Introduction

Prokaryotes are renowned for their metabolic adaptability and diversity. Many prokaryotes can switch with relative ease between aerobic and anaerobic conditions, for example, or can improvise the use of inorganic compounds, such as hydrogen or sulfur, to form respiratory chains under even the most extreme environmental conditions. This remarkable respiratory adaptability is thanks to an elaborate bank of complex, multi-subunit, cofactor-containing redox enzymes, many of which are embedded within, or located on the extracellular side of, the membrane. The biosynthesis of these enzymes therefore requires carefully controlled expression, assembly, targeting and transport processes. Here, focusing on the molybdenum-containing respiratory enzymes central to anaerobic respiration in *Escherichia coli*, recent descriptions of a chaperone-mediated ‘proofreading’ system involved in coordinating assembly and export of complex extracellular enzymes will be discussed. The paradigm proofreading chaperones are members of a large group of proteins known as the TorD family, and recent research in this area highlights common principles that underpin biosynthesis of both exported and non-exported respiratory enzymes.

The prokaryotic cytoplasmic membrane not only maintains cell integrity and forms a barrier between the cell and its outside environment, but is also the location for essential biochemical processes. Microbial model systems provide excellent bases for the study of fundamental problems in membrane biology including signal transduction, chemotaxis, solute transport and, as will be the topic of this review, energy metabolism. Bacterial respiration requires a diverse array of complex, multi-subunit, cofactor-containing redox enzymes, many of which are embedded within, or located on the extracellular side of, the membrane. The biosynthesis of these enzymes therefore requires carefully controlled expression, assembly, targeting and transport processes. Here, focusing on the molybdenum-containing respiratory enzymes central to anaerobic respiration in *Escherichia coli*, recent descriptions of a chaperone-mediated ‘proofreading’ system involved in coordinating assembly and export of complex extracellular enzymes will be discussed. The paradigm proofreading chaperones are members of a large group of proteins known as the TorD family, and recent research in this area highlights common principles that underpin biosynthesis of both exported and non-exported respiratory enzymes.

The well-studied Gram-negative γ-proteobacterium *Escherichia coli* is a facultative anaerobe and so is typically very flexible in its use of respiratory electron donors and acceptors (Fig. 1). In *E. coli*, respiratory chains are relatively short, with reducing power being shuttled between only two enzymes by lipid-soluble quinones (Fig. 1), namely ubiquinone, menaquinone and demethylmenaquinone (Soballe & Poole, 1999). The generation of energy by respiratory electron-transfer chains involves the ionically sealed cytoplasmic membrane and, in the context of this particular review, it is most notable that respiratory enzymes, or at least their catalytic domains, can be located on either side of the lipid bilayer (Fig. 1). Indeed, in terms of physiological function, the subcellular localization of all of these enzymes is absolutely critical. Some substrates, for example trimethylamine N-oxide (TMAO) and nicotinamide adenine dinucleotide (NADH), are membrane impermeable and therefore must be metabolized on one particular side of the membrane. More important, perhaps, is the application of the ‘redox loop mechanism’ by respiratory chains to develop a transmembrane proton-motive force (Mitchell, 1961; Mitchell & Moyle, 1967; recently reviewed by Jormakka et al., 2003). Redox loops are generated by membrane-associated respiratory enzymes in which the substrate and quinone/quinol-binding sites are juxtaposed on opposite sides of the lipid bilayer. Separation of the active sites in this way allows protons to be released, and taken up, on opposite sides of the membrane and thus contribute to a proton electrochemical gradient. Not all quinone/quinol-dependent respiratory systems are so ‘electrogenic’, but in *E. coli* the activities of two such enzymes are coupled to great effect under anaerobic conditions in the presence of nitrate (Enoch & Lester, 1975; Jormakka et al., 2003). Under these conditions formate oxidation in the periplasm is coupled, through a quinone mediator, to nitrate reduction in the cytoplasm. The formate dehydrogenase enzyme (often termed formate dehydrogenase-N as it is nitrate-inducible) is composed of three polypeptides, FdnG, FdnH and FdnL, with the formate-oxidation site being located on the FdnG subunit at the periplasmic side of the membrane (Jormakka et al., 2002, 2003). The quinone-reducing site is located on FdnL at the periplasmic side of the membrane (Jormakka et al., 2002, 2003). The topology and location of the substrate-binding sites are unequivocal since the 3D crystal structure of the intact complex has been solved to 1.6 Å.

**Constructing the wonders of the bacterial world: biosynthesis of complex enzymes**

Frank Sargent

Centre for Metalloprotein Spectroscopy and Biology, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

The prokaryotic cytoplasmic membrane not only maintains cell integrity and forms a barrier between the cell and its outside environment, but is also the location for essential biochemical processes. Microbial model systems provide excellent bases for the study of fundamental problems in membrane biology including signal transduction, chemotaxis, solute transport and, as will be the topic of this review, energy metabolism. Bacterial respiration requires a diverse array of complex, multi-subunit, cofactor-containing redox enzymes, many of which are embedded within, or located on the extracellular side of, the membrane. The biosynthesis of these enzymes therefore requires carefully controlled expression, assembly, targeting and transport processes. Here, focusing on the molybdenum-containing respiratory enzymes central to anaerobic respiration in *Escherichia coli*, recent descriptions of a chaperone-mediated ‘proofreading’ system involved in coordinating assembly and export of complex extracellular enzymes will be discussed. The paradigm proofreading chaperones are members of a large group of proteins known as the TorD family, and recent research in this area highlights common principles that underpin biosynthesis of both exported and non-exported respiratory enzymes.
resolution (Jormakka et al., 2002; Fig. 2A). The 3D crystal structure of the intact nitrate reductase located at the other end of this short respiratory chain has also been solved to near atomic resolution (Bertero et al., 2003; Fig. 2B). The crystal structure of this enzyme, termed nitrate reductase-A (i.e. the predominant nitrate reductase synthesized by *E. coli*) and comprising the NarG, NarH and NarI subunits, confirmed that the quinol-oxidizing site on NarI was located at the periplasmic side, and that the nitrate-reducing site on the NarG subunit was located at the cytoplasmic side, of the membrane (Fig. 2B). The determination of the structural bases for formate-to-nitrate respiration in *E. coli* has underscored the fundamental importance of enzyme topology in this type of bacterial energy metabolism.

In this review, advances in the understanding of the biosynthetic processes governing assembly and orientation of complex respiratory enzymes will be discussed. In particular, recent genetic, biochemical, biophysical and structural studies of a new family of molecular chaperones (the TorD family) will be used to highlight common principles that underpin biosynthesis of both exported and non-exported respiratory enzymes.

### Molybdenum-dependent respiratory enzymes

The *E. coli* respiratory formate dehydrogenases and nitrate reductases contain molybdenum as a key catalytic cofactor (Enoch & Lester, 1975; Jormakka et al., 2002, 2004; Bertero et al., 2003). The chemical properties of molybdenum, which can cycle between the MoIV and MoVI valence states, have been seized upon by nature and exploited in numerous 2-electron transfer reactions. *E. coli* formate dehydrogenase-N catalyses the dehydrogenation of formate (Khangulov et al., 1998):

\[
\text{HCOO}^- \rightarrow \text{CO}_2 + \text{H}^+ + 2e^{-}
\]

The active site that catalyses this reaction is located on the FdnG subunit and comprises one molybdenum ion ligated by two molybdopterin guanine dinucleotide (bis-MGD) cofactors and a selenocysteine side-chain (Jormakka et al., 2002). Following dehydrogenation of formate, the electrons are presumably passed via one of the MGD cofactors to an Fe–S cluster located in FdnG and then on to the FdnH subunit, which contains four Fe–S clusters, and finally on to FdnI, which has a proximal haem b close to FdnH and a distal haem b close to the quinone-binding site at the cytoplasmic side of the membrane (Jormakka et al., 2002, 2003).

The *E. coli* respiratory nitrate reductase-A catalyses the 2-electron reduction of nitrate to nitrite (an oxotransferase reaction), with the electrons usually being supplied by quinol (Berks et al., 1995; Bertero et al., 2003; Jormakka et al., 2004):
The pathway the electrons take through nitrate reductase-A is precisely the opposite to electron transfer in formate dehydrogenase-N. First, quinol binds close to a distal haem \(b\) within NarI at the periplasmic side of the membrane, and is reduced to quinone, with the electrons passing from the distal haem to a proximal haem \(b\) at the cytoplasmic side of the membrane. The four Fe–S clusters in NarH then channel the electrons to an Fe–S cluster located in the NarG subunit, and finally on to the catalytic molybdenum ion, which is ligated by two MGD cofactors and in this case an aspartate side-chain donated by the NarG polypeptide (Bertero et al., 2003; Jormakka et al., 2004).

When taken together with phylogenetic analyses, the alternative active-site structures of molybdoenzymes such as those of FdnGHI and NarGHI have proven useful in classifying this vast superfamily of enzymes. As a result, the \textit{bis}-MGD-containing catalytic proteins themselves can be grouped together into three different types (McDevitt et al., 2002; Jormakka et al., 2004). Type I \textit{bis}-MGD molybdoenzymes almost always have a cysteine or selenocysteine side-chain as a direct molybdenum ligand (McDevitt et al., 2002; Jormakka et al., 2004), and this group includes \textit{E. coli} FdnG, \textit{E. coli} NapA (the periplasmic nitrate reductase), \textit{Salmonella} TtrA (the catalytic subunit of tetrathionate reductase) (Hensel et al., 1999), and the assimilatory nitrate reductases, for example \textit{Synechococcus} NarB (Rubio et al., 2002). Type II \textit{bis}-MGD molybdoenzymes display an aspartate molybdenum ligand (Bertero et al., 2003) and include \textit{E. coli} NarG and its huge array of homologues such as \textit{E. coli} NarZ, \textit{Paracoccus} NarG, proteins from bacilli, amonoxycetes and the archaea (Richardson et al., 2001; van Keulen et al., 2005), and \textit{Rhodovulum sulfidophilum} DdhA, the catalytic subunit of dimethyl sulfide dehydrogenase (McDevitt et al., 2002). Finally, the type III \textit{bis}-MGD molybdoenzymes possess a serine ligand to the molybdenum and include the \textit{E. coli} TMAO reductases (TorA and TorZ), \textit{E. coli} DmsA (the catalytic subunit of dimethyl sulfoxide (DMSO) reductase), the TMAO/DMSO reductases of \textit{Rhodobacter} sp., and the \textit{E. coli} biotin sulfoxide reductase, BisC (Pierson & Campbell, 1990).

The type I, II and III classification of \textit{bis}-MGD molybdoenzymes, however, does not discriminate between exported and non-exported proteins. For example, under type I there are the periplasmic FdnG and NapA proteins together with the cytoplasmic FdhF and NarB enzymes. Under type II there is the cytoplasmically oriented \textit{E. coli} NarG together with the periplasmic \textit{Rhodobacter} sp., and the \textit{E. coli} biotin sulfoxide reductase, BisC (Pierson & Campbell, 1990).

The type I, II and III classification of \textit{bis}-MGD molybdoenzymes, however, does not discriminate between exported and non-exported proteins. For example, under type I there are the periplasmic FdnG and NapA proteins together with the cytoplasmic FdhF and NarB enzymes. Under type II there is the cytoplasmically oriented \textit{E. coli} NarG together with the periplasmic \textit{Rhodobacter} sp., and the \textit{E. coli} biotin sulfoxide reductase, BisC (Pierson & Campbell, 1990).

The \textit{twist-arginine transport system} Many eubacteria and archaea (and their chloroplast descendants) have the ability to transport pre-folded, very often oligomeric, and enzymically active respiratory...
enzymes across ionically sealed membranes (Palmer & Berks, 2003; Berks et al., 2003). The proteins so transported are usually synthesized with, or associated with folded partner proteins that possess, distinctive N-terminal signal peptides that bear a common ‘twin-arginine’ SRRxFLK amino acid sequence motif (Berks, 1996). Twin-arginine signal peptides have a common tripartite structure that includes a polar N-terminal (n-) region of variable length, a moderately hydrophobic (h-) region of 12–20 amino acids, and a C-terminal (c-) region that often contains basic residues (Table 1). The consensus motif is always located at the junction between the n- and h-regions (Table 1). These conserved common features have led to the development of algorithms designed to identify twin-arginine signal peptides from genomic information (Rose et al., 2002; Bendtsen et al., 2005). E. coli is predicted to produce ~26 such proteins and the vast majority of these are cofactor-containing respiratory enzymes (Berks et al., 2005). Indeed, all of the extracellular molybdoenzyme complexes synthesized by E. coli are either known or predicted to possess a subunit with an N-terminal twin-arginine signal peptide, including the type I molybdoproteins FdnG and NapA, and the type III molybdoenzymes TorA and DmsA (Fig. 1). Some other prokaryotes make more extensive use of the twin-arginine signal peptide; for example the genome of the Gram-positive actinomycete Streptomyces coelicolor is predicted to encode upwards of 250 diverse proteins bearing twin-arginine signal peptides (Dilks et al., 2003; Widdick et al., 2006), and almost all exported proteins expressed by some halophilic archaea are predicted to contain twin-arginine signal peptides (Bolhuis, 2002; Rose et al., 2002).

All proteins bearing bona fide twin-arginine signal peptides are transported by the twin-arginine translocation (Tat) system (Sargent et al., 1998a; Palmer & Berks, 2003). The Tat translocase is essentially a membrane-bound nanomachine dedicated to the transmembrane translocation of fully folded proteins. Depending on the biological model system under investigation, the complete Tat translocation machinery comprises two or three different proteins. In E. coli, three types of integral membrane proteins – TatA (and its homologue TatE), TatB, TatC – have been identified that are involved in the transport process (Bogsch et al., 1998; Sargent et al., 1998a, 1999; Weiner et al., 1998). TatA/E and TatB are predicted to be structurally related, each comprising a single N-terminal transmembrane α-helix with the extreme N-terminus on the periplasmic side of the membrane (Settles et al., 1997; Chanal et al., 1998; Lee et al., 2006). The TatC protein has six transmembrane helices, with both termini at the cytoplasmic side of the membrane (Drew et al., 2002; Behrendt et al., 2004; Ki et al., 2004). Some other bacteria, including the Gram-positive Bacillus subtilis, have minimalist Tat translocation systems that comprise only TatA and TatC (Jongbloed et al., 2006). Indeed, the Gram-positive Staphylococcus aureus has a genome that lacks a tatB homologue altogether (Pop et al., 2002). In E. coli TatB is essential for physiological Tat

Table 1. Classification of twin-arginine signal peptides

<table>
<thead>
<tr>
<th>Prokaryote</th>
<th>Protein</th>
<th>Peptide Sequence</th>
<th>Cleaved</th>
<th>Chaperone</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A-Class Signal Peptides (Targeting and Biosynthesis)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli</td>
<td>TorA</td>
<td>1-MMNNDIFQSSRRPKLAGLLGTVAGMGLGSLPIEPRPA-39</td>
<td>Yes</td>
<td>TorD</td>
</tr>
<tr>
<td>E. coli</td>
<td>DmsA</td>
<td>1-KKFIPDVALAAESRPGVLWTAIGGLANASSALTITPFISHRAA-45</td>
<td>Yes</td>
<td>DmsD</td>
</tr>
<tr>
<td>A. fulgidus</td>
<td>NarG</td>
<td>1-KKSRRFPLQASSAPAASGLGYESQSRG-32</td>
<td>unknown</td>
<td>NarJ*</td>
</tr>
<tr>
<td>H. mediterranei</td>
<td>NarG</td>
<td>1-MADGTYSSRTTLSGIGVALSLLGGTSAADGSLFG-36</td>
<td>unknown</td>
<td>NarJ*</td>
</tr>
<tr>
<td>P. aerophilum</td>
<td>NarG</td>
<td>1-LLHTIPPLAXCVAATISAAAMMALAANIQYQ-33</td>
<td>unknown</td>
<td>NarJ*</td>
</tr>
<tr>
<td>Ha. marismortui</td>
<td>NarG</td>
<td>1-DSNRLTDEDSAGISRDPFVRGLGAASLGLATGLSPDOGG-43</td>
<td>unknown</td>
<td>NarJ*</td>
</tr>
<tr>
<td><strong>B-Class Signal Peptides (Targeting Only)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli</td>
<td>AmiC</td>
<td>1-DGGNTAGSRPRLGQAGAWLWLLCVQYRSA-31</td>
<td>Yes</td>
<td>none to date</td>
</tr>
<tr>
<td>S. coelicolor</td>
<td>DagA</td>
<td>1-TNRRDLIKRSAVALGAGAGLACPAAHPH-30</td>
<td>Yes</td>
<td>none to date</td>
</tr>
<tr>
<td><strong>C-Class Remnant Signals (Biosynthesis Only)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli</td>
<td>NarG</td>
<td>1-KSFIPDRFYEFKKGETFADGHOQLTFNRRWEDGVY-36</td>
<td>No</td>
<td>NarJ</td>
</tr>
<tr>
<td>B. subtilis</td>
<td>NarG</td>
<td>1-KKKKRSFLPPRNYTSPETHHNSHKHHRTREDQGWIVY-40</td>
<td>No</td>
<td>NarJ</td>
</tr>
<tr>
<td>P. pantotrophus</td>
<td>NarG</td>
<td>1-MSLDDPMFLKSTKDVPSGEHGUHHTTENRRWEDDG-37</td>
<td>No</td>
<td>NarJ</td>
</tr>
<tr>
<td>E. coli</td>
<td>BiaC</td>
<td>1-MANSYRSLTAHWWGLVETDQTCSSRGAL-36</td>
<td>No</td>
<td>none to date</td>
</tr>
<tr>
<td><strong>N-tail</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T. thermophilus</td>
<td>NarG</td>
<td>1-MRDIKEVENPAPEPWEEK-20</td>
<td>No</td>
<td>none to date</td>
</tr>
</tbody>
</table>

*Chaperones marked ‘?’ are predicted to be peptide binding proteins.
transport activity and seems to form a link between TatA and TatC in the inner membrane (Sargent et al., 1999; Bolhuis et al., 2001).

Two distinct Tat complexes can be isolated from resting E. coli inner membranes. TatA can be purified as a large, heterogeneous complex of up to 600 kDa (McDevitt et al., 2005) that also contains trace amounts of TatB (Sargent et al., 2001; de Leeuw et al., 2002; Porcelli et al., 2002; Oates et al., 2005). The TatC protein is found in another large complex with an equimolar amount of TatB (Bolhuis et al., 2001; de Leeuw et al., 2002; Oates et al., 2003; McDevitt et al., 2005; Richter & Brüser, 2005). TatBC is the ‘signal recognition module’ of the Tat system and contains the recognition site for the twin-arginine motif on the signal peptide (Mori & Cline, 2002; Alami et al., 2003). Low-resolution structural analysis points to the TatA protein forming the ‘transport module’ of the Tat system – i.e. numerous bundles of TatA may form the protein-conducting channel (Gohlke et al., 2005; Dabney-Smith et al., 2006). During protein transport a transient TatABC complex is thought to form in the membrane, but then rapidly dissociate again once the transport step has been completed (Mori & Cline, 2002). The whole system is probably powered by the transmembrane proton-motive force (Alder & Theg, 2003), although this is controversial (Finazzi et al., 2003; Theg et al., 2005).

The establishment of the Tat pathway as an exciting, vibrant and fast-moving research area was preceded by a long gestation period. During that time microbiologists studying complex respiratory enzymes recognized the importance of the signal peptide and its twin-arginine motif (e.g. van Dongen et al., 1988; Nivière et al., 1992; Bernhard et al., 1996; Dreusch et al., 1997) and, with a hint of irony, some had already isolated tat mutant strains and even sequenced the genes (e.g. Friedrich et al., 1981; Yates et al., 1997). In parallel, plant scientists were characterizing biochemically a thylakoid protein transport system for folded proteins in which twin-arginine signal peptides were critical to the targeting process (e.g. Creighton et al., 1995; Chaddock et al., 1995). Mutant plants defective in this transport system had also already been isolated (Voelker & Barkan, 1995). Any possible commonality between the work of the microbial and plant research groups was originally dismissed because the substrate proteins transported by the plant thylakoid system were not present in cyanobacteria or any other prokaryotes. The seminal hypothesis of Berks (1996), however, raised the possibility that a prokaryotic protein transport system for folded proteins was indeed waiting to be unearthed. The race to identify components of the translocase was won by the plant scientists, who cloned and sequenced a gene subsequently shown to encode a homologue of E. coli TatB (Settles et al., 1997; Weiner et al., 1998; Sargent et al., 1998a; Chanal et al., 1998).

The central dogma of the Tat research field is that this system has evolved for the transmembrane translocation of fully folded proteins. That is not to say the Tat translocase cannot transport unfolded or misfolded proteins, as this has been clearly shown for the thylakoid Tat system (e.g. Hynds et al., 1998). Nevertheless, in the case of eubacteria such as E. coli where the majority of Tat substrates are cofactor-containing, multi-subunit, respiratory enzymes, and in halophilic archaea, where protein folding in the extracellular environment would be biophysically impossible, the ultimate challenge in the Tat research field is to understand the molecular mechanism of the transport of fully folded proteins.

**Biosynthesis of TorA, a paradigm Tat-targeted respiratory enzyme**

The ultimate aim for a bacterial physiologist would be to understand how a multi-subunit, multi-cofactor enzyme like formate dehydrogenase-N is assembled and transported across a membrane. In the first step, the Fe–S cluster and molybdenum cofactor must be loaded into FdnG (which also contains the twin-arginine signal peptide). At the same time the four Fe–S clusters must be loaded into FdnH. The FdnGH complex must then be formed in the cytoplasm before transport of the dimer via the Tat system to the periplasm. Next, part of FdnH must be integrated into the membrane (Hatzixanthis et al., 2003) before FdnGH associates with FdnI (which will have already followed a different biosynthetic pathway) and finally forms a ‘trimer-of-trimers’ in the membrane. All this must occur before a single molecule of formate is oxidized. The complexity of this ‘protein production line’ raises many questions about ‘quality control’ and coordination of different biological events. Why, for example, is the FdnG protein not exported without its FdnH partner? Because of the myriad of different processes that need to be considered when studying the biosynthesis of a large enzyme like formate dehydrogenase-N, most of the early research in this area focused on a simpler, more tractable system.

The E. coli TMAO reductase (TorA) enzyme is a type III molybdoenzyme (McDevitt et al., 2002) and is a water-soluble periplasmic protein that contains bis-MGD as its only cofactor. Crystal structures of TorA homologues from Shewanella massilias (Czjzek et al., 1998), Rhodobacter capsulatus (Schneider et al., 1996; McAlpine et al., 1998) and Rhodobacter sphaeroides (Schindelin et al., 1996) are available. The E. coli enzyme is abundant under anaerobic conditions in the presence of TMAO, an osmoprotectant found widely in marine animals (McCrindle et al., 2005), and the evidence for cytoplasmic assembly and folding of TorA prior to export is strong (reviewed by Berks et al., 2003).

The biosynthesis of the MGD cofactor itself has been extensively studied and crystal structures are now available for most of the biosynthetic enzymes. This entire process requires the elaborate use of nucleotide cofactors and so occurs exclusively in the cytoplasm (reviewed by Schwarz, 2005). To summarize, the biosynthesis of the cofactor is mediated by 10 proteins each given the prefix ‘Mo-‘.
First, the MoaAC complex rearranges GTP into a pterin called cyclic pyranopterin monophosphate (cPMP) in a process requiring S-adenosylmethionine. Thiol groups are next added to cPMP in an elaborate process driven by protein adenylation and involving MoaD, MoaE and Moeb. The MogA and MoaB proteins then catalyse the ligation of molybdenuim to the thiol groups to generate molybdopterin (MPT) and this process involves an adenylation/de-adenylation cycle of the cofactor itself. Finally, a GMP moiety is added to MPT by the MobA protein to generate MGD in a process that may be enhanced by MobB, a cytoplasmic GTPase. The take-home messages from this are that this elaborate cofactor must be constructed in the regulated environment of the cytoplasm, and that its extreme susceptibility to oxidation means that it is very unlikely to be secreted unprotected to the cell exterior.

The next pressing issue is the insertion of the completed MGD cofactor into the TorA apoenzyme. This event occurs before transport of TorA to the periplasm because if the above cofactor biosynthesis pathway is genetically blocked the *E. coli* TorA apoprotein remains in the cytoplasm (Santini *et al.*, 1998). The experiments of Santini *et al.* (1998) are compelling and helped to pave the way for the characterization of the bacterial Tat system (Sargent *et al.*, 1998a; Weiner *et al.*, 1998); however, it should be noted that the *Rhodobacter* TorA homologue behaves quite differently (Buchanan *et al.*, 2001 and references therein). In this case the enzyme is still transported to the periplasm in the absence of cofactor. The most likely explanation is that this particular enzyme folds around GMP instead, as also seen in the absence of cofactor. The most likely explanation is that this elaborate cofactor must be constructed in the regulated environment of the cytoplasm, and that its extreme susceptibility to oxidation means that it is very unlikely to be secreted unprotected to the cell exterior.

**The TorD family: structural analyses**

*E. coli* TorD is encoded by the final gene on the *torCAD* operon (Mejean *et al.*, 1994) and is a member of a vast family of related bacterial and archaeal proteins (Turner *et al.*, 2004; Ilbert *et al.*, 2004). The sequence databases currently contain hundreds of members of the TorD family which, after sequence analysis, broadly separate into three ‘clades’: the TorD clade, the DmsD clade and the NarJ clade (Turner *et al.*, 2004). The *E. coli* genome encodes a total of five members of the wider TorD family – TorD itself, DmsD and YcdY from the DmsD clade, and NarJ and NarW from the NarJ clade (Turner *et al.*, 2004). When overproduced and purified, TorD family proteins can adopt several oligomeric and/or conformational states (Tranier *et al.*, 2002; Sarfo *et al.*, 2004). While the monomeric form always dominates, homodimers and homotrimers have been clearly identified and the 3D crystal structure of a dimer form of the TorD homologue from *Shewanella massilia* has been solved by X-ray methods (Tranier *et al.*, 2003). The structure of the *Shewanella* TorD dimer is interesting in several respects, not least because it contains no extensive regions of β-sheet and, apart from some unstructured loops, is therefore made up entirely of 10 α-helices (Fig. 3A).

In addition, the *Shewanella* TorD dimer exhibits domain-swapping to the extreme (Fig. 3A; Tranier *et al.*, 2003). Each *Shewanella* TorD protomer consists of two distinct domains linked by a ‘hinge’ region. The N-terminal domain, comprising the first six α-helices, from one protomer interacts with the C-terminal domain, comprising the last four α-helices, of the second protomer, and Tranier *et al.* (2003) postulated that, based on small-angle solution X-ray scattering, the monomer form of TorD would involve more-or-less identical intra-protomer interactions. The 3D crystal structure of the *Salmonella typhimurium* LT2 DmsD monomer (Fig. 3B) corroborates this hypothesis completely [Y. Qiu, R. Zhang, V. Tereshko, Y. Kim, F. Collart, A. Joachimiak & A. Kossiakoff, unpublished; deposited in Protein DataBank in 2004 under 1S9U]. Although sharing only ~15% overall sequence identity with *Shewanella* TorD, *Salmonella* DmsD also displays an all-α-helical secondary structure (Fig. 3B) and has an analogous fold to a single ‘lobe’ of the domain-swapped *Shewanella* TorD dimer (Fig. 3A).

A third crystal structure of a member of the TorD family is also available (Fig. 3C). The crystal structure of the Af0173 monomer from the hyperthermophilic archaeon *Archaeoglobus fulgidus* has been solved in a structural genomics project [O. Kirillova, M. Chruszc, T. Skarina, E. Gorodichkenskaia, M. Cymborowski, I. Shumilin, A. Savchenko, A. Edwards & W. Minor, Midwest Center For Structural Genomics, unpublished; deposited in Protein DataBank in 2005 under 1ZE0]. The Af0173 protein is a peripheral member of the NarJ clade (Turner *et al.*, 2004). Despite some sequence similarity with DmsD proteins (Turner *et al.*, 2004), *A. fulgidus* Af0173 is probably best classified as a NarJ since it is encoded by a four-cistron...
cluster also containing the genes for a NarG-type nitrate reductase Af0176 (Richardson et al., 2001; Dridge et al., 2006), an Fe–S protein with limited homology with NarH (Af0175), and an unusual integral membrane protein, Af0174 (Dridge et al., 2006). Thus it is clear from the 3D structures available that a common fold is conserved across this entire family of proteins (Fig. 3).

The TorD family: twin-arginine signal peptide-binding proteins

The identification of \textit{E. coli} DmsD as a cytoplasmic twin-arginine signal peptide-binding protein was a watershed in the understanding of how TorD family proteins work. The \textit{E. coli} DMSO reductase is a trimeric complex comprising the DmsA catalytic subunit, which contains bis-MGD and an Fe–S cluster as cofactors, the DmsB protein, which contains four Fe–S clusters, and an unusual integral membrane protein termed DmsC (Bilous et al., 1988). The DmsA protein contains a twin-arginine signal peptide at its N-terminus (Berks, 1996; Sambasivarao et al., 2000) and this single signal peptide directs the export of the DmsAB dimer to the periplasmic side of the membrane, where it finally attaches to the DmsC membrane protein to form the holoenzyme (Stanley et al., 2002). Oresnik et al. (2001) developed an affinity chromatography method to detect proteins that bound tightly to the \textit{E. coli} DmsA twin-arginine signal peptide, and the primary interactor was identified as Ynfl (re-named DmsD). The DmsD protein is encoded by the \textit{ynfEFGHdmSC} operon for a second (essentially cryptic) DMSO reductase in \textit{E. coli} (Lubitz & Weiner, 2003). A study of the Tat transport activity of the DmsA signal peptide in the absence of \textit{dmsD} suggested that DmsD does not have an obligatory role in transport per se (Ray et al., 2003), and therefore does not operate as a kind of ‘signal recognition particle’. However, DmsD was certainly found to be essential for cellular DMSO respiration and DMSO reductase activity, which demonstrated that the signal-peptide-binding event was, in this case, critical for the biosynthesis of the native substrate (Ray et al., 2003).

The sequence similarity between DmsD and TorD suggested strongly that TorD could also be a twin-arginine signal peptide-binding protein (Sargent et al., 2002). This was initially confirmed genetically using a bacterial two-hybrid screen with \textit{torD} as ‘bait’ against a random genomic library as ‘prey’ (Jack et al., 2004). The strongest interacting clones identified by this completely blind screen encoded the Tat signal peptide of TorA (Jack et al., 2004). The powerful biophysical technique of calorimetry was subsequently used to characterize \textit{in vitro} the interaction between TorD and a synthetic TorA signal peptide. The isolated \textit{E. coli} TorD monomer binds to the TorA signal peptide with an apparent dissociation constant \((K_d)\) of \(\sim 1\ \mu M\) (Hatzixanthis et al., 2005). By comparison, recent calorimetric experiments have shown \textit{E. coli} DmsD to bind the DMSO reductase Tat signal peptide with an apparent \(K_d\) of 0.2 \(\mu M\) (Winstone et al., 2006). Optimum ligand binding to TorD requires a 27-residue fragment of the TorA signal peptide stretching from the twin-arginine motif, through the h-region, to another arginine dipeptide at the C-terminus (Hatzixanthis et al., 2005). The nature of the peptide-binding site has not yet been visualized for any TorD family member, and it is not immediately obvious from the available crystal structures where the peptide-binding site might be located. It is likely, however, that the TorA signal peptide is unstructured in aqueous solution, as this has been shown to be the case for the twin-arginine signal peptides of the \textit{Allochromatium vinosum} high-potential iron protein (HiPIP) (Kipping et al., 2003) and the \textit{E. coli} SufI protein (San Miguel et al., 2003). Note, however, that such an unstructured peptide would be excessively lengthy when compared to the folded, globular, TorD monomer. To be bound in its entirety by TorD, the TorA signal peptide will certainly need to adopt a high degree of secondary structure. Given the all-helical nature of TorD family proteins (Fig. 3), it is possible the signal peptide could also display an \(\alpha\)-helical conformation in the TorD-bound state, such as occurs when the signal peptide of the Sec substrate LamB is bound by the SecA protein (Chou & Gierasch, 2005).

![Fig. 3. Structures of TorD, DmsD and NarJ.](http://mic.sgmjournals.org)
What is TorD actually doing when it binds to the TorA signal peptide? One possibility is that the binding event simply protects the exposed, unstructured, signal peptide from degradation by cytoplasmic proteases prior to export (Li et al., 2006; Genest et al., 2006a). For example, when the native TorA enzyme was overproduced in the absence of TorD the signal peptide was prematurely cleaved within the c-region betwixt the arginine dipeptide at positions 35 and 36 (Table 1). However, co-expression of TorA with TorD prevented this degradation, presumably because chaperone-binding sterically hindered attack by a protease (Genest et al., 2006a). It is certainly true that premature degradation of the TorA precursor would undoubtedly be deleterious to the assembly process. The physiological relevance of this particular observation, however, remains in question since the protease cleavage site on TorA appears to fit exactly with the OmpT recognition sequence (R/K)↓(R/K)-A (Dekker et al., 2001), and OmpT is an outer-membrane protease that would not normally be exposed to cytoplasmic substrates (Vandeputte-Rutten et al., 2001). Moreover, experiments with the TorA signal peptide fused to a [NiFe] hydrogenase suggest a more sophisticated modus operandi for TorD (Jack et al., 2004).

E. coli hydrogenase-2 is a heterodimer of a nickel-containing α-subunit (HybC) and an Fe–S cluster-binding β-subunit (HybO) (Sargent et al., 1998b). Only the HybO protein bears a twin-arginine signal peptide and the αβ dimer is transported as a pre-formed active complex by the Tat translocase (Rodrigue et al., 1999; Sargent et al., 1998a). This observation raised many questions, one of the most compelling being: what prevents premature export of HybO without its HybC partner? One possible answer came some years later when Jack et al. (2004) swapped the native signal peptide of E. coli HybO for that of TorA. The assembly of the TorA::hydrogenase-2 chimera was disrupted such that the HybO subunit carrying the TorA signal peptide was targeted prematurely to the membrane without its HybC partner and as a result cellular hydrogenase activity dropped dramatically in this strain (Jack et al., 2004). Most remarkably, perhaps, this mutant phenotype could be rescued by overproduction of TorD – cellular hydrogenase activity returned since the HybOC dimer was once again correctly assembled and exported (Jack et al., 2004). In this case, protection of the TorA signal peptide from proteolysis was probably not an issue because the TorA::HybO fusion was not being degraded and was still being targeted in a Tat-dependent manner (Jack et al., 2004). Instead, TorD was somehow coordinating assembly and export of the hydrogenase. This result had wide-ranging connotations. First, it meant that what was previously assumed to be a private chaperone for a molybdenum enzyme (e.g. Pommier et al., 1998), could actually be employed in the maturation of a completely ‘alien’ enzyme with no molybdenum cofactors at all. Second, it meant that a significant part of TorD’s biological activity was focused solely on the TorA signal peptide with scant regard to the passenger protein to which the signal peptide was attached. Indeed, this finding was corroborated when Li et al. (2006) showed that extra copies of torD could enhance the levels of periplasmic jellyfish green fluorescent protein (GFP) when it was routed through the Tat pathway by the TorA signal peptide. Third, the action of TorD on the TorA signal peptide must mimic almost exactly a biochemical process normally central to the native hydrogenase assembly pathway (probably the pre-cursor-binding activity of the HybE protein: Dubini & Sargent, 2003; Jack et al., 2004; Butland et al., 2006). These papers, and others, led to the hypothesis that TorD was acting as a ‘Tat proofreading chaperone’, which is actually preventing protein export by shielding the signal peptide from the Tat translocase until all assembly processes are complete (Santini et al., 1998; Wu et al., 2000; Sargent et al., 2002; Dubini & Sargent, 2003; Jack et al., 2004, 2005).

The phrase ‘proofreading’ in this context of enzyme assembly is akin to a ‘quality control’ process (i.e. checking that folding is complete, as opposed to scanning for specific ‘mistakes’). However appropriate here, it should be noted well that the term ‘quality control’ in the Tat transport field had already been coined by Delisa et al. (2003) for a mechanism by which the TatABC translocase itself accepts or rejects proteins for export based on their folded state. As a result the term ‘Tat proofreading’ is preferred to distinguish the pre-export chaperone-mediated system described here (Jack et al., 2005).

Under laboratory conditions the stringency of different Tat proofreading systems is variable. The phenotype of an E. coli dmsD mutant is severe and results in complete disruption of DMSO reductase biosynthesis and blockage of all anaerobic respiration with DMSO as electron acceptor (Oresnik et al., 2001; Ray et al., 2003). This is different to that observed for an E. coli torD deletion mutant, in which respiration with TMAO is not significantly hampered and TorA levels are only reduced by around 50 % (Pommier et al., 1998; Jack et al., 2004). The differences are probably not due to the different structures of the TMAO and DMSO reductases in E. coli since disruption of the single dorD gene from Rhodobacter capsulatus, which encodes a TorD homologue, was enough to completely abolish activity of the cognate TorA-type reductase (Shaw et al., 1999). These discrepancies are not easy to explain, although it should be considered that another member of the TorD family may be participating in TorA maturation in E. coli in the absence of TorD.

The involvement of nucleotides in TorD function

Clearly, TorD binds tightly and specifically to the TorA signal peptide in an energetically favourable reaction, i.e. with a large negative Gibbs free energy change (Hatzixanthis et al., 2005). In the cell a physiological ‘Tord cycle’ could be envisaged in which the Tat proofreading chaperone would first bind, but then release, the TorA signal peptide once cofactor loading and protein folding were complete. What, if anything, triggers signal peptide release by TorD? It is certainly possible that, akin to the mitochondrial protein import system (Stan et al., 2000), the TorA signal peptide...
would be naturally ‘handed over’ from TorD to the TatBC signal-binding module if TatBC has a significantly lower $K_d$ (and thus a more negative Gibbs free energy for binding) for the signal peptide than TorD. To date, rigorous experiments to determine the binding affinity of twin-arginine signal peptides to TatBC have not been reported. Moreover, this particular model lacks any obvious control mechanism. An alternative hypothesis was put forward by Hatzixanthis et al. (2005) and was based on the known molecular mechanisms of other peptide-binding proteins: given the nucleotide-based mechanisms of the signal recognition particle (Doudna & Batey, 2004; Focia et al., 2004), the SecA protein (Vrontou & Economou, 2004) and the DnaK–DnaJ–GrpE system (Ellis & Hartl, 1999), TorD was also tested for its ability to interact with nucleotide triphosphates (Hatzixanthis et al., 2005).

By monitoring intrinsic tryptophan fluorescence, isolated TorD was shown to interact loosely ($K_d \sim 375 \mu M$) with GTP in vitro (Hatzixanthis et al., 2005). No GTP hydrolysis was observed, but the affinity for GTP was increased (to an apparent $K_d \sim 200 \mu M$) by pre-incubation of TorD with the TorA signal peptide (Hatzixanthis et al., 2005). This finding was controversial because TorD is an all-helical protein (Fig. 3) and as such does not display a classic nucleotide-binding Rossmann fold (Rossmann et al., 1975) or canonical ‘Walker-A’ (or ‘P-loop’) or ‘-B’ nucleotide-binding amino acid motifs (Walker et al., 1982). Nevertheless, inspection of the primary sequences of TorD family proteins does suggest some similarity with metal-dependent phosphohydrolases. The interdomain ‘hinge’ region of TorD family proteins contains a very highly conserved EPxDH amino acid motif where the aspartate is invariable and the histidine is only very occasionally substituted by tyrosine (Tranier et al., 2003; Ilbert et al., 2004; Turner et al., 2004). Indeed, in E. coli TorD the D-124 and H-125 residues within this signature have been shown to be essential in the in vitro Tat proofreading process (Jack et al., 2004). This motif is highly reminiscent of those identified in a range of phosphohydrolases including the GNH[DE] superfamily (Aravind & Koonin, 1998b), the HKD phospholipases (Koonin, 1996) and the DHH phosphoesterases (Aravind & Koonin, 1998a). Most notable of all, perhaps, is the HD superfamily of phosphohydrolases, which includes dGTPases, ppGpp hydrolases and cyclic-di-GMP phosphodiesterases (Aravind & Koonin, 1998c; Galperin et al., 2001; Yakunin et al., 2004; Ryan et al., 2006). The aspartate and histidine side-chains in these systems are usually the ligands for magnesium or manganese cofactors (e.g. Yamagata et al., 2002), ions that are also intimately linked with classical nucleotide-hydrolysing systems (e.g. Walker et al., 1982). It is notable, of course, that the available crystal structures (Fig. 3) do not show any opportunistic metal binding at the EPxDH motifs and it has also been shown that GTP binding by E. coli TorD is magnesium-independent (Hatzixanthis et al., 2005). Therefore, if the EPxDH motif is not a metal-binding site it is also worth considering that members of the J-domain co-chaperone family (including E. coli DnaJ) contain a functionally important conserved DPH tripeptide in their sequences (e.g. Mayer et al., 1999). As with the TorD EPxDH motif, the DnaJ DPH signature is located in a loop between two $\alpha$-helices, and, most importantly, has been shown to be involved in mediating protein–protein interactions with DnaK (e.g. Mayer et al., 1999).

It is worth considering at this point the alternative possibility that the observed GTP binding by E. coli TorD (Hatzixanthis et al., 2005) is more closely related to the role of TorD in assisting molybdenum cofactor insertion into TorA (Pommier et al., 1998) rather than Tat proofreading (Jack et al., 2004). Since the molybdenum cofactor in question (MGD) is a pyranopterin dinucleotide where the pyranopterin moiety is itself derived from rearranged GTP and is linked via a phosphate bridge to another guanosine (Schwarz, 2005), is it possible Hatzixanthis et al. (2005) have identified an MGD-binding site on TorD rather than a GTP-binding pocket? The available evidence argues against this. First, the $K_d$ for GTP binding would seem to be too high to account for efficient binding of MGD, of which there is apparently very little in the cell (Johnson et al., 1991). Second, the Tat proofreading activity of TorD has been shown to be biochemically distinct from the cofactor insertion activity (Jack et al., 2004). In contrast, Hatzixanthis et al. (2005) have shown that GTP binding by TorD is intimately linked to signal peptide binding since its presence served to increase the affinity of TorD for GTP about twofold. Third, molybdoenzymes purified in the absence of MGD cofactor often bind GMP or GDP in their active sites by default (e.g. Rothery et al., 2004; Temple & Rajagopalan, 2000). Opportunistic insertion of the more abundant GTP into vacant MGD-binding sites has never been reported. Finally, further molecular modelling suggests that the pyranopterin moiety of MGD does not fit into the hypothetical GTP-binding pocket described by Hatzixanthis et al. (2005). Molecular modelling can, of course, superimpose the guanosine moiety of MGD onto GTP in the putative binding site; however, numerous apparent steric clashes prevent binding of the pyranopterin half of the molecule (A. Oubrie & F. Sargent, unpublished).

The latest working model for the TorD cycle (as proposed by Hatzixanthis et al., 2005) takes into account the direct signal peptide-binding activity of this protein and its homologues, as well as the putative binding of GTP (Fig. 4). Since affinity for GTP is increased by pre-incubation of TorD with the TorA signal peptide (and pre-loading of TorD with GTP did not enhance signal peptide binding), it is proposed that TorD binds the TorA signal first, then GTP binds, then either GTP hydrolysis or nucleotide exchange drives a conformational change in TorD that results in signal peptide release (Fig. 4). The missing link in this model, however, is the trigger for signal release – how does TorD know when to release the signal peptide; how does it know TorA maturation is complete? It seems unlikely that TorD can specifically sense the TorA protein itself, otherwise the experiments with hydrogenase (Jack et al., 2004) and GTP
(Li et al., 2006) fusions to the TorA signal peptide would likely not work. It is possible, if TorD binds close enough to the junction between signal peptide and mature domain as some studies suggest (Genest et al., 2006a), that folding of the passenger protein could sequester part of the binding site and thus force off the chaperone, although it should be noted that the free energy released in most protein-folding events is relatively small. There is some evidence to suggest that the TorD family protein DmsD interacts with the Tat machinery at the cytoplasmic membrane (Papish et al., 2003). It is therefore conceivable that the action of a component of the Tat system could induce signal peptide release and/or GTP hydrolysis. One attractive alternative hypothesis, however, is that TorD operates as a ‘molecular clock’ (Berks et al., 2005), perhaps along similar lines as described for proteins involved in flagellar biosynthesis in Salmonella (Moriya et al., 2006). In this model the signal peptide release rate (effectively the dissociation rate constant, K_{off}) would coincide with the rate of assembly of the passenger protein.

**Dual functionality of twin-arginine signal peptides**

It is now clear from signal-swapping experiments (Sambasivarao et al., 2000; Jack et al., 2004), and from the recent studies of the chaperone-mediated Tat proofreading system described above, that some twin-arginine signal peptides have a separate biological role distinct from that of protein targeting: biosynthesis. As a result, twin-arginine signal peptides have been separated into two broad classes, ‘class 1’ containing those signal peptides with dual functions in biosynthesis (i.e. chaperone binding) and targeting, and ‘class 2’ containing those with a role in targeting only (Sargent et al., 2006). The terminology chosen to classify these signal peptides, however, was not altogether satisfactory and could conceivably cause some confusion with the different classes of Sec signal peptide (e.g. Tokuda & Matsuyama, 2004). It is therefore proposed here to dub instead the dual-function Tat signal peptides as ‘A-class’ and the targeting-only Tat signal peptides as ‘B-class’ (Table 1).

**Common principles in the assembly of molybdoenzymes...**

If dual-function A-class Tat signal peptides (transport and biosynthesis) can be identified, and mono-function B-class Tat signal peptides (transport only) have been identified, is it possible to close the loop and identify mono-function ‘C-class’ peptides that would have a biosynthetic function but have no role in protein transport? The answer is probably ‘yes’.

The N-terminal regions of non-exported eubacterial (type II) NarG proteins show a striking resemblance to twin-arginine signal peptides (Table 1) (Turner et al., 2004). The crystal structure of the non-exported NarG protein from E. coli shows a long N-terminal (‘N-tail’) domain (Bertero et al., 2003) that has amino acid signatures that follow so closely the canonical SRRxFLK Tat motif that it has been termed a ‘vestigial’ (or perhaps ‘remnant’ is a better description), twin-arginine signal peptide (Turner et al., 2004). Similar remnant signal peptides can be identified on the NarG homologues from Paracoccus pantotrophus and Bacillus subtilis (Table 1). Indeed, a remnant signal is also present on the type III molybdoenzyme biotin sulfoxide reductase (BisC) from E. coli, a close homologue of TorA.
(Table 1). The telling substitution of the second arginines by hydrophobic residues within the remnant twin-arginine motifs, and the peppering of the remnant ‘h-regions’ by acidic and polar residues, means that there is no argument that these N-terminal peptides are Tat inactive and that the enzymes in question are definitely cytoplasmic in location. So what functions do these remnant signals have?

As with the A-class Tat signal peptides, the remnant signal peptides are most likely the primary binding sites for biosynthetic accessory proteins. Both in vivo bacterial two-hybrid analysis (Vergnes et al., 2006) and in vitro calorimetry (Chan et al., 2006) have clearly demonstrated that the E. coli NarG remnant signal is recognized by, and bound directly to, the E. coli NarJ protein. The NarJ protein is essential for cellular NarG activity (Blasco et al., 1998) and, most strikingly, is distantly related to the bona fide twin-arginine signal peptide-binding proteins TorD and DmsD (Turner et al., 2004). Furthermore, NarJ binds to the NarG remnant signal with similar thermodynamics, and with a similar apparent dissociation constant (160 nM), as TorD and DmsD bind to their respective A-class signal peptides (Hatzixanthis et al., 2005; Chan et al., 2006; Winstone et al., 2006). It is proposed here that the remnant signals are perhaps so closely related in structure and function to dual-function (biosynthesis and targeting) A-class twin-arginine signal peptides that they should be grouped together as mono-function (biosynthesis only) C-class peptides (Table 1). Note that a biosynthetic chaperone for the remnant signal of the E. coli type III molybdoenzyme BisC has not yet been discovered. The genetic omission of this C-class peptide, however, led to a complete loss of cellular biotin sulfoxide reductase activity (Ezraty et al., 2005), suggesting strongly that it plays a central role in assembly of this enzyme.

Type II molybdoenzymes are distributed throughout the prokaryotic kingdoms, including the archaea. The halophilic archaeon Haloferax mediterranei, for example, encodes a NarG-type respiratory nitrate reductase (Lledo et al., 2004). Most notably, however, H. mediterranei NarG is synthesized with an apparent N-terminal twin-arginine signal peptide, pointing to the possible transport of NarGH to the extracytoplasmic side of the membrane in this archaeon (Table 1). Moreover, the NarG subunits of the archaea Pyrobaculum aerophilum, Archaeoglobus fulgidus and Haloarcula marismortui are also predicted to have N-terminal twin-arginine signal peptides (Table 1) (Afshar et al., 2001; Richardson et al., 2001; Yoshimatsu et al., 2007). It is possible that these archaeal signal peptides are not processed (Table 1), which may point to a role as N-terminal signal anchors as has been suggested for the Tat signals of the bacterial Rieske proteins (Berks et al., 2005; Bachmann et al., 2006). Note, however, that both the P. aerophilum and A. fulgidus genomes do encode proteins (Pae0264 and Af1655, respectively) bearing some similarity to signal peptidases from Gram-positive eubacteria, suggesting that signal peptide cleavage is at least feasible in these organisms. Nevertheless, the presence of what appear to be intact twin-arginine signal peptides on archaeal NarG proteins serves to reinforce the hypothesis that eubacterial NarG proteins do indeed display N-terminal remnants of Tat signal peptides. In addition, it is important to recognize that NarJ chaperone proteins are also encoded by each of the archaeal nar gene-clusters discussed here. It is therefore very likely that these NarJ homologues will recognize and bind to their cognate NarG Tat signal peptides, and as a result all of the archaeal NarG twin-arginine signal peptides should provisionally fall into the dual-function A-class (Table 1).

With the apparent relationship between A-class and C-class peptides, especially through their chaperone-binding activities, it is tempting to consider further the 3D crystal structure of the C-class remnant signal from the E. coli NarG protein (Bertero et al., 2003). In the alternative crystal structure of the water-soluble (i.e. NarJ-dissociated) NarGH dimer from E. coli, the remnant signal is exposed to solvent and completely disordered (Jormakka et al., 2004). This observation resonates with the limited structural analyses done on bona fide twin-arginine signal peptides, which show that these too are largely unstructured in aqueous solution (Kipping et al., 2003; San Miguel et al., 2003). In the NarGHI holoenzyme structure, however, the NarG remnant signal has some secondary structure comprising a 6 amino acid a-helix at the N-terminus, followed by a disordered region, and followed next by a twisted b-hairpin terminating at amino acid T-28 (Bertero et al., 2003; Fig. 5). This secondary structure must be induced in the holoenzyme by the extensive contacts made between the NarG remnant signal and NarI. Given that NarI is an integral membrane protein that contains an obvious binding site for a C-class remnant signal, it is tempting to speculate that the NarG::NarI complex could be providing a tantalizing glimpse of what a TatC::twin-arginine signal peptide complex could look like immediately prior to the transport event (Fig. 5).

**Fig. 5.** Structure of the E. coli NarG remnant C-class signal when bound to NarI. The crystal structure of NarG showing, in close-up, the structure of the initial 28 amino acids that correspond to the C-class remnant signal peptide.
Taken altogether, these observations point to an intimate evolutionary link between extracellular and cytoplasmic respiratory enzymes. Whether the ancient progenitor of the NarG proteins was originally located on the cell surface and subsequently relocated to the cytoplasm, or vice versa, during the course of evolution is a moot point, but one worth remembering if debating the evolution of the Tat pathway.

...and diversity in the assembly of molybdoenzymes

While the evidence for the existence of C-class remnant signals on some non-exported enzymes is compelling, this is not a universal feature amongst enzyme groups containing exported and non-exported homologues. The extremophilic deep-branched eubacterium *Thermus thermophilus*, for example, expresses a NarG-type nitrate reductase that is localized, together with its NarH partner, on the cytoplasmic face of the membrane attached to a complex comprising a NarI homologue and a membrane-bound cytochrome c (NarC) (Zafra *et al.*, 2005). The *T. thermophilus* NarG, however, contains a barely recognizable short N-tail (Table I), possibly just sufficient for the attachment of NarGH to NarIC, but not easily identifiable as a C-class remnant signal. Despite the lack of a clear C-class remnant signal, biosynthesis of *T. thermophilus* NarG remains strictly dependent upon a NarJ homologue (Zafra *et al.*, 2005). This observation either suggests that the remaining N-tail is sufficient for NarJ binding (note that the extreme N-terminal 15 amino acids of *E. coli* NarG was sufficient for NarJ binding *in vitro*; Chan *et al.*, 2006) or points strongly towards the existence of a second NarJ-binding site on NarG. Indeed, the fact that the *T. thermophilus* NarGH dimer is already membrane-localized (presumably via a NarG N-tail interaction with NarIC) before NarJ-dependent maturation occurs suggests strongly that the second chaperone-binding site is functionally important and that NarJ is not bound to the N-tail (Zafra *et al.*, 2005).

It is worth considering at this juncture that *E. coli* NarJ also binds to a second location on the *E. coli* NarG protein (Vergnes *et al.*, 2006). Moreover, *E. coli* TorD also binds to at least one other location on the *E. coli* TorA protein, distinct from the A-class signal peptide, and this alternative binding site seems to be important for efficient cofactor loading (Ilbert *et al.*, 2004; Jack *et al.*, 2004). The location and sequence of the secondary chaperone-binding sites have not yet been identified for either system. However, given that TorD family proteins recognize specific peptide sequences, it seems most logical to assume that the amino acid sequence of any secondary binding sites should show some identity with the sequences of the respective A-class and C-class peptides.

The intimate link between TorD family chaperones and the type II and type III *bis*-MGD molybdoenzymes is genetically and biochemically unequivocal. But what about the assembly of type I *bis*-MGD molybdoprotein subfamily contains the Tat-dependent membrane-bound formate dehydrogenases (FdgnGHI and FdoGHI in *E. coli*) and their non-exported cytoplasmic homologues (e.g. FdhF in *E. coli*). The type I subfamily also includes the periplasmic nitrate reductases (e.g. NapABC in *E. coli*) and, while a cytoplasmic homologue does not exist in *E. coli*, the related non-exported NapB-type assimilatory nitrate reductases which are widespread in other bacteria (Richardson *et al.*, 2001). In all cases the non-exported cytoplasmic members of the type I subfamily do not display N-tails or remnant signal peptides at all, which probably points both to a different evolutionary origin for these enzymes and governance by different biosynthetic principles. It is not yet known if TorD family proteins play any direct role in the assembly of type I molybdoenzymes. This is unlikely, however, since the genes encoding TorD homologues are only very rarely co-localized with those for type I molybdoenzymes. In the β-proteobacterium *Azorarcus*, for example, a gene encoding a TorD homologue (eaA2937) is found in a locus including various molybdoenzyme biosynthesis genes (*mobA, mobB, fdhD*) and the structural genes for a Tat-dependent respiratory formate dehydrogenase (eaA2936, eaA2935 and eaA2933) (Rabus *et al.*, 2003).

The common denominator in the biosynthesis of extracellular type I molybdoenzymes may be the requirement for an Fe–S cluster-containing ferredoxin accessory protein. The FdhE protein is consistently co-localized with gene clusters encoding Tat-dependent formate dehydrogenases. FdhE is known to be essential for the assembly of respiratory formate dehydrogenases and, most importantly, is not required for biosynthesis of the non-exported homologues (Mandrand-Berthelot *et al.*, 1988; Paveglio *et al.*, 1988; Schindwein *et al.*, 1990). While the FdnG and FdhD proteins appear to be completely absent from an *fdhE* mutant (Mandrand-Berthelot *et al.*, 1988), FdhE seems to be a specific chaperone for these Tat-dependent proteins rather than a transcriptional or translational regulator (Stewart *et al.*, 1991). It is not known whether FdhE is a twin-arginine signal peptide-binding protein, or indeed whether it binds to the formate dehydrogenase at all, but FdhE is certainly not present in the crystallized holoenzyme (Jormakka *et al.*, 2003). Sequence analysis suggests each FdhE polypeptide contains three conserved CxxC motifs (Turner *et al.*, 2004), possibly indicative of the presence of Fe–S clusters. Interestingly, this is a structural feature shared with an accessory protein (NapF) that is required for the biosynthesis of the type I periplasmic nitrate reductase, NapA. The NapF protein has been shown recently to bind to the Tat-dependent NapA precursor in the cell cytoplasm (Nilavongse *et al.*, 2006) and its biochemical role would seem to be to catalyse the insertion of the single Fe–S cluster into NapA prior to the export event (Olmo-Mira *et al.*, 2004). Again, NapF-type proteins have never been shown to be required for the assembly of the related non-exported assimilatory nitrate reductases. The presence of Fe–S clusters could point to roles in electron transfer for FdhE and NapF.
While it is certainly possible that extracellular and cytoplasmic isoenzymes could be operating at vastly different redox potentials, which may necessitate a degree of 'reductive activation' of the periplasmic protein prior to export (e.g. Field et al., 2005; Zajicek et al., 2004), it is certainly not clear why denial of this activity would completely destabilize the enzymes in question (e.g. Mandrand-Berthelot et al., 1988).

In addition to NapF, genetic evidence suggests that a second accessory protein (NapD) is required for the biosynthesis of some periplasmic nitrate reductases (Potter & Cole, 1999; Nilavongse et al., 2006). NapD is a very small protein (usually <100 amino acids) and, in E. coli, at least, has been found to be essential for the biosynthesis of NapA (Potter & Cole, 1999). NapD does not interact at all with NapF, but does bind tightly to NapA in the cytoplasm prior to export (Nilavongse et al., 2006). Again, the precise locations of NapD and NapF binding to NapA are not known; however, their physiological role must be, as with the FdhE protein, specific for the biosynthesis of the Tat-dependent type I enzymes, as genes encoding these proteins are never co-localized with non-exported type I homologues. Indeed, examples of accessory proteins required for the assembly of non-exported type I molybdenzymes are few and far between. The NarM protein is one such isolated example, and is required for the biosynthesis of assimilatory nitrate reductases in cyanobacteria (Maeda & Omata, 2004). This suggests that some type I molybdenzymes do not require the assistance of a biosynthetic chaperone at all, at least for the 'insertion' of the cofactor, and some heterologous expression studies of type I enzymes from Salmonella suggest that dedicated Tat proofreading chaperones may also be dispensed with (Hensel et al., 1999; Hinsley & Berks, 2002).

The type I, II and III bis-MGD enzymes are members of what was originally classified as 'the DMSO reductase family' of molybdenum and tungsten enzymes (Kisker et al., 1997). In addition to this group, a huge array of other molybdenum- and tungsten-dependent enzymes have been identified in nature and these can be sorted into another three large families: the sulfite oxidase (SO) family; the xanthine oxidase (XO) family; and the aldehyde : ferredoxin oxidoreductase (AOR) family (Kisker et al., 1997).

Members of the SO family bind a single molecule of the MPT form of the molybdenum cofactor, as opposed to the dinucleotide form, and include the cytoplasmic NADH-dependent assimilatory nitrate reductases from plants and fungi (e.g. Fischer et al., 2005), and the periplasmic YedY protein from E. coli (Loschi et al., 2004). YedY is synthesized with a (long) twin-arginine signal peptide and is thought to associate with an unusual membrane-bound haemoprotein (YedZ) following translocation (von Rozycki et al., 2004; Brokx et al., 2005). YedY is apparently not a sulfite oxidase nor a nitrate reductase but a very unusual TMAO reductase (Loschi et al., 2004), and there are no known signal peptide-binding proteins required for its biosynthesis.

Members of the XO family are found in all kingdoms of life and bind a single molecule of MPT or, perhaps more commonly in the prokaryotic world, a single molecule of molybdopterin cytosine dinucleotide (MCD) (Kisker et al., 1997). Examples include the crystallographically defined carbon monoxide dehydrogenase from the x-proteobacterium Oligotropha carboxidovorans (Dobbeck et al., 1999), which contains MCD, and the xanthine dehydrogenase from Rhodobacter capsulatus (Leimkuhler et al., 1998; Truglio et al., 2002), which contains MPT as cofactor. The genes encoding the three-subunit O. carboxidovorans carbon monoxide dehydrogenase are encoded on a megaplasmid by a large gene cluster containing 12 genes – coxBCMSLDEFGHIK (Santiago et al., 1999). Of these, the CoxF protein is probably the most important in terms of biosynthesis since it belongs to the XdhC family (Leimkuhler & Klipp, 1999). XdhC is an accessory protein which binds to the apo-form of the molybdoenzyme subunit (XdhB) of R. capsulatus xanthine dehydrogenase (Neumann et al., 2006), and its role seems to be akin to that originally described for TorD and NarJ (Blasco et al., 1998; Pommier et al., 1998) in that it may maintain XdhB in an ‘open conformation’ during cofactor insertion. Arguably the most important feature of XdhC, however, is that this accessory protein does actually bind the molybdenum cofactor directly (Neumann et al., 2006).

Both the O. carboxidovorans and R. capsulatus XO family enzymes described above are cytoplasmically located and have no hints even of remnant C-class peptides to which XdhC proteins might attach. However, the E. coli genome does encode XO family members that may be transported to the periplasm via the Tat pathway. The E. coli yagPQRST cluster encodes a putative MCD-binding subunit (YagR) that, by analogy with carbon monoxide dehydrogenases, should associate with a flavoprotein (YagS) and an Fe–S protein (YagT) to form the holoenzyme (Berks et al., 2003). In this case the YagT Fe–S protein, rather than the molybdoprotein, bears a twin–arginine signal peptide at its N-terminus (Berks et al., 2003). As a result, if YagRST is indeed transported to the periplasm it must be via a single signal peptide and as a fully assembled trimer – surely such a scenario would require a Tat proofreading chaperone?

The YagQ protein is a member of the XdhC family and is therefore dedicated to the insertion of the MCD cofactor into YagR (Neumann et al., 2006), which leaves YagP as a possible candidate, a 136 amino acid water-soluble protein that has been annotated as a putative transcriptional regulator in the databases but whose function clearly awaits further experimentation. The E. coli genome encodes two other YagR homologues that have been implicated in purine catabolism, XdhA (b2866) and XdhB (b2881) (Xi et al., 2000). Neither contains a twin–arginine signal peptide and neither is predicted to associate with partner proteins that possess Tat signal peptides. Moreover, there are also no obvious XdhC homologues encoded at these loci.
Finally, members of the aldehyde:ferredoxin reductase (AOR) family were originally identified in hyperthermophilic archaea, where tungsten is often substituted for molybdenum in the active site (Kisker et al., 1997). The biosynthesis of this class of molybdoprotein has not yet been studied; however, the enzymes themselves have proven tractable to structural projects (Kisker et al., 1997). AOR family enzymes are usually monomeric proteins (or oligomers of identical subunits) that bind a single Fe–S cluster as well as a single tungsten atom ligated by two MPT (not dinucleotide) cofactors (Kisker et al., 1997). As the name suggests, these enzymes usually couple the oxidation of aldehydes (or related compounds) to the reduction of a ferredoxin in the cell cytoplasm. Interestingly, the E. coli genome encodes a member of the AOR family, termed YdhV, which is encoded in a putative operon containing six genes – ydhY,V,W,X,UT. YdhV does not bear a twin-arginine signal peptide. The roles of the YdhW and YdhT proteins are difficult to predict from sequence alone; however, both YdhY and YdhX are predicted to be ferredoxins which could, in principle, receive electrons from the YdhV subunit. However, YdhX does, in fact, contain an N-terminal Tat targeting signal (Berks et al., 2005). Moreover, YdhU is predicted to be an integral membrane cytochrome of the PhsC class normally associated with periplasmically oriented, Tat-dependent and molybdenum-dependent thiosulfate reductases (e.g. Hinsley & Berks, 2002). This is certainly suggestive that YdhX could associate with YdhU at the periplasmic side of the membrane, but is the YdhV AOR family protein also periplasmic and translocated in tandem with YdhX? Only experimentation will reveal the answer, though it is tempting to speculate that this enzyme could be located on both sides of the membrane depending on which ferredoxin partner is attached, as reported for a Tat-targeted catalase in Helicobacter pylori (Harris & Hazell, 2003).

Concluding remarks

While this review has focused on recent advances in the understanding of how some prokaryotic molybdoenzymes are assembled, it is fair to say that this research just scratches the surface in terms of the overall diversity and complexity of post-translational biosynthetic processes. Studies of Fe–S cluster assembly, which can almost be classified as a ‘housekeeping’ biochemical process, form a vibrant research area that straddles both prokaryotic (Böck et al., 2006; Johnson et al., 2005) and eukaryotic systems (Balk & Lobreaux, 2005; Lill & Mühlenhoff, 2005). Of course, the assembly and insertion of Fe–S clusters must be coordinated with many of the molybdoenzyme-specific processes outlined above. Moreover, cytochrome c biogenesis (Stevens et al., 2004) and nickel-dependent enzyme assembly (Böck et al., 2006; Mulrooney & Hausinger, 2003) are another two large research fields that in some cases must also dovetail considerably with the molybdoenzyme systems described here. Finally, it is possible that pressure from the biotechnology and biomedical sectors will inspire research into biosynthesis of overtly unconsidered, but potentially extremely important, complex enzyme systems such as the cobalamin-dependent reductive dehalogenases (Janssen et al., 2001), and the pyrroloquinoline quinone (PQQ)-dependent dehydrogenases (Anthony, 2004).

ACKNOWLEDGEMENTS

I dedicate the 2006 Fleming Prize Lecture to my wife Tracy Palmer and our boys James and Jack. I would like to thank the University of East Anglia, and in particular David J. Richardson, for supporting and mentoring me in my early years as a group leader. I would also like to thank Ben Berks (Oxford) for guidance, support and many useful discussions. For stimulating collaborations I should acknowledge Arthur Oubrie (NV Organon), Ray Turner (Calgary), Lynne Macaskie (Birmingham), Geerten Vuister (Nijmegen) and Chris Spronk (Nijmegen). The postdocs, graduate students and technical staff who have worked with me over the last 6 years have all contributed something to this prize award and deserve special mention here – Judith Mayne, Verity Lyall, AlexDubini, Kostas Hatziathanthos, Kev Moore, Julien Maillard, Karolina Fornvik, Iris Lüke, Bérengère Ize, Isabelle Caldelari and Dave Guymet. Thanks also to everyone in Tracy Palmer’s and Ben Berks’ research teams, especially Grant Buchanan and Rachael Jack, for discussions, ideas and contributions to original research papers. I would finally like to acknowledge The Royal Society, the BBSRC and the John and Pamela Salter Charitable Trust for financial support. I look forward now to new challenges at the University of Dundee.

REFERENCES


Ilbert, M., Mejean, V. & Iobb-Nicol, C. (2004). Functional and structural analysis of members of the TorD family, a large chaperone family dedicated to molybdoproteins. Microbiology 150, 935–943.


FEBS Lett 534, 156–160.


Comm Biol Chem 275, 22526–22531.


Comm Biol Chem 275, 22526–22531.


