polyamine/organocation (APC) superfamily members were identified. Two of the subfamilies in the APC superfamily are represented. These porters are predicted to transport a range of zwitterionic and basic amino acids.

The CDF family and the ZIP family of heavy metal divalent cation transporters are represented with three and one members, respectively. All three CDF proteins have been characterized in detail (Anton *et al*., 2004; Munkelt *et al*., 2004). They belong to different clusters of the CDF protein family (Nies, 2003) and transport Cd^{2+}, Co^{2+}, Zn^{2+}, Fe^{2+} and Ni^{2+}. A single member of the NiCoT family (TC #2.A.53), probably a Ni^{2+} transporter, was also identified. A related protein is involved in nickel uptake for synthesis of the hydrogenases in the related bacterium *Ralstonia eutropha* (Degen and Eitinger, 2002; Eberz *et al*., 1989; Eitinger and Friedrich, 1991, 1994; Eitinger *et al*., 1997; Wolfram *et al*., 1991, 1995).

Copyright © 2005 John Wiley & Sons, Ltd.

Table 4. Family associations including subfamilies within the MFS, APC, RND, DMT, MOP and ABC superfamilies of transporter constituents

Family	Abbreviation	Typical substrates	No. of members (%)
1.A.1	VIC	Na^+, K^+, Ca^{2+}, multiple cations	2 (0.2)
1.A.8	MIP	H_2O, glycerol, urea, polyols, NH_3, CO_2	2 (0.2)
1.A.11	ClC	Cl^-, anions	4 (0.4)
1.A.20	CytB	H^+	1 (0.1)
1.A.22	MscL	Proteins, ions (slightly cation-selective)	1 (0.1)
1.A.23	MscS	Ions (slight anion selectivity)	9 (1)
1.A.30	Mot/Exb-Mot	H^+, Na^+	2 (0.2)
1.A.33	Hsp70	Ions, polypeptides	2 (0.2)
1.A.35	MIT	Heavy-metal ions, Mg^{2+}, Mn^{2+}, Co^{2+}, Ni^{2+}, Fe^{2+}, Al^{3+}, Mn^{2+}	4 (0.4)
1.B.1	GBP	Ions, small (M_r < 1000 Da) molecules	29 (3.1)
1.B.6	OOP	Ions, small molecules	3 (0.3)
1.B.9	FadL	Fatty acid, toluene, *m*-xylene and benzyl alcohol	1 (0.1)
1.B.11	FUP	Protein folding and subunit assembly	3 (0.3)
1.B.12	AT	N-terminal protein domains	1 (0.1)
1.B.14	OMR	Iron–siderophore complexes, vitamin B_{12}, Cu^{2+}, colicin, DNA of various phages	17 (1.8)
1.B.17	OMF	Heavy metal cations, drugs, oligosaccharides, proteins, etc.	28 (3)
1.B.18	OMA	Exo- or capsular polysaccharide	2 (0.2)
1.B.19	OprB	Ions, small molecules	1 (0.1)
1.B.20	TPS	Proteins	2 (0.2)
1.B.22	Secretin	Proteins	6 (0.6)
1.B.39	OmpW	Methyl viologen and benzyl viologen	2 (0.2)
1.C.1	Colicin	Ions, small molecules	1 (0.1)
1.E.14	LrgA Holin	Zn^{2+}, Fe^{2+}	1 (0.1)
2.A.1	MFS	Various small molecules	Total 83 (8.9)
	-SP (1)	Sugars	1 (0.1)
	-DHA1 (12 spanner) (2) drugs	Drugs	16 (1.7)
	-DHA2 (14 spanner) (3) drugs	Drugs	15 (1.6)
	-OPA (4)	Sugars, glycerol	1 (0.1)
	-MHS (6)	Dicarboxylates, tricarboxylates	15 (1.6)
	-NNP (8)	Nitrate, nitrite	3 (0.3)
	-OFA (11)	Oxalate, formate	3 (0.3)
	-SHS (12)	Sialate, lactate, pyruvate	1 (0.1)
	-ACS (14)	Organic acids	9 (1)
	-AAHS (15)	Aromatic acids	5 (0.5)
	-CP (17)	Cyanate	1 (0.1)
	-OCT (19)	Organic cations	1 (0.1)
	-SET (20)	Sugars	1 (0.1)
	-DHA3 (12 spanner) (21) drugs	Drugs	1 (0.1)
	-VNT (22)	Neurotransmitter	1 (0.1)
	-BST (23)	Unknown	1 (0.1)
	-PAT (25)	Peptides, AcCoA	1 (0.1)
	-UMC-terminal fragment (26)	Unknown	1 (0.1)
	-PPP (27)	Phenylpropionate	1 (0.1)
	-ADT (30)	Abietane diterpenoid	1 (0.1)
	-Nre (31)	Ni^{2+}	1 (0.1)
	-Fsr (35)	Fosmidomycin	1 (0.1)
	-AtoE (37)	Short chain fatty	2 (0.2)
2.A.3	APC	Amino acids, polyamines, choline	Total 6 (0.6)
	-AAA (1)	Amino acids	5 (0.5)
	-CAT (3)	Cationic amino acids	1 (0.1)
2.A.4	CDF	Cd^{2+}, Co^{2+}, Zn^{2+}	3 (0.3)
2.A.5	ZIP	Zn^{2+}, Fe^{2+}	1 (0.1)

Copyright © 2005 John Wiley & Sons, Ltd.

Table 4. Continued

Family	Abbreviation	Typical substrates	No. of members (%)
2.A.6	RND	Heavy metal ions, multiple drugs, oligosaccharides, organic solvents, fatty acids, phospholipids, cholesterol	Total 30 (3.2)
	-HME (1)	Heavy metals	17 (1.8)
	-HAE1 (2)	Hydrophobe/amphiphiles	9 (1)
	-SecDF(4)	Sec secretory accessory proteins	2 (0.2)
	-HAE2 (5)	Hydrophobe/amphiphiles	1 (0.1)
	-ORF4 (8)	Hydrophobe/amphiphiles	1 (0.1)
2.A.7	DMT	Multiple drugs and dyes (mostly cationic)	Total 18 (1.9)
	-SMR (1)	Drugs	2 (0.2)
	-BAT (2)	Unknown	2 (0.2)
	-DME (3)	Drugs, metabolites	12 (1.3)
	-RarD (7)	Chloramphenicol	2 (0.2)
2.A.9	Oxa1	Proteins	1 (0.1)
2.A.10	KDGT	2-Keto-3-deoxygluconate	1 (0.1)
2.A.11	CitMHS	Citrate	1 (0.1)
2.A.12	AAA	ATP, ADP	1 (0.1)
2.A.14	LctP	Lactate	1 (0.1)
2.A.19	CaCA	Ca^{2+}	1 (0.1)
2.A.20	PiT	Inorganic phosphate	1 (0.1)
2.A.21	SSS	Sugars, amino acids, vitamins, nucleosides, inositols, iodide, urea	5 (0.5)
2.A.23	DAACS	C_4-dicarboxylates, acidic and neutral amino acids	5 (0.5)
2.A.24	CCS	Mono-, di-, and tricarboxylates	1 (0.1)
2.A.36	CPA1	Na^+/H^+, Na^+ or K^+/H^+	1 (0.1)
2.A.37	CPA2	Na^+/H^+ or K^+/H^+	6 (0.6)
2.A.40	NCS2	Nucleobases, urate	3 (0.3)
2.A.45	ArsB	Arsenite, antimonite	1 (0.1)
2.A.46	BenE	Benzoate	1 (0.1)
2.A.47	DASS	Dicarboxylates, phosphate, sulphate	4 (0.4)
2.A.49	Amt	Ammonium	2 (0.2)
2.A.51	CHR	Chromate, sulphate (uptake or efflux)	4 (0.4)
2.A.52	NiCoT	Ni^{2+}, Co^{2+}	1 (0.1)
2.A.53	SulP	Sulphate	5 (0.5)
2.A.56	TRAP−T	C_4-dicarboxylates, acidic amino acids, sugars?	6 (0.6)
2.A.58	PNaS	Inorganic phosphate	2 (0.2)
2.A.59	ACR3	Arsenite	1 (0.1)
2.A.64	Tat	Redox proteins	4 (0.4)
2.A.66	MOP	Drugs, lipid-linked oligosaccharide precursors	Total 5 (0.5)
	-MATE (1)	Drugs	3 (0.3)
	-PST (2)	Polysaccharides	1 (0.1)
	-MVF (4)	Unknown	1 (0.1)
2.A.67	OPT	Peptides	2 (0.2)
2.A.69	AEC	Auxin (efflux)	2 (0.2)
2.A.72	KUP	K^+ (uptake)	1 (0.1)
2.A.75	LysE	Basic amino acids	1 (0.1)
2.A.76	RhtB	Neutral amino acids and their derivatives	11 (1.2)
2.A.78	LIV−E	Carboxylates, amino acids, amines (efflux)	1 (0.1)
2.A.80	TTT	Tricarboxylate	74 (8)
2.A.81	AAE	Aspartate, alanine	2 (0.2)
2.C.1	TonB	H^+?, drives solute uptake across outer bacterial membranes	5 (0.5)
3.A.1	ABC	All sorts of inorganic and organic molecules of small, intermediate, and large sizes, from simple ions to macromolecules	Total 213 (23)
	-CUT1(1)	Sugars, metabolites	15 (1.6)

Copyright © 2005 John Wiley & Sons, Ltd.

Table 4. Continued

Family	Abbreviation	Typical substrates	No. of members (%)
	-CUT2 (2)	Sugars, metabolites	4 (0.4)
	-PAAT (3)	Polar amino acids	14 (1.5)
	-HAAT (4)	Hydrophobic amino acids	45 (4.8)
	-PepT (5)	Peptide, opine, nickel	33 (3.5)
	-SulT (6)	Sulphate, tungstate	6 (0.6)
	-PhoT (7)	Phosphate	6 (0.6)
	-MolT (8)	Molybdate	4 (0.4)
	-PhnT (9)	Phosphonate	6 (0.6)
	-POPT(11)	Polyamine, opine, phosphonate	8 (0.9)
	-QAT (12)	Quaternary amine	5 (0.5)
	-VB12T(13)	Vitamin B_{12}	1 (0.1)
	-FeCT (14)	Iron chelate	5 (0.5)
	-MZT (15)	Manganese, zinc, iron chelate	1 (0.1)
	-NitT (16)	Nitrate, nitrite, cyanate	14 (1.5)
	-TauT (17)	Taurine	13 (1.4)
	-BIT (20)	Fe^{3+}	1 (0.1)
	-CPSE (101)	Capsular polysaccharides	2 (0.2)
	-LOSE (102)	Lipo-oligosaccharide	7 (0.8)
	-DrugE1(105)	Drugs	2 (0.2)
	-HemeE(107)	Heme	6 (0.6)
	-Prot1E (109)	Proteins	1 (0.1)
	-Prot2E (110)	Proteins	1 (0.1)
	-Drug RA1 (120)	Drugs	5 (0.5)
	-Drug RA2 (121)	Drugs	1 (0.1)
	-MacB (122)	Macrolide	2 (0.2)
	-LPT (125)	Lipoproteins	3 (0.3)
	-HMT (210)	Heavy metals	2 (0.2)
3.A.2	F-ATPase	H^+, Na^+	8 (0.9)
3.A.3	P-ATPase	Na^+, H^+, K^+, Ca^{2+}, Mg^{2+}, Cd^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, Co^{2+}, Ni^{2+}, Ag^+, phospholipids (flipping)	12 (1.3)
3.A.5	IISP	Proteins	5 (0.5)
3.A.6	IIISP	Proteins	10 (1.1)
3.A.7	IVSP	Proteins, protein–DNA complexes	20 (2.2)
3.A.11	DNA-T	Single-stranded DNA	1 (0.1)
3.A.12	S-DNA-T	DNA, DNA–protein complexes	2 (0.2)
3.A.13	FPhE	Viruses	1 (0.1)
3.A.15	MTB	Pilin/fimbrilin	18 (1.9)
3.B.1	NaT-DC	Na^+	2 (0.2)
3.D.1	NDH	H^+ or Na^+ (efflux)	19 (2)
3.D.2	PTH	H^+ (efflux)	6 (0.6)
3.D.3	QCR	H^+ (efflux)	3 (0.3)
3.D.4	COX	H^+ (efflux)	23 (2.5)
4.A.6	Man	Glucose, mannose, fructose, sorbose, etc.	2 (0.2)
5.A.1	DsbD	2 e^-	2 (0.2)
5.A.2	DsbB	2 e^-	1 (0.1)
5.A.3	PMO	Proton translocation	7 (0.8)
8.A.1	MFP	Proteins, peptides, lipopolysaccharides, drugs, dyes, signalling molecules, heavy metal ions, etc.	25 (2.7)
8.A.3	MPA1	Complex polysaccharides	3 (0.3)
8.A.4	MPA2	Complex polysaccharides	1 (0.1)
8.A.7	EI	Sugars	1 (0.1)
8.A.8	HPr	Sugars	1 (0.1)
9.A.2	MerTP	Hg^{2+} (uptake)	3 (0.3)
9.A.8	FeoB	Fe^{2+} (uptake)	1 (0.1)
9.A.10	OFeT	Fe^{2+} (uptake)	3 (0.3)
9.A.17	PbrT	Lead resistance	2 (0.2)
9.A.21	ComC	DNA, proteins	1 (0.1)

Copyright © 2005 John Wiley & Sons, Ltd.

Table 4. Continued

Family	Abbreviation	Typical substrates	No. of members (%)
9.B.3	MPE	Lipid-linked murein precursors, such as NAG–NAM–pentapeptide pyrophosphoryl undecaprenol (lipid II)	2 (0.2)
9.B.4	PET	Unknown	4 (0.4)
9.B.10	MarC	Multiple antibiotic resistance	1 (0.1)
9.B.14	HEP	Heme	2 (0.2)
9.B.17	FAT	Fatty acyl CoA ligases (fatty acyl CoA synthases), carnitine CoA ligases, and putative transporters	30 (3.2)
9.B.20	MgtC	Mg^{2+}	2 (0.2)
9.B.22	PerM	Unknown	1 (0.1)
9.B.24	TEGT	Glucose (and fructose?) uptake or metabolism, cell death	1 (0.1)
9.B.26	PC-terminal fragment (7)	Unknown	1 (0.1)
9.B.30	Hly III	Unknown	1 (0.1)
9.B.32	VGP	Polysaccharides	2 (0.2)
9.B.37	HCC	Ions?	3 (0.3)
9.B.40	DotA/TraY	Unknown	1 (0.1)
9.B.42	ExeAB	Secretin	1 (0.1)
9.B.43	YedZ	Unknown	1 (0.1)
9.B.45	YnfA	Unknown	1 (0.1)
9.B.53	UIT6	Unknown	1 (0.1)
	Unclassified	Unknown	1 (0.1)
Total			932 (100)

The RND superfamily of export pumps is well represented, with 30 members. Of these, over half (17) in subfamily 1 are predicted to function in heavy metal efflux. Another nine (in subfamily 2) probably export drugs and other hydrophobic and amphipathic substances. The RND proteins of Rme have been compared to those from other bacteria recently (Nies, 2003). The two SecDF system components (subfamily 4), facilitate protein secretion via the general secretory pathway (Sec; 3.A.5). Lipid (subfamily 5) and pigment (subfamily 8) exporters may also be present.

Another well-represented superfamily encoded within the genome of Rme is the drug/metabolite transporter (DMT) superfamily, with 18 members within four of the families of this superfamily. Most of these transporters (families 1, 2 and 3) probably function in drug and metabolite efflux, but one (family 7) may be a sugar uptake permease.

A single putative 2-keto-3-deoxygluconate uptake permease was identified. Additionally, one member of the CitMHS (citrate uptake) family and one member of the LctP lactate uptake family were found. One system may export Ca^{2+} (CaCA family) while another may import phosphate (PiT family). A surprise was the identification of a member of the ATP : ADP antiporter (AAA) family, because such transporters were previously predominantly identified in intracellular pathogenic organisms and rarely in other bacteria (until now in *Ralstonia eutropha* strain JMP134, *Pseudomonas fluorescens, Pirella, Rhodopirellula baltica* and *Magnetospirillum magnetotacticum*). However, what it could be doing in a free-living organism remains to be determined.

Five members of the SSS family most resemble characterized permeases for organoanions and cations as well as a putative nitrogen sensor. All of the five members of the DAACS family are predicted to transport dicarboxylates. These may include the two dicarboxylate amino acids, aspartate and glutamate. A putative CCS family member is also predicted to take up dicarboxylates. The four DASS family members probably serve similar functions but may also take up tricarboxylate compounds.

Both the CPA1 and CPA2 monovalent cation antiporter families are represented, with one and six members, respectively. CPA1 family members are predicted to be $Na^{+} : H^{+}$ antiporters, while CPA2 family members may be K^{+} efflux systems. Three NCS2 nucleobase/nucleoside uptake systems and

Copyright © 2005 John Wiley & Sons, Ltd.

two Amt ammonia/ammonium transporters were identified.

Two putative arsenite exporters (one of the ArsB-type and one of the Acr3-type) were found. Four potential chromate resistance (CHR) pumps and five putative sulphate uptake permeases (SulP) may be involved in chromate and sulphate metabolism, respectively. The CHR and SulP porters may be functionally related, since chromate is a sulphate analogue.

Six constituents of the tripartite TRAP-T family (2.A.56) may comprise three distinct systems for dicarboxylate uptake. However, studies indicate that members of this family may transport substrates of diverse structure, rendering substrate identification difficult. Only two TRAP-T receptors but at least three large and one small integral membrane constituents of these systems were identified. Because of rapid sequence divergence of the small integral membrane constituents, some of these proteins may have been missed. This situation can be contrasted with the superficially similar tripartite TTT family (2.A.80), where 74 potential constituents were found. Interestingly, about five proved to resemble the large and 11 the small integral membrane constituents of these systems, while 58 proved to be homologous to TTT family receptors. The occurrence of multiple probable receptors for TTT family systems in some bacteria has been noted before (Antoine *et al*., 2003).

Several additional families of transporters are probably involved in nutrient uptake (BenE, OPT and AAE) and metabolite efflux (AEC, LysE, RhtB and LIV-E). All of these are concerned with transport of peptides, amino acids and their derivatives. The largest of these families is the RhtB family, with 11 members. Additionally, constituents of a TonB–ExbBD system, which probably functions primarily to energize transport across the outer membrane by a proton electrophoretic mechanism, were identified.

A complete twin arginine targeting (TatABC) system, as well as a single Oxa1 homologue, is encoded within the genome of Rme. These two independently acting systems function in the secretion of a subset of extracellular proteins and in the insertion of integral membrane proteins, including redox enzymes, respectively (Yen *et al*., 2002). Genome analyses of the leader sequences of potential secretory proteins should reveal which are substrates of the Tat system and which are exported via the Sec system.

Primary active transporters

The vast majority of protein constituents of primary active transporters encoded within the Rme genome are members of the ABC superfamily; 213 proteins in Rme belong to this superfamily, 181 putative uptake system proteins and 32 putative efflux system proteins. Most ABC systems consist minimally of two membrane protein (M) and two ATP hydrolysing cytoplasmic protein (C) subunits which may be fused in various combinations. Consequently, the basic unit of an ABC transporter may be encoded by a single gene or up to four distinct genes. Additionally, extracytoplasmic receptors are associated with all uptake systems, and there may be several of these per system. Therefore, it is not possible to estimate accurately the number of intact ABC transporters present. The problem is exacerbated by the fact that the constituents of ABC systems are often encoded within multiple, non-adjacent operons.

Table 4 summarizes the family associations of the various ABC transporter constituents. The ratio of sugar uptake system constituents (CUT1 + CUT2) to amino acid plus peptide uptake systems (PAAT + HAAT + PepT) is 15 : 52 or about 1 : 4. This fact, together with the corresponding analyses of secondary carriers discussed above, reveals the much greater dependency of Rme on amino acid metabolism than carbohydrate metabolism (see also Table 2). Values for numbers of sugar and amino acid transporter constituents can be compared with the total number of organic and inorganic anion and cation uptake transporter constituents (about 20 of each). ABC-type efflux systems are concerned with the export of drugs (10), complex carbohydrates (5), heme (6), proteins (5) and heavy metals (7) (Tables 3 and 4).

Rme has a single multicomponent F-type ATPase for the interconversion of chemical and chemiosmotic energy. It also possesses a dozen paralogous cation transporting P-type ATPases. Three of them have been characterized in detail (Borremans *et al*., 2001; Legatzki *et al*., 2003a) and all of them have been compared to P-type ATPases from other bacteria (Nies, 2003). Recently, the ongoing annotation work (**http://genome.ornl.gov/microbial/rmet/**)

identified another P-type ATPase (ZP_00 273 867) that was not included here.

A complete multicomponent general protein secretory (Sec) system (TC #3.A.5) was found in Rme, and this system undoubtedly serves as the primary protein export system for transport of proteins from the cytoplasm to the periplasm (Cao and Saier, 2003). However, Rme also has types II (MTB), III and IV macromolecular export systems. The first of these functions exclusively to export proteins across the outer membrane, but the latter two transport their substrates across both membranes. Type IV systems may also function in conjugation, and, in plant pathogens, in DNA export to the host cell. Additional potential DNA translocation proteins of the DNA-T, S-DNA-T and FphE families were also identified (Table 3). However, assignment of their specific functional roles must await experimental studies.

The Na^+ transporting carboxylate decarboxylases (TC #3.B.1) are multicomponent systems where the β-subunit catalyses Na^+ export in response to cytoplasm substrate decarboxylation catalysed by the α-subunit. These systems minimally require the presence of α-, β- and γ-subunits (Dimroth *et al*., 2001). One such system may be present in Rme.

Proton pumping electron carriers

Rme has a single member of each of the three proton- or sodium-translocating electron transfer complexes of the NADH dehydrogenase (NDH), quinol : cytochrome *c* reductase (QCR) and cytochrome oxidase (COX) families. It also has at least two multicomponent transhydrogenases (PTH family). Rme therefore has a complete electron transfer chain for oxidizing NADH, using molecular oxygen as electron acceptor. All four electron carrier complexes cited above have the potential to generate an ion motive force as a primary source of energy. These coupled systems probably function together under aerobic conditions. Other transmembrane electron flow systems that can influence cellular energetics (class 5A and 5B) were also identified.

Group translocators

The complete phosphoenolpyuvate–sugar phosphotransferase system (PTS; TC #4.A) is present in Rme. It includes, however, just one mannose (Man)-type PTS permease (Zhang *et al*., 2003). Only one Enzymes I and one HPr were identified. It is clear that Rme possesses a minimal PTS, in agreement with the earlier conclusion, based on secondary and primary active transporter analyses, that Rme is not strongly dependent on sugar metabolism as a source of energy.

Poorly-defined transporters

Among the poorly characterized permeases of TC class 9.A, Rme has systems that probably transport heavy metal ions: mercury, iron, lead and magnesium. Several putative permeases of TC class 9.B were also identified (Table 3), but their functions are not known.

Perspectives and conclusions

We have analysed transporters in the heavy metal-resistant organism, *R. metallidurans* (Rme). This organism possesses several α-type channel proteins. Some are concerned specifically with monovalent or divalent inorganic cation or anion transport, but several non-specific stress response channels also appear to be present. Rme also has a huge repertoire of outer membrane β-barrel porins involved in transport of small molecules as well as macromolecules across the outer membrane. Many (e.g. OMRs) are probably specific for uptake, while others (e.g. OMFs) mediate efflux.

Regarding secondary carriers for sugars, Rme seems to have a very limited repertoire of such systems relative to most other sequenced Gram-negative bacteria, such as *E. coli* and other enteric bacteria. Thus, Rme has only one MFS carbohydrate transporter in the sugar porter family. It has no putative glycoside transporters of the GPH family (TC #2.A.2). It does have a putative 2-keto-3-deoxygluconate transporter of the KDGT family, and it has a few ABC uptake transporters specific for monosaccharides and small oligosaccharides of the CUT1 and CUT2 subfamilies, as well as a complete phosphotransferase system. Rme may only transport hexoses via the one PTS permease identified.

The capacity of Rme to transport carboxylic acids and their derivatives as sources of carbon appears to be fairly extensive. Thus, several families of secondary mono- and dicarboxylate carriers

Copyright © 2005 John Wiley & Sons, Ltd.

(MFS, DAACS, DASS and TRAP-T) were identified. It also possesses members of the tricarboxylate transporting CitMHS, CCS and TTT families (Winnen *et al*., 2003). ABC-type carboxylate transporters were also found. Thus, the results point to a strong respiratory-type metabolism, with greater dependency on exogenous organic acids than carbohydrates.

Our genome analyses revealed several transporters that are probably specific for amino acids, peptides and their derivatives. Thus, for the uptake of amino acids, three families of secondary carriers were represented [MFS (MHS), APC and SSS], while members of two ABC families with this specificity (PAAT and HAAT) were found. For the uptake of peptides, two potential families of secondary carriers (OPT, MPE) and one ABC family (PepT) were represented. Finally, for amino acid efflux, members of five potential families were identified (DMT, AEC, LysE, RhtB and LIV-E). It seems clear that the transport and metabolism of amino acids and their derivatives is of considerable importance to the lifestyle of Rme.

Our analyses also revealed a large number of potential drug/hydrophobe/amphiphile export systems. Many of these belong to the DHA1, -2 and -3 families of the MFS. While a few of these efflux pumps may be involved in sugar export (Table 3; Saier, 2000), it is possible that some export amino acids and their derivatives, particularly those of a hydrophobic nature. It should be noted, however, that this has not yet been established for any member of the three MFS DHA families.

Other families, including transporters that probably export hydrophobic substances, include the HAE1 family in the RND superfamily, and the DME family of the DMT superfamily. At least some of these are probably concerned with drug export. Members of the MATE family within the MOP superfamily and several putative drug exporters of the ABC superfamily may serve similar functions. All of these families are represented in Rme. The diversity of substrates exported by these systems has yet to be studied.

As noted in Table 2 and further exemplified in Tables 3 and 4, over 220 transporters in Rme are probably concerned with inorganic ion transport. The following families are represented (see Table 3): (1) for monovalent cations: VIC, CytB, MscL, MscS, CPA1, CPA2, Amt, KUP, F-ATPase, P-ATPase and four proton-translocating electron carriers (NDH, PTH, QCR and COX); (2) for di- or trivalent cations: MIT, NNP(MFS), CDF, ZIP, RND, CaCA, NiCoT, FeCT(ABC), MZT(ABC), P-ATPase and MgtC; and (3) for anions: MFS, Pit, ArsB, DASS, CHR, SulP, PNaS, ACR3, SulT(ABC), PhoT(ABC), MolT(ABC) and NitT(ABC).

Inspection of Table 3 reveals possible transporters for a variety of additional interesting metabolites, such as organic anions (benzoate, phenylacetate, cyanate, phosphonates, sulphonates). Transporters specific for osmolytes, both purine and pyrimidine bases and nucleosides, quaternary ammonium compounds and possibly nucleotides (ADP/ATP in the AAA family), were also identified.

An extensive repertoire of macromolecular exporters was found. Protein secretion and membrane protein insertion systems include the Sec, Tat, Oxa1 and types I–IV systems. Complex carbohydrates can probably be exported via MOP, ABC and VGP family transporters. Possible lipid exporters of the RND superfamily have been identified, and several MFS and ABC systems may similarly catalyse lipid ‘flip-flop’, which is equivalent to export from the inner leaflet of the cytoplasm membrane bilayer to the outer leaflet. Some of these transporters may also export lipids from the inner membrane to the outer membrane.

Finally, several of the identified transporters could not be assigned even a tentative function. It should also be kept in mind that transporters that belong to functionally uncharacterized families may not be included in the TC system and therefore may not be identified using the computer approaches used here. Although our studies have revealed a disproportionate number of transporters concerned with inorganic ion transport, particularly with heavy metal resistance, and while these studies clearly point to the dominant types of metabolic activity upon which Rme depends for energy, it is clear that we are only at the beginning of an understanding of the scope of molecular transport processes in *Ralstonia metallidurans*.

Acknowledgements

This work was supported by NIH Grant No. GM64368. We thank the Joint Genomic Institute (JGI) for permitting us use of the preliminary genomic sequence of *R. metallidurans*, without which this study would not have been

Copyright © 2005 John Wiley & Sons, Ltd.

possible. We thank Mary Beth Hiller for her assistance in the preparation of this manuscript.

References

Andres Y, Thouand G, Boualam M, Mergeay M. 2000. Factors influencing the biosorption of gadolinium by microorganisms and its mobilization from sand. *Appl Microbiol Biotechnol* **54**: 262–267.

Antoine R, Jacob-Dubuisson F, Drobecq H, *et al*. 2003. Overrepresentation of a gene family encoding extracytoplasmic solute receptors in *Bordetella*. *J Bacteriol* **185**: 1470–1474.

Anton A, Weltrowski A, Haney JH, *et al*. 2004. Characteristics of zinc transport by two bacterial cation diffusion facilitators from *Ralstonia metallidurans* and *Escherichia coli*. *J Bacteriol* **186**: 7499–7507.

Arispe N, De Maio A. 2000. ATP and ADP modulate a cation channel formed by Hsc70 in acidic phospholipid membranes. *J Biol Chem* **275**: 30 839–30 843.

Borremans B, Hobman JL, Provoost A, Brown NL, van Der Lelie D. 2001. Cloning and functional analysis of the pbr lead resistance determinant of *Ralstonia metallidurans* CH34. *J Bacteriol* **183**: 5651–5658.

Boucher C, Genin S, Arlat M. 2001. Current concepts on the pathogenicity of phytopathogenic bacteria. *C R Acad Sci III* **324**: 915–922.

Busch W, Saier MH Jr. 2002. The transporter classification (TC) system, 2002. *CRC Crit Rev Biochem Mol Biol* **37**: 287–337.

Cao TB, Saier MH Jr. 2003. The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions. *Biochim Biophys Acta* **1609**: 115–125.

Degen O, Eitinger T. 2002. Substrate specificity of nickel/cobalt permeases: insights from mutants altered in transmembrane domains I and II. *J Bacteriol* **184**: 3569–3577.

Dimroth P, Jockel P, Schmid M. 2001. Coupling mechanism of the oxaloacetate decarboxylase Na^+ pump. *Biochim Biophys Acta* **1505**: 1–14.

Eberz G, Eitinger T, Friedrich B. 1989. Genetic determinants of a nickel-specific transport system are part of the plasmid-encoded hydrogenase gene cluster in *Alcaligenes eutrophus*. *J Bacteriol* **171**: 1340–1345.

Eitinger T, Friedrich B. 1991. Cloning, nucleotide sequence, and heterologous expression of a high-affinity nickel transport gene from *Alcaligenes eutrophus*. *J Biol Chem* **266**: 3222–3227.

Eitinger T, Friedrich B. 1994. A topological model for the high-affinity nickel transporter of *Alcaligenes eutrophus*. *Mol Microbiol* **12**: 1025–1032.

Eitinger T, Wolfram L, Degen O, Anthon C. 1997. A Ni^{2+} binding motif is the basis of high affinity transport of the *Alcaligenes eutrophus* nickel permease. *J Biol Chem* **272**: 17 139–17 144.

Goris J, De Vos P, Coenye T, *et al*. 2001. Classification of metal-resistant bacteria from industrial biotopes as *Ralstonia campinensis* sp. nov., *Ralstonia metallidurans* sp. nov. and *Ralstonia basilensis*. *Int J Syst Evol Microbiol* **51**: 1773–1782.

Juhnke S, Peitzsch N, Hubener N, Grosse C, Nies DH. 2002. New genes involved in chromate resistance in *Ralstonia metallidurans* strain CH34. *Arch Microbiol* **179**: 15–25.

Kimball R, Saier MH Jr. 2002. Voltage-gated H^+ channels associated with human phagocyte superoxide-generating NADPH oxidases: sequence comparisons, structural predictions, and phylogenetic analyses. *Mol Membr Biol* **19**: 137–147.

Krogh A, Larsson B, von Heijne G, Sonnhammer E. 2001. Predicting transmembrane protein topology with a hidden Markov model. Application to complete genomes. *J Mol Biol* **305**: 567–580.

Legatzki A, Grass G, Anton A, Rensing C, Nies DH. 2003a. Interplay of the Czc system and two P-type ATPases in conferring metal resistance to *Ralstonia metallidurans*. *J Bacteriol* **185**: 4354–4361.

Legatzki A, Franke S, Lucke S, *et al*. 2003b. First step towards a quantitative model describing Czc-mediated heavy metal resistance in *Ralstonia metallidurans*. *Biodegradation* **14**: 153–168.

Mergeay M, Houba C, Gerits J. 1978. Extrachromosomal inheritance controlling resistance to cadmium, cobalt and zinc ions: evidence from curing in a *Pseudomonas*. *Arch Int Physiol Biochem* **86**: 440–441.

Mergeay M, Monchy S, Vallaeys T, *et al*. 2003. *Ralstonia metallidurans*, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. *FEMS Microbiol Rev* **27**: 385–410.

Mergeay M, Nies D, Schlegel HG, *et al*. 1985. *Alcaligenes eutrophus* CH34 is a facultative chemolithotroph with plasmid-bound resistance to heavy metals. *J Bacteriol* **162**: 328–334.

Mouz S, Merlin C, Springael D, Toussaint A. 1999. A GntR-like negative regulator of the biphenyl degradation genes of the transposon Tn4371. *Mol Gen Genet* **262**: 790–799.

Munkelt D, Grass G, Nies DH. 2004. The chromosomally encoded CDF proteins DmeF and FieF are transporters of broad specificity and DmeF is necessary for cobalt detoxification through the CzcCBA efflux system. *J Bacteriol* **186**: 8036–8043.

Nies DH. 2003. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol Rev* **27**: 313–339.

Nottebrock D, Meyer U, Kramer R, Morbach S. 2003. Molecular and biochemical characterization of mechanosensitive channels in *Corynebacterium glutamicum*. *FEMS Microbiol Lett* **218**: 305–309.

Paulsen IT, Nguyen L, Sliwinski MK, Rabus R, Saier MH Jr. 2000. Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J Mol Biol* **301**: 75–100.

Pivetti CD, Yen M-R, Miller S, *et al*. 2003. Two families of prokaryotic mechanosensitive channel proteins. *Microbiol Mol Biol Rev* **67**: 66–85.

Roux M, Sarret G, Pignot-Paintrand I, Fontecave M, Coves J. 2001. Mobilization of selenite by *Ralstonia metallidurans* CH34. *Appl Environ Microbiol* **67**: 769–773.

Saier MH Jr. 2000. A functional–phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev* **64**: 354–411.

Saier MH Jr. 2003. Answering fundamental questions in biology with bioinformatics. *ASM News* **69**: 175–181.

Salanoubat M, Genin S, Artiguenave F, *et al*. 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415**: 497–502.

Copyright © 2005 John Wiley & Sons, Ltd.

Silver S. 2003. Bacterial silver resistance: molecular biology and uses and misuses of silver compounds. *FEMS Microbiol Rev* **27**: 341–353.

Springael D, Ryngaert A, Merlin C, Toussaint A, Mergeay M. 2001. Occurrence of Tn4371-related mobile elements and sequences in (chloro)biphenyl-degrading bacteria. *Appl Environ Microbiol* **67**: 42–50.

Taghavi S, Mergeay M, van der Lelie D. 1997. Genetic and physical maps of the *Alcaligenes eutrophus* CH34 megaplasmid pMOL28 and its derivative pMOL50 obtained after temperature-induced mutagenesis and mortality. *Plasmid* **37**: 22–34.

Toussaint A, Merlin C, Monchy S, *et al*. 2003. The biphenyl- and 4-chlorobiphenyl-catabolic transposon Tn4371, a member of a new family of genomic islands related to IncP and Ti plasmids. *Appl Environ Microbiol* **69**: 4837–4845.

Tran CV, Yang NM, Saier MH Jr. 2003. TC-DB: An architecture for membrane transport protein analysis. Proceedings of the 2nd International IEEE Computer Society Computational Systems Bioinformatic Conference (CSB 2003), Stanford, CA; 658.

Vandamme P, Coenye T. 2004. Taxonomy of the genus *Cupriavidus*: a tale of lost and found. *Int J Syst Evol Microbiol* **54**: 2285–2289.

Vaneechoutte M, Kämpfer P, De Baere T, *et al*. 2004. *Wautersia* gen. nov., a novel genus accommodating the phylogenetic lineage including *Ralstonia eutropha* and related species, and proposal of *Ralstonia* [*Pseudomonas*] *syzygii* (Roberts, *et al*., 1990) comb. nov. *Int J Syst Evol Microbiol* **54**: 317–327.

Winnen B, Hvorup RN, Saier MH Jr. 2003. The tripartite tricarboxylate transporter (TTT) family. *Res Microbiol* **154**: 457–465.

Wolfram L, Eitinger T, Friedrich B. 1991. Construction and properties of a triprotein containing the high-affinity nickel transporter of *Alcaligenes eutrophus*. *FEBS Lett* **283**: 109–112.

Wolfram L, Friedrich B, Eitinger T. 1995. The *Alcaligenes eutrophus* protein HoxN mediates nickel transport in *Escherichia coli*. *J Bacteriol* **177**: 1840–1843.

Yen M-R, Tseng YH, Nguyen EH, Wu LF, Saier MH Jr. 2002. Sequence and phylogenetic analyses of the twin arginine targeting (Tat) protein export system. *Arch Microbiol* **177**: 441–450.

Zhai Y, Saier MH Jr. 2001. A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence. *J Mol Microbiol Biotechnol* **3**: 501–502.

Zhang Z, Aboulwafa M, Smith M, Saier MH Jr. 2003. The ascorbate transporter of *Escherichia coli*. *J Bacteriol* **185**: 2243–2250.

Copyright © 2005 John Wiley & Sons, Ltd.

Antonie van Leeuwenhoek
DOI 10.1007/s10482-008-9284-5

REVIEW PAPER

Cupriavidus metallidurans: evolution of a metal-resistant bacterium

Torsten von Rozycki · Dietrich H. Nies

Received: 1 September 2008 / Accepted: 17 September 2008
© Springer Science+Business Media B.V. 2008

Abstract *Cupriavidus metallidurans* CH34 has gained increasing interest as a model organism for heavy metal detoxification and for biotechnological purposes. Resistance of this bacterium to transition metal cations is predominantly based on metal resistance determinants that contain genes for RND (resistance, nodulation, and cell division protein family) proteins. These are part of transenvelope protein complexes, which seem to detoxify the periplasm by export of toxic metal cations from the periplasm to the outside. Strain CH34 contains 12 predicted RND proteins belonging to a protein family of heavy metal exporters. Together with many efflux systems that detoxify the cytoplasm, regulators and possible metal-binding proteins, RND proteins mediate an efficient defense against transition metal cations. To shed some light into the origin of genes encoding these proteins, the genomes of *C. metallidurans* CH34 and six related proteobacteria were investigated for occurrence of orthologous and paralogous proteins involved in metal resistance. Strain CH34 was not much different from the other six bacteria when the total content of transport proteins was compared but CH34 had significantly more putative transition metal transport systems than the other bacteria. The genes for these systems are located on its chromosome 2 but especially on plasmids pMOL28 and pMOL30. Cobalt–nickel and chromate resistance determinants located on plasmid pMOL28 evolved by gene duplication and horizontal gene transfer events, leading to a better adaptation of strain CH34 to serpentine-like soils. The *czc* cobalt–zinc–cadmium resistance determinant, located on plasmid pMOL30 in addition copper, lead and mercury resistance determinants, arose by duplication of a *czcICAB* core determinant on chromosome 2, plus addition of the *czcN* gene upstream and the genes *czcD, czcRS, czcE* downstream of *czcICBA*. *C. metallidurans* apparently evolved metal resistance by horizontal acquisition and by duplication of genes for transition metal efflux, mostly on the two plasmids, and decreased the number of uptake systems for those metals.

Keywords *Cupriavidus metallidurans* · *Ralstonia* · *Wautersia* · *Alcaligenes* · CH34 · H16 · JMP134 · Transport proteins · Metal resistance · RND · CDF · CHR · P-type ATPases · Evolution

This paper is dedicated to Dr. Max Mergeay for a long time of cooperation, constructive competition and friendship.

T. von Rozycki · D. H. Nies (✉)
Institute for Biology/Microbiology, Molecular Microbiology, Martin-Luther University Halle-Wittenberg, Kurt-Mothes-Str. 3, 06099 Halle, Germany
e-mail: d.nies@mikrobiologie.uni-halle.de

Introduction

What makes a metal-resistant bacterium metal resistant? First answers were given (Mergeay et al. 2003;

Nies 2003; von Rozycki et al. 2005) when the genome of *Cupriavidus metallidurans* strain CH34 was analyzed: more genes and better genes involved in metal homeostasis (Nies et al. 2006). In this initial analysis, all *C. metallidurans* proteins related to each other were simply counted as "paralogs" but would better addressed as members of the same protein family. A more detailed analysis of paralogs plus examination of the expression of their genes in combination with mutagenesis and biochemical work (Große et al. 2007; Legatzki et al. 2003a; Munkelt et al. 2004; Nies 2004; Nies et al. 2006) now provides a better picture, which is presented here.

Cupriavidus metallidurans strain CH34

Cupriavidus (*Wautersia, Ralstonia, Alcaligenes*) *metallidurans* strain CH34 (Vandamme and Coenye 2004) is a model system for a mesophilic heavy metal resistant bacterium (Mergeay et al. 1978, 1985; Nies 1999, 2003; Nies and Silver 1995). Related strains are widespread around the world, present for instance in Belgium, Germany, New Zealand and Japan (Diels and Mergeay 1990; Dressler et al. 1991; Goris et al. 2001; Kunito et al. 1996). First observations indicated that metal resistance of *C. metallidurans* is linked to two plasmids residing in this bacterium, pMOL28 and pMOL30 (Mergeay et al. 1978, 1985), but later the essential contributions of chromosomal genes was recognized (Legatzki et al. 2003a; Munkelt et al. 2004). Two component regulatory proteins (Nies and Brown 1998; van der Lelie et al. 1997) and extracytoplasmic function (ECF) sigma factors are required for the fine-tuning between metal resistance and other physiological functions (Große et al. 2007; Nies 2004). Resistance is based on metal efflux at a first glance (Nies et al. 1989b, 1990; Nies and Silver 1989). However, the determinants encoding those metal efflux systems do not function properly when transferred into other proteobacteria (Nies et al. 1987), although they were functionally expressed (Nies 1995). This indicates an essential contribution of the "host background" for metal resistance. Moreover, some other proteobacteria reach even higher levels of resistance to some metals without using metal resistance systems as complicated as those from *C. metallidurans* (Marrero et al. 2007). So, what defines *C. metallidurans* metal resistance in toto?

Bacteria related to *C. metallidurans*

Genomic sequences are available for *C. metallidurans* and some related bacteria: *Cupriavidus eutrophus* = *Cupriavidus necator* = *Ralstonia eutropha* strain H16, *C. eutrophus* strain JMP134, *Ralstonia solanacearum* and several *Burkholderia* species. This should allow defining of a core inventory of genes present in all or most free-living members of this group, and identification of genes specifically required for the different life styles as a facultative hydrogen-oxidizing bacterium (*C. eutrophus* H16), plant-pathogen (*R. solanacearum*), xenobiotica degrader (*C. eutrophus* JMP134) or metal resistant bacterium (*C. metallidurans*). There is confusion about the names for some of these bacteria. In this publication, we will use the currently valid species names *"Cupriavidus metallidurans"* (although some of the gene names in the databases are still starting with "Rme" for "*Ralstonia metallidurans*") and "*Cupriavidus eutrophus*" (Tindall 2008) [although the genome sequence of this bacterium was published as "*Ralstonia eutropha*" (Pohlmann et al. 2006)].

Seven bacteria will be compared in this publication (Fig. 1), all belonging to the β-proteobacterial family *Burkholderiaceae* of the order *Burkholderiales*. Most closely related are the two *C. eutrophus* strains, the hydrogen-oxidizing strain H16 (Pohlmann et al. 2006) and the xenobiotica degrader strain JMP134. Both are able to degrade recalcant aromatic compounds and to produce polyhydroxyalkalonates (Don and Pemberton 1981; Johnson and Stanier 1971; Pohlmann et al. 2006). *C. metallidurans* is closely related to these two bacteria, also able to degrade xenobiotics and to live facultatively as hydrogen-oxidizing bacterium (Mergeay et al. 1985). In strain CH34 an unusual large percentage of its transport protein repertoire is directed towards metal ion homeostasis (von Rozycki et al. 2005) fitting to its role as a model bacterium for heavy metal resistance. *C. metallidurans* CH34 has also gained increasing interest for the development of biosensors and for bioremediation purposes (Diels et al. 1995; Mergeay et al. 2003; Saier 2005).

In contrast to the *Cupriavidus* strains, *R. solanacearum* is a plant pathogen and two strains have been sequenced. In *R. solanacearum* strain GMI1000 (Salanoubat et al. 2002) a high genomic flexibility seems to be one aspect for its versatility, saprophytic fitness and pathogenic flexibility (Grover et al. 2006).

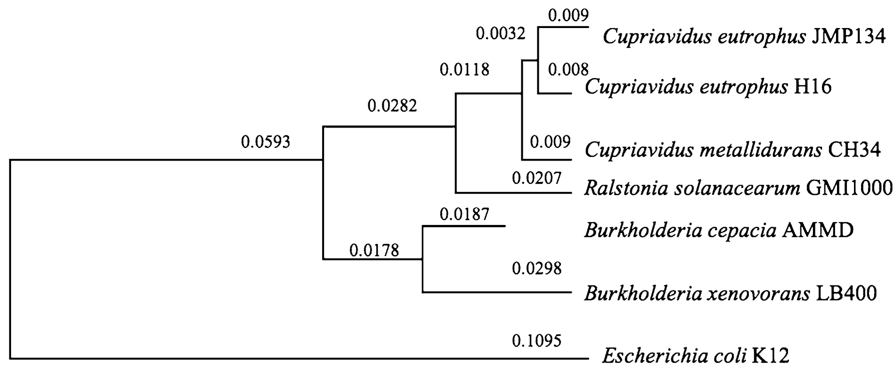


Fig. 1 Relationship of six of the seven investigated beta proteobacteria. The Tree Builder (RDPII) program was used to create a phylogenetic tree derived from 16S rRNA sequences (Neighbor-Joining with bootstrapping). It shows the evolutionary relationship between six of the seven investigated proteobacteria. *R. solanacearum* strain UW551 is not shown but should be close to strain GMI1000. To find the root for this tree the 16S rRNA sequence of *E. coli* K12 was used as the taxonomic out-group

As a highly heterogeneous species it is a devastating plant pathogen for crops as tomato, tobacco and banana (Guidot et al. 2007). The other strain, *R. solanacearum* strain UW551 (Race 3 Biovar 2; R3B2) is a cold-tolerant organism and is capable of infecting crops in temperate climates. Its genomic sequence (NZ_AAKL0000000) has not yet been fully annotated and published. So, it can be included in some parts of this analysis, but not in all.

The two *Burkholderia* strains are taxonomically distant from the *Ralstonia–Cupriavidus* cluster. *Burkholderia xenovorans* strain LB400 contains the second largest bacterial genome having been sequenced so far (9.73 Mb). It shows a high degree of genomic plasticity and 20% recently acquired genes (Chain et al. 2006), enabling *B. xenovorans* strain LB400 to take up and utilize complex organic compounds from soil including single-carbon (C1) groups, isoflavonoids, diterpenoids, and sulfonates (Chain et al. 2006). This organism is also an aerobic biphenyl (PCB) degrader and able to consume about twenty different PCB congeners of this problematic environmental pollutant by at least two independent benzoate pathways (Denef et al. 2004).

As in *B. xenovorans* strain LB400 broad metabolic capabilities can also be observed in *Burkholderia cepacia* strain AMMD. Its unusual large genome (twice that of *E. coli*) is distributed to four circular replicons (NC_008390, NC_008391, NC_008392, NC_008385) and contains numerous insertion sequences. The complex of five strains of *B. cepacia* is phenotypical very similar (Vandamme et al. 2003). One strain (Genomovar III) is an opportunistic human pathogen found among cystic fibrosis patients, involved in about 90% of cystic fibrosis-related deaths. On the other hand, *B. cepacia* also occurs as a plant-growth promoting rhizosphere bacterium with the ability to degrade polychlorinated aromatic compounds (Vandamme et al. 2003).

How the genomes were compared

Transport protein classification

Since the genomic sequences for a growing number of organisms have been released within the past decade, many specific tools and databases were developed to make use of this information. Computational phylogenetic and functional investigations mostly rely on sequence similarity (evidence of orthology) of proteins of known function (Tatusov et al. 1997). Proteins can be assigned to protein families on the basis of common structural, evolutionary, and functional aspects. This information can be derived from the protein sequences. The classification of transport protein families is the objective of databases as ABCdb (http://www-abcdb.biotoul.fr/), TransporterDB (http://www.membranetransport.org/), and TCDB (http://www.tcdb.org/). An international standard for categorizing corresponding proteins is the transporter protein classification system (TC; Saier et al. 2006; Zhou et al.

2003). This database is basically organized around the mode of transport, bioenergetics, molecular phylogeny, and substrate specificity of transport proteins (Saier et al. 2006). Those features have proven to be conserved among transport protein families (Saier 2000). Since transport proteins can be involved in nutrient uptake, sensing of environmental conditions, efflux of waste products, and other toxic compounds as drugs and heavy metal ions, the knowledge of the transporter repertoire gives important insights into the lifestyle and physiology of organisms (Saier 2000).

This report compares the orchestration of transport and regulatory proteins of *C. metallidurans* strain CH34 with that of six related beta proteobacteria to elucidate which proteins are unique for *C. metallidurans* and are defining its specific ecological niche. We will examine the differences between the genomes without a priori assumptions but will focus more and more on proteins involved in transition metal transport. By trying to estimate when gene duplication events leading to paralogs have occurred, we will shed some light on the evolution of *C. metallidurans* CH34.

Genomes

The latest releases of the genomic protein sequences and 16S rRNA for *C. metallidurans* strain CH34, *C. eutrophus* strain JMP134, *B. cepacia* strain AMMD, and *B. xenovorans* strain LB400 were obtained from JGI (DOE Joint Genome Institute, http://www.jgi.doe.gov/). Corresponding sequences of *R. solanacearum* strain GMI1000, *R. solanacearum* strain UW551, *C. eutrophus* strain H16, a draft version of *R. solanacearum* strain UW551, and *E. coli* were obtained from NCBI (National Center for Bioinformatics, http://www.ncbi.nlm.nih.gov/).

Treatment of the transport proteomes

The predicted proteins were collected for the seven bacteria using the data bases mentioned above and blasted against the protein sequences of representatives of TCDB database (Busch and Saier 2002). Proteins were considered homologs to members of the TC-database when the expectation value by chance was below e^{-10} (Saier et al. 2006). In these cases, the membrane topology, protein size, and functional assignments of the queries and the respective TC homologs were compared for additional verification. Membrane topology was determined by charged bias analysis based on TMHMM (http://www.cbs.dtu.dk/services/TMHMM/). SWISSPROT and TrEMBL databases (both at http://www.ebi.ac.uk/swissprot/) were used for further functional analysis. The transportDB (http://www.membranetransport.org) was also used.

Phylogenetic relatedness of the characterized bacteria

16S rRNA sequences were used to create a phylogenetic tree of the seven characterized bacteria (Fig. 1) using the neighbor joining method and Tree Builder (RDPII, http://rdp.cme.msu.edu/treebuilder/treeing.spr). *E. coli* strain K12 was used as the out-group root.

Comparison of the proteome of the seven investigated bacteria

The proteins of all investigated bacteria were blasted (Altschul et al. 1997) vice versa. To identify orthologs, reverse BLAST was carried out and the proteins were compared to those of *E. coli* strain K12 as taxonomic out-group (Mushegian et al. 1998). In cluster comparisons of related sequences from the seven bacterial strains sequence similarities and the amino acid identities were used to distinguish candidate orthologs from closely related paralogs. To investigate the evolution of paralogs the organisms were step wisely removed from consideration in an iteration procedure, starting with the strain showing the lowest degree of phylogenetic relatedness and ending with the most related strain (Koonin et al. 1995; Raymond et al. 2002). Regulatory proteins were assigned according to the information and sequences from recent publications and corresponding supplementary material (Große et al. 2007; Nies et al. 2006; von Rozycki et al. 2005). Keyword searches in SWISSPROT and NCBI were additionally performed. Moreover, two-component regulatory systems and sigma factors serving as possible regulators of expression of RND-driven efflux systems were identified by the proximity of their genes to those coding for the efflux system.

Transport proteome of strain CH34 in general

Transport proteins in the seven compared bacteria

The predicted proteomes of *C. metallidurans* strain CH34 and the six related bacteria was blasted against protein sequences of representatives for transport protein families at TCDB (Busch and Saier 2002). This procedure sorted putative transporter proteins into nine major transport protein classes (Table 1), which stand for different transport modes and energy coupling mechanisms (Saier 2000; Saier et al. 2006). The three *Cupriavidus* strains CH34, H16, JMP134, and *B. cepacia* contained about 1,000 putative transport proteins, *B. xenovorans* 50% more, and the two phytopathogenic *R. solanacearum* strains about 30% less. Distribution of proteins into the nine major classes was similar. Most transport proteins were primary active transporters (TC#3, 45 ± 5%), followed by electrochemical potential-driven transporters (TC#2, 27 ± 4%) and channels/pores (TC#1, 11.5 ± 2.5%) equaling incompletely characterized transport proteins (TC#9, 11 ± 2%). Thus, the seven bacteria were not much different when the distribution of transport proteins into the nine major classes was considered but the total number of transport proteins was more than two-fold higher in *B. xenovorans* compared to *R. solanacearum* strain UW551 (chromosomal DNA, Table 1).

However, the genome size of *B. xenovorans* was also much larger (9.73 Mb; Chain et al. 2006) than that of the other six bacteria. When the number of transport proteins was plotted against the genome sizes, this number increased in general with the genome sizes (Fig. 2). Two different groups of points were roughly visible. In one, mostly composed of bacteria with genome sizes larger than 4 Mb, the number of transport proteins increased in a linear fashion with a slope of about 167 transport proteins per Mb (Fig. 2). In the second type, mostly composed out of bacteria with smaller genome sizes, the slope was only about 125 transport proteins per Mb. All seven characterized bacteria similarly displayed a transporter:genome size ratio that was typical for bacteria with a genome larger than 4 Mb.

The number of transport proteins encoded by the various replicons in six of the characterized bacteria (all except *R. solanacearum* strain UW551) was also compared (data not shown). As a mean value, 151 ± 49 transport proteins were encoded per Mb, which matches the result obtained with total genomes. Plasmid pMOL28 from *C. metallidurans* contained 65 transporters/Mb and thus an unusually low content. The transport protein content encoded

Table 1 Predicted number of transport proteins in the seven characterized bacteria[a]

TC#	Transporter class	*Cme* CH34	*Ceu* H16	*Ceu* JMP134	*Bxe* LB400	*Bce* AMMD	*Rso* UW551	*Rso* GMI1000
1	Channels/Pores	116	98	113	209	148	95	105
2	Electrochemical potential-driven transporters	308	317	330	350	327	169	188
3	Primary active transporters	396	457	426	759	485	308	330
4	PTS-group translocators	1	2	1	2	3	3	4
5	Transport electron carriers	19	21	19	17	14	13	15
8	Accessory factors involved in transport	31	28	31	48	34	17	21
9	Incompletely characterized transport systems	131	130	135	145	118	71	77
1–9	Total number	1,002	1,053	1,055	1,530	1,129	676	740
	Total number of predicted proteins	6,351	6,573	6,446	8,635	6,545	–[b]	5,051

[a] The predicted proteomes of *C. metallidurans* (*Cme* CH34), *C. eutrophus* strains H16 (*Ceu* H16) and JMP134 (*Ceu* JMP134), *B. xenovorans* strain LB400 (*Bxe* LB400), *B. cepacia* strain AMMD (*Bce* AMMD), *R. solanacearum* strains UW551 (*Rso* UW551), and GMI1000 (*Rso* GMI1000) were BLASTed against the protein sequences of the representatives for transport protein families at TCDB (Busch and Saier 2002) to obtain the number of transport proteins and group them into the nine major classes of transport proteins. Only contigs representing chromosomal DNA were used in case of *R. solanacearum* stain UW551

[b] The genome of this bacterium has not yet been fully annotated

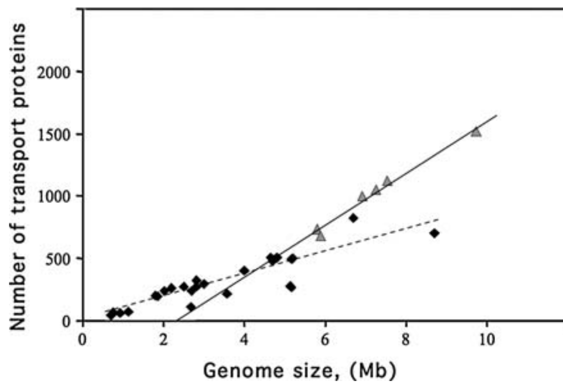


Fig. 2 Relationship between the number of transport proteins and the genome sizes in sequenced proteobacteria. The putative number of transport proteins of sequenced proteobacteria in the TCDB (*diamonds*) and in the seven characterized bacteria (*triangles*) was plotted against the respective genome sizes. The two *C. eutrophus* strains are represented by one point only (middle one of the three triangles in line). The dashed and the solid lines indicate groups of bacteria with a low increase of the number of transport proteins per Mb (*dashed line*) or a higher increase per Mb (*solid line*)

by plasmid pJP4 from *C. eutrophus* strain JMP134 was 233/Mb and that of the plasmid of *B. cepacia* 300/Mb, both rather high values. However, these three plasmids were also the smallest of the replicons compared, ranging in size from 0.04 to 0.17 Mb, and these numbers could be stochastical fluctuations. Again but with the exception of the mentioned three plasmids, the number of transport proteins per Mb was similar in the seven analyzed bacteria and this number was also similar to that of other proteobacteria. Metal resistance of *C. metallidurans* was therefore not solely the result of a higher total number of transport proteins. Therefore, the differences in representatives of various transport protein families was exploited.

Number of transport proteins in representative protein families or transport family classes

There was a very low number (1–4) of phosphotransferase-driven group translocators (PTS systems, TC#4, 0.1–0.4%, Table 1) in the seven bacteria, the most in the phytopathogenic *R. solanacearum* strains. Group translocators are mainly transport proteins for carbohydrates (Lengeler 1990; Reizer et al. 1991; Saier et al. 1985). Thus, all seven characterized bacteria have a low capacity for import of sugars by PTS systems.

For a further characterization of the sugar transport capacity, the number of proteins belonging to families of other sugar-specific transporters (CUT1 and CUT2 families) was counted. This number (data not shown) varied widely from 18 (*C. metallidurans, C. eutrophus* strain H16) to 81 (*B. xenovorans*). It was compared to that of putative amino acid transporters (PAAT, HAAT, PepT families), which varied between 91 (*R. solanacearum* strain GMI1000) and 257 (*B. xenovorans*). Thus, *B. xenovorans* with its large genome has the highest number of putative sugar and amino acid transport systems, which are in a ratio of 1:3. The two *R. solanacearum* strains with their smaller genomes have fewer transporters of these families, but the same ratio of 1:3. *B. cepacia* has a higher relative number of sugar transporters (1:2) but the three *Cupriavidus* strains a much lower relative number of sugar transport systems, decreasing from 1:6 (*C. metallidurans*, *C. eutrophus* strain JMP134) to 1:9 (*C. eutrophus* strain H16). Thus, after separation of the genera *Cupriavidus* and *Ralstonia* (Fig. 1), the *Cupriavidus* strains might have shifted their nutrition from sugars to amino acids, and strain H16 made the strongest evolutionary move into this direction.

Paralogs

To learn more about the origin of the additional, metal resistance associated genes in the genome of *C. metallidurans*, paralogs and orthologs of the predicted gene products were assigned by BLAST comparisons of the predicted proteomes of six of the seven investigated bacteria (Strain UW551 with the unfinished genome was omitted). Paralogs arose in the past by a gene duplication event within an organism and exhibit a higher sequence similarity to other proteins within the same proteome than to proteins of another organism (Tatusov et al. 1997). In general, the frequency of paralogs in the genomes of the compared bacteria was about 9% (Table 2). *B. xenovorans* with its large genome contained 25% of paralogs. This indicated that most of its genome increase might have been evolved by gene duplication events during speciation of this bacterium.

The frequencies of paralogs of transport proteins were also within the 9% range for most bacterial strains (Table 2). It was only considerable lower for *B. cepacia* (5%). This bacterium evolved its genome

Table 2 Number of paralogs of proteins and transport proteins in the compared bacteria[a]

Bacterial strain	Replicons	All proteins		Transport proteins	
		Total number	Paralogs (%)	Total number	Paralogs (%)
C. metallidurans CH34	All	6,351	643 (10%)	1,002	93 (9%)
	plasmids	416	141 (34%)	40	22 (52%)
C. eutrophus H16	All	6,573	564 (8%)	1,053	84 (8%)
	plasmid	416	130 (31%)	45	13 (28%)
C. eutrophus JMP134	All	6,446	418 (6%)	1,055	71 (6%)
	plasmids	600	124 (20%)	104	15 (14%)
R. solanacearum GMI1000	All	5,051	564 (9%)	740	51 (6%)
	plasmid	1,483	190 (12%)	232	6 (2%)
B. cepacia AMMD	All	6,545	741 (11%)	1,129	52 (4%)
	plasmid	44	1 (2%)	12	0
B. xenovorans LB400	All	8,635	2,137 (25%)	1,530	129 (8%)

[a] All predicted proteins from the respective bacteria were BLASTP-compared to all other proteins of all strains with a cut-off value of a by chance expectation $<e^{-10}$. Paralogs were more closely related to another protein in the same genome than to any other protein in one of the other bacterial strains

mainly by duplication of genes determining soluble proteins. A low degree of conservation to related bacteria has already been reported for this organism (Chain et al. 2006).

A very high percentage of paralogous protein-encoding genes (soluble and membrane-bound) was located on plasmids in strain CH34 (34%), H16 (31%) and JMP134 (21%). Concerning the plasmid in *B. cepacia*, the contrary was the case (2%). This plasmid may thus have been acquired by conjugation. When only plasmid-encoded paralogs of transport proteins were considered, half of those in strain CH34 were paralogs, but only 29% in strain H16 and 14% in strain JMP134. Therefore, an important process in the evolution of *C. metallidurans* strain CH34 could have been to duplicate genes for transport proteins, to locate them to plasmids pMOL28 and pMOL30, and maybe alter the substrate range and expression pattern of their products. Paralogs harbored by plasmids were probably also important in the evolution of strains H16 and JMP134.

Orthologs

A protein A in proteome 1 can be considered to be a candidate ortholog of a protein B in proteome 2 if (i) protein B shows highest sequence identity to protein A of all proteins in proteome 2 and (ii), vice versa, protein A shows highest sequence identity to protein B of all proteins in proteome 1 (Koonin et al. 1996; Mushegian et al. 1998). Additionally, proteins A and B should be more closely related to each other than to any other protein in the proteome of a taxonomic out-group (Tatusov et al. 1997).

Using *E. coli* strain K12 as taxonomic out-group, the numbers of orthologs in the predicted proteomes of *C. metallidurans* and the closely related strains *C. eutrophus* H16 and JMP134 were identified (Table 3). The total genomes of these three bacterial strains contain about 6,400 genes and 76–82% had orthologs in at least one of the other six investigated β-proteobacteria. The number of genes having orthologs in all other investigated β-proteobacteria was between 15 and 26%. This number was very low in strain H16.

Comparing the replicons, chromosome 1 in all three *Cupriavidus* strains was the most conserved replicon with 84–89% genes having orthologs in at least one of the other investigated bacteria. Thus, chromosome 1 is the "housekeeping" chromosome of these bacteria. Genes on the three chromosomes 2 had less frequently orthologs in at least one of the other six bacteria and only a small frequency of genes with orthologs in all other six. On the other hand, the number of genes having orthologs exclusively in the *Cupriavidus* strains was higher when chromosome 2 was compared to chromosome 1 (Table 3). Thus, if chromosome 1 is a general "housekeeping" chromosome, chromosome 2 is more a "genus-specific" chromosome.

Table 3 Orthologs and paralogs in the genomes and replicons of the investigated *Cupriavidus/Ralstonia* strains

Replicon	Number of proteins				
	Total number of genes	With orthologs in the analyzed bacteria			With paralogs
		In $\geq$1	In all other six	Only in H16/JMP/CH34	
C. metallidurans CH34					
Chromosome 1	3,593	3,006 (84%)	1,468 (41%)	281 (8%)	247 (7%)
Chromosome 2	2,342	1,709 (73%)	178 (8%)	310 (13%)	255 (11%)
Plasmid pMOL30	251	43 (17%)	0	4 (2%)	103 (41%)
Plasmid pMOL28	165	76 (46%)	0	1 (1%)	38 (23%)
Total genome	6,351	4,834 (76%)	1,646 (26%)	596 (9%)	643 (10%)
C. eutrophus JMP134					
Chromosome 1	3,439	3,069 (89%)	859 (25%)	276 (8%)	105 (3%)
Chromosome 2	2,407	1,913 (79%)	130 (5%)	323 (13%)	189 (8%)
Megaplasmid	512	248 (48%)	1 (0%)	5 (1%)	107 (21%)
Plasmid pJP4	88	61 (69%)	0	0	0
Total genome	6,446	5,291 (82%)	990 (15%)	604 (9%)	418 (6%)
C. eutrophus H16					
Chromosome 1	3,623	3,127 (86%)	1,424 (39%)	277 (8%)	224 (6%)
Chromosome 2	2,534	2,016 (80%)	168 (7%)	317 (13%)	210 (8%)
Plasmid pHG1	416	199 (48%)	0	26 (6%)	130 (31%)
Total genome	6,573	5,342 (81%)	1,592 (24%)	620 (9%)	564 (9%)

The megaplasmid from strain JMP134, plasmids pHG1 from H16 and pMOL28 from CH34 showed similar features (Table 3). Nearly half of their genes had orthologs in at least one of the other six bacteria but none (1 exception) had orthologs in all of them. On the other hand, the genes harbored by these plasmids had 21–31% paralogs that may have "recently" evolved by gene duplication events. Thus, these three plasmids may carry a part of the common gene pool of the *Burkholderia/Cupriavidus/Ralstonia* cluster that is nevertheless specific for the individual bacterium.

Plasmids pMOL30 and pJP4 were rather unique. Plasmid pMOL30-encoded genes had only 17% orthologs in at least one of the other bacteria, none with orthologs in all, but 41% with paralogs. Thus, pMOL30 may carry a collection of genes acquired rather "recently" by horizontal gene transfer and gene duplications. About 2/3rd of the genes on plasmid pJP4 had orthologs in at least one of the other six bacteria, but none had orthologs in all of them or in the *Cupriavidus* strains only, and none had paralogs (Table 3). This makes pJP4 a plasmid that might have specifically transferred genes from *Burkholderia* strains or *R. solanacearum* into strain JMP134 by horizontal gene transfer.

Orthologs and paralogs of transport proteins

Zooming specifically into transport proteins, about 2/3rd of the predicted transport proteins from the three *Cupriavidus* strains CH34, H16, and JMP134 had orthologs in the other two strains from this clade (data not shown). This defines a core transport proteome of about 700 proteins common to *Cupriavidus* strains. Half of the transport proteins in these three strains had also orthologs in the two *Burkholderia* strains LB400 and AMMD. So, about 500 proteins were common to all five strains. This number was lower (about 450, in one case only 316) when the *R. solanacearum* strains with their smaller genomes were also considered, which may be a result of the genome reduction of these plant pathogenic strains. The ratio of transport protein paralogs encoded by the genomes of the three *Cupriavidus* strains was 7% (JMP134), 8% (H16), and 9% (CH34). Thus, the transport proteome of the three *Cupriavidus* strains has a size of slightly more than 1,000 proteins

(Table 1), about 700 of them belonging to a common or core transport proteome of the three *Cupriavidus* strains, about 500 of them may be common to all seven investigated strains, and less than 10% were paralogs.

The core transport proteome

The predicted transport proteins in the core transport proteome of *C. metallidurans* were assigned to the families of the transport protein database and the number of orthologs in the other six bacteria were counted (Table 4). In 20 protein families the number of orthologs was identical in all seven bacteria. The 43 proteins in these 20 families are thus defining the inner core transport proteome. Some members of this group are the constituents of the F_1F_0 ATP synthase, the twin arginine protein export system TAT, the common modules EI and HPr of the PTS phosphotransferase system, DNA uptake systems, and the uptake systems for inorganic phosphate, potassium, and lysine. The second group of core transport proteins has also orthologs in all six bacteria but strain CH34 contains up to twice the number of family members that is present in all six strains. Examples are the Mot/Exb complex components that energize active transport across the outer membrane, CDF proteins, an uptake system for ammonium, ABC transport systems, and MIT (metal inorganic transport) systems. The last group contains protein families with at least twice as many members in strain CH34 compared to the minimum number that is present in all six other compared bacteria. Examples are members of the RND, MFP and OMF protein families, P-type ATPases, proteins of the major facilitator superfamily MFS, and components of the type III (TTS) and the type IV (TFS) secretion system (Table 4). This shows again that the unusual nature of the *C. metallidurans* CH34 genome with the possession of many export proteins for metal cations.

Rare transport proteins

Two transport protein families (FPhe, MPA2) contained members present in *C. metallidurans* but not in the other six bacteria (Table 5). Similarly, *C. eutrophus* strains H16 and JMP134 contained a few proteins without orthologs in the other bacterial strains compared. Three protein families had members only in the *Cupriavidus* strains while a couple of protein families had none (Table 5).

Summary of this part

Strain CH34 was not much different from the other six bacteria when the distribution of transport proteins into the nine major classes, the transporter:genome size ratio, the number of PTS systems, and the percentage of paralogs in the total genome was compared. Differences became visible when the numbers of RND, MFP, OMF proteins, P-type ATPases, proteins of the major facilitator superfamily MFS, components of the type III (TTS) and the type IV (TFS) secretion system was considered. Moreover, many of the specific features of strain CH34 are encoded by the plasmids pMOL28 and pMOL30.

Transport proteins for transition metals

Metal homeostasis systems in *C. metallidurans* CH34

The number of proteins that might be involved in transition metal transport was further analyzed (CDF, MerTP, MFP, MIT, NiCoT, OMF, OMR, P-type ATPase, CHR, HME/RND, and ZIP protein families; names defined at http://www.tcdb.org). The number of putative transition metal transport systems was higher in CH34 (83 proteins) than in the other bacteria, which harbored between 44 (*R. solanacearum* strain UW551 chromosomal DNA) and 69 (*B. xenovorans*) proteins. When these numbers were related to the genome sizes, strain CH34 had 12 transition metal transport proteins/Mb while all the other six bacteria had 6–8/Mb. Thus, strain CH34 had significantly more putative transition metal transport systems than the other bacteria, which is consistent with the molecular data concerning metal resistance in this bacterium.

When the replicons of the three *Cupriavidus* strains were considered, the two chromosomes from strains JMP134 and H16, the megaplasmid from strain JMP134, and chromosome 1 from strain CH34 had average contents (7.1 ± 1/Mb) of putative transition metal transporters (between 6.6 and 8.9 proteins/Mb). The contents of the other replicons

Table 4 Core transport proteome of *C. metallidurans* CH34[a]

TC#	TC family	Description or possible substrates	# CH34	# All
Transporters with orthologs in all six other strains				
1.A.33	Hsp70	Ions, polypeptides	3	3
1.B.20	TPS	Proteins	2	2
1.B.39	OmpW	Methyl viologen and benzyl viologen	2	2
2.A.20	PiT	Inorganic phosphate	1	1
2.A.64	Tat	Folded proteins	2	2
2.A.72	KUP	Potassium uptake	1	1
2.A.75	LysE	LysE Basic amino acids	1	1
2.A.9	Oxa1	Proteins	1	1
2.C.1	OMR	TonB-driven uptake across outer membrane	3	3
3.A.11	DNA-T	Single-stranded DNA	1	1
3.A.12	S-DNA-T	DNA, DNA–protein complexes	2	2
3.A.2	F-ATPase	H^+, Na^+ (all subunits counted)	7	7
3.A.5	Sec	Proteins	7	7
3.D.3	QCR	Proton efflux	2	2
5.A.2	DsbB	Electrons	1	1
8.A.7	EI	Sugars, central component of PTS system	2	2
8.A.8	HPr	Sugars, central component of PTS system	1	1
9.B.24	TEGT	Hexose uptake	1	1
9.B.3	MPE	Lipid-linked murein precursors	2	2
9.B.53	UIT6	Unknown	1	1
Transporters possessing 33–67% orthologs in all other six bacteria				
1.A.30	Mot/Exb	Active transport across outer membrane	6	4
2.A.4	CDF	Divalent transition metal cations	3	2
3.D.1	NDH	Components of NADH oxidase	24	13
1.A.23	MscS	Mechanosensitive channel, Ions	6	3
1.B.18	OMA	Auxiliary protein for outer membrane	2	1
1.B.6	OOP	Ions, small molecules	4	2
2.A.49	Amt	Ammonium	2	1
2.A.66	MOP	Drugs, lipid-linked oligosaccharide precursors	4	2
3.A.15	MTB	Pilin/fimbrilin	16	8
5.A.1	DsbD	Electrons	2	1
8.A.3	MPA1	Complex polysaccharides	2	1
3.B.1	NaT-DC	Sodium	9	4
2.A.23	DAACS	C4-dicarboxylates, some amino acids	5	2
3.A.1	ABC	Inorganic and organic molecules	226	81
1.A.35	MIT	Transition metal cations	3	1
1.B.22	Secretin	Proteins	3	1
Transporters possessing 6–33% orthologs in all other six bacteria				
2.A.40	NCS2	Nucleobases, urate	3	1
3.D.2	PTH	Proton efflux	6	2
3.A.3	P-ATPase	Cations	13	4
3.D.4	COX	Proton efflux	24	7
2.A.21	SSS	Sugars, amino acids, vitamins, nucleosides, inositols	44	12

Table 4 continued

TC#	TC family	Description or possible substrates	# CH34	# All
2.A.1	MFS	Various small molecules	72	18
9.B.37	HCC	Ions?	4	1
2.A.7	DMT	DMT Multiple drugs and dyes	13	3
5.A.3	PMO	Proton translocation	13	3
3.A.6	IIISP	Proteins (TTS)	11	2
1.B.17	OMF	Part of transenvelope efflux complexes	17	3
2.A.3	APC	Amino acids, polyamines, choline	6	1
2.A.37	CPA2	Na^+/H^+ or K^+/H^+ antiport	6	1
2.A.6	RND	Heavy metal ions, multiple drugs, proteins	49	7
3.A.7	IVSP	Proteins, protein–DNA complexes	21	3
9.B.17	FAT	Fatty acyl CoA ligases	38	5
1.B.14	OMR	Iron–siderophore complexes, vitamin B12	17	2
2.A.53	SulP	Sulfate	10	1
1.B.1	GBP	Ions, small (Mr <1,000 Da) molecules	31	3
8.A.1	MFP	Heavy metal ions, multiple drugs	18	1

[a] Transport protein families with members found as orthologs of the *C. metallidurans* CH34 protein in the other investigated proteobacteria. "# CH34" gives the number of members in the genome of *C. metallidurans* and "# All" the number of proteins with orthologs in all investigated bacterial strains. The names of the protein families and the TC classification number was taken from http://www.tcdb.com

differed from this range, increasing from chromosome 2 of strain CH34 (12/Mb), plasmid pJP4 of strain JMP134 (22/Mb) and plasmid pMOL28 of strain CH34 (29–74/Mb) in plasmid pMOL30 of strain CH34. Plasmid pMOL30 harbors an average relative number of transport proteins (126/Mb) and plasmid pMOL28 a low number (65/Mb), but 17 of the 29 transport proteins of pMOL30 and 5 of the 11 transporters of pMOL28 are probably involved in transition metal transport. Therefore, the genes for the extraordinary high number of transition metal transport systems in strain CH34 are located on its chromosome 2 but especially on plasmids pMOL28 and pMOL30. Accordingly, deletion of these plasmids leads to a sharp decrease in metal resistance (Mergeay et al. 1985).

Orthologs and paralogs of transport proteins for transition metals

Transport proteins were considered that could be involved in transition metal transport. *C. metallidurans* (as a metal-resistant bacterium) showed in a high number of components for RND-driven efflux systems, which are composed of a large inner membrane protein RND (resistance, nodulation, cell division protein family), a membrane fusion protein (MFP) and an outer membrane factor (OMF) (Table 6), especially those involved in heavy metal efflux, HME-RND (Tseng et al. 1999). Enhanced numbers of P-type ATPases, CHR chromate transporters, and MerT mercury transporters were also evident. Identified paralogs belong also to these transport protein families. In contrast, the number of metal uptake systems metal inorganic transport (MIT), ZIT/IRP family (ZIP), and nickel cobalt transport family (NiCoT) was not higher in CH34 or even lower than in the other bacterial strains. Strain CH34 contains no NiCoT (Table 6) and no NRAMP (data not shown) transporter for uptake of divalent transition metal cations. Thus, *C. metallidurans* apparently evolved metal resistance by horizontal acquisition and by duplication of genes for transition metal efflux, mostly on the two plasmids, and decreased the number of uptake systems for those metals.

Two types of paralogs

A protein was only designated as "paralog" here when it was more closely related to a protein in the

Table 5 Transport protein families with unique members[a]

TC#	Transporter family	Family description	*Cme* CH34	*Ceu* H16	*Ceu* JMP134	*Bxe* LB400	*Bce* AMMD	*Rso* GMI	*Rso* UW
Unique to *C. metallidurans* strain CH34									
3.A.13	FPhE	Filamentous phage exporter	1	0	0	0	0	0	0
8.A.4	MPA2	Cytoplasmic membrane-periplasmic auxiliary-2	1	0	0	0	0	0	0
Unique to *C. eutrophus* strain H16									
1.B.20	TPS	Two-partner secretion	0	1	0	0	0	0	0
2.A.27	ESS	Glutamate:Na^+ symporter	0	1	0	0	0	0	0
2.A.33	NhaA	Na^+:H^+ antiporter	0	1	0	0	0	0	0
2.A.38	Trk	K^+ transporter	0	1	0	0	0	0	0
9.A.30	TerC	Tellurium ion resistance	0	1	0	0	0	0	0
Unique to *C. eutrophus* strain JMP134									
1.C.2	ICP	Channel-forming δ-endotoxin insecticidal crystal proteins	0	0	1	0	0	0	0
Only occurring in the *Cupriavidus* strains CH34, H16 and JMP134									
2.A.10	KDGT	2-Keto-3-deoxygluconate transporter	1	1	0	0	0	0	0
2.A.36	CPA1	Monovalent cation:proton antiporter-1	2	1	2	0	0	0	0
2.A.63	CPA3	Monovalent cation (K^+ or Na^+):proton antiporter-3	0	5	5	0	0	0	0
Not occurring in the *Cupriavidus* strains CH34, H16 and JMP134									
1.A.2	IRK-C	Inward rectifier K^+ channels	0	0	0	1	1	1	0
1.B.19	OprB	Glucose-selective OprB porins	0	0	0	1	1	1	1
1.B.25	OPr	Outer membrane porins	0	0	0	1	1	0	0
1.C.65	HrpF	Type III secretion system plant host cell membrane pore-forming	0	0	0	0	0	2	2
1.E.3	P2 Holin	P2 holins	0	0	0	0	1	0	0
2.A.17	POT	Proton-dependent Oligopeptide Transporter	0	0	0	1	1	0	0
2.A.25	AGCS	Alanine or glycine:cation symporter	0	0	0	0	1	0	0
2.A.28	BASS	Bile acid:Na^+ symporter	0	0	0	1	0	0	0
2.A.41	CNT	Concentrative nucleoside transporter	0	0	0	2	1	2	2
2.A.69	AEC	Auxin efflux carrier (AEC)	0	0	0	1	0	0	0
3.A.17	T7 Injec.	Phage T7 injectisomes	0	0	0	0	0	1	1
4.A.6	Man	PTS mannose–fructose–sorbose	0	0	0	0	0	1	0
4.B.1	PnuC	Nicotinamide ribonucleoside (NR) uptake permeases	0	0	0	0	1	0	0
9.A.33	Pyocin R2	Pyocin R2 phage P2 tail fiber proteins	0	0	0	0	0	1	0
9.B.25	YbbM	The YbbM family	0	0	0	1	1	1	1
9.B.26	PF27	The PF27 family	0	0	0	0	0	1	0
9.B.44	YiaAB	The YiaA-YiaB family	0	0	0	0	1	1	1

[a] Names and TC classification taken from http://www.tcdb.com

same proteome than to any other protein in the predicted proteomes of the other six bacteria plus the outside reference *E. coli* strain K12. Paralogous states under these conditions indicate gene duplication events during speciation and the respective proteins should be precisely designated "strain-specific paralogs".

Strain-specific paralogs for CH34 were five P-type ATPases, two of the four Hg^{2+}-transporting MerT-like proteins, 2 HME-RND, 3 MFP and 2 OMF

Table 6 Heavy metal-specific transport proteins in *C. metallidurans* CH34 with the corresponding number of paralogs and orthologs in related proteobacteria[a]

TC number	TC family	*Cme* CH34		Orthologs[d] #					
		Abs[b] #	Paral[c] #	*Ceu* JMP134	*Ceu* H16	*Rso* GMI1000	*Rso* UW551 Chrom	*Bxe* LB400	*Bce* AMMD
2.A.6.1	HME-RND[e]	12	2	5	6	6	7	1	2
8.A.1	MFP	18	3	10	7	7	5	8	7
1.B.17	OMF	17	2	12	9	8	6	9	9
3.A.3	P-type ATPases	13	5	7	8	6	6	7	6
2.A.4	CDF	3	0	2	2	2	2	2	2
2.A.51	CHR	4	0	2	2	1	1	2	2
9.A.2	MerT	4	2	0	0	0	0	3	0
2.A.5	ZIP	1	0	1	1	1	1	0	0
2.A.52	NiCoT	0	0	0	0	0	0	0	0
1.A.35	MIT	3	0	3	2	2	2	2	2

[a] The number of representatives of predicted proteins from *C. metallidurans* belonging to the indicated transport protein families (indicated by the TC numbers and family names) is shown

[b] The absolute number

[c] The number of these proteins that are paralogs in strain CH34

[d] Orthologs in the indicated bacterial strains

[e] Heavy-metal efflux protein family of the RND superfamily. Names and TC classification taken from http://www.tcdb.com

proteins (Tables 6, 7). The genes for these proteins should have been duplicated during speciation of strain CH34. In contrast, no paralog was found for the three CDF proteins, which represent the three subfamily's of this protein family (Nies 2003), three of the four CHR proteins, the four P-type export ATPases apparently for Cu^{+}/Ag^{+}, Zn^{2+}, Cd^{2+}, and Pb^{2+}-exporting P-type ATPase ZntA, and three predicted K^{+}-transporting P-type ATPases (Table 7).

However, two predicted proteins in the same proteome may not be identified as paralogs although they are closely related, because one or both might have at least one more much more closely related protein in the predicted proteome of another bacterium. To identify such constellations (Koonin et al. 1995; Raymond et al. 2002), the predicted sequences of metal efflux proteins from strain CH34 were compared to the total proteomes of the other six bacteria. After the initial comparison, which measured similarity by the expectation values and BLAST scores, the proteome of the bacterium with the most distant relationship to *C. metallidurans* (Fig. 1), *B. cepacia* strain AMMD, was removed, and the comparison was repeated using the predicted proteomes of the remaining five bacteria. In the next step, the proteome of *B. xenovorans* strain LB400 was removed, and the comparison was repeated. In the following steps, the proteomes of *R. solanacearum* strain UW551, *R. solanacearum* strain GMI1000 and finally of *C. eutrophus* strain JMP134 were removed, leading to a mere comparison of the respective gene product to the proteomes of strain H16 and *E. coli* K12.

This comparison identified another type of paralog in addition to the species-specific paralogs (Table 7). This type were proteins with paralogs appearing not in the first overall comparison but in one of the later steps, always after removal of *C. eutrophus* strain JMP134 from consideration (Table 7). This indicated the existence of at least one closely related protein in strain JMP134 but not in strain H16 or *E. coli* K12. Paralogs of this type were two HME-RND and two MFP proteins possibly involved in copper/silver detoxification, a chromate efflux protein and the remaining two of the four MerT-like proteins (Table 7). This type of paralog could be designated "old paralog". Several evolutionary scenarios including gene duplication events, gene losses and

Table 7 Paralogs of putative heavy metal-specific proteins in *C. metallidurans* strain CH34[a]

Predicted protein	Replicon	Transporter	TC number	vs. All 6	After removal from consideration				
					Bce AMMD	*Bxe* LB400	*Rso* UW551	*Rso* GMI1000	*Ceu* JMP134
CDF-Proteins									
Rmet5979	pMOL30	CzcD	2.A.4.1.1	–	–	–	–	–	–
Rmet0198	chrom.1	DmeF	2.A.4.1.2	–	–	–	–	–	–
Rmet3406	chrom.1	FieF	2.A.4.1.5	–	–	–	–	–	–
RND/MFP-Proteins									
Rmet6136	pMOL30	RND (SilA)	2.A.6.1.3	–	–	–	–	–	+
Rmet5033	chrom.2	RND (CusA)	2.A.6.1.3	–	–	–	–	–	+
Rmet6135	pMOL30	MFP (CusB)	2.A.6.1.4	–	–	–	–	–	+
Rmet5032	chrom.2	MFP (CusB)	2.A.6.1.4	–	–	–	–	–	+
CHR-Proteins									
Rmet6202	pMOL28	$ChrA_1$	2.A.51.1.1	–	–	–	–	–	–
Rmet3865	chrom.2	$ChrA_2$	2.A.51.1.1	–	–	–	–	–	+
Rmet4831	chrom.2	$ChrA_3$	2.A.51.1.2	–	–	–	–	–	–
Rmet2518	chrom.1	$ChrA_{4N}$/ $ChrA_{4C}$	2.A.51.1.1	–	–	–	–	–	–
P-type ATPases									
Rmet5396	chrom.2	Ca^{2+}-type	3.A.3.2.4	+	+	+	+	+	+
Rmet2211	chrom.1	Ca^{2+}-type	3.A.3.2.4	+	+	+	+	+	+
Rmet3524	chrom.1	Cu^{+}/Ag^{+}-type	3.A.3.5.1	–	–	–	–	–	–
Rmet2046	chrom.1	Cu^{+}/Ag^{+}-type	3.A.3.5.10	–	–	–	–	–	–
Rmet6119	pMOL30	Cu^{+}/Ag^{+}-type	3.A.3.5.4	–	–	–	–	–	–
Rmet2379	chrom.1	Cu^{+}/Ag^{+}-type	3.A.3.5.7	–	–	–	–	–	–
Rmet5970	pMOL30	CzcP	3.A.3.6.1	+	+	+	+	+	+
Rmet5947	pMOL30	PbrA	3.A.3.6.4	+	+	+	+	+	+
Rmet4594	chrom.2	ZntA	3.A.3.6.4	–	–	–	–	–	–
Rmet2303	chrom.1	CadA	3.A.3.6.4	+	+	+	+	+	+
Rmet0038	chrom.1	K^{+}-type	3.A.3.7.1	–	–	–	–	–	–
Rmet0039	chrom.1	K^{+}-type	3.A.3.7.1	–	–	–	–	–	–
Rmet0040	chrom.1	K^{+}–type	3.A.3.7.1	–	–	–	–	–	–
MerT-Proteins									
Rmet6345	pMOL28	$MerT_1$	9.A.2.1.1	+	+	+	+	+	+
Rmet5991	pMOL30	$MerT_2$	9.A.2.1.1	–	–	–	–	–	+
Rmet6172	pMOL30	MerT4	9.A.2.1.1	+	+	+	+	+	+
Rmet2313	chrom.1	$MerT_3$	9.A.2.1.1	–	–	–	–	–	+

[a] After stepwise removal of organisms from consideration according to the phylogenetic distance of the proteobacteria: +, occurrence of paralogs; –, absence of paralogs; chrom., chromosome

horizontal gene transfer might explain the origin of an old paralog. For instance, a gene in the *Cupriavidus* ancestral strain got lost in H16, was kept in JMP134 and duplicated in CH34 (with both genes being more related to the ancestral gene than to each other). So, the time of the gene duplication event leading to an old paralog is not well defined (but might have happened before speciation) while a species-specific paralog probably has duplicated during speciation of the respective bacterium.

RND-driven efflux systems

Overview

The RND proteins are a superfamily of proteins that are part of multi subunit proteins complexes involved in efflux reactions (Tseng et al. 1999). The best-characterized example is the AcrB protein from *E. coli* that belongs to the hydrophobe amphiphile efflux protein family HAE1-RND, which transport hydrophobic and amphophilic substances including many antibiotics. HME-RND (heavy metal efflux) is a second RND family with many representatives in Gram-negative bacteria (Tseng et al. 1999). These members can be subcategorized in Zn^{2+}/Co^{2+}/Cd^{2+} transporters (HME1), Ni^{2+}/Co^{2+} transporters (HME2), Zn^{2+} (HME3a), Co^{2+} (HME3b), and Cu^{+}/Ag^{+} (HME4) transporters (Nies 2003). RND proteins are trimeric proteins located in the cytoplasmic membrane (Murakami et al. 2002; Rensing et al. 1997). They form protein complexes with a trimeric outer membrane factor OMF (Koronakis et al. 2000; Paulsen et al. 1997) and a bridging periplasmic adapter protein (Akama et al. 2004; Andersen et al. 2001; Higgins et al. 2004) of the MFP (Saier et al. 1994) family. This protein complex is supposed to "vacuum-clean" the periplasma and interior of the cytoplasmic membrane by export of toxic substances to the outside (Lomovskaya and Totrov 2005). The RND proteins drives this process as a kind of peristaltic pump (Dastidar et al. 2007; Murakami et al. 2006), which is driven by the proton-motive force (Goldberg et al. 1999). Periplasmic cations are also likely the substrates for HME-RND efflux pumps (Franke et al. 2003; Legatzki et al. 2003a; Munkelt et al. 2004).

Cupriavidus metallidurans contains 12 operons coding for HME-RND-driven transenvelope efflux systems that may remove transition metals by transport from the periplasm to the outside (Nies 2003). These 12 systems could be grouped in pairs or triples into the 5 clusters introduced above. Their expression profiles as well as contribution to metal resistance of these systems has been analyzed (Nies et al. 2006). Ten of these RND proteins have orthologs in the predicted proteomes of the seven investigated bacteria (Table 6). The number of orthologs increased from 1 to 2 orthologs in the most distant CH34 relatives (Fig.1), the two *Burkholderia* species, to 7 orthologs in *R. solanacearum*. However, none of the other strains contains 12 predicted HME-RND proteins. This gives the impression that during evolution of *C. metallidurans* from an ancestral beta proteobacterium the number of HME-RND proteins has steadily increased, maybe by horizontal gene transfer, and that *C. metallidurans* contains the full orchestra while the related bacteria posses only a more limited assortment. However, most of the resulting acquired genes are silent, inactivated by mutation, constitutively expressed on a low level, or inducible only to a moderate degree (Nies et al. 2006). That leaves 3 of the 12 systems that were vigorously expressed and are central to transition metal resistance in strain CH34: the chromosomal copper/silver HME4-RND system, *cnr* for cobalt/nickel resistance on plasmid pMOL28, and *czc* for cobalt/zinc/cadmium resistance on plasmid pMOL30. In each case, the respective RND protein was product of a gene duplication during evolution of strain CH34.

The Czc system

The HME1 group contains pMOL30-encoded CzcA, the RND protein central to cobalt, zinc, cadmium resistance in *C. metallidurans* (Nies 2003; Nies et al. 1989b), and a putative protein encoded by chromosome 2 (Hme468, Rmet_4468), which is not expressed (Nies et al. 2006). Both RND proteins share 80% sequence identity but have more closely related orthologs encoded by the chromosomes 2 of the two *C. eutrophus* strains: AcrB10 from strain H16 is 82% identical to CzcA but only 80% identical to Hme468, while Reut_B3968 from JMP134 is 82% identical to Hme469 but only 78% identical to CzcA. The genes for the RND proteins in strains H16 and JMP134 have upstream genes for a MFP, OMF and a putative periplasmic metal binding protein. The products of these genes are highly related to CzcB (MFP), CzcC (OMF) and CzcI, respectively. On the other hand, the silent gene for Hme468 has only a MFP gene upstream and seems to be truncated (Nies et al. 2006). Thus, all three strains might have inherited a *czcICBA*-like operon on chromosome 2 from an ancestral *Cupriavidus* strain. In CH34, this operon was duplicated onto plasmid pMOL30 while a truncated, silent copy remained on the chromosome.

The two proteins of the two-component regulatory system of Czc (Große et al. 1999; Nies and Brown

1998; van der Lelie et al. 1997), CzcS, and CzcR, had no orthologs in the six related strains. In contrast, genes for a two-component regulatory system (histidine kinase Rmet_4466, response regulator Rmet_4465) reside adjacent to the silent gene for Hme468 (Rmet_4468). The putative histidine kinase Rmet4466 has orthologs in all six other 6 strains, the response regulator Rmet4465 in JMP134, H16, and *R. solanacearum*. CzcR is highly specific and cannot be complemented by Rmet_4465 while Rmet_4466 is partially able to complement a missing CzcS (Scherer and Nies, unpublished observation). Thus, Rmet_4465/Rmet_4466 might have been the original regulatory system of the truncated Hme465-encoding operon while the *czc* determinant on plasmid pMOL30 is now under control of the CzcRS system, which may have been acquired by horizontal gene transfer.

CzcN is encoded upstream of *czcI* and contains a CzcR binding site (Große et al. 2004). The protein has no orthologs in the other six bacteria but a related protein, NccN, is part of the nickel–cobalt–cadmium resistance determinant *ncc* of *Achromobacter* sp. 31A (*Alcaligenaceae*, *Burkholderiales*; Schmidt and Schlegel 1994). CzcI has orthologs encoded by the *czcICBA*-like operons located on chromosomes 2 of strain H16 and JMP134 (AcrD and Reut_B3971, respectively). Moreover, CzcI has a paralog in CH34, Rmet_4595. Downstream of this second *czcI* gene is a *czcC*-like (Rmet_4596) and the 5′ end of a truncated *czcB*-like gene (Rmet_4697). Upstream and in the other direction of transcription is the gene for the P-type ATPase ZntA (Rmet_4594, see below). This arrangement is similar to that in strain JMP134 with a *zntA*-like gene on one DNA-strand and the *czcICBA*-like operon on the other strand. The two ZntA proteins are orthologs (74% identity). The ZntA ortholog from strain H16 was named CadA (77% identity to ZntA from CH34) and is encoded as a single gene of chromosome 2 of H16, but not adjacent to the *czcICBA*-like operon.

On pMOL30, the *czcDRSE* regulatory region downstream of *czcNICBA* (Große et al. 1999, 2004) encodes the CDF protein CzcD (see below), which does not have orthologs in JMP134, H16 or *R. solanacearum*, the two-component system CzcRS and the small periplasmic metal-binding protein CzcE (Große et al. 2004). CzcE has no orthologs in the other six bacteria but is a strain-specific paralog of the CopH protein, which is part of the copper resistance system of strain CH34 (Monchy et al. 2006). Interestingly, this protein still binds copper when isolated as native protein (Zoropogui et al. 2008). Thus, taking strain JMP134 as a model of the ancestral *Cupriavidus* strain that might have possessed a *zntA*↔*czcICBA* resistance determinant, *czcICBA* was duplicated in strain CH34 onto plasmid pMOL30, while the copy on chromosome 2 of CH34 got separated into a *zntA*↔*czcICB*′ and a *czc*′*BA* part, which is silent because the promoter *czcIp* (Große et al. 2004) is missing. On plasmid pMOL30, *czcN*, acquired by horizontal gene transfer, became fused to *czcICBA* yielding a *czcNICBA* operon. In different steps, *czcD* and *czcRS*, also acquired by horizontal gene transfer, were added downstream of *czcA*. Finally, the *copH* gene of the copper resistance cluster of plasmid pMOL30 was duplicated, yielding the gene *czcE* downstream of *czcS*. Since CzcE binds copper (Zoropogui et al. 2008) and is involved in control of *czc* expression by CzcS (Große et al. 2004), it might act as regulatory link between the copper and cobalt/zinc/cadmium resistance systems of strain CH34.

The Cnr system

The HME2 protein CnrA, central to nickel/cobalt resistance and encoded by plasmid pMOL28 (Liesegang et al. 1993), and NccA, encoded by pMOL30, form a paralogous couple with both proteins most closely related to each other. The RND proteins CnrA and NccA are strain-specific paralogs, also are the respective MFPs CnrB/NccB and the OMFs CnrC/NccC. Neither RND protein has orthologs in one of the other bacterial strains. Both are related to the NccA protein of the nickel–cobalt–cadmium determinant *nccYXHCBAN* of *Achromobacter* sp. 31A (Schmidt and Schlegel 1994). Probably, CH34 acquired a *nccYXHCBAN* determinant by horizontal gene transfer. The *nccYXHCBA* part was copied to plasmid pMOL28 yielding *cnrYXHCBA* while *nccN* was fused to *czc* leading to *czcN*. Expression of the *cnr* determinant is under control of the ECF sigma factor (Lonetto et al. 1994; Missiakas and Raina 1998) CnrH and the membrane-bound anti-sigma factor complex CnrYX (Grass et al. 2000, 2005a; Tibazarwa et al. 2000). On the other hand, the *nccCBA* determinant on plasmid pMOL30 does not contain *nccYXH* regulatory genes upstream, and the

determinant is silent (Nies et al. 2006). CnrH had neither orthologs in the other 6 bacterial strains nor paralogs in *C. metallidurans* but is 65% identical to NccH from *Achromobacter* sp. 31A.

Thus, the most prominent efflux systems of *C. metallidurans*, Czc and Cnr, show signatures of gene duplication events during speciation of the bacterium, with an active operon on plasmid pMOL30 or pMOL28 that is respectively, essential for the high degree of metal resistance (Mergeay et al. 1985), plus an inactive operon on chromosomal DNA and, interestingly, also plasmid pMOL30. The difference between Czc and Cnr is that the core of the Czc system may have been inherited from an ancestral, JMP134-like *Cupriavidus* strain while Cnr was acquired by horizontal gene transfer from outside of the *Burkholderia*/*Ralstonia*/*Cupriavidus* cluster.

Other HME-RND proteins

Of the two possible copper/silver-exporting HME4 proteins, the protein encoded by chromosomal DNA is inducible at high Ag^{+} concentrations while its homolog, encoded by plasmid pMOL30 as part of a huge copper resistance operon cluster (Monchy et al. 2006), is constitutively expressed (Nies et al. 2006). Both RND proteins are old paralogs and thus maybe the result of a gene duplication event before speciation of strain CH34.

The situation is more complicated when the three HME3a and the three HME3b-encoding genes were considered, which are all encoded by chromosomal DNA and were not paralogs of each other. Expression of one of the HME3a-encoding genes is induced by zinc but the expression level is low (Nies et al. 2006). The second gene is constitutively expressed but contains a frame shift mutation, which inactivates it. The third gene is silent. Similarly, two HME3b-encoding genes were silent, expression of the third is strongly induced by cobalt, however, this gene is inactivated by a transposon insertion. Accordingly, these HME-RND systems do not contribute to metal resistance in strain CH34 (Nies et al. 2006).

P-type ATPases

P-type ATPases form a family of membrane-bound primary transport systems (Fagan and Saier 1994), mostly for metal cations from H^{+} to Pb^{2+}. Strain CH34 contains a high number (Table 3) of 13 predicted P-type ATPases, the other two *Cupriavidus* strains 7 or 8 orthologs (Table 6). Five of the 13 P-type ATPases in the proteome of strain CH34 were strain-specific paralogs (Table 7). These were the two presumed Ca^{2+}-exporting enzymes and three of the four $Zn^{2+}/Cd^{2+}/Pb^{2+}$-exporting proteins, CadA, PbrA, and CzcP. The fourth protein, ZntA, had orthologs in strain H16 and JMP134, and its gene may have been the ancestor of the other three. Thus, strain CH34 has duplicated the number of calcium exporters and increased the number of efflux systems for heavy metals fourfold during its evolution. The latter are forming the second line of defense ahead of the RND-driven efflux pumps (Nies 2003) because they might be able to export heavy metal cations even when they were previously bound to thiol compounds (Rensing et al. 1999). The contribution of P-type ATPases to metal resistance in strain CH34 has been examined (Legatzki et al. 2003a) and also the interaction of P-type ATPases and glutathione in *E. coli* (Helbig et al. 2008).

Interestingly, the two other *Cupriavidus* strains contained four potential Cu^{+}/Ag^{+}-exporting P-type ATPases in addition to two RND-driven efflux pumps with the same predicted substrate range, indicating competence in removal of copper or silver already in the ancestral *Cupriavidus* strain. The *E. coli* ortholog of these P-type ATPases is CopA. Together with the periplasmic copper-containing Cu^{+}-oxidase CueO and the RND-driven efflux system Cus, CopA is responsible for copper and silver resistance in *E. coli* (Franke et al. 2001, 2003; Grass and Rensing 2001a, b; Grass et al. 2004; Roberts et al. 2002). The CueO function might be performed by other proteins in *C. metallidurans* that were encoded by the plasmid pMOL30 (Monchy et al. 2006).

CDF proteins (cation diffusion facilitators)

CDF proteins (Paulsen and Saier 1997) are a family of membrane-bound secondary transport systems for divalent transition metal cations. They form the third line of defense against excess transition metal cations (Nies 2003). *Cupriavidus* strain CH34 contains three CDF proteins, one of each of the three subfamilies of CDF proteins (Nies 2003). These are the proteins

CzcD (Nies 1992) mentioned above, DmeF and FieF (Munkelt et al. 2004). FieF has an ortholog in *E. coli* (YiiP) and proteins from this subfamily may be export systems for Fe^{2+} and/or Zn^{2+}. The *E. coli* protein is a dimer in a roughly Y-like structure with the two transmembrane domains of each monomer being the arm of the letter Y while the C-terminal cytoplasmic domains together form a metallo-chaperone-like fold (Lu and Fu 2007). The latter domain binds Zn^{2+} cations, and binding of these cations may control transport activity of the protein for an efficient control of cytoplasmic metal homeostasis (Nies 2007b). Thus, metal cations that were imported by unspecific routes like the CorA (Lunin et al. 2006) or the MgtE (Hattori et al. 2007) magnesium uptake systems, can be selected from the range of cytoplasmic divalent metal cations by efflux systems like CDF proteins, which are switched-on when the concentration of their substrates becomes too high.

FieF/YiiP transports Zn^{2+} in vitro (Chao and Fu 2004) but the in vivo substrate of this protein seems to be Fe^{2+} in *E. coli* (Grass et al. 2005b) and in magnetotactic bacteria (Grünberg et al. 2001, 2004). CzcD, its *E. coli* ortholog ZitB (Anton et al. 2004) and related proteins from eukaryotes like ZnT proteins transport Zn^{2+}, CzcD additionally Co^{2+} and Cd^{2+} (Anton et al. 1999). DmeF represents the third subfamilies of CDF proteins and transports cations with a broad substrate specificity, predominantly Co^{2+} (Munkelt et al. 2004). Thus, with these three CDF proteins, *C. metallidurans* CH34 contains the complete set of CDF efflux systems that allows flux-controlled homeostasis of transition metal cations Fe^{2+}, Co^{2+}, Ni^{2+}, Zn^{2+} and Cd^{2+} with overlapping substrate ranges of the individual transporters.

Proteins of the chromate-efflux-related (CHR) protein family

Cupriavidus metallidurans contains four CHR proteins (Nies 2003; Nies et al. 1998) related to the first described chromate efflux pumps ChrA from *C. metallidurans* (Nies et al. 1989a, 1990) and *Pseudomonas aeruginosa* (Cervantes et al. 1990). These systems are driven by the proton-motive force (Alvarez et al. 1999) and remove chromate by efflux in cooperation with other proteins (Juhnke et al. 2002). Two CH34 proteins, $ChrA_1$ and $ChrA_2$, are indeed involved in chromate detoxification by efflux, the other two probably not (Nies et al. 2006). $ChrA_1$, $ChrA_3$ and $ChrA_{4N}$/Chr_{AC} (two polypeptides that form together a full-length CHR protein) all had no paralogs in the strain CH34 predicted proteome while $ChrA_2$ was an old paralog. Thus, the ancestral *Cupriavidus* strain might have duplicated its chromate detoxification competence.

Mercury detoxification

The highly toxic Hg^{2+} cation is bound in the periplasm by MerP, imported into the cytoplasm by MerT, and reduced to metallic mercury by the NADPH-dependent flavoprotein MerA. MerR regulates expression of the *mer* operon (Silver and Phung 1996). Three *mer* operons, located on plasmids pMOL28, pMOL30 and chromosome 1, are each complete *merA*-containing operons (Diels et al. 1985; Dressler et al. 1991). A fourth is an incomplete *merRPT* fragment located also on plasmid pMOL30. As examined with the MerT proteins, $MerT_4$ (encoded by plasmid pMOL30) is a strain-specific paralog of $MerT_1$ (encoded by plasmid pMOL28). $MerT_2$ and $MerT_3$ are old paralogs (Table 7). Thus, duplication of *mer* determinants was also part of both processes, evolution of the ancestral *Cupriavidus* strain and subsequent evolution of strain CH34.

Regulatory genes

Sigma factors

Most sigma factors from *C. metallidurans* had orthologs in all of the other 6 bacterial strains (data not shown). This was true for RpoN, the major sigma factor RpoD, the heat-shock factor RpoH, the starvation factor RpoS, and the unusual (Nies 2004) putative sigma factor RpoG (previously Sig345) from *C. metallidurans* CH34. Interestingly, both *R. solanacearum* strains and *B. xenovorans* contained 2 RpoN sigma factors. Such an outfit has been previously noted only for rhizobia (Nies 2004). A new sigma factor that had not been identified in the previous analysis (Nies 2004) was Rmet3844. This sigma factor belongs to the RpoD family but did not belong

to one of the main subfamilies intracytoplasmic functions, motility and differentiation or ECF. It had orthologs in *C. eutrophus* JMP134, both strains of *R. solanacearum*, *B. xenovorans*, and *B. cepacia*. Following the nomenclature of the other sigma factors from *C. metallidurans*, this factor might be named "RpoT". All of these sigma factors had no paralogs but related proteins belonging to the same protein family.

The ECF sigma CnrH was mentioned above. About 5 of the remaining 10 ECF sigma factors from *C. metallidurans* also had orthologs in all other six bacterial strains. These were RpoE, the RpoE-like RpoP (Sig588), the distant ECF factor RpoR (Sig605), and RpoL (Sig679) from the ECF:FecI2-cluster (Nies 2004). The second distant ECF factor, RpoQ (Sig538) had only an ortholog in *B. xenovorans*. The two other ECF:FecI2 cluster proteins had orthologs in all six strains except strain H16 (RpoO, Sig562) or in all strains except the two *R. solanacearum* strains (RpoM, Sig681). None of these seven ECF sigma factors were paralogs (data not shown). With the exception of RpoE no function could be assigned to them yet (Große et al. 2007).

The three related ECF proteins RpoI (Sig680), RpoJ (Sig611) and RpoK (Sig697) are involved in iron homeostasis with RpoI being the most important of these three (Große et al. 2007). RpoI had an ortholog in *R. solanacearum* strain UW551, and was an old paralog maybe appearing after the separation of *Ralstonia* and *Cupriavidus*. RpoK and RpoJ had no orthologs and were species-specific paralogs. Thus, *C. metallidurans* triplicated during its' speciation the main sigma factor for iron homeostasis RpoI, and this sigma factor might have originated from a FecI-like ancestor during evolution of the genus *Cupriavidus*.

Two-component regulatory systems

Cupriavidus metallidurans contained 63 predicted response regulators and 62 histidine kinases of two-component regulatory systems (Table 8). Both related strains, *C. eutrophus* H16 and JMP134, contained comparable numbers of two-component systems but strain JMP134 10 histidine kinases less and strain H16 nearly 20 response regulators more. In both strains, "lone" response regulators may exist that need a histidine kinase of another operon for phosphorylation. Each of these three strains contained 2–6 paralogs of the respective proteins. In cross comparisons among the *Cupriavidus* strains, the respective two other *Cupriavidus* strains contained 40–62 orthologs of the response regulators and 35–40 orthologs of the histidine kinases (Table 8). These

Table 8 Two-component regulatory systems in *C. metallidurans* CH34 and its closest relatives with the corresponding number of paralogs and orthologs in *Cupriavidus* strains[a]

Bacterial strain	Abs[b] #	Paral[c] #	Orthologs[d] #						
			Cme CH34	*Ceu* JMP134	*Ceu* H16	*Rso* GMI1000	*Rso* UW551 chrom	*Bxe* LB400	*Bce* AMMD
Response regulators (R)									
Cme CH34	63	5	–	49	60	29	24	24	23
Ceu JMP134	62	3	40	–	62	26	29	30	29
Ceu H16	80	5	43	50	–	40	33	37	32
Histidine kinases (S)									
Cme CH34	62	6	–	35	40	27	27	28	25
Ceu JMP134	52	2	39	–	39	21	25	22	21
Ceu H16	62	6	40	37	–	31	29	28	23

[a] The number of representatives is shown

[b] The absolute numbers

[c] The number of these proteins that are paralogs in strain CH34

[d] Orthologs in the indicated bacterial strains

numbers were smaller in the four non-*Curpiavidus* strains, 23–40 and 21–31, respectively. Thus, the three *Cupriavidus* strains contained about 60 two-component regulatory systems, <10% of these were paralogs, 2/3rd occurred also in the respective two other strains, and at least 1/3rd seemed to form a kind of common core of these systems present in all seven bacterial strains.

Plasmids

The genes on the chromosomes indicate that the life-style of the ancestral *Cupriavidus* strain might have been that of a facultatively hydrogen-oxidizing, moderately metal-resistant degrader of aromatic compounds and organic acids rather than a dweller on sugars. The plasmids in the three strains were involved in a better adaptation of the *Cupriavidus* strains to their specific ecological niches. The mega-plasmid from strain JMP134, plasmids pHG1 from H16, and pMOL28 from CH34 are related replicons (Schwartz et al. 2006; Taghavi et al. 1996). These plasmids encoded a high percentage of paralogs and many "rare" orthologs that occurred only in very few of the other investigated bacteria. While plasmid pHG1 carried duplications of the genes for auto-trophic growth as a hydrogen-oxidizing bacterium and for nitrate respiration (Schwartz et al. 2006), these genes were absent in pMOL28 (NC_006525). Instead, pMOL28 contained the *cnr* Ni^{2+}/Co^{2+}, the *chr* chromate (Nies et al. 1989a) and a *mer* mercury resistance determinant (Dressler et al. 1991). Both plasmids are conjugative plasmids (Friedrich et al. 1981; Mergeay et al. 1985) with an extensive „junkyard" (Schwartz et al. 2006) of various mobile elements and fragments thereof. The pMOL28-encoded RND protein CnrA, the chromate resistance protein $ChrA_1$ and the $MerT_1$ protein were all strain-specific or old paralogs. This indicates that pHG1 and pMOL28 may have played an important function in the evolution of strains H16 and CH34, respectively, by facilitating horizontal acquisition of genes, maybe even retrotransfer (Szpirer et al. 1999), recombination events including gene duplications, deletions, and other rearrangements.

While *C. eutrophus* H16 has only one plasmid, pHG1, strains JMP134, and CH34 have two, pJP4 in addition to the megaplasmid (Don and Pemberton 1981) and pMOL30 in addition to pMOL28 (Mergeay et al. 1985), respectively. Plasmids pJP4 and pMOL30 are rather unique and not related. Plasmid pJP4 (which replicates readily in *C. metallidurans*, data not shown) carries genes for degradation of xenobiotics that probably originated from the *Burkholderia* relatives. Plasmid pMOL30 replication protein ParA is related (data not shown) to that of the 1.5 Mb chromosome three of *B. xenovorans* (Chain et al. 2006), which points into the origin of this plasmid. It also seems to contain genes that originated "recently" by horizontal gene transfer and duplication events. Best example for this process is the *czc* gene region (Nies et al. 1987, 1989b; Nies and Silver 1989) as mentioned above. The combined Czc proteins form a resistance system that is central to the outstanding Co^{2+}, Zn^{2+} and Cd^{2+} resistance of *C. metallidurans* (Legatzki et al. 2003b; Nies et al. 2006).

Summary and conclusions

Old paralogs indicate possible gene duplication events before speciation of the species *C. metallidurans*, maybe on the level of the ancestral *Cupriavidus* strain. Genes for detoxification of mercury and export of chromate were doubled, however, this strain contained already all genes for efficient detoxification of divalent heavy metal cations, from the cytoplasm by CDF proteins and further on to the outside by RND-mediated outer membrane efflux driven by HME1/HME2/HME3 proteins. Genes were especially available allowing efficient efflux of Cu^{+}/Ag^{+} in a similar two-step manner. The genus name "*Cupriavidus*" is indeed well taken!

From this ancestral strain, enhancement of the ability to degrade xenobiotics by horizontal acquisition of plasmid pJP4 led to *C. eutrophus* JMP134. Thus, the plasmid pJP4-free strain JMP222 (Don and Pemberton 1981) could serve as a model for the ancestral *Cupriavidus* strain. On the other hand, copying genes for hydrogen-oxidation and nitrate respiration to the megaplasmid plus increased ability to obtain nutrients leads to *C. eutrophus* H16. During speciation of *C. metallidurans* CH34, two MerT proteins duplicated into four, yielding three active mercury-detoxification systems (Dressler et al. 1991). The *zntA* gene for a P-type ATPase of the $Zn^{2+}/Cd^{2+}/Pb^{2+}$-exporting group, which was inherited as part of

a *zntA*↔*czcICBA* determinant on chromosome 2, multiplicated into four genes *zntA*, *pbrA*, *cadA*, and *czcP* (Borremans et al. 2001; Legatzki et al. 2003a; Scherer and Nies, unpublished). These proteins export divalent metal cations from the cytoplasm to the periplasm and seem to be able to pump efficiently even cations that are bound to thiol groups (Rensing et al. 1999, 2000).

With respect to their ionic diameters and binding affinities, Co^{2+} and Ni^{2+} are potential inhibitors of the metabolism of the bioelement Fe^{2+} (Nies 2007a). Strain-specific paralogs are the two RND proteins CnrA and NccA, CnrA being involved in export of cobalt and nickel by the CnrCBA efflux complex. Neither CnrA nor NccA had orthologs in the other investigated bacterial strains. An ancestral RND-encoding gene might thus have been acquired from outside the *Burkholderia*/*Ralstonia* genus cluster by the ancestor of *C. metallidurans* and duplicated. Nevertheless, one of the respective genes, *nccA* from plasmid pMOL30, is not expressed. Neither are the three RND (HME3b) proteins that may have the same substrate specificity (Nies et al. 2006). The remaining gene, *cnrA*, is under control of an ECF sigma factor, CnrH (Grass et al. 2000, 2005a; Tibazarwa et al. 2000), which has also apparently been acquired rather "recently". On the other hand, *C. metallidurans* does not contain a Ni^{2+}/Co^{2+} uptake system of the NiCoT protein family (Eberz et al. 1989; Eitinger and Friedrich 1991; Eitinger et al. 2005; Rodionov et al. 2006). Finally, FecI-like (Braun et al. 1998; Mahren and Braun 2003) ECF sigma factors that are in control of iron homeostasis in *C. metallidurans* triplicated (Große et al. 2007; Nies 2004). Thus, *C. metallidurans* decreased its cobalt/nickel uptake ability, shifted instead to homeostasis by controlled outer membrane efflux and enhanced its' iron uptake set of instruments. This indicates a stepwise evolutionary adaptation from the ancestral *Cupriavidus* strain that was already well able to deal with toxic heavy metals, especially copper, to a bacterium, which is extraordinarily adapted to Ni^{2+}/Co^{2+} and chromate-containing environments, e.g., in serpentine soils (Baker 1987). Due to its important contribution to nickel, cobalt, and chromate resistance (Mergeay et al. 1985; Nies et al. 1989a), this adaptation process seem to be accompanied by evolution of plasmid pMOL28. Moreover, a combination of horizontal gene transfer events plus gene duplication and re-arrangements formed plasmid pMOL30, which led to adaptation to high levels of other transition metal cations, that of copper, zinc, and cadmium. This may have been the evolutionary road that *C. metallidurans* has traveled.

Acknowledgments This work was supported by Graduiertenkolleg "Stress" of the *Deutsche Forschungsgemeinschaft* and a research fund from the *Land Sachsen-Anhalt.*

References

Akama H, Matsuura T, Kashiwagi S, Yoneyama H, Narita SI, Tsukihara T et al (2004) Crystal structure of the membrane fusion protein, MexA, of the multidrug transporter in *Pseudomonas aeruginosa*. J Biol Chem 279:25939–25942. doi:10.1074/jbc.C400164200

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. doi:10.1093/nar/25.17.3389

Alvarez AH, Moreno-Sanchez R, Cervantes C (1999) Chromate efflux by means of the ChrA chromate resistance protein from *Pseudomonas aeruginosa*. J Bacteriol 181:7398–7400

Andersen C, Hughes C, Koronakis V (2001) Protein export and drug efflux through bacterial channel-tunnels. Curr Opin Cell Biol 13:412–416. doi:10.1016/S0955-0674(00)00229-5

Anton A, Große C, Reißman J, Pribyl T, Nies DH (1999) CzcD is a heavy metal ion transporter involved in regulation of heavy metal resistance in *Ralstonia* sp. strain CH34. J Bacteriol 181:6876–6881

Anton A, Weltrowski A, Haney JH, Franke S, Grass G, Rensing C et al (2004) Characteristics of zinc transport by two bacterial cation diffusion facilitators from *Ralstonia metallidurans* and *Escherichia coli*. J Bacteriol 186:7499–7507. doi:10.1128/JB.186.22.7499-7507.2004

Baker AJM (1987) Metal tolerance in plants. New Phytol 106:93–111

Borremans B, Hobman JL, Provoost A, Brown NL, Van der Lelie D (2001) Cloning and functional analysis of the *pbr* lead resistance determinant of *Ralstonia metallidurans* CH34. J Bacteriol 183:5651–5658. doi:10.1128/JB.183.19.5651-5658.2001

Braun V, Hantke K, Koster W (1998) Bacterial iron transport: mechanisms, genetics, and regulation. Met Ions Biol Syst 35:67–145

Busch W, Saier MHJ (2002) The transporter classification (TC) system. Crit Rev Biochem Mol Biol 37:287–337. doi:10.1080/10409230290771528

Cervantes C, Ohtake H, Chu L, Misra TK, Silver S (1990) Cloning, nucleotide sequence, and expression of the chromate resistance determinant of *Pseudomonas aeruginosa* plasmid pUM505. J Bacteriol 172:287–291

Chain PS, Denef VJ, Konstantinidis KT et al (2006) Burkholderia xenovorans LB400 harbors a multi-replicon,

9.73-Mbp genome shaped for versatility. Proc Natl Acad Sci USA 103:15280–15287. doi:10.1073/pnas.0606924103

Chao Y, Fu D (2004) Thermodynamic studies of the mechanism of metal binding to the *Escherichia coli* zinc transporter YiiP. J Biol Chem 279:17173–17180. doi:10.1074/jbc.M400208200

Dastidar V, Mao W, Lomovskaya O, Zgurskaya HI (2007) Drug-induced conformational changes in multidrug efflux transporter AcrB from Haemophilus influenzae. J Bacteriol 189:5550–5558. doi:10.1128/JB.00471-07

Denef VJ, Park J, Tsoi TV, Rouillard J-M, Zhang H, Wibbenmeyer JA et al (2004) Biphenyl and benzoate metabolism in a genomic context: outlining genome-wide metabolic networks in *Burkholderia xenovorans* LB400. Appl Environ Microbiol 40:4961–4970. doi:10.1128/AEM.70.8.4961-4970.2004

Diels L, Mergeay M (1990) DNA probe-mediated detection of resistant bacteria from soils highly polluted by heavy metals. Appl Environ Microbiol 56:1485–1491

Diels L, Faelen M, Mergeay M, Nies D (1985) Mercury transposons from plasmids governing multiple resistance to heavy metals in *Alcaligenes eutrophus* CH34. Arch Int Physiol Biochim 93:B27–B28. doi:10.3109/13813458509080622

Diels L, van Roy S, Somers K, Willems I, Doyen W, Mergeay M et al (1995) The use of bacteria immobilized in tubular membrane reactors for heavy metal recovery and degradation of chlorinated aromatics. J Membr Sci 100:248–258. doi:10.1016/0376-7388(94)00253-U

Don RA, Pemberton JM (1981) Properties of six pesticide degradation plasmids isolated from *Alcaligenes paradoxus* and *Alcaligenes eutrophus*. J Bacteriol 145:681–686

Dressler C, Kües U, Nies DH, Friedrich B (1991) Determinants encoding multiple metal resistance in newly isolated copper-resistant bacteria. Appl Environ Microbiol 57:3079–3085

Eberz G, Eitinger T, Friedrich B (1989) Genetic determinants of a nickel-specific transport system are part of the plasmid-encoded hydrogenase gene cluster in *Alcaligenes eutrophus*. J Bacteriol 171:1340–1345

Eitinger T, Friedrich B (1991) Cloning, nucleotide sequence, and heterologous expression of a high-affinity nickel transport gene from *Alcaligenes eutrophus*. J Biol Chem 266:3222–3227

Eitinger T, Suhr J, Moore L, Smith JAC (2005) Secondary transporters for nickel and cobalt ions: theme and variations. Biometals 18:399–405. doi:10.1007/s10534-005-3714-x

Fagan MJ, Saier MH Jr (1994) P-type ATPases of eukaryotes and bacteria: sequence comparisons and construction of phylogenetic trees. J Mol Evol 38:57–99. doi:10.1007/BF00175496

Franke S, Grass G, Nies DH (2001) The product of the *ybdE* gene of the *Escherichia coli* chromosome is involved in detoxification of silver ions. Microbiology 147:965–972

Franke S, Grass G, Rensing C, Nies DH (2003) Molecular analysis of the copper-transporting CusCFBA efflux system from *Escherichia coli*. J Bacteriol 185:3804–3812. doi:10.1128/JB.185.13.3804-3812.2003

Friedrich B, Hogrefe C, Schlegel HG (1981) Naturally occurring genetic transfer of hydrogen-oxidizing ability between strains of *Alcaligenes eutrophus*. J Bacteriol 147:198–205

Goldberg M, Pribyl T, Juhnke S, Nies DH (1999) Energetics and topology of CzcA, a cation/proton antiporter of the RND protein family. J Biol Chem 274:26065–26070. doi:10.1074/jbc.274.37.26065

Goris J, De Vos P, Coenye T, Hoste B, Janssens D, Brim H, Diels L, Mergeay M, Kersters K, Vandamme P (2001) Classification of metal-resistant bacteria from industrial biotopes as *Ralstonia campinensis* sp. nov., *Ralstonia metallidurans* sp. nov., *Ralstonia basilensis* Steinle et al. 1998 emend. Int J Syst Evol Microbiol 51:1773–1782

Grass G, Rensing C (2001a) CueO is a multi-copper oxidase that confers copper tolerance in *Escherichia coli*. Biochem Biophys Res Commun 286:902–908. doi:10.1006/bbrc.2001.5474

Grass G, Rensing C (2001b) Genes involved in copper homeostasis in *Escherichia coli*. J Bacteriol 183:2145–2147. doi:10.1128/JB.183.6.2145-2147.2001

Grass G, Große C, Nies DH (2000) Regulation of the *cnr* cobalt/nickel resistance determinant from *Ralstonia* sp. CH34. J Bacteriol 182:1390–1398. doi:10.1128/JB.182.5.1390-1398.2000

Grass G, Thakali K, Klebba PE, Thieme D, Müller A, Wildner GF et al (2004) Linkage between catecholate siderophores and the multicopper oxidase CueO in *Escherichia coli*. J Bacteriol 186:5826–5833. doi:10.1128/JB.186.17.5826-5833.2004

Grass G, Fricke B, Nies DH (2005a) Control of expression of a periplasmic nickel efflux pump by periplasmic nickel concentrations. Biometals 18:437–448. doi:10.1007/s10534-005-3718-6

Grass G, Otto M, Fricke B, Haney CJ, Rensing C, Nies DH et al (2005b) FieF (YiiP) from *Escherichia coli* mediates decreased cellular accumulation of iron and relieves iron stress. Arch Microbiol 183:9–18. doi:10.1007/s00203-004-0739-4

Große C, Grass G, Anton A, Franke S, Navarrete Santos A, Lawley B et al (1999) Transcriptional organization of the *czc* heavy metal homoeostasis determinant from *Alcaligenes eutrophus*. J Bacteriol 181:2385–2393

Große C, Anton A, Hoffmann T, Franke S, Schleuder G, Nies DH (2004) Identification of a regulatory pathway that controls the heavy metal resistance system Czc via promoter *czcNp* in *Ralstonia metallidurans*. Arch Microbiol 182:109–118

Große C, Friedrich S, Nies DH (2007) Contribution of extracytoplasmic function sigma factors to transition metal homeostasis in *Cupriavidus metallidurans* strain CH34. J Mol Microbiol Biotechnol 12:227–240. doi:10.1159/000099644

Grover A, Azmi W, Gadewar AV, Pattanayak D, Naik PS, Shekhawat GS et al (2006) Genotypic diversity in a localized population of *Ralstonia solanacearum* as revealed by random amplified polymorphic DNA markers. J Appl Microbiol 101:798–806. doi:10.1111/j.1365-2672.2006.02974.x

Grünberg K, Wawer C, Tebo BM, Schüler D (2001) A large gene cluster encoding several magnetosome proteins is

conserved in different species of magnetotactic bacteria. Appl Environ Microbiol 67:4573–4582. doi:10.1128/AEM.67.10.4573-4582.2001

Grünberg K, Müller E-C, Otto A, Reszka R, Linder L, Kube M et al (2004) Biochemical and proteomic analysis of the magnetosome membrane in *Magnetospirillum gryphiswaldense*. Appl Environ Microbiol 70:1040–1050. doi:10.1128/AEM.70.2.1040-1050.2004

Guidot A, Prior P, Schoenfeld J, Carrere S, Genin S, Boucher C (2007) Genomic structure and phylogeny of the plant pathogen *Ralstonia solanacearum* inferred from gene distribution analysis. J Bacteriol 189:377–387. doi:10.1128/JB.00999-06

Hattori M, Tanaka Y, Fukai S, Ishitani R, Nureki O (2007) Crystal structure of the MgtE Mg^{2+} transporter. Nature 448:1072–1076. doi:10.1038/nature06093

Helbig K, Bleuel C, Krauss GJ, Nies DH (2008) Glutathione and transition metal homeostasis in *Escherichia coli*. J Bacteriol 190:5431–5438. doi:10.1128/JB.00271-08

Higgins MK, Bokma E, Koronakis E, Hughes C, Koronakis V (2004) Structure of the periplasmic component of a bacterial drug efflux pump. Proc Natl Acad Sci USA 101:9994–9999. doi:10.1073/pnas.0400375101

Johnson BF, Stanier RY (1971) Dissimilation of aromatic compounds by *Alcaligenes eutrophus*. J Bacteriol 107:468–475

Juhnke S, Peitzsch N, Hübener N, Große C, Nies DH (2002) New genes involved in chromate resistance in *Ralstonia metallidurans* strain CH34. Arch Microbiol 179:15–25. doi:10.1007/s00203-002-0492-5

Koonin EV, Tatusov RL, Rudd KE (1995) Sequence similarity analysis of *Escherichia coli* proteins: functional and evolutionary implications. Proc Natl Acad Sci USA 92:11921–11925. doi:10.1073/pnas.92.25.11921

Koonin EV, Tatusov RL, Rudd KE (1996) Protein sequence comparison at genome scale. Methods Enzymol 266:295–322. doi:10.1016/S0076-6879(96)66020-0

Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. Nature 405:914–919. doi:10.1038/35016007

Kunito T, Kusano T, Oyaizu H, Senoo K, Kanazawa S, Matsumoto S (1996) Cloning and sequence analysis of *czc* genes in *Alcaligenes* sp. strain CT14. Biosci Biotechnol Biochem 60:699–704

Legatzki A, Anton A, Grass G, Rensing C, Nies DH (2003a) Interplay of the Czc-system and two P-type ATPases in conferring metal resistance to *Ralstonia metallidurans*. J Bacteriol 185:4354–4361. doi:10.1128/JB.185.15.4354-4361.2003

Legatzki A, Franke S, Lucke S, Hoffmann T, Anton A, Neumann D et al (2003b) First step towards a quantitative model describing Czc-mediated heavy metal resistance in *Ralstonia metallidurans*. Biodegradation 14:153–168. doi:10.1023/A:1024043306888

Lengeler JW (1990) Molecular analysis of the enzyme II complexes of the bacterial phosphotransferase system (PTS) as carbohydrate transport system. Biochim Biophys Acta 1018:155–159. doi:10.1016/0005-2728(90)90238-Y

Liesegang H, Lemke K, Siddiqui RA, Schlegel H-G (1993) Characterization of the inducible nickel and cobalt resistance determinant *cnr* from pMOL28 of *Alcaligenes eutrophus* CH34. J Bacteriol 175:767–778

Lomovskaya O, Totrov M (2005) Vacuuming the periplasm. J Bacteriol 187:1879–1883. doi:10.1128/JB.187.6.1879-1883.2005

Lonetto MA, Brown KL, Rudd KE, Buttner MJ (1994) Analysis of the *Streptomyces coelicolor sigF* gene reveals the existence of a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions. Proc Natl Acad Sci USA 91:7573–7577. doi:10.1073/pnas.91.16.7573

Lu M, Fu D (2007) Structure of the zinc transporter YiiP. Science 317:1746–1748. doi:10.1126/science.1143748

Lunin V, Dobrovetsky E, Khutoreskaya G, Zhang R, Joachimiak A, Doyle DA et al (2006) Crystal structure of the CorA Mg^{2+} transporter. Nature 440:833–837. doi:10.1038/nature04642

Mahren S, Braun V (2003) The FecI extracytoplasmic-function sigma factor of *Escherichia coli* interacts with the beta′ subunit of RNA polymerase. J Bacteriol 185:1796–1802. doi:10.1128/JB.185.6.1796-1802.2003

Marrero J, Auling G, Coto O, Nies DH (2007) High-level resistance to cobalt and nickel but probably no transenvelope efflux: metal resistance in the Cuban *Serratia marcescens* strain C-1. Microb Ecol 53:123–133. doi:10.1007/s00248-006-9152-7

Mergeay M, Houba C, Gerits J (1978) Extrachromosomal inheritance controlling resistance to cadmium, cobalt and zinc ions: evidence for curing a *Pseudomonas*. Arch Int Physiol Biochim 86:440–441

Mergeay M, Nies D, Schlegel HG, Gerits J, Charles P, van Gijsegem F (1985) *Alcaligenes eutrophus* CH34 is a facultative chemolithotroph with plasmid-bound resistance to heavy metals. J Bacteriol 162:328–334

Mergeay M, Monchy S, Vallaeys T, Auquier V, Benotmane A, Bertin P et al (2003) *Ralstonia metallidurans*, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. FEMS Microbiol Rev 27:385–410. doi:10.1016/S0168-6445(03)00045-7

Missiakas D, Raina S (1998) The extracytoplasmic function sigma factors: role and regulation. Mol Microbiol 28:1059–1066. doi:10.1046/j.1365-2958.1998.00865.x

Monchy S, Benotmane MA, Wattiez R, van Aelst S, Auquier V, Borremans B et al (2006) Transcriptomics and proteomic analysis of the pMOL30-encoded copper resistance in *Cupriavidus metallidurans* strain CH34. Microbiology 152:1765–1776. doi:10.1099/mic.0.28593-0

Munkelt D, Grass G, Nies DH (2004) The chromosomally encoded cation diffusion facilitator proteins DmeF and FieF from *Wautersia metallidurans* CH34 are transporters of broad metal specificity. J Bacteriol 186:8036–8043. doi:10.1128/JB.186.23.8036-8043.2004

Murakami S, Nakashima R, Yamashita R, Yamaguchi A (2002) Crystal structure of bacterial multidrug efflux transporter AcrB. Nature 419:587–593. doi:10.1038/nature01050

Murakami S, Nakashima R, Yamashita E, Matsumoto T, Yamaguchi A (2006) Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. Nature 443:173–179. doi:10.1038/nature05076

Mushegian AR, Garey JR, Martin J, Liu LX (1998) Large-scale taxonomic profiling of eukaryotic model organisms: a

comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. Genome Res 8:590–598

Nies DH (1992) CzcR and CzcD, gene products affecting regulation of resistance to cobalt, zinc and cadmium (*czc* system) in *Alcaligenes eutrophus*. J Bacteriol 174:8102–8110

Nies DH (1995) The cobalt, zinc, and cadmium efflux system CzcABC from *Alcaligenes eutrophus* functions as a cation–proton-antiporter in *Escherichia coli*. J Bacteriol 177:2707–2712

Nies DH (1999) Microbial heavy metal resistance. Appl Microbiol Biotechnol 51:730–750. doi:10.1007/s002530051457

Nies DH (2003) Efflux-mediated heavy metal resistance in prokaryotes. FEMS Microbiol Rev 27:313–339. doi:10.1016/S0168-6445(03)00048-2

Nies DH (2004) Incidence and function of sigma factors in *Ralstonia metallidurans* and other bacteria. Arch Microbiol 181:255–268. doi:10.1007/s00203-004-0658-4

Nies DH (2007a) Bacterial transition metal homeostasis. In: Nies DH, Silver S (eds) Molecular microbiology of heavy metals. Springer-Verlag, Berlin, pp 118–142

Nies DH (2007b) How cells control zinc homeostasis. Science 317:1695–1696. doi:10.1126/science.1149048

Nies DH, Brown N (1998) Two-component systems in the regulation of heavy metal resistance. In: Silver S, Walden W (eds) Metal ions in gene regulation. Chapman Hall, London, pp 77–103

Nies DH, Silver S (1989) Plasmid-determined inducible efflux is responsible for resistance to cadmium, zinc, and cobalt in *Alcaligenes eutrophus*. J Bacteriol 171:896–900

Nies DH, Silver S (1995) Ion efflux systems involved in bacterial metal resistances. J Ind Microbiol 14:186–199. doi:10.1007/BF01569902

Nies D, Mergeay M, Friedrich B, Schlegel HG (1987) Cloning of plasmid genes encoding resistance to cadmium, zinc, and cobalt in *Alcaligenes eutrophus* CH34. J Bacteriol 169:4865–4868

Nies A, Nies DH, Silver S (1989a) Cloning and expression of plasmid genes encoding resistances to chromate and cobalt in *Alcaligenes eutrophus*. J Bacteriol 171:5065–5070

Nies DH, Nies A, Chu L, Silver S (1989b) Expression and nucleotide sequence of a plasmid-determined divalent cation efflux system from *Alcaligenes eutrophus*. Proc Natl Acad Sci USA 86:7351–7355. doi:10.1073/pnas.86.19.7351

Nies A, Nies DH, Silver S (1990) Nucleotide sequence and expression of a plasmid-encoded chromate resistance determinant from *Alcaligenes eutrophus*. J Biol Chem 265:5648–5653

Nies DH, Koch S, Wachi S, Peitzsch N, Saier MHJ (1998) CHR, a novel family of prokaryotic proton motive force-driven transporters probably containing chromate/sulfate transporters. J Bacteriol 180:5799–5802

Nies DH, Rehbein G, Hoffmann T, Baumann C, Grosse C (2006) Paralogs of genes encoding metal resistance proteins in *Cupriavidus metallidurans* strain CH34. J Mol Microbiol Biotechnol 11:82–93. doi:10.1159/000092820

Paulsen IT, Saier MH Jr (1997) A novel family of ubiquitous heavy metal ion transport proteins. J Membr Biol 156:99–103. doi:10.1007/s002329900192

Paulsen IT, Park JH, Choi PS, Saier MHJ (1997) A family of gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from gram-negative bacteria. FEMS Microbiol Lett 156:1–8

Pohlmann A, Fricke WF, Reinecke F et al (2006) Genome sequence of the bioplastic-producing “Knallgas” bacterium *Ralstonia eutropha* H16. Nat Biotechnol 24:1257–1262. doi:10.1038/nbt1244

Raymond J, Zhaxybayeva P, Gogarten JP, Gerdes SY, Blankenship RE (2002) Whole-genome analysis of photosynthetic prokaryotes. Science 298(5598):1616–1620. doi:10.1126/science.1075558

Reizer A, Pao GM, Saier MH Jr (1991) Evolutionary relationships among the permease proteins of the bacterial phosphoenolpyruvate: sugar phosphotransferase system. Construction of phylogenetic trees and possible relatedness to proteins of eukaryotic mitochondria. J Mol Evol 33:179–193. doi:10.1007/BF02193633

Rensing C, Pribyl T, Nies DH (1997) New functions for the three subunits of the CzcCBA cation–proton-antiporter. J Bacteriol 179:6871–6879

Rensing C, Ghosh M, Rosen BP (1999) Families of soft-metal-ion-transporting ATPases. J Bacteriol 181:5891–5897

Rensing C, Fan B, Sharma R, Mitra B, Rosen BP (2000) CopA: an *Escherichia coli* Cu(I)-translocating P-type ATPase. Proc Natl Acad Sci USA 97:652–656. doi:10.1073/pnas.97.2.652

Roberts SA, Weichsel A, Grass G, Thakali K, Hazzard JT, Tollin G et al (2002) Crystal structure and electron transfer kinetics of CueO, a multicopper oxidase required for copper homeostasis in *Escherichia coli*. Proc Natl Acad Sci USA 99:2766–2771. doi:10.1073/pnas.052710499

Rodionov DA, Hebbeln P, Gelfand MS, Eitinger T (2006) Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. J Bacteriol 188:317–327. doi:10.1128/JB.188.1.317-327.2006

Saier MHJ (2000) A functional-phylogenetic classification system for transmembrane solute transporters. Microbiol Mol Biol Rev 64:354–411. doi:10.1128/MMBR.64.2.354-411.2000

Saier MH (2005) Beneficial bacteria and bioremediation. J Mol Microbiol Biotechnol 9:63–64. doi:10.1159/000088836

Saier MH Jr, Grenier FC, Lee CA, Waygood EB (1985) Evidence for the evolutionary relatedness of the proteins of the bacterial phosphoenolpyruvate: sugar phosphotransferase system. J Cell Biol 27:43–56

Saier MH Jr, Tam R, Reizer A, Reizer J (1994) Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. Mol Microbiol 11:841–847. doi:10.1111/j.1365-2958.1994.tb00362.x

Saier MHJ, Tran CV, Barabote RD (2006) TCDB: the transporter classification database for membrane transport protein analyses and information. Nucleic Acids Res 34:D181–D186. doi:10.1093/nar/gkj001

Salanoubat M, Genin S, Artiguenave F et al (2002) Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature 415:497–502. doi:10.1038/415497a

Schmidt T, Schlegel HG (1994) Combined nickel–cobalt–cadmium resistance encoded by the *ncc* locus of *Alcaligenes xylosoxidans* 31A. J Bacteriol 176:7045–7054

Schwartz E, Henne A, Cramm R, Eitinger T, Friedrich B, Gottschalk G (2006) Complete nucleotide sequence of pHG1: a *Ralstonia eutropha* H16 megaplasmid encoding key enzymes of H-2-based lithoautotrophy and anaerobiosis. J Mol Biol 332:369–383. doi:10.1016/S0022-2836(03)00894-5

Silver S, Phung LT (1996) Bacterial heavy metal resistance: new surprises. Annu Rev Microbiol 50:753–789. doi:10.1146/annurev.micro.50.1.753

Szpirer C, Top E, Couturier M, Mergeay M (1999) Retrotransfer or gene capture: a feature of conjugative plasmids, with ecological and evolutionary significance. Microbiology 145:3321–3329

Taghavi S, Provoost A, Mergeay M, van der Lelie D (1996) Identification of a partition and replication region in the *Alcaligenes eutrophus* megaplasmid pMOL28. Mol Gen Genet 250:169–179

Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278:631–637. doi:10.1126/science.278.5338.631

Tibazarwa C, Wuertz S, Mergeay M, Wyns L, van der Lelie D (2000) Regulation of the *cnr* cobalt and nickel resistance determinant of *Ralstonia eutropha* (*Alcaligenes eutrophus*) CH34. J Bacteriol 182:1399–1409. doi:10.1128/JB.182.5.1399-1409.2000

Tindall BJ (2008) Rule 15 of the international code of nomenclature of bacteria: a current source of confusion. Int J Syst Evol Microbiol 58:1775–1778. doi:10.1099/ijs.0.2008/005314-0

Tseng T-T, Gratwick KS, Kollman J, Park D, Nies DH, Goffeau A et al (1999) The RND superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. J Mol Microbiol Biotechnol 1:107–125

Vandamme P, Coenye T (2004) Taxonomy of the genus *Cupriavidus*: a tale of lost and found. Int J Syst Evol Microbiol 54:2285–2289. doi:10.1099/ijs.0.63247-0

Vandamme P, Holmes B, Coenye T, Goris , Mahenthiralingam E, LiPuma JJ, Govane JRW (2003) *Burkholderia cenocepacia* sp. nov.—a new twist to an old story. Res Microbiol 154:91–96. doi:10.1016/S0923-2508(03)00026-3

van der Lelie D, Schwuchow T, Schwidetzky U, Wuertz S, Baeyens W, Mergeay M et al (1997) Two component regulatory system involved in transcriptional control of heavy metal homoeostasis in *Alcaligenes eutrophus*. Mol Microbiol 23:493–503. doi:10.1046/j.1365-2958.1997.d01-1866.x

von Rozycki T, Nies DH, Saier MHJ (2005) Genomic analyses of transport proteins in *Ralstonia metallidurans*. Comp Funct Genomics 6:17–56. doi:10.1002/cfg.454

Zhou X, Hvorup RN, Saier MHJ (2003) An automated program to screen databases for members of protein families. J Mol Microbiol Biotechnol 5:7–10

Zoropogui A, Gambarelli S, Coves J (2008) CzcE from *Cupriavidus metallidurans* CH34 is a copper-binding protein. Biochem Biophys Res Commun 365:735–739. doi:10.1016/j.bbrc.2007.11.030