Analysis of the Influence of the N- and C-Terminal Domains of a Signal Sequence on SRP Dependence

AND

Investigating the Essentiality of Genes in the Presence of Azide

An MRes thesis submitted by Rachael Chandler

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

September 2014

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Acknowledgements

I was only the first of a growing number of students to be lucky enough to find themselves under Damon's supervision at the University of Birmingham. I have benefited from his patience, friendliness and his ability to create enthusiasm in others. He gave me the opportunity to stay in his lab beyond the MRes as a research technician thereby enabling me to work on the most exciting projects for my PhD.

I have passed many happy hours with other students in and out of the lab during the course of my MRes.
Myfirstclassbiochemistrygraduatefriendformerlymyundergraduateprojectstudentfriendinthelab Yousra (aka Yousra Djouider),Tamar Cranford-Smith, Alexander Osgerby, Steve Peake and Laura Harford have all been great friends. A particular highlight has to be the Secretion Meetings which involved as much cake as actual science!

I would like to thank Ian Henderson for enabling me to continue working on Sec translocation during my second MRes project; Ash Robinson, a super-friendly Slytherin, for showing me how to do the genomic DNA preparation and for actually carrying out the high-throughput sequencing and data processing; and Peter Winn for letting me onto the MRes course when I applied at the last minute!

Analysis of the Influence of the N- and C-Terminal Domains of a Signal Sequence on SRP Dependence

An MRes project

by Rachael Chandler

Supervised by Dr. Damon Huber

Abstract

In bacteria most extracytoplasmic proteins are secreted by the SecYEG translocon, which forms a protein conducting channel in the cytoplasmic membrane. Proteins are delivered to SecYEG by one of two targeting pathways: Some are recognised early in translation by the signal recognition particle (SRP), which delivers them to SecYEG for cotranslational export. In the alternative pathway proteins are targeted later in translation by SecA, the motor ATPase that drives translocation through SecYEG, and are exported posttranslationally. Proteins destined for export have cleavable N-terminal signal sequences of around 21 amino acids, consisting of a hydrophobic core flanked by shorter N- and C-terminal domains. These signal sequences are thought to determine the targeting pathway through recognition by SRP or SecA. Hydrophobicity is thought to be the determining factor within the signal sequence for SRP recognition. However, some SecA-dependent signal sequences have cores of equal or greater hydrophobicity- which suggests that other features of a signal sequence, such as the N- and C-terminal domains, can affect SRP dependence. In this study, the N- or C-terminal portions of an SRP-dependent signal sequence were replaced by those of SecA-dependent signal sequences that have similarly hydrophobic cores. In order to detect the targeting pathway of choice the hybrid signal sequences were fused to the cytoplasmic protein thioredoxin, which can be used as a reporter for cotranslational export. The results suggest a role for the C-terminal region of the signal sequence in SRP dependence and show a possible effect of the signal sequence on protein expression.

Introduction

The Sec Translocation System

Bacteria have a number of secretion systems used for protein export across or insertion into the cytoplasmic membrane. The Sec pathway, also present in archaea and eukaryotes, is the most widely used of these secretion systems and, in Gram negative bacteria, is required for the export of most periplasmic proteins as well as outer membrane proteins. Along with the membrane protein insertase YidC, the Sec translocon is also required for the insertion of proteins into the cytoplasmic membrane. Other export pathways include secretion systems that traverse the inner and outer membranes such as the type III secretion system. Other secretion systems in Gram negative and Gram positive bacteria only transport proteins across the outer membrane from the periplasm and so require the Sec translocon for the inner membrane transport step. The twin-arginine translocation (Tat) complex is an alternative inner membrane transport system found in most bacteria and archaea that is responsible for the export of a relatively small number of proteins. While the Sec translocon exports only unfolded proteins, the Tat complex exports fully folded proteins and is used for the translocation of cofactor-bound proteins-although it is not essential in many species (Palmer and Berks, 2012).

The core Sec translocon consists of the complex SecYEG, which associates with cytoplasmic factors, such as the SRP and SecA, as well as other membrane proteins such as YidC and SecDF-YajC (Schulze *et al.*, 2014).

The SecY subunit forms an hourglass shaped, protein conducting channel through which only fully unfolded proteins can pass (Van den Berg, *et al*., 2004). It has a ring of hydrophobic amino

acid residues at its centre, which are thought to function as a barrier to maintain membrane impermeability.

SecA is an essential motor ATPase that drives the translocation of substrates through SecYEG, but is also thought to be involved in the targeting of preproteins to the translocon (Gelis *et al.*, 2007; Huber *et al.*, 2011).

In bacteria there are two targeting pathways that deliver preproteins to the SecYEG translocon: The cotranslational or SRP-dependent pathway is undertaken by inner membrane proteins and around 10% of periplasmic proteins in *E. coli* (Huber *et al.*, 2005). In this pathway, the signal recognition particle (SRP) recognises the transmembrane domains of inner membrane proteins or the hydrophobic N-terminal signal sequences of periplasmic proteins and directs the ribosome-nascent chain complex (RNC) to the translocon by interaction with its receptor, FtsY. In this pathway, translocation is thought to be driven by the translation of the protein through the channel by the ribosome.

The alternative posttranslational or SecA-dependent pathway is responsible for the targeting of most soluble exported proteins and begins with the recognition of the signal sequence of the translocation substrate later in translation by SecA, which then delivers it to SecYEG to be translocated posttranslationally. Because proteins must be in an unfolded conformation in order to pass through the SecYEG channel, folding of these substrates must be prevented in the cytoplasm following their release from the ribosome. Some periplasmic proteins, such as the alkaline phosphatase PhoA, require the formation of intramolecular disulfide bonds in order to fold-so this only takes place in the periplasm because of the reducing environment of the cytoplasm. Other proteins are kept in the unfolded conformation by chaperones such as the translocation-specific chaperone SecB, which interacts with SecA.

The SecA-dependent pathway

Most periplasmic and outer membrane proteins in *E. coli* are translocated through SecYEG posttranslationally. In this pathway, substrates begin translocation when they are nearly or fully synthesised. SecA is the essential targeting factor for this pathway. While cotranslationally translocated presecretory proteins are exported before they can fold, posttranslationally targeted substrates must be prevented from folding in the cytoplasm in order for translocation through SecYEG to take place.

A number of chaperones are known to prevent the cytoplasmic folding of periplasmic proteins. SecB is a homotetrameric Sec system-specific chaperone that binds presecretory proteins with low specificity and is non-essential for cell growth but is required for the export of a subset of posttranslational substrates (Baars *et al.*, 2006). It is also capable of acting as a general chaperone: overexpression of SecB can suppress the phenotypes of a strain lacking the chaperones trigger factor (TF) and DnaK (Ullers *et al.*, 2004). DnaK and its cochaperone DnaJ have also been shown to play a role in secretion of SecB-dependent and -independent translocation substrates (Wild *et al.*, 1992).

A tetramer of SecB specifically binds the homodimeric ATPase SecA, which is required for the targeting and translocation of preproteins. Only the C-terminal 22 amino acid residues of SecA, which form a positively charged zinc-binding domain, are required for SecB binding (Fekkes *et al.*, 1997). For a translocation-active complex, however, a second binding site is required: the C-terminal α-helices of SecB bind the interfacial region of SecA (Randall *et al.*, 2005).

The SRP-dependent pathway

The bacterial SRP consists of the ~50 kDa protein Ffh and a 4.5S RNA molecule. The Ffh protein has an N-terminal 'NG' domain comprising a ribosome-binding 4-helix bundle and a GTPase; it has a methionine-rich 'M' domain, which binds signal sequences as well as the 4.5S RNA, at its C terminus. The 4.5S RNA forms a hairpin structure capped by a GGAA tetraloop. As well as acting as a platform for interaction between Ffh and FtsY, the 4.5S RNA also has an active role in the recruitment of the RNC to the membrane (Jagath *et al.*, 2001; Voigts-Hoffmann *et al.*, 2013).

The SRP targets ribosome-nascent chain complexes (RNCs) to the Sec translocon by interacting with its receptor, FtsY, which can associate with the cytoplasmic membrane and SecYEG. FtsY, like Ffh, has an NG GTPase domain. The interaction between Ffh and its receptor is mediated by their NG domains, which together form an active site that can bind two GTP molecules (Zhang, X. *et al.*, 2010).

Signal Sequences

Proteins destined for export through the SecYEG translocon are synthesised with N-terminal signal sequences. Despite a lack of sequence similarity between signal peptides they have a common domain structure consisting of a short, positively charged N terminus, a hydrophobic core of ~12 amino acids and a hydrophilic C terminus (box 1).

Box 1. The typical domain structure of a signal peptide

Signal sequences are recognised by the factors that target proteins for translocation, SecA or the SRP. They therefore determine which targeting pathway a preprotein undergoes. One of the challenges in understanding the targeting of preproteins for translocation is the apparent absence of a feature within the signal sequence that stringently determines whether it is targeted by SecA or the SRP. It is currently believed that recognition by the SRP (and therefore the targeting pathway) is determined by the degree of hydrophobicity of the signal sequence core (Lee and Bernstein, 2001; Huber et al., 2005(a)). The SRP has long been known to target inner membrane proteins for insertion into the membrane (Macfarlane & Müller, 1995).

Hydrophobicity is known to play a role in SRP recognition. For example, it was found that some posttranslationally translocated proteins (OmpA, MBP and LamB) could be diverted to the

SRP-dependent pathway by increasing the hydrophobicity of their signal sequences (Lee and Bernstein, 2001; Bowers et al., 2003). It was later found that when signal sequences were ranked by the hydrophobicity of their core, all of the SRP-dependent signal sequences were very hydrophobic. However, it was also noted that some non-SRP-dependent signal sequences were above the hydrophobicity threshold for SRP dependence (Huber et al., 2005 (a)). This suggested that another feature of the signal sequence, apart from the hydrophobicity of the core, has a role in SRP dependence. To investigate this further, the similarly hydrophobic cores of SRP- and SecA-dependent signal sequences were swapped and the SRP dependence of the hybrids was tested. None of these hybrid signal sequences were found to be SRP dependent (Huber, 2006). This suggested that the hydrophobic core alone could not determine SRP dependence and that the core may contain another necessary determinant besides hydrophobicity. Or, it could suggest that a single SRP recognition interface extends beyond the length of the hydrophobic core that was defined in this study.

The secondary structure, conformation or orientation of a signal sequence could be an important factor in SRP dependence; for example, the position of the nascent signal sequence on the surface of the ribosome could determine whether the SRP is able to interact with it. A study on the SRP dependence of signal sequences in yeast (which, like bacteria, have a posttranslational Sec pathway in addition to the SRP-dependent one) found that a proline-to-methionine substitution two amino acids C terminal to the signal peptidase cleavage site of an SRP-dependent signal sequence resulted in the loss of SRP dependence and a resulting decrease in translocation efficiency (Matoba and Ogrydziak, 1998). In addition to suggesting that SRP dependence is not determined by hydrophobicity alone, this study highlights the possibility that the amino acid sequence of the 'mature' region of the preprotein adjacent to the signal peptide could affect SRP dependence by influencing conformation or orientation. Perhaps this should be beared in mind when using signal sequence-reporter protein fusions for

investigating SRP dependence: If, for example, a proline residue is present in the N terminus of the native mature protein but absent in the reporter, or vice versa, the orientation of the signal sequence region of the construct could differ from that in the native preprotein due to a helix break and this might affect the apparent SRP dependence of the signal sequence.

Thioredoxin as a Reporter for SRP-Dependent Export

The SRP dependence of signal sequences has previously been investigated using thioredoxin-1 (TrxA) as a reporter for cotranslational export (Huber *et al.*, 2005 (a)). TrxA is a cytoplasmic oxidoreductase that, when fused to a signal sequence, can only be efficiently exported in its unfolded state (Huber *et al.*, 2005 (b)). This is thought to be because of its rapid folding in the cytoplasm, which prevents it from passing through the SecYEG translocon. In these studies, the export of thioredoxin was investigated by dividing the cells into periplasmic and cytoplasmic fractions by spheroplasting and then western blotting against TrxA.

While thioredoxin export can be detected directly using antibodies, it can also be indirectly detected in a mutant strain deficient in disulfide bond formation in the periplasm by utilising its interaction with the flagellar protein, FlgI (Hizukuri *et al.*, 2006).
In this study, we used both of these methods to investigate the export of TrxA when fused to hybrid signal sequences. To investigate the roles of the N- and C-termini of a signal sequence in SRP dependence, we swapped either the N- or C-terminus of the SRP-dependent DsbA signal sequence (DsbAss) with that of a SecA-dependent signal sequence (AppAss or TraUss). The results suggest a role for the C terminus of the signal sequence in SRP dependence and also show that the signal sequence could affect protein expression.

Materials and Methods

E. coli strains

Name	Description/Genotype
DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1
BW25113	*F-, DE(araD-araB)567, lacZ4787(del)::rrnB-3, LAM-, rph-1, DE(rhaD-rhaB)568, hsdR514*
DRH657	DHB4 *ΔtrxA::kan*
JW3832	BW25113 *ΔdsbA::kan* Keio collection (Baba *et al.*, 2006)
RAC1	JW3832 kan^S (this study)
RAC2	RAC1 *ΔdsbC::kan* (this study)
JW2861	BW25113 *ΔdsbC::kan* (Keio)
JW5734	BW25113 *ΔdsbD::kan* (Keio)
RAC5	RAC1 *ΔdsbD::kan* (this study)
RAC13	RAC5 kan^S (this study)
RAC14	RAC13 *ΔdsbC::kan* (this study)

Plasmids

Name	Description	Reference
pDSW204	pTRC99a derivative with an attenuated promoter, amp^R	Weiss *et al.*, 1999
pMO2	pDSW204 + *trxA*	Huber *et al.*, 2005(a)
pCFS123	pDSW204 + *dsbAss* (DDD) + *trxA*	Huber *et al.*, 2005(a)
pCFS126	pDSW204 + *phoAss* + *trxA*	Huber *et al.*, 2005(a)
pDH897	pDSW204 + *βlacss* + *trxA*	
pRC7	pMO2 + TDT	This study
pRC8	pMO2 + TDD	This study
pRC9	pMO2 + DDT	This study
pRC10	pMO2 + ADA	This study

pRC11	pMO2 + ADD	This study
pRC12	pMO2 + DDA	This study
pRC16	pMO2 + TD-'SPAASA'	This study
pRC17	pMO2 + TD-'SAAASA'	This study
pRC18	pMO2 + TD-'SASASA'	This study
pRC22	pMO2 + TD-'VAAASA'	This study
pRC20	pMO2 + DD-'TAQSAFA'	This study
pRC21	pMO2 + DD-'SAQSAFA'	This study
pRC25	pRC12 + *phoA*	This study
pRC26	pCFS126 + *phoA*	This study
pCP20	amp^R, Cm^R, temperature-sensitive replication, temperature-induced Flp recombinase synthesis	Cherepanov and Wackernagel, 1995

Growth conditions

Strains were grown at 37°C, unless otherwise stated, in LB medium. Liquid cultures were shaken at 200 rpm. The following antibiotic concentrations were used when appropriate: 200 µg/ml ampicillin, 30 µg/ml kanamycin.

Fractionation experiments

Strains were subcultured and grown to OD_{600} ~0.5 in the presence of 10 µM (with DDT and DDA signal sequences) or 25 µM IPTG to induce *SS-trxA* expression. Cells were fractionated by spheroplasting as in Huber et al., 2005(a) and the whole-cell, periplasmic and cytoplasmic extracts were subjected to SDS-PAGE. Western blotting was carried out using Biorad wet blot apparatus. Blots were blocked overnight using a solution of 5% skimmed milk powder in TBS and then incubated with α-TrxA and α-β-lactamase antibodies diluted 1:6,000 and 1:10,000, respectively, in TBS and immunodetection was carried out using the Amersham Enhanced Chemiluminescence reagents and a horseradish peroxidase-coupled α-rabbit IgG antibody (1:5,000 in TBS).

Genetic Manipulation

Deletions of the *dsb* genes were carried out using P1 transduction of P1 lysates of kan^R Keio deletion strains (Datsenko and Wanner, 2000): 500 µl of an overnight culture of the recipient

strain was pelleted and resuspended in MC (100 mM $CaCl_2$, 10 mM $MgCl_2$). Recipient cells were mixed with the P1 lysate and incubated at 30°C for 20 min. 1M sodium citrate and LB broth were added, followed by incubation with shaking at 30°C for 1 hr. Cells were pelleted, resuspended in 1M sodium citrate, spread onto LB agar plates containing kanamycin and incubated overnight at 30°C.

Motility tests

Low agar (0.25-0.35%) LB plates were pricked with the appropriate *E. coli* strains and incubated for 1-3 days at 30°C. Plates were photographed after each night of incubation.

Alkaline phosphatase (PhoA) activity assays

Overnight cultures were diluted 1/100 in LB broth containing ampicillin and 25 μM IPTG and grown to OD_{600} 0.5-0.7 at 37°C with shaking. 100-200 μl culture was added to 1M Tris acetate (pH8) to a total volume of 1 ml. The alkaline phosphatase reactions were started by adding 4 mg/ml PNPP during incubation at 37°C. Reactions were stopped using 1M Na_2HPO_4, cell debris was pelleted and OD_{420} readings of the supernatant were taken. Units of PhoA activity were then worked out using the formula:

$$\frac{1000 \times \mathrm{OD420}}{\mathrm{t} \times \mathrm{v} \times \mathrm{OD600}}$$

where t = reaction time (min) and v = volume of culture used in the reaction (ml).

Results

The C-terminus of the DsbA signal sequence affects SRP dependence

The DsbA signal sequence was previously shown to be capable of exporting thioredoxin because it directs it to the SRP-dependent pathway (Schierle *et al.*, 2003). In addition, a 2005 study by Huber *et al.* classified a group of 36 native *E. coli* signal sequences as SRP-dependent or non-SRP-dependent based on their ability to export thioredoxin. These signal sequences were ranked by the hydrophobicity of the core and it was found that some non-SRP dependent signal sequences were above the hydrophobicity threshold for SRP dependence. This suggested that another feature of the signal sequence, apart from the hydrophobicity of the core, plays a role in SRP dependence.

In order to investigate a possible role for the N- or C-terminus of the signal sequence in SRP dependence, we constructed a set of hybrid signal sequences (Box 2). In each novel hybrid, the N- or C-terminus of the DsbA signal sequence (DsbAss or DDD), was replaced by that of a non-SRP-dependent signal sequence. We chose two non-SRP-dependent signal sequences,TraUss and AppAss, which have cores that fall above the threshold of hydrophobicity for SRP-dependent signal sequences.

Native Signal Sequences

DsbAss (DDD) MKKIWLALAGLVLAFSASA

TraUss (TTT) MKRRLWLLMLFLFAGHVPAASA

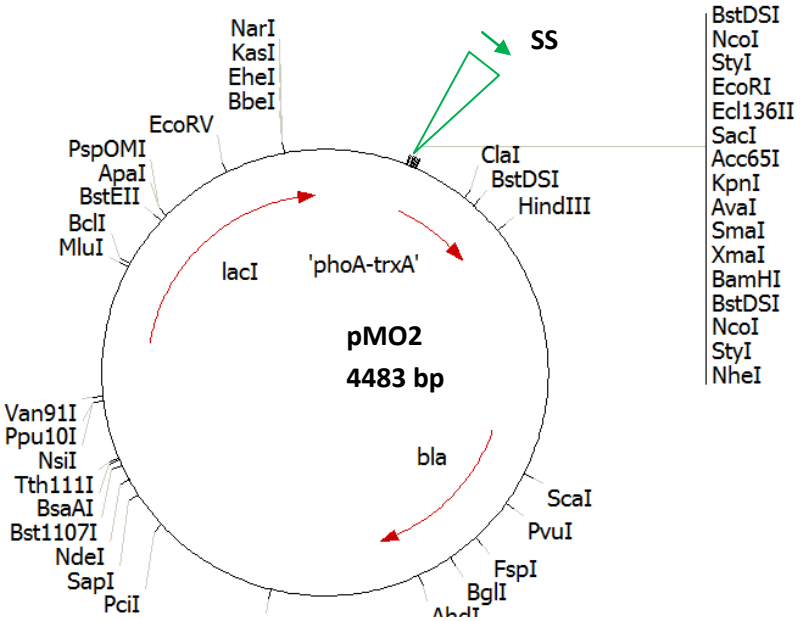
AppAss (AAA) MKAILIPFLSLLIPLTPQSAFA

Hybrid Signal Sequences

TDT	KRRIWLALAGLVLAFVPAASA
TDD	MKRRIWLALAGLVLAFSASA
DDT	MKKIWLALAGLVLAFVPAASA
ADA	MKAIWLALAGLVLAFTPQSAFA
ADD	MKAIWLALAGLVLAFSASA
DDA	KKIWLALAGLVLAFTPQSAFA

Box 2. Hybrid signal sequences used to investigate the role of the N- and C-termini.

In order to determine SRP dependence, we fused the hybrid signal sequences to thioredoxin-1 (TrxA) by inserting the signal sequence DNA into the plasmid pMO2 (Box 3).



Box 3. pMO2

Hybrid signal sequences were fused to the *trxA* gene using *NcoI* and *NheI* restriction sites.

We transformed a Δ*trxA* strain with plasmids carrying the new *SS-trxA* fusions. We then carried out subcellular fractionation to determine whether thioredoxin was exported. This was followed by SDS-PAGE and western blotting against thioredoxin (Fig. 1). As expected DDD-trxA, which has the full SRP-dependent DsbA signal sequence, was exported to the periplasm while the two hybrids with only the core of DsbAss were not exported (Fig. 1A and 1B). Both hybrids consisting of DsbAss with only its N-terminus replaced by that of a non-SRP dependent signal sequence appeared to export thioredoxin, while the hybrids with the TraUss (Fig. 1A) or AppAss Fig. 1B) C terminus did not.

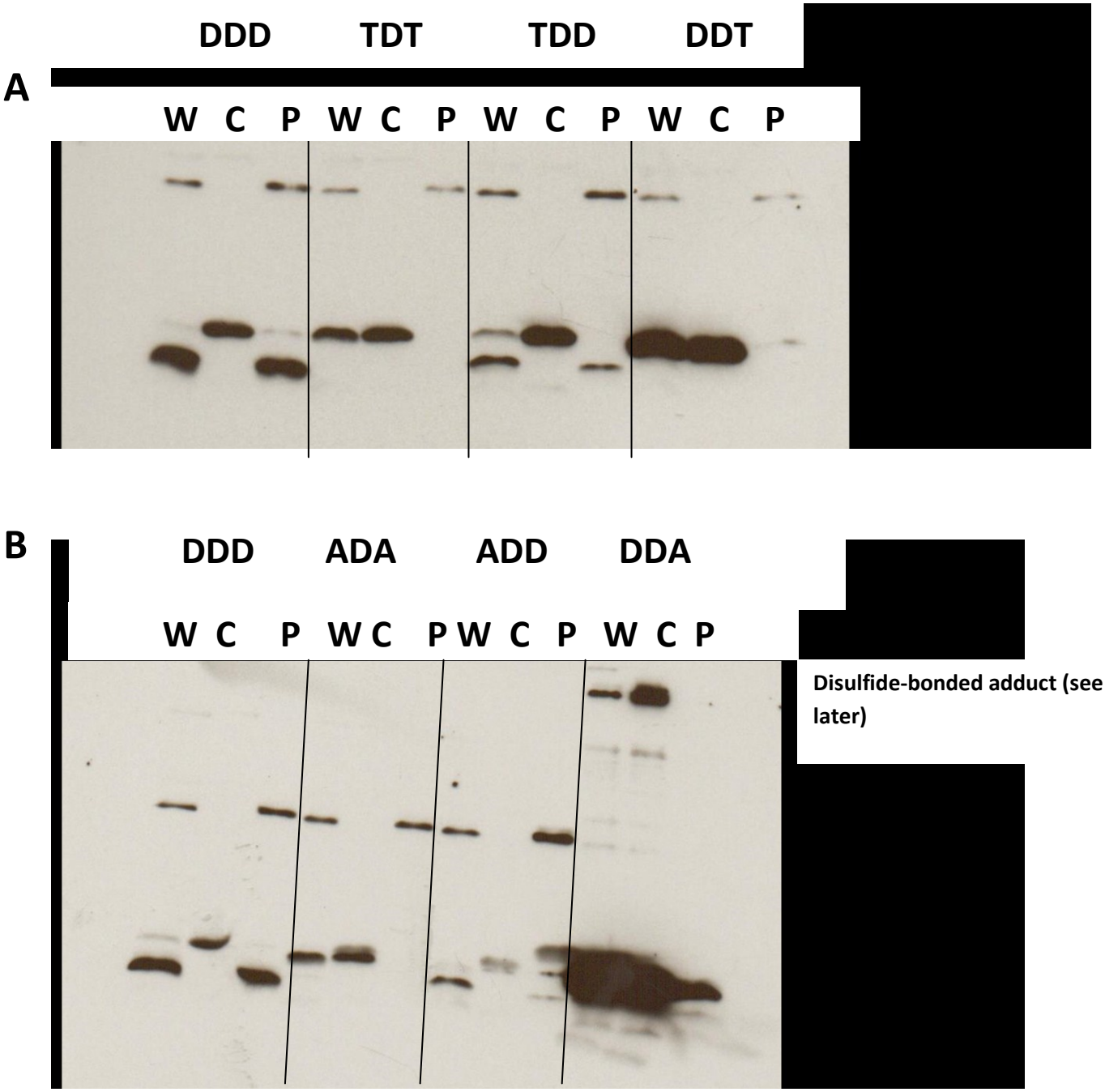


Fig. 1. The C-terminal domain of the DsbA signal sequence has a role in SRP dependence. *E. coli* DHB4 (Δ*trxA*) cultures carrying the SS-trxA fusion plasmids were induced with 25 μM IPTG, followed by subcellular fractionation, SDS-PAGE and western blotting using antibodies against thioredoxin-1 and β-lactamase, which is included as a periplasmic control. W, whole-cell extract; C, cytoplasmic fraction plus membranes; P, periplasmic fraction.

A C-terminal proline residue appears to determine the SRP dependence of the signal sequence hybrids

In order to narrow down the C-terminal determining factor of SRP dependence, the first four residues of the C terminus of the non-SRP-dependent TDT signal sequence hybrid were successively replaced by those of the SRP-dependent DsbAss C terminus (Box 4). The overlapping primers that encode the hybrid signal sequences were designed so that the replacement codons were identical to those encoding the amino acid residues of the DsbAss C terminus.

TDT MKRRIWLALAGLVLAFVPAASA

MKRRIWLALAGLVLAFSPAASA

MKRRIWLALAGLVLAFSAAASA

MKRRIWLALAGLVLAFSASASA

DDD MKKIWLALAGLVLAFSASA

Box 4. Hybrid signal sequences used to investigate the role of individual amino acids within the TDT C terminus in determining the targeting pathway

As before, subcellular fractionation was carried out to determine the location of the signal sequence hybrid-thioredoxin fusions. The first hybrid, "SPAASA", in which the valine of the TDT C terminus (VPAASA) was exchanged for a serine residue, did not export thioredoxin (Figure 2A). The other two hybrids, which both lack the proline residue of the TraUss C terminus, did export thioredoxin (Figure 2B). The third fusion, which has the signal sequence C terminus: SASASA, appeared to have a lower-molecular weight unprocessed form compared to the other signal sequence fusions run on the same gel (Figure 2B). A repeat of the fractionation

experiment showed the same lower-molecular weight band, which suggests that it is not a gel artefact. The "SASASA" C terminus could contain two alternative signal sequence processing sites (the signal peptidase is thought to recognise the sequence "AXA"). However signal peptidase cleavage is thought to only take place when proteins are exported (Auclair *et al.*, 2012). Furthermore, if either of these ASA sites were cleaved in the cytoplasm, thioredoxin would not be exported. It seems unlikely, but if there were a cytoplasmic protease that recognised a site that includes more of the 'SASASA' C terminus than is present in the DsbAss C terminus ('SASA'), which cleaves before the ASA site that is required for signal peptidase cleavage *and* if this protease could cleave as much of the fusion protein as is required for no 'uncleaved' band to show up on the blot-then this result could have occurred by protease cleavage. However TD-SASASA-TrxA is exported, the 'lower molecular weight' unprocessed form of the 'SASASA' fusion is in the cytoplasmic fraction and there is no detectable higher molecular weight population of the preprotein. It might also be possible that the different electrophoretic mobility of the SASASA precursor is due to a difference in its conformation.

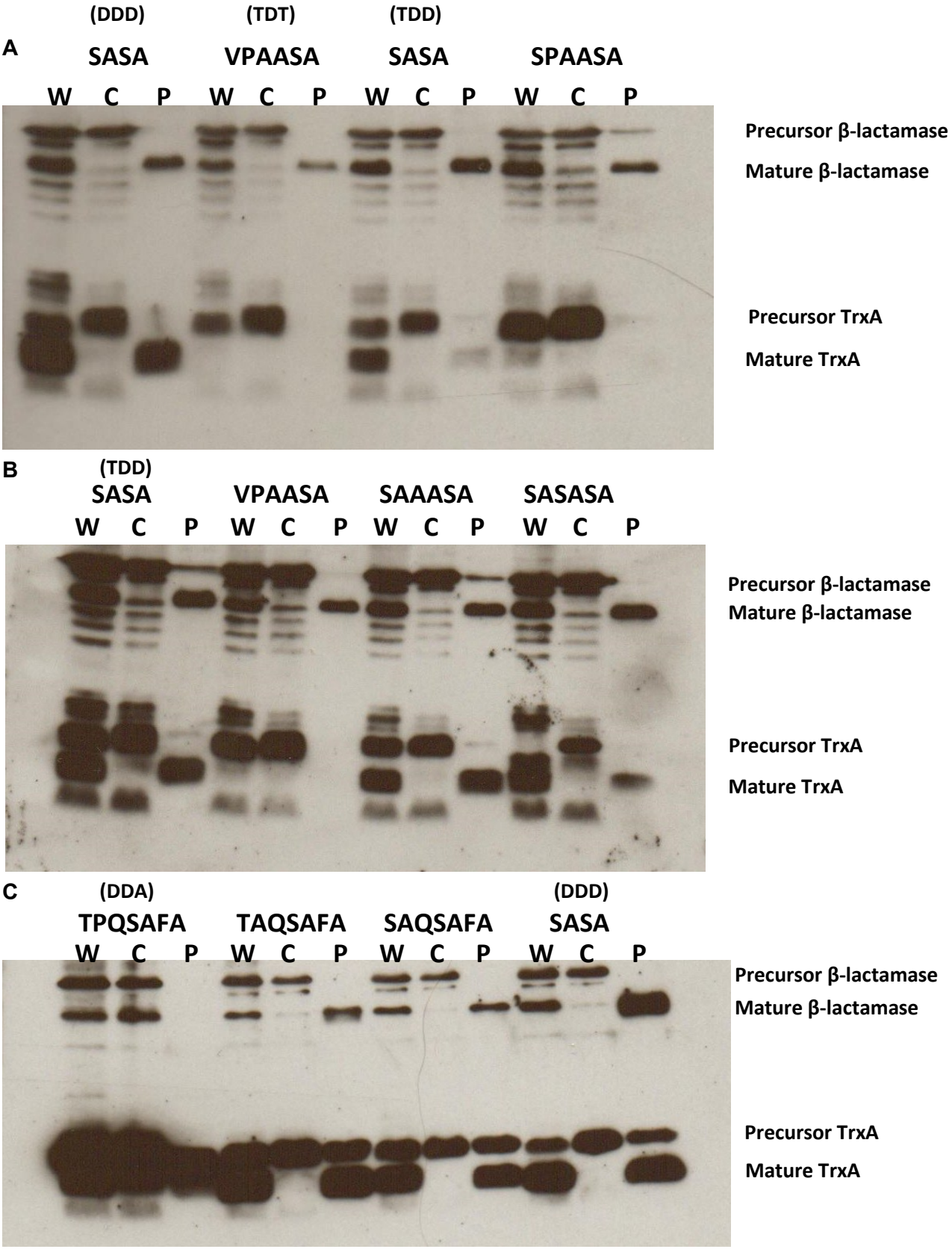


Fig. 2. The role of individual amino acids on SRP dependence. Subcellular fractionation and western blotting was carried out on strains expressing new hybrid signal sequence-thioredoxin fusions. Amino acids of the C terminus of the TDT hybrid were progressively exchanged for DsbAss C terminal residues (A and B). Residues of the DDA C terminus were replaced by the first two amino acids of the DsbAss C terminus to investigate the role of the proline and serine in SRP dependence (C).

Of the three new C termini of the TDT hybrid signal sequence, only the one with the proline remaining, SPAASA, failed to export thioredoxin. We therefore replaced the proline in the same position in the DDA C terminus (TPQSAFA) with the alanine in the equivalent position in the C terminus of the SRP-dependent DsbAss (Box 5). We constructed another variant in which the threonine residue was replaced by the first serine of the DsbAss C terminus to test the importance of a serine in that position (Box 5).

DDA MKKIWLALAGLVLAFTPQSAFA

MKKIWLALAGLVLAFTAQSAFA

MKKIWLALAGLVLAFSAQSAFA

DDD MKKIWLALAGLVLAFSASA

Box 5. Hybrid signal sequences used to investigate the role of the proline residue within the DDA C terminus in determining the targeting pathway

Both of the DDA derivatives exported thioredoxin and expression also returned to normal levels (Figure 2C). This showed that removal of a single amino acid residue of the DDA hybrid signal sequence, a proline, diverts the thioredoxin fusion protein to the SRP-dependent pathway. To confirm this, we constructed a final signal sequence hybrid in which the proline of the TDT C terminus (VPAASA) was replaced by the DsbAss alanine. The result was unclear because the expression of the construct was dramatically affected by the substitution (see below), but at very high levels of induction a band matching the processed form of a control fusion (TDD-TrxA) could be distinguished-suggesting that the signal sequence does export thioredoxin (Figure 4).

Changing the C-terminus of a signal sequence hybrid affects protein expression

The western blots in Figure 1 also show differences in expression of the various SS-TrxA fusions. Most noticeably, the *DDA-trxA* fusion is very highly expressed compared to the other constructs (Fig. 1B). In addition, this level of induction of *DDA-trxA* results in the appearance of a higher molecular weight band in the cytoplasmic fraction. To investigate the nature of this band and to confirm differences in expression between the fusions, we carried out western blotting on whole-cell protein extracts of all strains (Fig 3).

A comparison of thioredoxin expression in all strains (Fig. 3A) showed that DDT-TrxA and DDA-TrxA are overexpressed relative to the other strains. The level of DDA-TrxA expression was approximately 5-fold higher than the level of DDT-TrxA expression. These results suggest that the C-terminus of the signal sequence can affect protein expression.

It would be conceivable that the level of overexpression of DDA-TrxA could affect the results of the thioredoxin experiment, causing TrxA to accumulate in the cytoplasm-making it appear as though the DDA signal sequence is not SRP-dependent (for example, by obstructing/fully occupying the SecYEG translocon). However, we previously induced expression of *DDA-trxA* using 10 μM IPTG and found that the signal sequence did not export thioredoxin to the periplasm.

In order to investigate the appearance of higher molecular weight bands in the whole-cell and cytoplasmic DDA-TrxA lanes on the western blot of the fractionation experiment, we then subjected whole-cell lysates of the strains expressing DDA-TrxA and DDD-TrxA to denaturing conditions in the presence or absence of the reducing agent β-mercaptoethanol (Fig. 3B). The high-molecular weight band was diminished in the presence of β-mercaptoethanol, suggesting that there is a disulfide bonded adduct between the overexpressed thioredoxin and a larger protein that interacts with it.

In the western blot of the fractionation experiment, Fig. 3B, the β-lactamase band appears to be absent from the DDA-TrxA lanes, which suggested that β-lactamase might form the adduct with thioredoxin. However, separate incubation with thioredoxin and β-lactamase antibodies showed that this disulfide-bonded adduct does not involve β-lactamase (Fig. 3B). It is therefore unclear why β-lactamase is not visible in these fractions. It could be the result of proteolytic activity of a stress-related enzyme in response to the toxic effects of the overexpressed thioredoxin. In Fig. 3A the processed form, at least, seems to be absent from the whole cell fraction of the *DDA-trxA* strain. It is worth noting that the western blots in Figs. 1 and 3A were made using the same whole-cell samples. A repeat of this fractionation experiment (inducing *DDA-trxA* with 25 μM IPTG) is therefore necessary in order to investigate the vanishing β-lactamase bands. Fig. 3B shows that β-lactamase is still present when the *DDA-trxA* construct is induced with the same concentration of IPTG, in independent whole-cell extracts, so the absent β-lactamase is likely an artefact of the experiment.

To determine whether a mutationally increased plasmid copy number was the cause of overexpression, we streaked out all *SS-trxA* strains on plates with ampicillin concentrations in the range of 0.2-5 mg/ml, as the plasmids carry a β-lactamase gene and the level of ampicillin resistance should be proportional to plasmid copy number. All strains grew at ampicillin concentrations up to 2 mg/ml and none grew at 4 or 5 mg/ml (results not shown). As there were no visible differences in ampicillin resistance between the strains encoding the overexpressed constructs (DDT-TrxA and DDA-trxA) and the others, an increase in plasmid copy number is probably not the cause of overexpression.

To show that the overexpression wasn't a result of a *lacI* mutation in the plasmid(s), future experiments should include β-galactosidase assays in a *lacI-* strain transformed with each of the plasmids. Plasmid mutations as the cause of overexpression could also be ruled out by repeating the cloning of the signal sequences into the parent vector.

The overexpression of the DDT or DDA signal sequence is probably not wholly due to its C terminus (*e.g.* it could be the result of some secondary structural feature) as the ADA and TDT signal sequences do not cause overexpression.

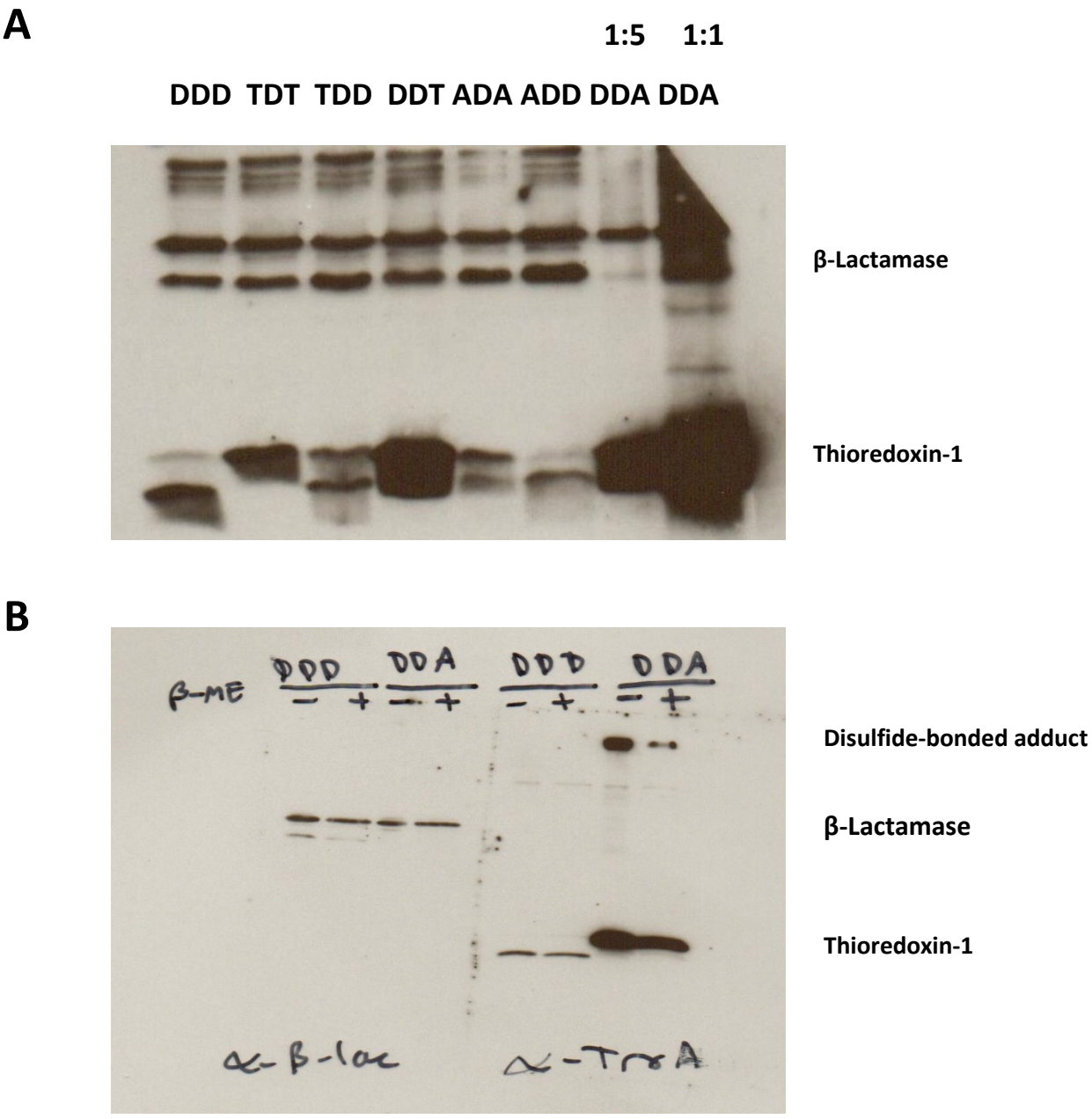


Fig. 3. A) DDT and DDA are overexpressed. Whole-cell samples of all strains were electrophoresed, followed by western blotting against β-lactamase and thioredoxin. An additional DDA sample, which was diluted 1:5 in SDS sample buffer, was included. **B) DDA-TrxA forms a disulfide-bonded adduct.** In order to test if the high molecular weight band that appears in DDA lanes when expression is induced with 25 μM IPTG is an adduct between thioredoxin and another protein, whole-cell extracts were boiled either with or without β-mercaptoethanol and left overnight. This was followed by SDS-PAGE and western blotting with β-lactamase and thioredoxin antibodies separately, in order to determine whether the adduct includes β-lactamase.

The expression of another signal sequence-thioredoxin fusion was greatly affected by a change in the C terminus of the signal sequence: The replacement of the proline residue of the TDT C terminus (VPAASA) with an alanine resulted in barely visible expression. To confirm that this was an effect of the signal sequence fusion and not a cloning problem, expression of VAAASA-TrxA was tested by increasing the concentration of inducer and comparing expression to that of TDD-TrxA, which is exported-which means that both unprocessed and processed forms can be seen for comparison by size (Figure 4).

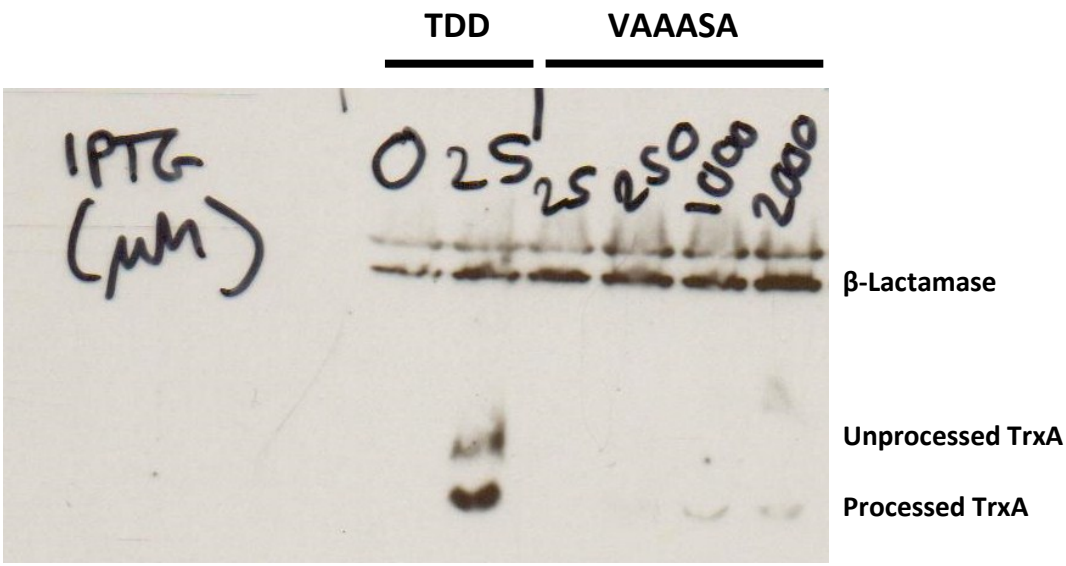


Fig. 4. VAAASA-TrxA expression is very low. 1/100 subcultures of *ΔtrxA* expressing TDD-TrxA or TD-VAAASA-TrxA were grown to OD_{600} ~0.5-0.6 and cell pellets were resuspended in SDS sample buffer, followed by SDS-PAGE and western blotting against β-lactamase and thioredoxin-1.

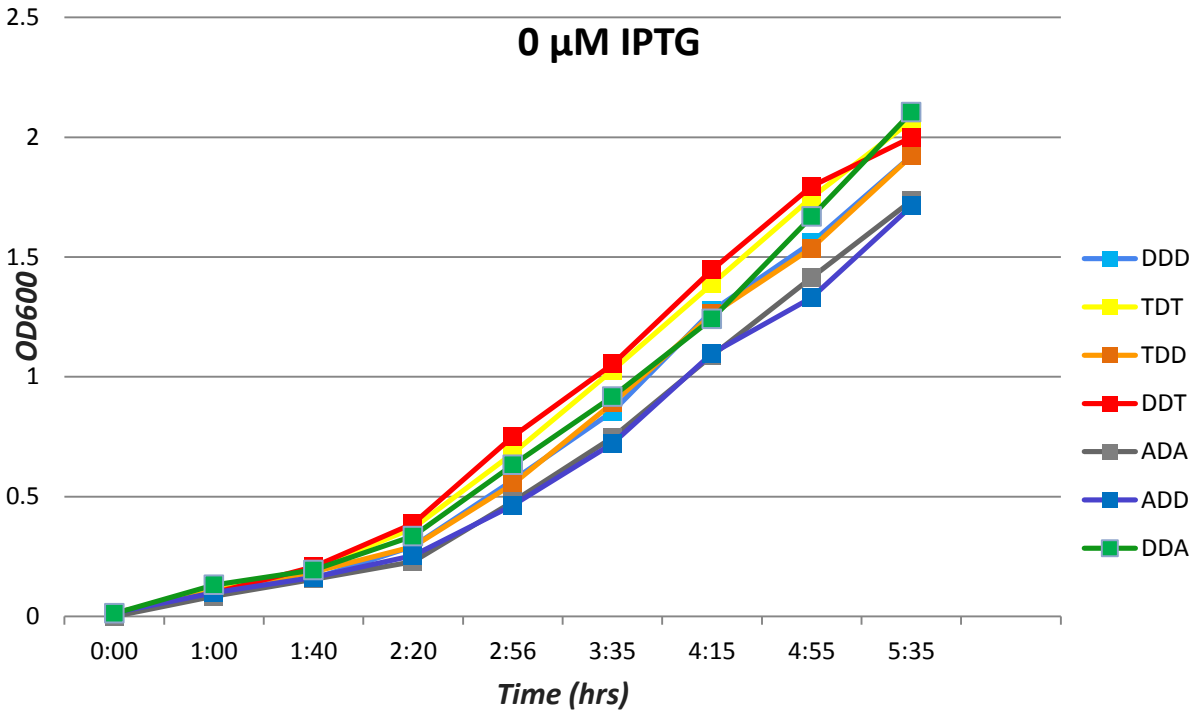
Faint expression could be seen at maximum induction (Figure 4), which suggests that the cloning of the signal sequence was successful and that the construct is expressed. These faint bands appear to be equivalent in molecular weight to processed thioredoxin (fig. 4), suggesting that this signal sequence hybrid is SRP dependent.

Overexpression of DDA-TrxA is toxic

During the fractionation experiments we found that induction with 25 μM IPTG was necessary to see clear expression of some of the constructs. Initially, we used 10 μM IPTG to induce expression. There was no noticeable toxic effect of the DDA-TrxA construct at this IPTG concentration. However, when induced with 25 μM IPTG the strain expressing DDA-TrxA grew more slowly than the others. This was also the condition under which the disulfide-bonded adduct appeared. If the overexpressed thioredoxin forms a disulfide-bonded complex with an essential cytoplasmic protein, for example, ribonucleotide reductase, this could be the cause of its toxic effect. However, we later found that the DDA signal sequence is also toxic when fused to alkaline phosphatase-although this could cause toxicity through a different mechanism.

Before this, we found that DDA strains did not grow on agar plates containing 250 μM IPTG, whereas the other strains did. To confirm the toxic effect of DDA-TrxA overexpression, we constructed growth curves in the presence or absence of IPTG (Figure 5).

A



B

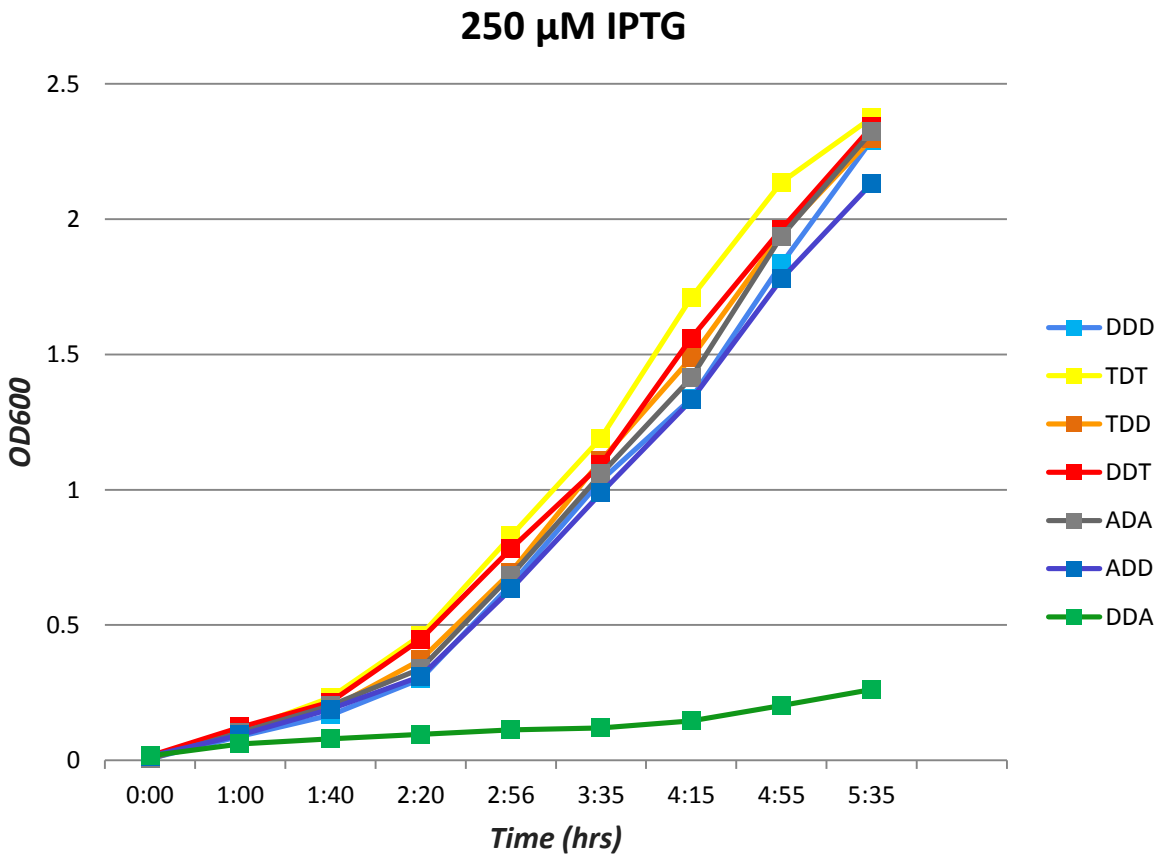


Fig. 5. Toxicity of DDA-TrxA.

ΔtrxA strains with each of the SS-*trxA* plasmids were grown overnight. These were diluted 1/100 without (A) or with (B) 250 µM IPTG and grown for 6 hours, during which OD_{600} readings were taken approximately every 40 min.

DDA is a functional signal sequence

The construction of hybrid signal sequences could prevent them from functioning as signal sequences altogether, in which case they would appear to be SecA dependent when testing for thioredoxin export. To determine whether the hybrid signal sequences are capable of targeting proteins via the SecA-dependent pathway, we inserted the alkaline phosphatase gene *phoA* into the *trxA* gene of each *SS-trxA* plasmid.

PhoA is a periplasmic enzyme with a SecA-dependent native signal sequence. It is only catalytically active in the periplasm as its folding requires disulfide bond formation. We carried out PhoA assays on cultures of DH5α transformed with the *SS-phoA* plasmids to determine whether the hybrid signal sequences were capable of exporting PhoA by the SecA-dependent pathway (Fig. 6).

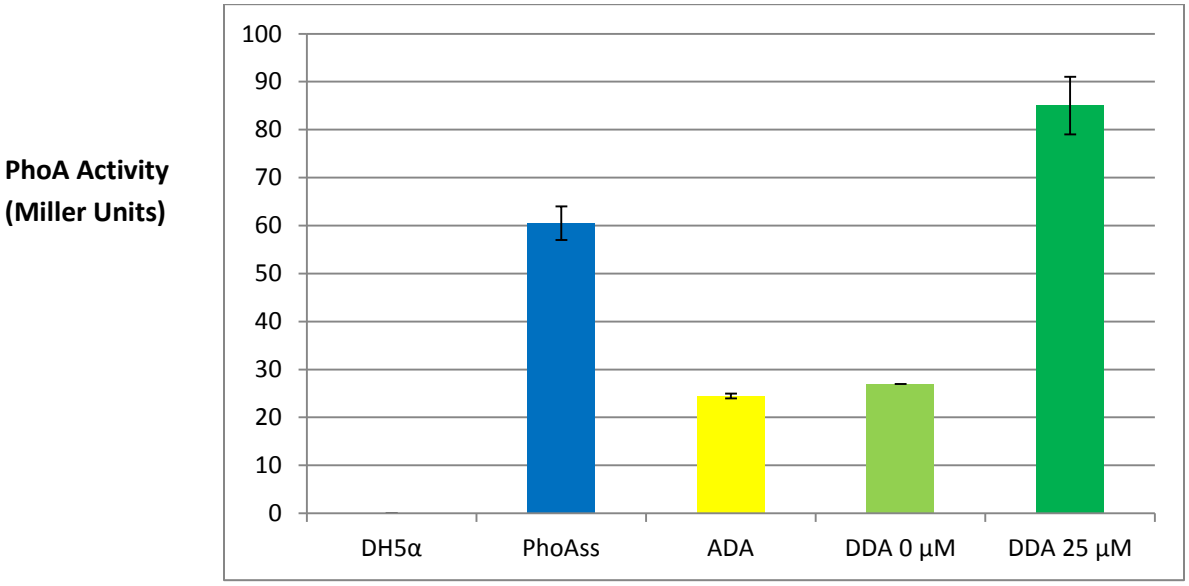


Fig. 6. DDA is a functional SecA-dependent signal sequence. *E. coli* strain DH5α was transformed with the *SS-phoA* plasmids and the resulting strains were assayed for PhoA activity. All strains except "DDA 0 µM" were induced with 25 µM IPTG. Error bars show the range of values.

When its expression was induced the DDA signal sequence gave the highest PhoA activity while DH5α alone, which lacks the *phoA* gene, gave no alkaline phosphatase activity. These results strongly suggest that DDA is capable of exporting PhoA and, assuming that the thioredoxin experiment is robust, is therefore a SecA-dependent signal sequence.

PhoA activity in the strain expressing the DDA-PhoA fusion was higher than in any other strain, notably including the PhoAss-PhoA construct (in which the DNA sequence of the native PhoA signal sequence was inserted into TrxA, as in the other PhoA fusions). Also, PhoA activity was higher in the uninduced DDA-PhoA strain than in the induced ADA-TrxA strain (Fig. 6). This is consistent with the known overexpression of *DDA-trxA*.

Thioredoxin can partially restore motility in a *dsbAdsbCdsbD* mutant when fused to the TDD and ADD signal sequences

As described earlier, thioredoxin export can be indirectly detected by testing motility in a strain deficient in the periplasmic disulfide bond-forming enzyme DsbA. With the intention of cloning randomly mutagenised hybrid signal sequences into pMO2, transforming these into a *dsbA* mutant and selecting for motility we set up a motility test in this strain using control plasmids encoding known signal sequences.

We found that in the *dsbA* mutant alone, the background level of motility was too high to distinguish between signal sequences that could or could not export thioredoxin. This was probably due to suppressor mutations in the membrane-bound protein DsbD, which normally reduces disulfide bonds in the periplasmic disulfide bond isomerase DsbC. When DsbD cannot function, DsbC can be oxidised by the membrane-bound, disulfide-forming protein DsbB-which normally forms disulfide bonds in DsbA. DsbC can then substitute for DsbA, forming disulfide bond in other periplasmic proteins including FlgI.

We therefore proceeded to transform a *dsbAdsbC* mutant with the control plasmids and tested for motility in this strain. To optimise the test, we also tried induction with different IPTG concentrations in the range 0-1 mM, different agar concentrations (0.25-0.35%), different temperatures (30°C and 37°C) and constructed the triple mutant strain *dsbAdsbCdsbD.* The triple mutant provided clearer differences in motility due to thioredoxin export and these differences became clear more quickly (after incubation at 30°C for 1 day instead of 2 days).

The control motility plates showed that TrxA fused to an SRP-dependent signal sequence (DsbAss) can partiallyrestore motility in the *dsbAdsbCdsbD*mutant strain, while it cannot when it is fused to the SecA-dependent signal sequences β-lacss and PhoAss (Fig. 7A).

Of the TraUss hybrids, only TDD-TrxA was able to restore some motility to the mutant strain (Fig. 7B). This is consistent with the result of the fractionation experiment with the same signal sequences, which showed that TDD, but not TDT or DDT, can export TrxA. However, the degree to which TDD-TrxA restores motility appears to be less than that by TrxA fused to the full DsbA signal sequence (DDD). A similar result can be seen with the AppAss hybrids: ADD-TrxA appears to increase the motility of the mutant by a much smaller amount than DDD-TrxA, while ADA-TrxA and DDA-TrxA do not increase motility (DDA-TrxA appears to decrease the motillity of the strain, because of its toxicity) (Fig. 7C). This could be due to differences in expression of the signal sequences, or they could have different levels of affinity for the SRP.

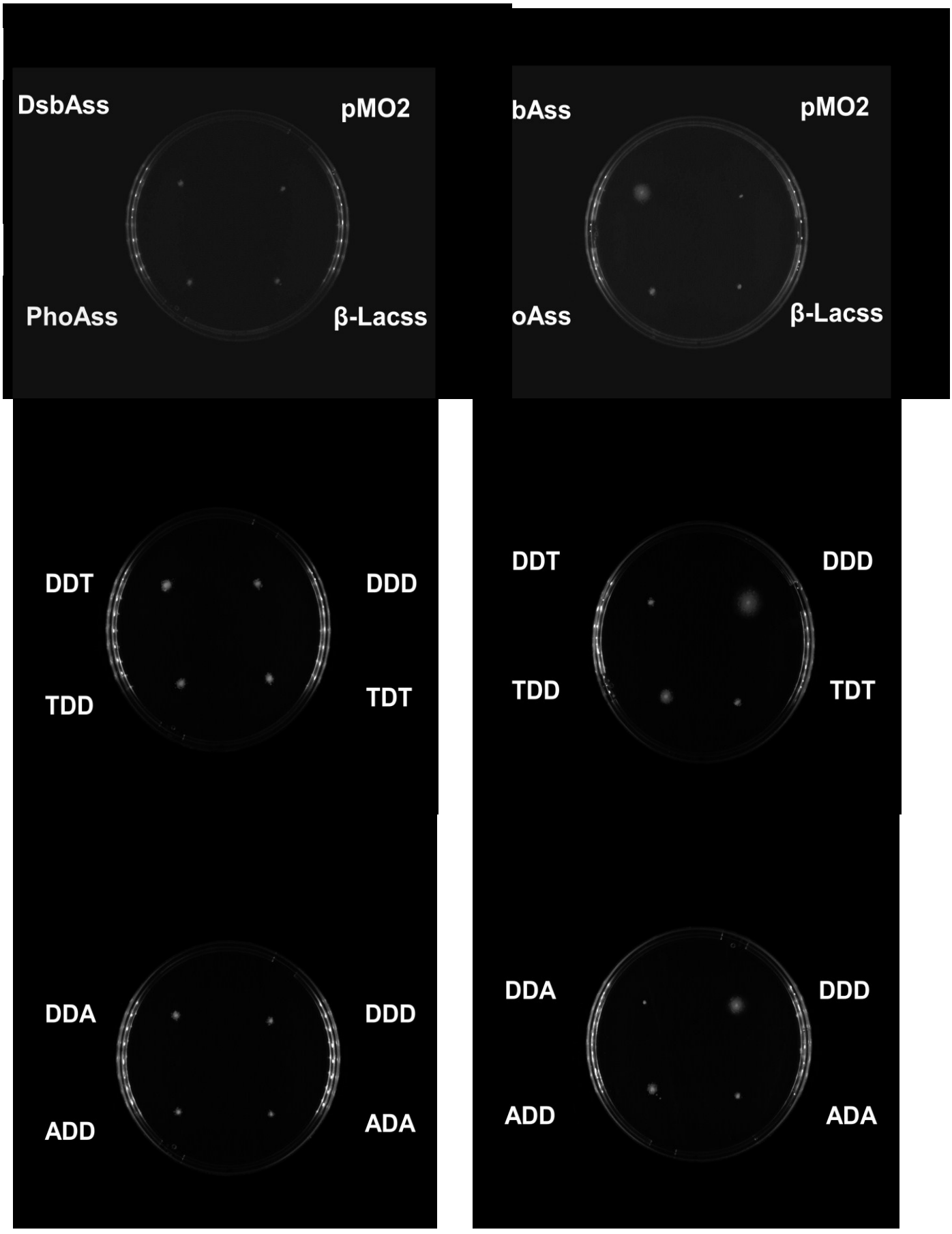


Fig. 7. Motility tests. Low agar LB plates were pricked with the appropriate *E. coli* strains and incubated for 1-3 days at 30°C.

Deletion of both *dsbA* and *dsbC* results in temperature sensitivity

During the construction of the triple *dsb* mutant strain for the motility tests, we found that the *dsbAdsbC* mutant barely grew at 42°C, even though the *dsbA* single mutant did grow at this temperature. To determine whether this combination of mutations alone leads to temperature sensitivity (*e.g.* as opposed to a single *dsbC* mutation) and to try to find out more about the cause, we tested a set of *dsb* mutant strains (and the wild type strain BW25113) for temperature sensitivity at 42°C (Fig. 8).

Of all the dsb mutant strains tested, only the dsbAdsbC and dsbAdsbCdsbD mutants were unable to grow at 42°C (Fig. 8). When incubated at 37°C, both strains had a mucoidal phenotype.

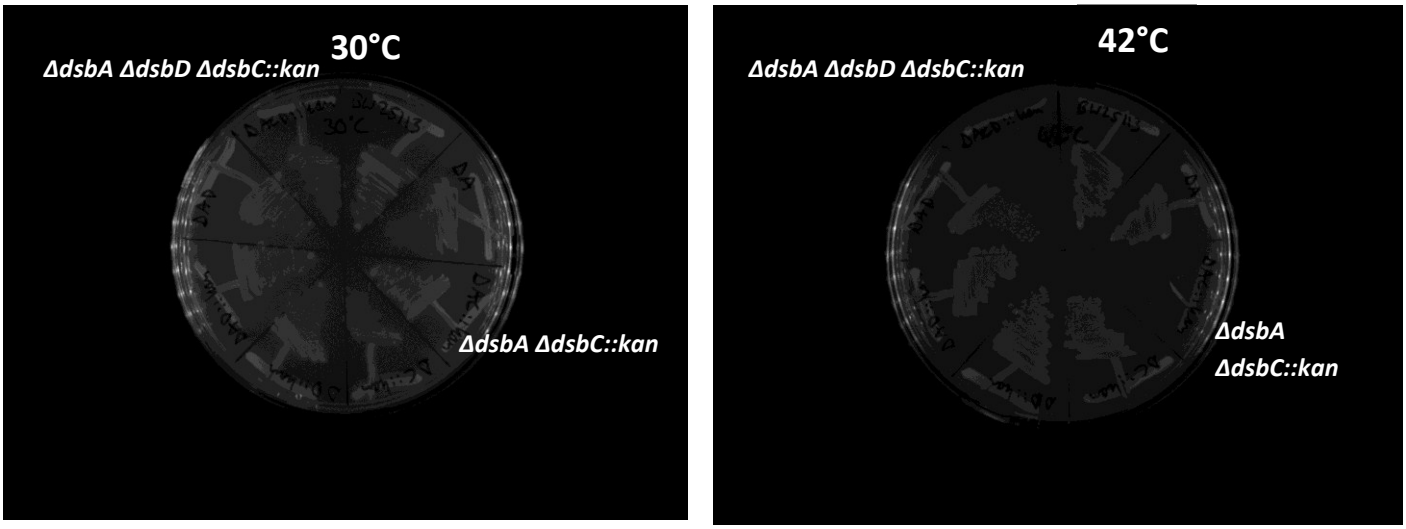


Fig. 8. Temperature senstivity of *dsb* mutants Clockwise from top right: BW25113, *ΔdsbA*, *ΔdsbAΔdsbC::kan*, *ΔdsbC::kan*, *ΔdsbD::kan*, *ΔdsbA ΔdsbD::kan*, *ΔdsbA ΔdsbD*, *ΔdsbA ΔdsbD ΔdsbC::kan*. Single mutants were obtained from the Keio collection. Double and triple mutants were constructed using P1 transduction with lysates of Keio mutant strains. Strains were restreaked onto LB agar and incubated overnight at 30°C or 42°C.

Discussion

It is currently believed that SRP dependence is determined by signal sequence hydrophobicity. While signal peptides lack sequence similarity, when ranked by the hydrophobicity of their cores all of the SRP-dependent signal sequences are above a certain threshold (Huber et al. 2005(a)). The SRP was previously thought to only be involved in the targeting of inner membrane proteins to the Sec translocon in bacteria. Inner membrane proteins do not have cleavable N-terminal signal sequences and the SRP recognises their highly hydrophobic transmembrane domains. So, it is not surprising that the cleavable N-terminal signal sequences now known to be recognised by the SRP are all very hydrophobic. Importantly, two studies have shown that a SecA-dependent signal sequence can be diverted to the SRP pathway by increasing its hydrophobicity. However, some SecA-dependent signal sequences have similarly hydrophobic cores (Huber *et al.*, 2005(a)). Previously, the hydrophobic cores of SRP- and SecA-dependent signal sequences were swapped in order to find out whether the hydrophobic core alone or the N- and C-termini are responsible for SRP dependence (Huber, 2006). None of the resulting hybrids were SRP-dependent, showing that a) the N- and C-termini play a role in determining SRP dependence and b) there might be another feature of the core, other than its hydrophobicity, that determines SRP dependence. In this study, we constructed hybrid signal sequences to test individually the role of the N- or C-terminus in SRP dependence. The results suggest that the C terminus of a signal sequence plays a role in determining which branch of the Sec pathway it is targeted to, presumably by SRP recognition.

It is important to remember that when the similarly hydrophobic cores of SRP- and SecA-dependent signal sequences were swapped, none of the resulting hybrids were SRP-dependent (Huber, 2006). This shows that even with a core of sufficient hydrophobicity (according to a

particular hydrophobicity scale), the N- and C-termini alone cannot make a signal sequence SRP dependent. It could be that specific features within the sequence of the hydrophobic core are required, or a secondary-structural characteristic is required for some part in the SRP or translocation pathway. Alternatively, SRP recognition could depend on one (e.g. secondary structural) feature/interface that extends beyond the hydrophobic core. Although the determination of the targeting pathway probably begins very early in translation with the recognition of the signal sequence by the SRP, there are thought to be multiple 'checkpoints' during the SRP pathway where regulation could occur (Zhang *et al.*, 2010). The signal sequence could play a role in pathway determination at later stages of the pathway, up to and even including its interaction with the Sec translocon. If this was the case, perhaps different features of the signal sequence could be important/affect different parts of this targeting pathway. For example, the SRP could initially recognise the signal sequence based on its highly hydrophobic core. The C-terminal domain characteristic could affect another part of the pathway. This could be investigated (partially) by finding out if SRP can bind RNCs with non-SRP dependent hybrid signal sequences that have very hydrophobic/SRP-dependent hydrophobic cores, compared with signal sequences that do not have SRP-dependent/very hydrophobic cores as well as native SRP-and non-SRP-dependent signal sequences.

As we have identified a possible role for the C terminus in SRP recognition of a signal sequence, the next step should be to precisely identify the determinant within this region by constructing intermediate signal sequences with single-base changes from a non-SRP dependent C-terminus (for example, that of TDT) to an SRP-dependent one (that of TDD, which has the DsbAss C-terminus). These would be tested in the same way, by fractionation of cells expressing the signal sequences fused to thioredoxin. A C-terminal determinant, such as the proline residue in the AppAss and TraUss C termini, could be tested more broadly by finding out if other SecA-dependent signal sequences that naturally have this C-terminal feature can be

converted to SRP-dependent signal sequences by swapping their cores with those of SRP-dependent ones. It is also important to find out whether any C-terminal 'determinant' of SRP dependence is present in other naturally SRP-dependent signal sequences at the same position, and more generally, if the C-terminus is important in the SRP dependence of natural signal sequences other than DsbAss.

It is also possible that the mRNA sequence of the signal sequence is itself recognised by the RNA component of the SRP. This could be tested by substituting synonymous codons that would alter or abolish base-pairing interactions and testing the effect on SRP dependence and protein expression.

The influence of signal sequences on protein expression

A contributing factor to the effects of signal sequences seen on protein expression could be the nature of SRP targeting, as SRP-dependent preproteins are generally thought to have lower expression than SecA-dependent substrates. Although the *E. coli* SRP lacks the subunit of the eukaryotic SRP required for translational pausing, if the bacterial SRP pathway could pause translation by other means then this could explain this difference in expression between the two pathways. SRP-dependent effects on protein expression are perhaps not surprising given its proximity with translation of the substrate. This relationship between SRP dependence and protein expression is consistent with the results seen with the two signal sequence fusions that showed the most dramatic expression levels: DDA-TrxA and TD-VAAASA-TrxA. Swapping the C terminus of the SRP-dependent DsbA signal sequence (DDD) to make DDA, abolished SRP dependence and caused overexpression of the fusion. Substitution of the proline from the DDA C terminus (creating the C-terminal sequence TAQSAFA) led to both a reduction in expression

and SRP dependence. Likewise, substitution of the proline from the non-SRP-dependent TDT signal sequence (creating the C terminus VAAASA) caused SRP dependence along with a massive decrease in expression.

The first few codons encoding the signal sequence could also have an effect on translation. While the DDA signal sequence greatly overexpresses, the non-SRP-dependent ADA signal sequence-which only differs by its C terminus-does not. It has been found that signal sequences show a bias for lysine codons at codon positions 2 and 3, which affects the efficiency of translation initiation (Zalucki *et al.*, 2007). The DDA and ADA N termini both have a lysine encoded by the optimal AAA at the second position, but only DDA also has a lysine (AAG) at the third amino acid position.

Other potential contributors of the extreme expression differences seen between some of the signal sequences could be the formation of mRNA secondary structure that lowers expression, or the recognition of the VAAASA fusion by proteases.

References

Auclair, S., Bhanu, M. & Kendall, D., 2012. Signal peptidase I: cleaving the way to mature proteins. Protein science : a publication of the Protein Society, 21(1), pp.13–25. Available at: http://dx.doi.org/10.1002/pro.757.

Baars, L. et al., 2006. Defining the role of the Escherichia coli chaperone SecB using comparative proteomics. The Journal of biological chemistry, 281(15), p.10024–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16352602.

Baba, T. et al., 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular systems biology, 2, p.2006.0008. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16738554.

Bowers, C., Lau, F. & Silhavy, T., 2003. Secretion of LamB-LacZ by the signal recognition particle pathway of Escherichia coli. Available at: http://jb.asm.org/content/185/19/5697.short.

Cherepanov, P. & Wackernagel, W., 1995. Gene disruption in Escherichia coli: Tc R and Km R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Available at: http://www.sciencedirect.com/science/article/pii/037811199500193A.

Datsenko, K. & Wanner, B., 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America, 97(12), p.6640–5. Available at: http://www.pnas.org/cgi/pmidlookup?view=long&pmid=10829079.

Fekkes, P., van der Does, C. & Driessen, A., 1997. The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. The EMBO journal, 16(20), p.6105–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9321390.

Gelis, I. et al., 2007. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell, 131(4), p.756–69. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092-8674(07)01269-X.

Hizukuri, Y. et al., 2006. Role of the intramolecular disulfide bond in FlgI, the flagellar P-ring component of Escherichia coli. Journal of bacteriology, 188(12), p.4190–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16740925.

Huber, D. et al., 2005(a). Use of thioredoxin as a reporter to identify a subset of Escherichia coli signal sequences that promote signal recognition particle-dependent translocation. Journal of bacteriology, 187(9), p.2983–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1583802

Huber, D. et al., 2005(b). A selection for mutants that interfere with folding of Escherichia coli thioredoxin-1 in vivo.Proceedings of the National Academy of Sciences of the United States of America, 102(52), p.18872–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16357193.

Huber, D., 2006. A Genetic Analysis of Protein Folding and Protein Translocation in Escherichia coli. PhD Thesis. Harvard University

Huber, D. et al., 2011. SecA interacts with ribosomes in order to facilitate posttranslational translocation in bacteria.Molecular cell, 41(3), p.343–53. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097-2765(10)01013-0.

Jagath, J. et al., 2001. Important role of the tetraloop region of 4.5S RNA in SRP binding to its receptor FtsY. RNA (New York, N.Y.), 7(2), pp.293–301. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11233986.

Lee, H. & Bernstein, H., 2001. The targeting pathway of Escherichia coli presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. Proceedings of the National Academy of Sciences of the United States of America, 98(6), p.3471–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11248102

Macfarlane, J. & Müller, M., 1995. The functional integration of a polytopic membrane protein of Escherichia coli is dependent on the bacterial signal-recognition particle. European Journal of Biochemistry, 233(3), p.766-771. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1995.766_3.x/abstract

Matoba, S. & Ogrydziak, D.M., 1998. Another Factor Besides Hydrophobicity Can Affect Signal Peptide Interaction with Signal Recognition Particle. Journal of Biological Chemistry, 273. Available at: http://dx.doi.org/10.1074/jbc.273.30.18841.

Palmer, T. & Berks, B., 2012. The twin-arginine translocation (Tat) protein export pathway. Nature reviews. Microbiology, 10(7), p.483–96. Available at: http://dx.doi.org/10.1038/nrmicro281

Randall, L. et al., 2005. Asymmetric binding between SecA and SecB two symmetric proteins: implications for function in export.Journal of molecular biology, 348(2), p.479–89. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15811382.

Schierle, C., Berkmen, M. & Huber…, D., 2003. The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the Escherichia coli periplasm via the signal recognition particle pathway. Available at: http://jb.asm.org/content/185/19/5706.short.

Schulze, R. et al., 2014. Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. Proceedings of the National Academy of Sciences of the United States of America. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24550475.

Ullers, R. et al., 2004. SecB is a bona fide generalized chaperone in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America, 101(20), p.7583–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15128935

Van den Berg, B. et al., 2004. X-ray structure of a protein-conducting channel. Nature, 427(6969), pp.36–44. Available at: http://dx.doi.org/10.1038/nature02218.

Voigts-Hoffmann, F. et al., 2013. The structural basis of FtsY recruitment and GTPase activation by SRP RNA. Molecular cell, 52(5), p.643–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24211265.

Weiss, D. et al., 1999. Localization of FtsI (PBP3) to the septal ring requires its membrane anchor, the Z ring, FtsA, FtsQ, and FtsL. Journal of bacteriology, 181(2), p.508–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9882665.

Wild, J. et al., 1992. DnaK and DnaJ heat shock proteins participate in protein export in Escherichia coli. Genes & Development, 6. Available at: http://dx.doi.org/10.1101/gad.6.7.1165

Zalucki, Y., Power, P. & Jennings, M., 2007. Selection for efficient translation initiation biases codon usage at second amino acid position in secretory proteins. Nucleic acids research, 35(17), p.5748–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17717002.

Zhang, X. et al., 2010. Sequential checkpoints govern substrate selection during cotranslational protein targeting. Science (New York, N.Y.), 328(5979), p.757–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20448185.

Investigating the Essentiality of Genes in the Presence of Azide

An MRes project

by Rachael Chandler

Supervised by Prof. Ian Henderson

Abstract

Sodium azide is thought to cause toxicity in *E. coli* primarily by targeting SecA, an essential motor ATPase that drives protein export (it also inhibits other proteins such as the F_0F_1 ATPase). SecA was first found to be involved in azide toxicity when mutations that resulted in azide resistance were found to map to the *secA* gene. It was also shown that azide inhibited SecA-dependent protein translocation *in vivo* and *in vitro*. Azide has since been used in studies of SecA function and it is generally accepted to be a cause of translocation stress.

As azide causes translocation stress, it could be used to identify the functions of uncharacterised secretion proteins-particularly ones that become more important during translocation stress (for example, because they function to increase the efficiency of translocation).

The purpose of this study was to identify genes that become essential during azide stress, using transposon directed insertion site sequencing (TraDIS)-a high throughput sequencing method used to compare gene essentiality under different conditions.

Introduction

Sodium azide is a universally toxic compound, known to inhibit essential enzymes such as cytochrome oxidases and the F_0F_1 ATPase (Fortin *et al.*,1990). Azide-resistant *E. coli* mutants were first obtained in 1950 (Lederberg, 1950). The azide resistance, or *azi*, mutations were later found to be within the *secA* gene, and *azi* mutants are resistant to azide concentrations up to 4 mM (Huie and Silhavy, 1995).

SecA is a component of the Sec translocation system. Sec substrates are exported through a channel in the cytoplasmic membrane that is formed by the complex SecYEG. SecA is an essential motor ATPase that associates with the SecYEG translocon and drives translocation of substrates into the periplasm. It is also thought to recognise these substrates cotranslationally and deliver them to SecYEG (Huber *et al.*, 2011; Huber *et al.*, unpublished manuscript).

The fact that all of the mutations causing resistance to azide have been found in *secA* shows that SecA and Sec translocation in general is the most important target of azide at 1-4 mM.It strongly suggests that SecA is directly affected by azide, which is easy to believe as azide is known to inhibit ATPases. Most of the mutations map to NBD1 and NBD2, the domains of SecA that are required for ATP hydrolysis (Huie and Silhavy, 1995).

Interestingly, many of the *azi* mutations were found to be *prlD* suppressor mutations also (Huie and Silhavy, 1995). The *prlD* suppressors are mutations in *secA* that restore the export of substrates with defective signal peptides. While these mutations appear to allow the binding of a wider variety of signal sequences, only one *prlD* suppressor is located near the peptide

crosslinking domain, amino acids 267-340, that was proposed by Kimura *et al*. (1991) as a result of crosslinking N- and C-terminal SecA fragments of a range of sizes to proOmpF-Lpp. The others are located within or near the ATP-binding domains of SecA (Huie and Silhavy, 1995). This might suggest that they enable the translocation of defective preprotein substrates by changing the activation state of SecA. It has been suggested that signal peptides are mainly required for the 'activation' of wild-type SecA (Gouridis *et al.*, 2009), so if SecA were activated by the *prlD* mutation translocation could take place independent of correct signal sequence binding.

Also, the *azi* phenotype was found to be recessive. One possibility is that azide inhibits a step in the SecA ATPase reaction cycle that is required for SecA release from the translocon and this prevents free, azide-resistant SecA from associating with the substrate or the translocon .

However, the recessive *azi* phenotype seen in this experiment could be due to the higher copy number of the wildtype *secA* allele, which could outcompete the azide resistant SecA for binding sites at the inner membrane translocon or for other interacting components of the Sec machinery. Oliver *et al.* (1990) tested this and found that the recessiveness of the *azi* allele was dependent on gene dosage.

The fact that azide resistance is recessive might tell us that azide does irreversibly affect some component of Sec translocation, in a way that could possibly interfere with specific translocation experiments. While the above mechanism of inhibition of translocation would not be a problem in many circumstances, as azide is often intended to be used for the straightforward inhibition of Sec translocation, experiments investigating finer/mechanistic details of SecA or translocation using azide could give misleading results.

Oliver *et al.* (1990) found that azide inhibited the export of SecA-dependent substrates, determined by the processing of pulse-labelled precursors, *in vivo* in strains with wildtype SecA, but didn't affect export in *azi* strains at 1 mM and these strains still showed export at 3 mM sodium azide. These export effects also matched the growth effects of azide on these strains (Oliver *et al.*, 1990).

They tested the inhibition of translocation into SecA-depleted inverted inner membrane vesicles with purified wild-type SecA or Azi4-SecA added, at different concentrations of sodium azide, and found that Azi4-SecA-dependent translocation was significantly less inhibited by azide. They also tested whether sodium azide inhibits the ATPase activity of SecA that is normally stimulated when it binds precursor proteins and found that the ATPase activity of wildtype SecA is more sensitive to azide than Azi4-SecA (Oliver *et al.*, 1990). Sodium azide did not appear to inhibitthe precursor-independent ATPase activity of SecA.

Although azide has been shown to affect SecA-dependent translocation and the ATPase activity of SecA, the direct binding of azide to SecA hasn't been investigated.

Oliver *et al.* also found that the translocation-specific ATPase activity of SecA was less sensitive to azide inhibition than the SecA-dependent *in vitro* or *in vivo*translocation.

It is possible that the mechanism of the azide sensitivity of SecA-dependent translocation isn't simply the inhibition of ATPase activity. Translocation itself seems to be specifically affecteddespite other cellular stresses caused by treatment with azidesince translocation*in vivo* and *in vitro* are similarly affected. It does not, however, preclude the possibility that components of the translocation apparatus are affected by azide.

Direct binding of azide to SecA has not yet been demonstrated, but azide is a strong oxidant that causes multiple effects including DNA damage, oxidative stress and the inhibition of cytochrome oxidases. In order to identify the cellular processes affected by azide and to

investigate how the *E. coli* cell responds to translocation stress, we decided to use TraDIS (transposon-directed insertion site sequencing). This technique is used to identify gene essentiality under certain conditions. It involves the growth in the condition of interest of a library of transposon insertion mutants (which contain insertions in every gene), followed by genomic DNA preparation and high-throughput sequencing primed from the transposons (Langridge *et al.*, 2009). In this method, genes that are essential will contain no (or very few) transposon insertions. However,in addition to essentiality, TraDIS could also be used to identify fitness effects of genes during the inherently competitive growth of the insertion mutant library, by comparing the numbers of reads obtained within each gene in each condition (Langridge *et al.*, 2009). Changes in the fitness effects of genes in the condition of interest could therefore be useful for identifying nonessential genes that affect survival in the condition of interest.

As azide is thought to affect the *E. coli* cell primarily by targeting SecA, we thought that there might be nonessential translocation genes that become more important in the presence of azide-for example, genes that increase the efficiency of translocation. In addition to providing evidence that azide affects the cell at least partly by causing translocation stress, the TraDIS experiment could also allow the investigation of the roles of nonessential Sec proteins that might increase translocation efficiency under certain conditions.

Materials and Methods

Strains used in this study

Strain	Description/Genotype
E. coli BW25113	F-, DE(araD-araB)567, lacZ4787(del)::rrnB-3, LAM-, rph-1, DE(rhaD-rhaB)568, hsdR514
RAC23	BW25113 *ΔyecA::kan*
JW3883	BW25113 *ΔcpxR::kan* (Keio collection)
JW5856	BW25113 *ΔtrxA::kan* (Keio collection)
JW3832	BW25113 *ΔdsbA::kan* (Keio collection)

Genomic DNA preparation, Sequencing and Data Analysis

A library of ~1.1 million *E. coli* BW25113 insertion mutants was created as described in Langridge *et al.*, 2009. In short, transposomes were electroporated into cells. This library was grown to OD_{600} ~1 in LB with three different concentrations of sodium azide: 0 mM, 0.25 mM and 0.5 mM. Genomic DNA was extracted and then processed using a two-step PCR method (Christen *et al.*, 2011) which results in Illumina-compatible products. The PCR products were purified using the Agencourt AMPure XP system by Beckman Coulter. These were then sequenced using an Illumina MiSeq sequencer and the reads were mapped to the *E. coli* reference genome by Ashley Robinson. The number of insertions per gene was then worked out and sorted computationally by Ashley Robinson and Dr. Damon Huber. Full data set is available upon request by emailing Dr. Damon Huber (d.huber@bham.ac.uk).

Growth Curves

Overnight cultures were diluted 1/100 in LB broth containing the appropriate concentration of sodium azide and grown at 37°C with shaking at 180 or 200 rpm. OD_{600} measurements were taken approximately every 30 min over 8 hrs.

Spot Assays

Overnight cultures of BW25113 and either of two kanamycin-resistant deletion mutants were subcultured in a mixture of 1:1 in LB broth with a sodium azide concentration of 0 mM, 0.25 mM or 0.5 mM and left to grow overnight, shaking, at 37°C. Serial ten-fold dilutions of the mixed cultures were then spotted onto LB agar plates with or without 30 µg/ml kanamycin and incubated at 37°C overnight.

Results

Optimisation of TraDIS growth conditions

We wanted to establish concentrations of azide that were subinhibitory, to allow us to use TraDIS to investigate the cellular response to severe translocation stress. We compared the growth of *E. coli* cultures with a range of sodium azide concentrations (0-10 mM).

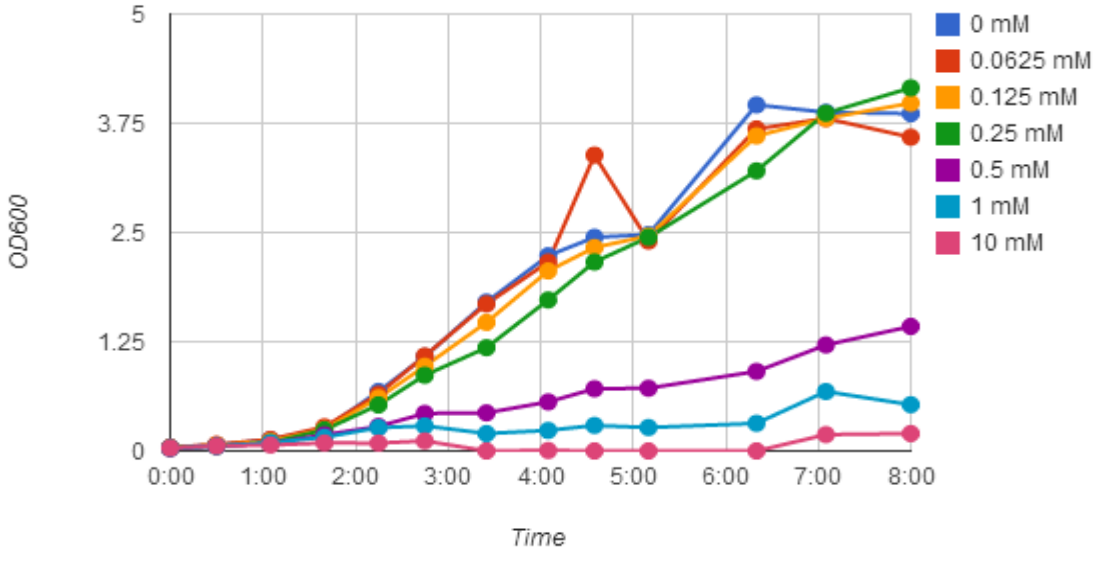


Fig. 1.Inhibition of cell growth by sodium azide. Overnight cultures of *E. coli* strain BW25113 were diluted 1/100 in LB with different concentrations of sodium azide. OD_{600} readings were taken over 8 hours.

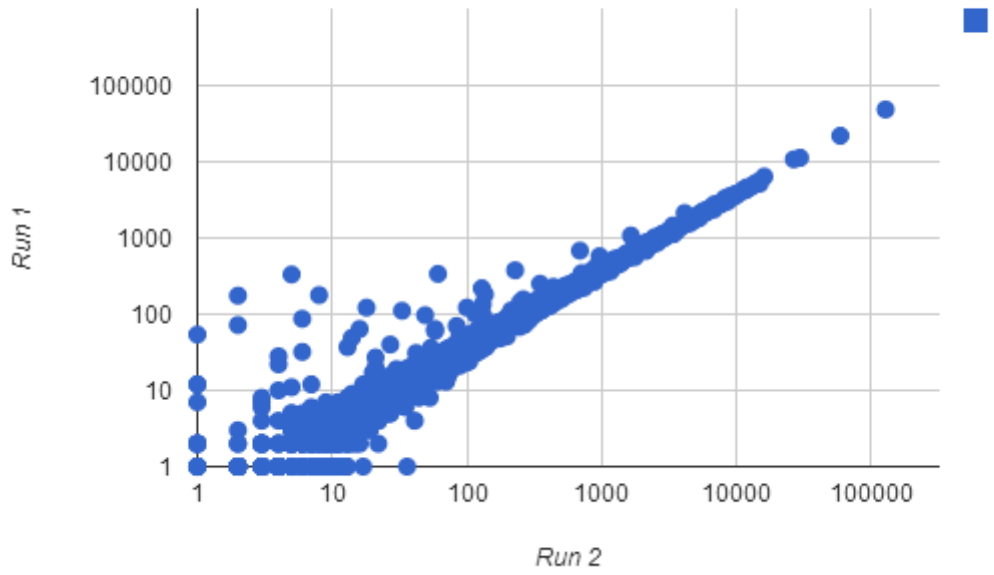
The highest sublethal concentration of azide used in this experiment was 0.5 mM (Figure 1). In order to determine whether there were any dose-dependent effects of azide, two concentrations of sodium azide were used in the screening conditions for the TraDIS experiment, 0.25 mM and 0.5 mM, along with a control culture with no sodium azide added.

The 0.5 mM azide condition might result in the essentiality of a larger number of genes during competitive growth, due to indirect or general effects of stress on the cell, so results gained from this condition might be less relevant than those gained from the 0.25 mM azide condition. However, the higher azide concentration would presumably allow the detection of a higher number of conditionally essential gene candidates and the higher sensitivity would be useful if the 0.25 mM condition yielded no differences in essentiality. The 0.25 mM condition might result in more accurate azide-specific effects rather than bringing about a state of general cell stress. The use of both azide concentrations could enable dose dependent effects to be seen. These effects could more convincingly show that the fitness effects on certain genes are azide-specific.

TraDIS was carried out to investigate gene essentiality in the presence of azide

A TraDIS library of approximately one million insertion mutants of *E. coli* strain BW25113 was grown to an OD_{600} ~1 in three different conditions: 0, 0.25 mM and 0.5 mM sodium azide. Genomic DNA was extracted and PCR-amplified from the site of each transposon insertion using transposon-specific and arbitrary primers and TraDIS was carried out using an Illumina MiSeq sequencer.

The results of the sequencing data were examined using two different methods: 1) manually, using a genome browser (Artemis) and 2) by counting the number of insertions in each gene, regardless of their locations within the gene.



There was little variability in insertion number data between the two runs, with most variability occurring in genes with very few insertions (Figure 2).

Fig. 2. Number of insertions per gene in run 1 vs run 2. Genes were ranked by the number of insertions detected within them in each sequencing run. The graph shows the ranked insertion number values from run 1 vs those from run 2.

Many genes showing significant changes in insertion number in the presence of 0.5 mM sodium azide can be grouped into functional categories

In order to identify genes that showed a substantial increase or decrease in number of insertions in 0.5 mM azide compared to the 0 mM control, the genes were ranked in order of the insertion number change found in the library grown in 0.5 mM azide. When the ratios of the number of insertions in each gene was plotted in a rank-ordered list (Figure 3), most genes showed a ratio of ~1, indicating that most genes did not not affect the ability of BW25113 to grow in the presence of azide However,the plot tapered at either end with inflections at ratios of ~0.5 and ~2.we therefore chose to select those genes that showed a change of two-fold or

greater compared to the 0 mM control. 297 genes lay below our cut-off point for decreases in the number of insertions in 0.5 mMazide, while 326 genes lay above the cut-off point for significant increases in insertion number. However when genes with a total number of insertions in both sequencing runs of less than 30 were eliminated to avoid including stochastic changes in read number amongst genes with very few insertions, 229 genes remained below the threshold for 'significant' insertion number decreases and 264 remained above the threshold for insertion number increases.

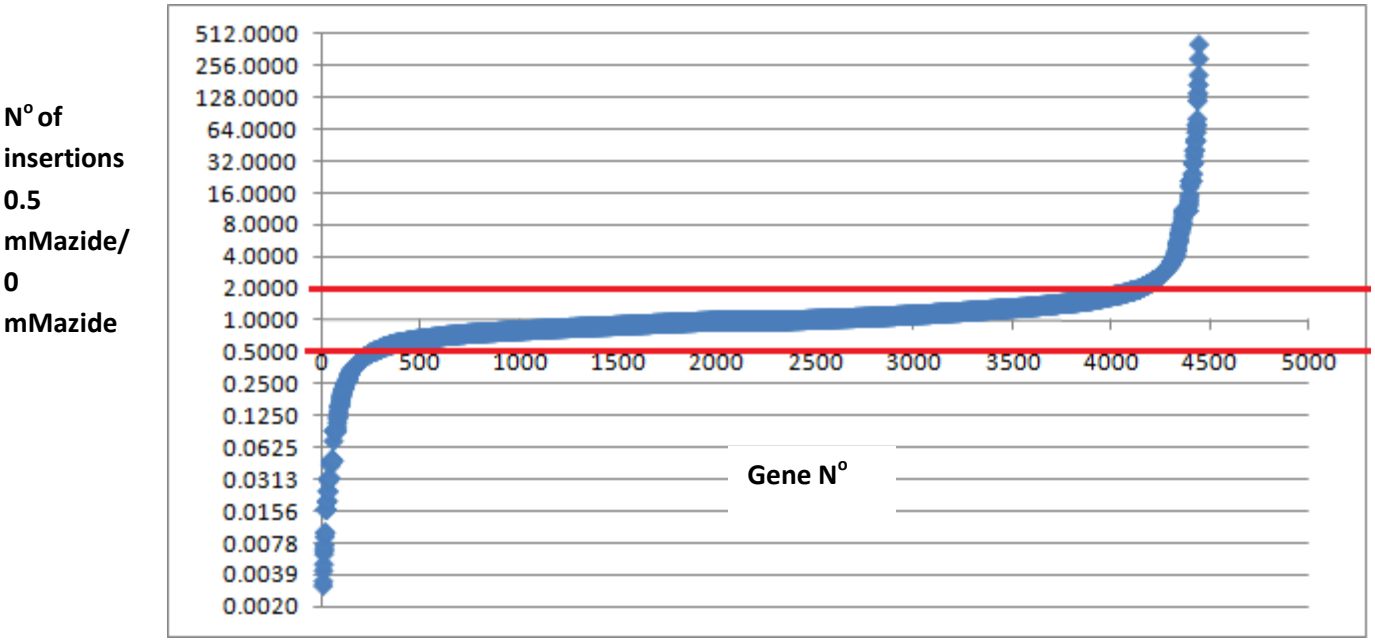


Fig. 3. Genes were ranked by the fraction of the 0 mM insertions when in the presence of 0.5 mM azide. All genes were ranked in order of the change in the number of insertions in 0.5 mM azide compared to 0 mM azide. The red lines indicate the cut-off points used to make a selection of genes that showed a significant change in insertion number in 0.5 mM azide compared with 0 mM azide. These data are from sequencing run 2.

Within the groups of genes that increased or decreased significantly in insertion number when exposed to 0.5 mM azide, we looked for examples of trends in gene function in order to investigate the specific effects of azide on the cell.

Genes involved in protein translocation

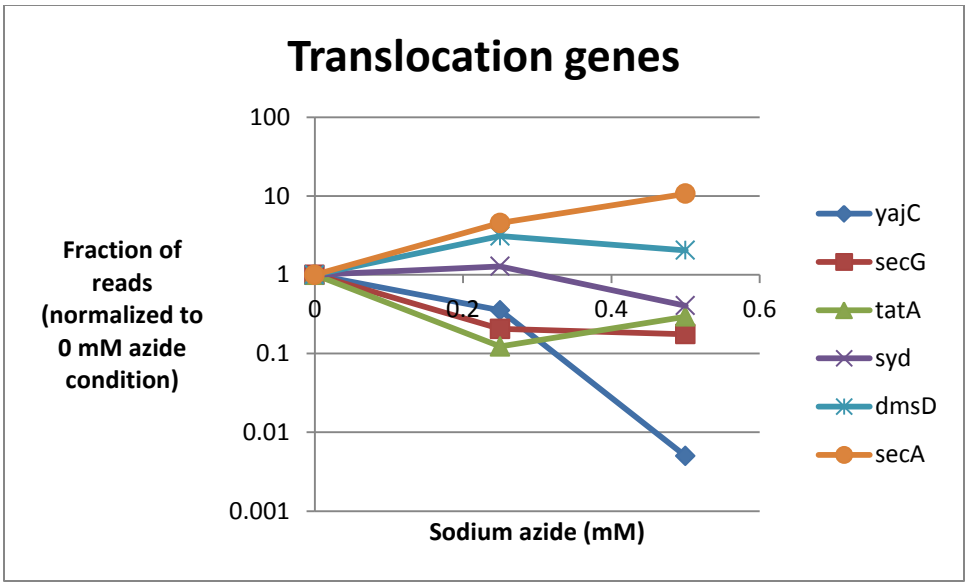


Fig. 4. Translocation genes showing significant changes in insertion number in the presence of azide. (Run 2)

Of the genes showing a change in insertion number of at least two-fold in 0.5 mM azide, there were six genes known or thought to be involved in translocation (figure 4). Of these, *yajC* showed the largest decrease in the number of reads in the presence of 0.5 mM azide. In both runs its decrease progressed with increasing azide concentration, to 0 reads in 0.5 mM. YajC forms a complex with SecD and SecF, although its exact role in secretion has not been characterised.

It is possible that the decrease in *yajC* insertions is due to insertions in this gene having a polar effect on *secD* and *secF*, which are downstream of *yajC* in the same operon. Although *secD* and *secF* are not generally considered essential, they appear to be essential in the results of the TraDIS experiment (Figure 5). *secDF* mutants usually have severe growth defects, so their essentiality in this case is probably due to the competitive growth conditions used for TraDIS. However, polar effects are unlikely as the TraDIS transposon has outward- facing promoters.

Alternatively, this could possibly show that *yajC* mutants in particular have a competitive disadvantage in the presence of azide (and maybe also in the absence of azide, although to a lesser extent, as the total number of *yajC* reads obtained was fairly low and near our 'cut-off' point for each run) which might not be surprising as YajC is thought to function in a complex with SecD and SecF.

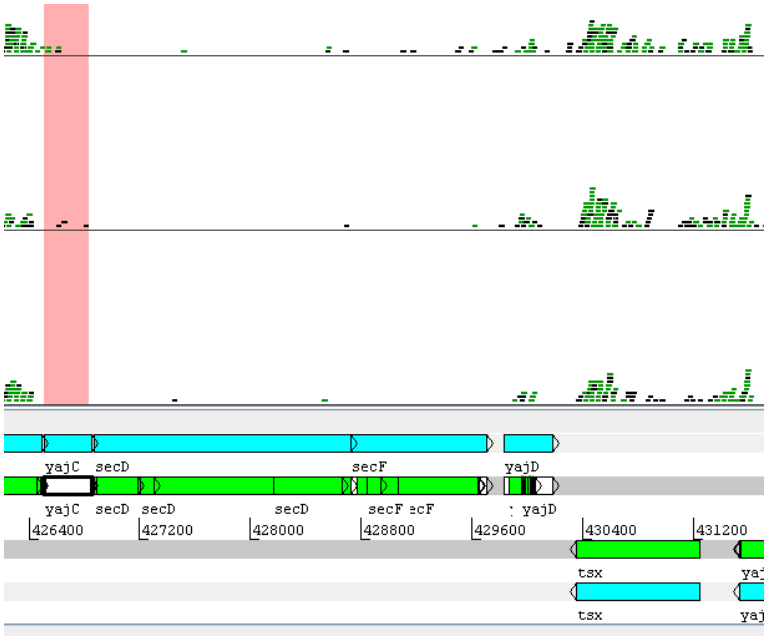


Fig. 5. Possible polar effects of *yajC* insertions. Screenshot taken from genome browser, showing number and location of insertions. Insertions in *yajC,* which is directly upstream of *secD* and *secF*, decrease significantly in the presence of 0.5 mM azide.

Surprisingly, the gene encoding SecA shows an increased number of reads with increasing azide concentration (Figure 4). The insertions appear to be localised to a C terminal area of SecA (Figure 6), starting from ~818aa. The significance of the increased number of insertions is discussed in the “Discussion” section.

Fig. 6. C-terminal insertions in secA increase with azide concentration.

Genome browser screenshot. A dose-dependent increase in insertions at the 3' end of the essential gene *secA* can be seen in the presence of azide

Another *sec* gene showing a decrease in insertion number is *secG* (figure 4), which encodes a nonessential component of the SecYEG channel complex that has been found to be required for translocation at cold temperatures and in the absence of a proton motive force.

Also, genes encoding components of the nonessential Tat (twin-arginine translocation) pathway showed a decrease in insertion number in the presence of azide. The Tat genes *dmsD* and *tatA* both showed significant changes in insertion number (figure 4).

Genes involved in cell envelope integrity

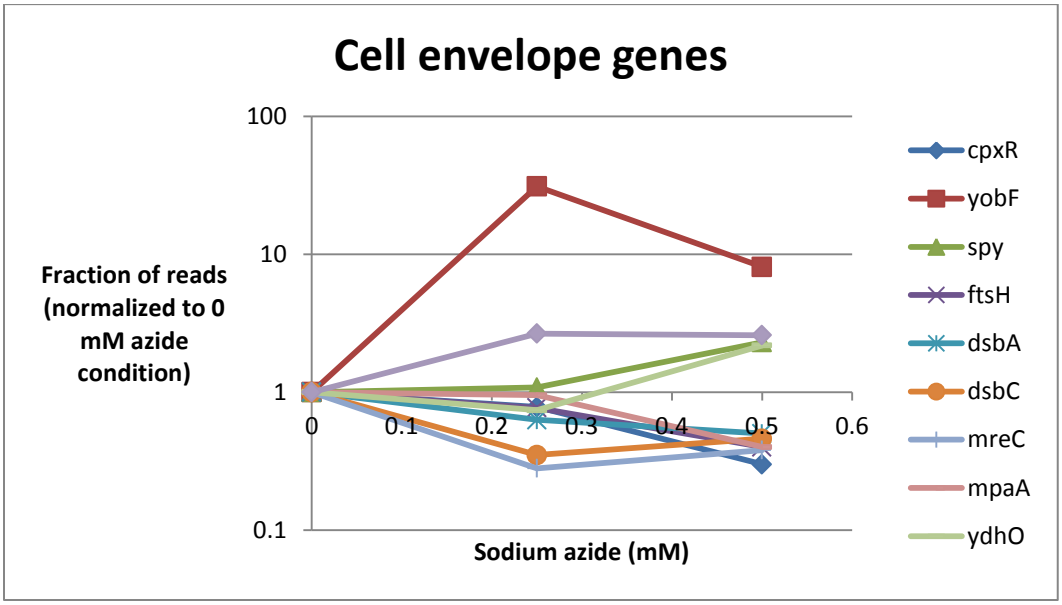


Fig. 9. Genes involved in cell envelope biogenesis or stress. Many genes with roles in dealing with cell envelope stress show increases or decreases in insertion number of two-fold or greater in the presence of azide.

There were also a lot of insertion number changes found in genes involved in peptidoglycan synthesis and turnover, such as ***mpaA*** (mureintripeptideamidase A), ***mreC*** (encoding a part of the longitudinal peptidoglycan synthesis/chromosome segregation-directing complex), ***ydhO*** (murein DD-endopeptidase MepH), ***oppD*** (mureintripeptide ABC transporter) and ***pbl*** (a predicted peptidoglycan-binding enzyme) (Figure 9). ***cpxR***, which encodes a response regulator that activates envelope stress response genes, also appears to be important for fitness in the presence of azide-showing a dose dependent decrease in insertion number, dropping 70% in 0.5 mM azide. CpxR is a response regulator that is activated by phosphorylation by CpxA, a sensor kinase that detects conditions such as/associated with cell envelope stress. However, *cpxA* insertions did not change significantly in the presence of azide (Figure 10).

Fig. 10.Conflicting importance of the CpxAR two-component system. *cpxR* shows a dose-dependent decrease in reads in the presence of azide, but the *cpxA* read count does not change significantly.

Although CpxA is thought to be the only kinase that interacts with CpxR, CpxR can be activated by other means. The signalling molecule acetyl phosphate, which is produced by the phosphotransacetylase-acetate kinase (Pta-AckA) pathway, can also phosphorylate CpxR. This is thought to be a way in which diverse signals, for example-related to carbon metabolism, are integrated by CpxR (Wolfe *et al.*, 2008). This could mean that, in the presence of azide, it is not envelope stress factors that are required for CpxR activation but other, potentially metabolic, conditions that require its activity and stimulate it by other means. It is also possible that CpxR has a basal level of activity even if CpxA is nonfunctional. Within the 38 operons regulated by phosphorylated CpxR are 8 genes that showed significant insertion number changes in the presence of azide. Among their respective proteins are some that are likely to be involved in dealing with envelope stress: the periplasmic chaperone Spy increases about two-fold in insertion number in 0.5 mM azide, whereas DsbA and DsbC-periplasmic enzymes that catalyse

the formation and rearrangement, respectively, of disulfide bonds-both show a two-fold decrease in insertion number in 0.5 mM azide. DsbC has also been found to have a role in dealing with oxidative stress in the periplasm, where it can reduce sulfenic acid groups that result from attack by reactive oxygen species on the sulfhydryl groups of proteins that have a single (free) cysteine residue (Denoncin *et al.*, 2014). If envelope stress is an issue in the presence of azide, the increase in *spy* insertion number is surprising. However, this increase only occurs in 0.5 mM azide: the number of insertions does not change in 0.25 mM azide. Increases in insertions in periplasmic chaperones such as Spy, which assist in protein refolding and prevent aggregation, could possibly be due to an advantage of having free misfolded or aggregated proteins that could induce envelope stress response pathways (for example, by the Cpx pathway) at a basal level so that they are already active during a condition of envelope stress, which might occur in 0.5 mM azide (figure 9).

A similar pattern can be seen with ***yobF*** (figure 9), which is thought to be involved in stress responses and the deletion mutant of which was found in a competition growth assay to have an increased sensitivity to envelope stress induced by SDS and EDTA (Hobbs *et al.*, 2010). In the presence of azide, *yobF* insertions increase (figures 9 and 11).

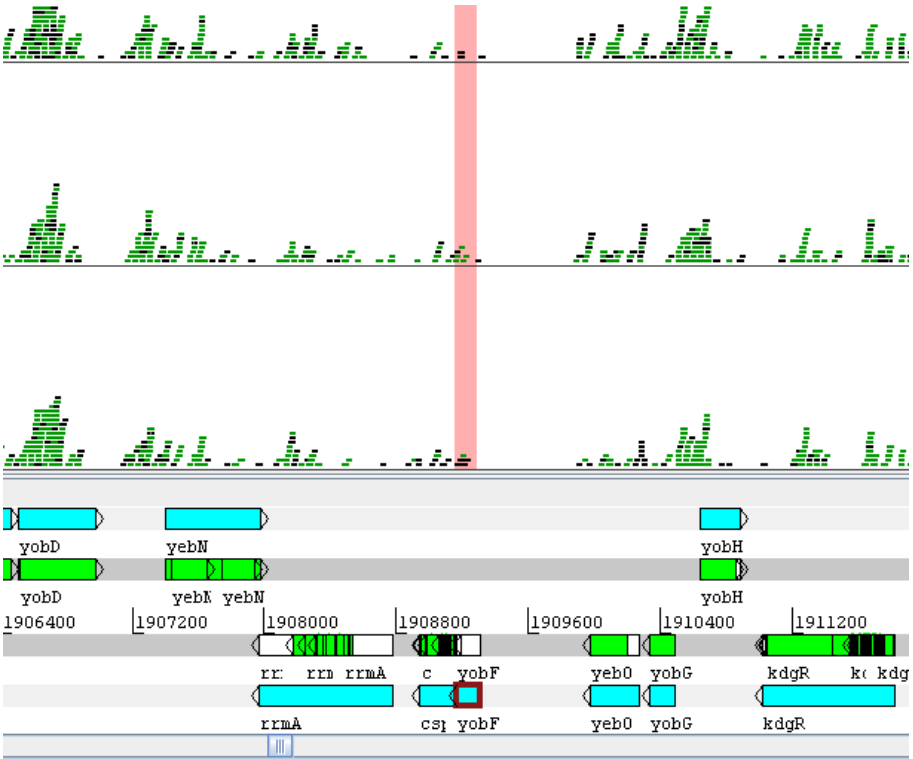


Fig. 11. Genome browser screenshot showing an increase in insertion number in the potential envelope stress-related gene *yobF* in the presence of azide.

Another important envelope stress factor is the membrane-bound protease **FtsH**, which is thought to degrade incorrectly inserted inner membrane proteins.

Fig. 12. Azide concentration-dependent, C-terminal insertions in FtsH. Screenshot taken from genome browser, showing number and location of insertions in 0 mM (top section), 0.25 mM (middle) and 0.5 mM (bottom section) sodium azide. A dose-dependent decrease in insertions can be seen in *ftsH*, an essential gene.

ftsH is an essential gene but has a significant number of insertions, which reduce in number with increasing azide concentration, at the 3' end of the gene. As FtsH is essential the insertions probably do not prevent it from functioning, although they appear to have a detrimental effect in the presence of azide (Figure 9; Figure 12). The part of the protein that would be affected by these insertions, C-terminal to around amino acid 580, is cytoplasmic and does not include characterised catalytic or otherwise functional domains. The dose-dependent decrease in these C-terminal insertions in the presence of azide might suggest that the insertions do affect function in these conditions. In addition to its role in inner membrane protein biogenesis, FtsH is also thought to degrade jammed SecY channels (van Stelten et al., 2009) so it could become more important during azide-induced translocation stress (the C terminal deletion could have a mild effect on function that becomes stronger in the presence of azide).

Azide affects the respiratory electron transport chain

Azide is known to inhibit cytochrome oxidases and there are lots of significant insertion number changes in genes required for the aerobic respiratory electron transport chain, as well as for alternative electron transport chains. Affected genes include *nuoE, nuoF, nuoI* and *nuoJ*, which encode subunits of the NADH:ubiquinone oxidoreductase; *cyoC*, encoding a subunit of cytochrome *bo* terminal oxidase; genes encoding subunits of enzymes required for formate-dependent nitrate reduction, such as *fdoI* and *fdnH;* and genes encoding enzymes involved in the synthesis of ubiquinone and menaquinone (such as *ubiD, ubiE* and *menB*)-showing that components of both aerobic and anaerobic respiratory pathways appear to have a greater effect on fitness in the presence of 0.5 mM azide. It is possible that the inactivation of cytochrome oxidases by azide increases the requirement for alternative electron transport chains.

Genes involved in oxidative stress responses

Several genes showing substantial changes in insertion number could be related to oxidative stress. For example, *yebR*- encoding free methionine sulfoxide reductase- showed an 80% decrease in read number from 0 mM to 0.5 mM azide (although there was no significant change in 0.25 mM azide). The gene encoding chain B of dimethyl sulfoxide reductase, *dmsB,* decreased in read number by around 50% in both azide conditions. However *dmsD*, encoding a maturation protein required for the Tat translocation of DMSO reductase, increased two-fold in read number in 0.5 mM azide (Figure 7).

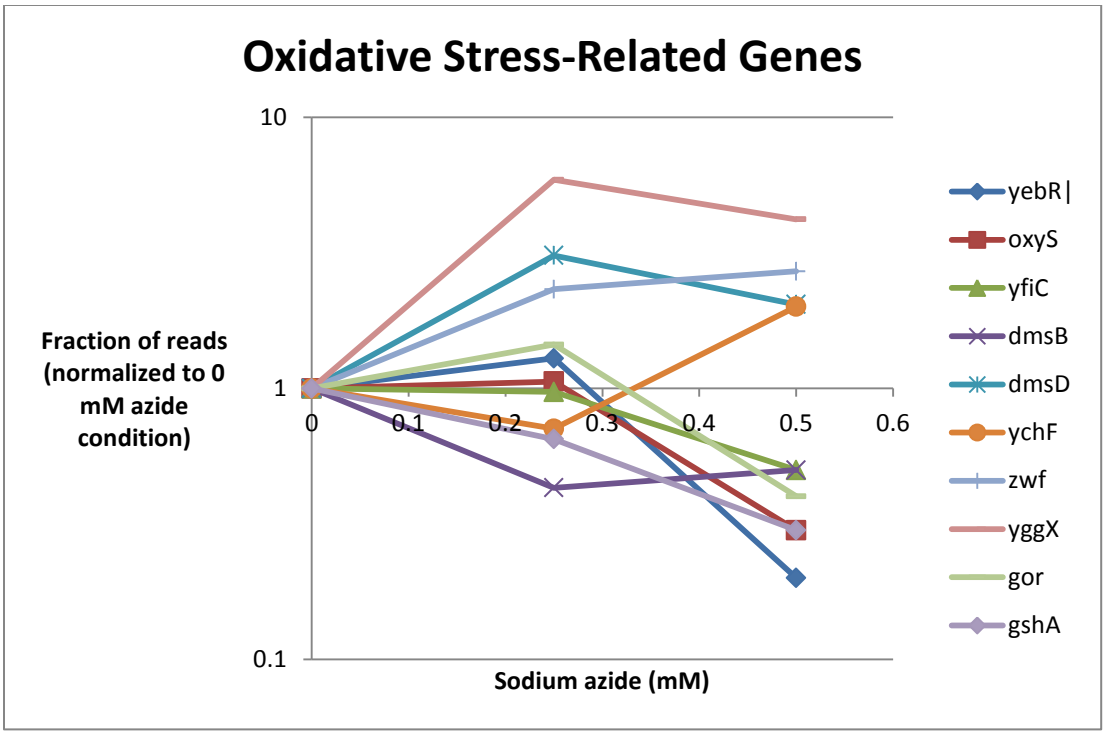


Fig. 7. Genes involved in oxidative stress

Also, methionine sulfoxide reductase B shows a slight increase in insertion number and methionine sulfoxide reductase A shows no change. Insertions in *oxyS*, a small regulatory RNA involved in oxidative stress response, strongly affected fitness in the presence of azide (Figure 7). As with free methionine sulfoxide reductase, OxyS only seems to strongly affect survival at the higher concentration of azide. Another gene that showed a significant decrease in read number in the presence of 0.5 mM azide, *yfiC,* encodes a tRNA methyltransferase previously found to be important for surival under osmotic or oxidative stress (Golovina *et al.*, 2009). Confusingly, some genes with an increase in insertions in 0.5 mM azide also appear to belong to oxidative stress response pathways. Glucose 6-phosphate 1-dehydrogenase (Zwf) is a metabolic enzyme that has a role in a pathway that responds to superoxide stress. Also, *yggx* encodes a protein that prevents oxidative stress effects on iron-sulfur cluster-containing proteins. However, the increase in *ychF* insertions in 0.5 mM azide makes sense in the light of the importance of genes such as *oxyS* (transcription of which is induced by hydrogen peroxide;

OxyS is also known to regulate hydrogen peroxide production) as this gene encodes a catalase inhibitor.

In the presence of azide there appears to be an increased requirement for reduced glutathione, a thiol-containing tripeptide that reduces disulfide bonds in the cytoplasm via four glutaredoxin proteins. This cysteine reducing activity is particularly important when dealing with reactive oxygen species. Several key genes required for the glutathione pathway show significant insertion number decreases in the presence of azide. Importantly *gor*, which encodes glutathione reductase, shows an approximate 60% decrease in insertion number in 0.5 mM azide, although it shows an increase in insertion number in 0.25 mM azide. Also the genes required for glutathione biosynthesis, *gshA* and *gshB,* each show a dose-dependent decrease in insertion number (by approximately 70% and 40% from 0mM to 0.5mM azide, respectively). It is surprising, then, that none of the four glutaredoxins showed significant insertion number changes in 0.5 mM azide-although *grxB* showed a ~65% decrease in 0.25 mM azide. The general lack of an effect of glutaredoxin insertions on fitness in the presence of azide might be difficult to explain, but it is possible that they share substrates that are important under this condition-as they are known to have some overlapping functions.

Several genes whose products utilise glutathione show significant changes in insertion number including *cydC*, which showed an 80% decrease in 0.5 mM azide and encodes a component of an ABC transporter required for the export of glutathione to the periplasm. An enzyme that degrades organic hydroperoxides, encoded by *btuE*, appears to become important in the presence of azide. There are other genes encoding glutathione-dependent enzymes that show insertion number decreases, including *yqjG* (encoding S-glutathionyl-(chloro)hydroquinone reductase) and *frmB*, encoding S-formylglutathione hydrolase-which oxidises formaldehyde to formate (Figure 8). Genes required for the thioredoxin pathway also decrease in insertion number, but by less than two fold in 0.5 mM azide.

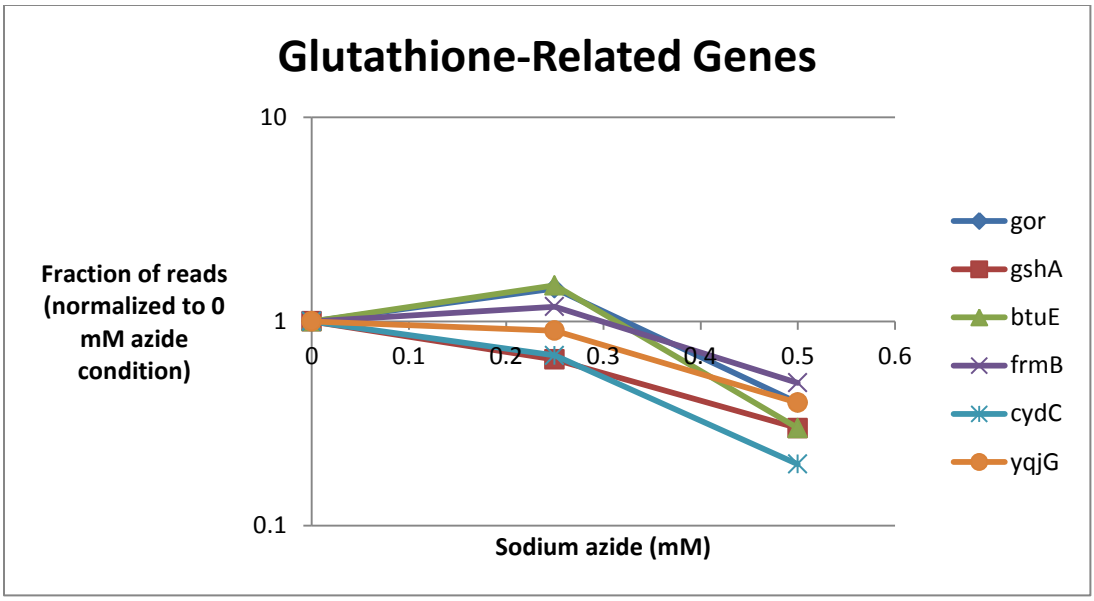


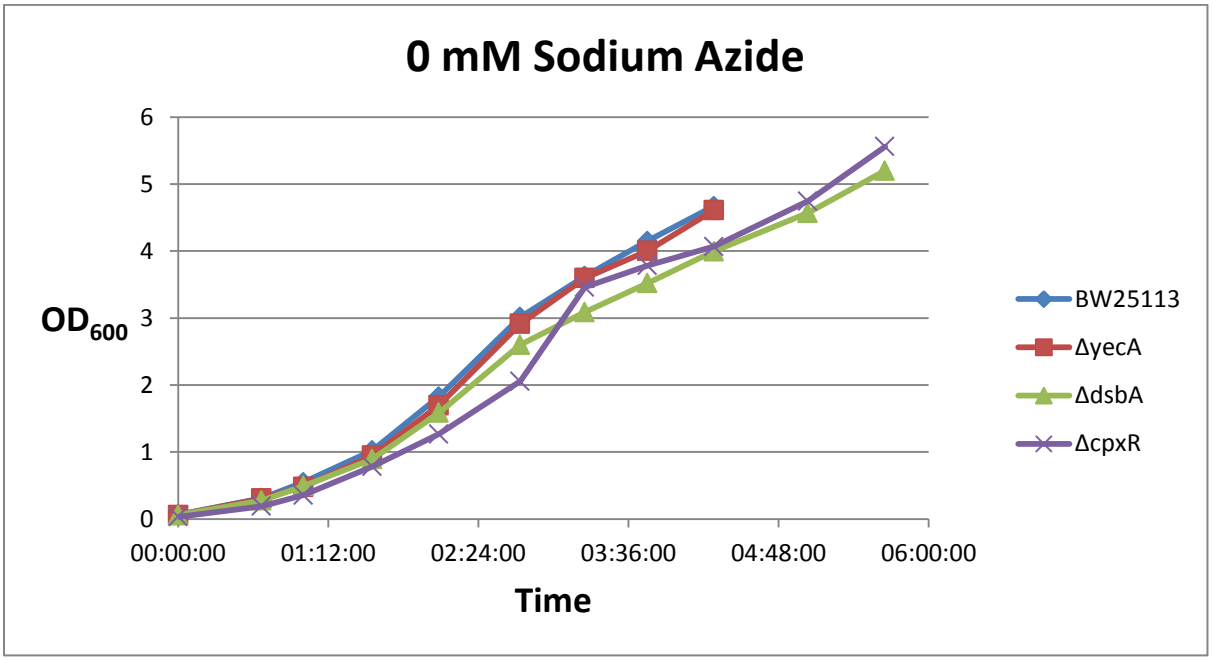
Fig. 8. Genes involved in the production of reduced glutathione and in glutathione-dependent pathways

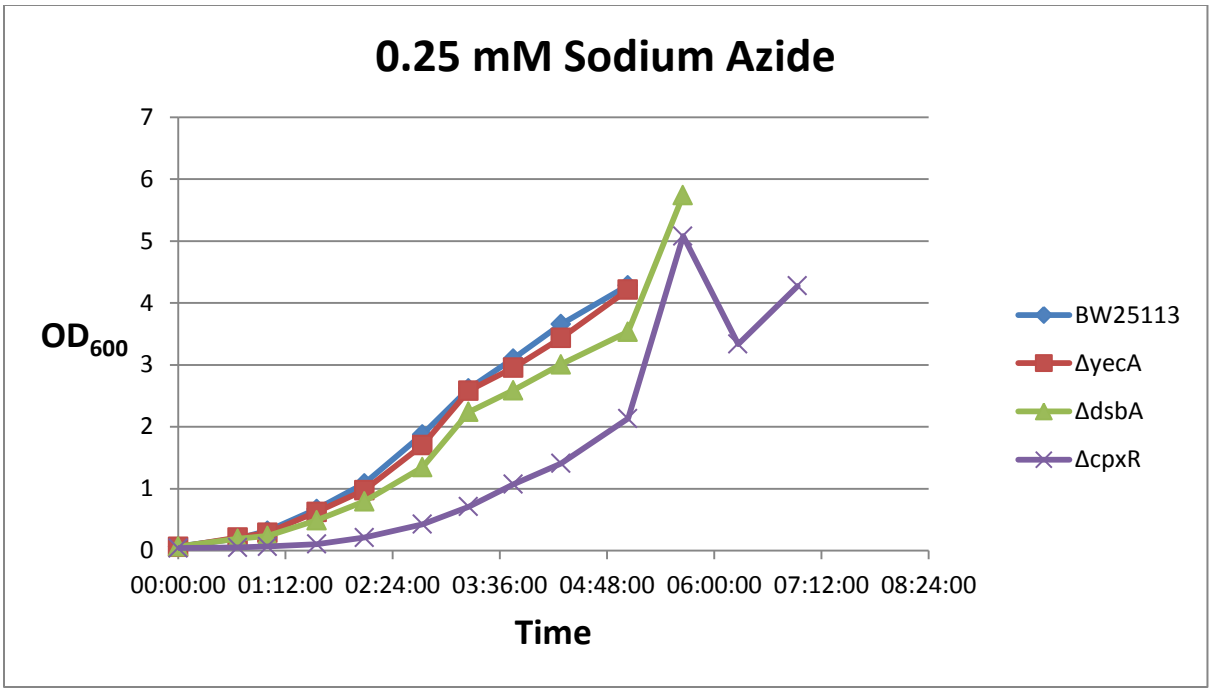
Testing the growth of knockout mutants in the presence of azide

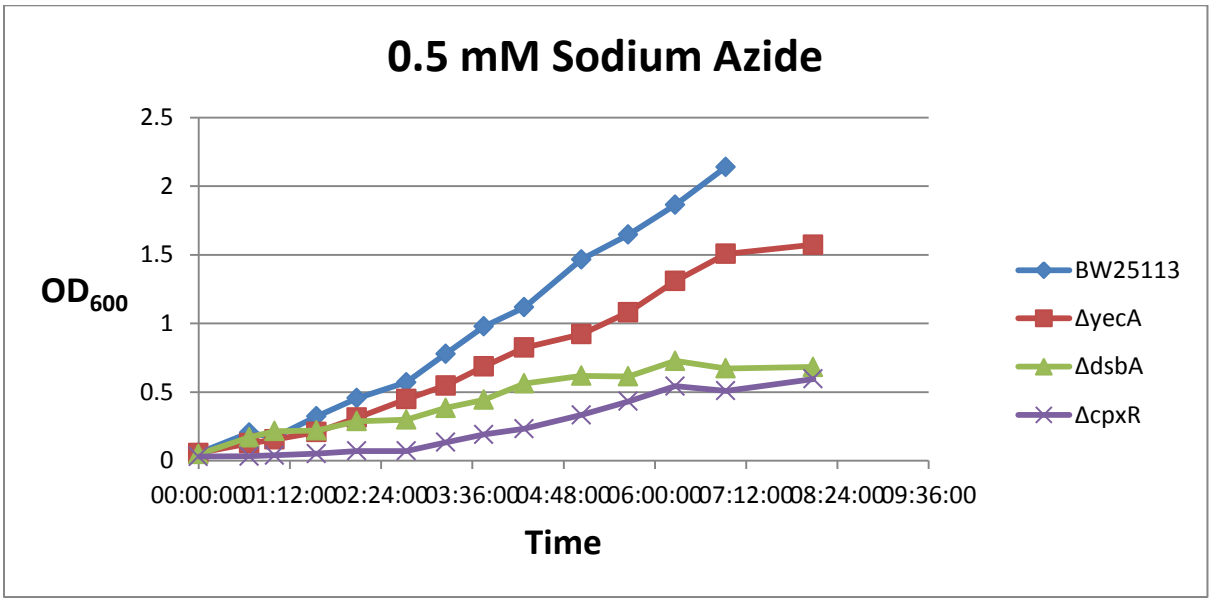
To confirm the effect of selected gene insertions on survival in the presence of azide, knock-out strains were constructed by transducing P1 lysates of Keio deletion strains into wild type BW25113 cells. The mutants used were *ΔyecA::kan*, *Δdsb::kan* and *ΔcpxR::kan*. The *ΔyecA::kan* strain was chosen primarily because of its interest as an uncharacterised gene likely to be involved in Sec translocation (as it has a high degree of similarity to the C-terminal zinc-binding domain of SecA) and because its insertion number decreased by approximately 40% in the presence of 0.5 mM sodium azide (Figure 13). The mutants *ΔdsbA::kan* and *ΔcpxR::kan* were chosen because of their significant insertion number decreases in the presence of azide and their importance in responding to envelope stress.

Unlike the effects suggested by the insertion data, the growth of *ΔyecA::kan* is comparable to the wild type strain in 0.25 mM azide (figure13). However, its growth does seem to be

moderately affected relative to the wild type at 0.5 mM azide (figure 13), and as the TraDIS data would suggest, it is the least affected of the mutant strains tested. Similarly, *Δdsb::kan* shows little relative growth defects in 0.25 mMazide and is noticeably affected by 0.5 mM azide. The *ΔcpxR::kan* mutant appears to be the most strongly affected at all concentrations of sodium azide, which is consistent with the TraDIS data. Growth of *ΔcpxR::kan* is also much more severely affected by 1mM sodium azide than the other strains (figure 13), which suggests that there is strong envelope stress under this condition.




0.25 mM Sodium Azide
OD600
7
6
5
4
3
2
1
0
00:00:00
01:12:00
02:24:00
03:36:00
04:48:00
06:00:00
07:12:00
08:24:00
Time
BW25113
ΔyecA
ΔdsbA
ΔcpxR


0.5 mM Sodium Azide
OD600
2.5
2
1.5
1
0.5
0
00:00:00
01:12:00
02:24:00
03:36:00
04:48:00
06:00:00
07:12:00
08:24:00
09:36:00
Time
BW25113
ΔyecA
ΔdsbA
ΔcpxR

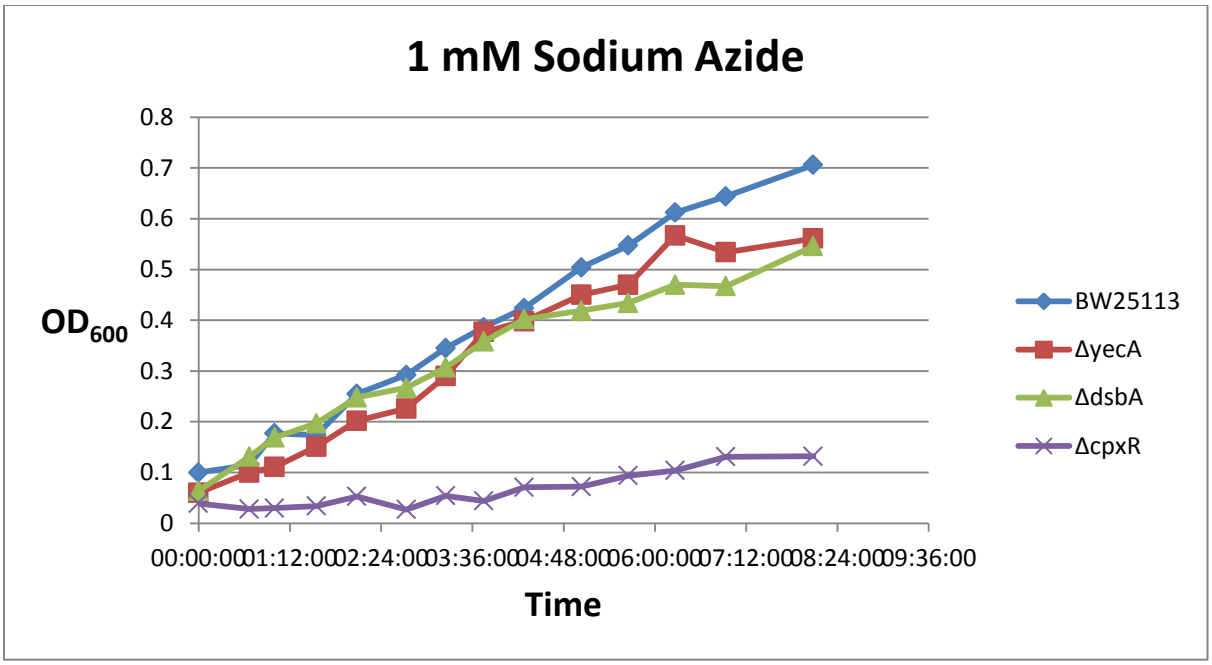
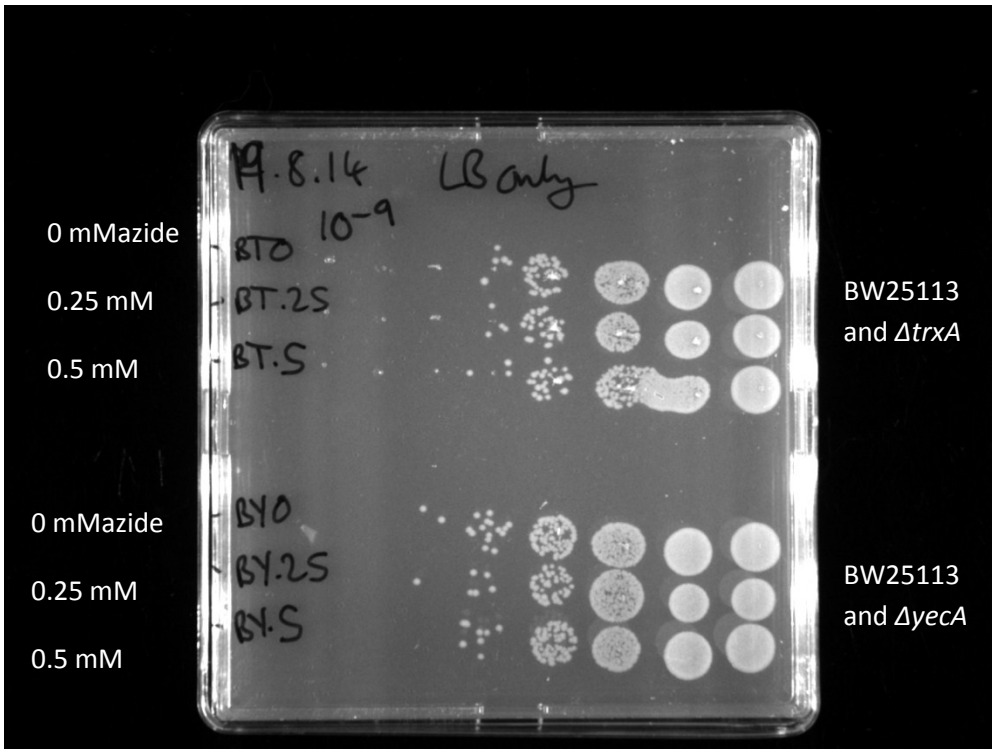


Fig. 13.Comparing the growth of knockout mutants in sodium azide. Overnight cultures of wild type BW25113, *ΔyecA::kan*,*ΔdsbA::kan* and *ΔcpxR*::*kan*were diluted 1:100 into LB broth containing sodium azide at concentrations from 0-1 mM and grown at 37°C, 180 RPM over 8 hours. OD_{600} readings were taken approximately every 30 min.

Spot Assays

One of the characteristics of the TraDIS technique is that a library of insertion mutants is subjected to competitive growth in the conditions of interest. This means that certain mutations might have more severe fitness costs in a competitive growth experiment than when grown independently in the condition of interest. In order to test the competitive growth of the *ΔyecA::kan* mutant in the presence of azide, the strain was grown in competition with the wild type BW25113 and serially diluted in a spot assay in order to compare its plating efficiency with that of a different knock out mutant (in this case, *ΔtrxA::kan* was used).

A


19.8.14 LB only
10⁻⁹
BTO
BT.25
BT.5
BYO
BY.25
BY.5
0 mMazide
0.25 mM
0.5 mM
0 mMazide
0.25 mM
0.5 mM
BW25113 and ΔtrxA
BW25113 and ΔyecA

B

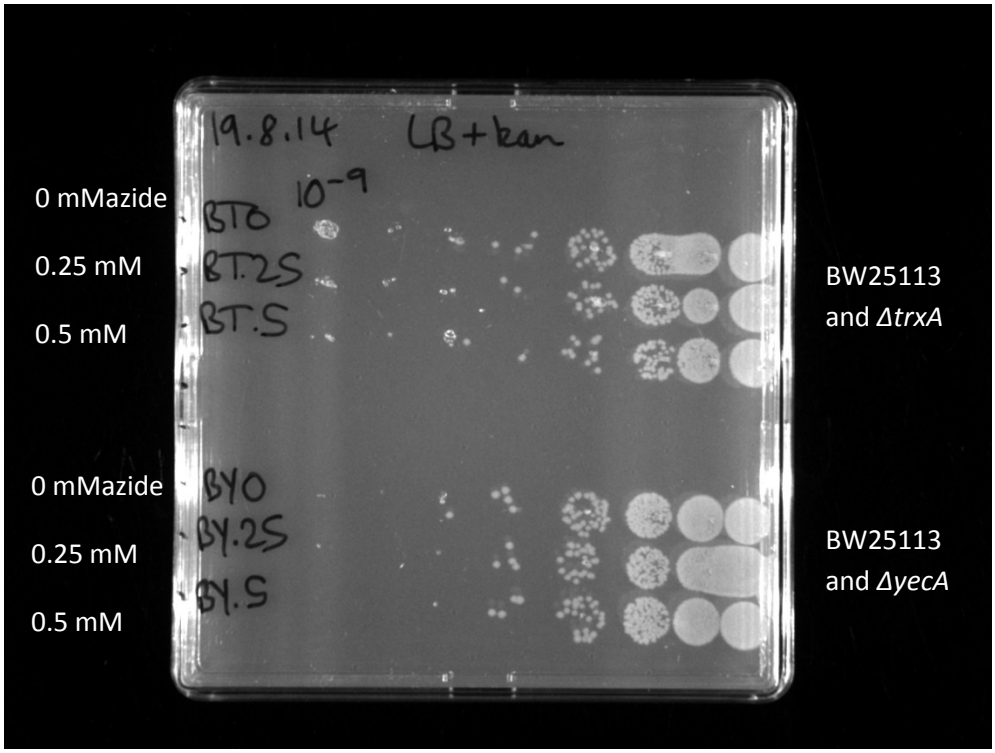

19.8.14 LB+kan
10⁻⁹
BTO
BT.25
BT.5
BYO
BY.25
BY.5
0 mMazide
0.25 mM
0.5 mM
0 mMazide
0.25 mM
0.5 mM
BW25113 and ΔtrxA
BW25113 and ΔyecA

Fig. 14. Competition growth spot assay. Overnight cultures of wild type BW25113, BW25113 *ΔtrxA::kan* and BW25113 *ΔyecA::kan* were subcultured in 1:1 mixtures into LB broth containing either 0 mM, 0.25 mM or 0.5 mM sodium azide. After overnight growth, 10-fold serial dilutions of each culture were spotted onto LB plates with (B) or without (A) kanamycin.

For the purpose of comparing the ability of the two knockout strains to compete with the wildtype strain in the presence of azide by working out plating efficiencies, the spot assay experiment appeared to be unsuitable (figure 14). One problem could be the choice of deletion mutant used for comparison with *ΔyecA*: *trxA*, which encodes thioredoxin-1 (an oxidoreductase that reduces disulfide bonds in the cytoplasm and would be expected to play a role in oxidative stress management) itself showed a decrease in insertion number of 20% in both azide conditions. The spot assay experiment possibly shows that *ΔtrxA* has a stronger growth defect than *ΔyecA*, although the azide dependence of this effect is not clear. This could be tested by comparing growth curves of the two mutants and wild type strain at different concentrations of sodium azide. If the competition experiment were repeated, the knockout strain used for comparison should be one that is expected to be unaffected by the presence of azide based on TraDIS or deletion mutant growth data. It might be useful to select a knockout mutant that also has a similar growth defect to *ΔyecA* in the absence of azide.

Discussion

The purpose of this study was to investigate the effects of sodium azide on *E. coli*, in order to determine which genes become essential under conditions of severe translocation stress. The experiment could also provide insight into the mechanisms of inhibition by azide (*e.g.* which cellular processes are specifically targeted by azide) and allow the investigation of genes of

unknown function. The TraDIS experiment involved the competitive growth of a library of transposon insertion strains in three different conditions: 0, 0.25 and 0.5 mM sodium azide.

Surprisingly *secA* itself, which is essential, was one of the genes of interest in the TraDIS experiment, as it showed the biggest azide-dependent increase in number of reads of all the translocation genes (figure 4). The reads all map to the C terminal domain of SecA, starting around amino acid 818. This suggests that when the C terminus of SecA is deleted, the bacterium becomes more competitive and capable of surviving in the presence of azide. This could be explained by the proposed autoinhibitive role of the ~ 50 aa flexible linker domain within the C terminus of SecA, which connects the "C-terminal domain" to the zinc-binding domain that is comprised of the C-terminal 20 aa of the protein. This linker domain is thought to bind the substrate-binding groove between the first ATPase domain and the domain thought to bind preprotein substrates. Autoinhibition is thought to be relieved by the binding of the C-terminal zinc-binding domain to SecB or the ribosome (Gelis *et al.*, 2007; Huber *et al.*, unpublished manuscript). It is possible that in the presence of azide the removal of the autoinhibiting domain could become more of an advantage. This could be investigated by comparing the effects of azide on growth with wildtype SecA and SecA N95, which lacks flexible linker domain and zinc-binding domain. A SecA variant lacking only the zinc-binding domain (N880) could also be included to make sure that it is loss of the linker domain and not the zinc-binding domain that is advantageous, although this is might be apparent from the TraDIS data as the insertions first appear just before the start of the flexible linker domain. Also, the variant lacking only the zinc-binding domain might be expected to cause growth defects in all conditions if the linker domain can still bind the peptide-binding groove in this structure. It would be interesting to see if the growth of this strain is more severely affected by azide than the other strains.

At least two non-essential *sec* genes showed significant decreases in insertion number in the presence of azide. In this experiment, although *secG* didn't become essential in the presence of azide, *secG* insertion strains appeared to show a specific sensitivity to azide as the number of reads in *secG* in both 0.25 and 0.5 mM azide was decreased to ~20% of that in the absence of azide. As SecG is a component of the Sec translocon itself, is known to be essential at lower temperatures and is sensitive to insertions in the presence of 0.25 mM azide, the less selective condition, this result might provide evidence of translocation-specific effects of azide.

Other future experiments should also focus on YajC, a nonessential protein involved in Sec translocation with an undefined role-but believed to form a heterotrimeric complex with SecD and SecF. In this TraDIS experiment, *yajC* showed the largest decrease in reads from 0 to 0.5 mM azide, despite no translocation phenotype having been found so far in a *yajC* mutant. This could suggest that YajC somehow functions to increase the efficiency of translocation in the presence of azide, which could mean that it has a role in survival in translocation stress. Perhaps this could be tested by inducing other conditions of general and translocation stress and checking whether there is a phenotype for a *yajC* mutant. The effect of a *yajC* mutant in the presence of azide should be checked by comparing growth in the presence of azide, if not other translocation stress-inducing conditions also-if they are known. Deletions in well-known nonessential Sec genes cause cold sensitivity (Pogliano and Beckwith, 1993), so perhaps the cold sensitivity of a *yajC* deletion strain should be tested.

To distinguish between polar effects and effects specific to YajC function, growth of a *yajC* deletion mutant (in addition to a strain complemented with *yajC* alone if there is a chance of polar effects of the deletion mechanism, or anyway as a control) in competition with the wild type strain should be tested in the presence or absence of azide. Although, if *yajC* does function in a complex with SecDF, expression of *yajC* alone from a plasmid for complementation

could still hinder complex formation and hence SecDF function, causing a growth defect in the presence of azide.

Oxidative stress was one of the main themes that arose in the TraDIS experiment and might provide a clue to the effects of azide on the cell. Some DNA repair genes also showed significant changes in insertion number, although DNA damage might result from oxidative stress or mutagenic effects of azide itself.

One of the conclusions that could be made from this experiment is the dependence of the type of cellular response on the concentration of sodium azide used. While many genes of interest showed a dose-dependent change in insertion number in the presence of azide, in some cases 0.25 mM had the opposite effect of 0.5 mM azide on insertion number change while, unsurprisingly, in some cases 0.25 mM appeared to be simply too low a concentration of azide to cause a change in insertion number. Conversely, for some genes a significant effect could only be seen in the presence of 0.25 mM azide. It appears that the two concentrations of azide used could be affecting distinct as well as overlapping groups of genes. It is unclear why this might be. Also within each functional group of genes, even genes encoding subunits of the same complex, there was great variability in the effect of azide on insertion number.

In future studies the concentration of azide used should be taken into account. Concentrations up to 4 mM have also been used in experiments to select for azide-resistant mutants. It is possible that these concentrations do not display the same variability in cellular responses as those used in this study, so these results have little relevance to those uses of azide. However, these concentrations are incompatible with TraDIS since they completely inhibit growth.

It is possible that 0.5 mM azide, a significantly more inhibitory concentration than 0.25 mM, might cause nonspecific stress related effects. If we get general stress effects in the presence of 0.5 mM azide, this could be an issue in translocation experiments using higher concentrations of azide. However, *azi* mutants are known to show normal growth at higher azide concentrations (Fortin *et al.*, 1990), so it seems likely that translocation is the most important process affected by these concentrations of azide. Although this might be distinguished from translocation-specific effects by *in vitro* translocation experiments.

One problem with our TraDIS experiment is the possibility that many of the fitness effects on certain pathways that we are attributing to direct targeting by sodium azide are actually the result of a state of general cell stress brought on by the azide-thereby not actually telling us the mode of action of azide. A future experiment to investigate which pathways are 'targeted' by azide should include a control condition that would activate pathways associated with general cell stress, such as sublethal concentrations of ethanol that are beforehand found to affect growth rate at levels comparable to the concentrations of azide used.

In summary, one thing that the azide concentration-dependent variability in the results suggests is that the cellular responses seen in our experiment could be very different to those caused by the concentrations of azide used in translocation experiments. Unfortunately, the severe growth effects of azide concentrations of this magnitude probably make mechanistic study of the azide inhibition too difficult.

However it can't be ignored that all of the azide resistance mutations, selected using up to 4mM sodium azide, have been found in the translocation ATPase SecA. This strongly suggests that Sec translocation is central to the lethal effects of azide because mutations in *secA* can allow the cell to survive in very high concentrations of azide.

References

Christen, B. et al., 2011. The essential genome of a bacterium.Molecular systems biology, 7, p.528. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21878915.

Denoncin, K. et al., 2014. A new role for Escherichia coli DsbC protein in protection against oxidative stress. The Journal of biological chemistry, 289(18), p.12356–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24634211.

Fortin, Y., Phoenix, P. & Drapeau, G., 1990. Mutations conferring resistance to azide in Escherichia coli occur primarily in the secA gene. Journal of bacteriology, 172(11), p.6607–10. Available at: http://jb.asm.org/cgi/pmidlookup?view=long&pmid=2146254.

Gelis, I. et al., 2007. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell, 131(4), p.756–69. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092-8674(07)01269-X

Golovina, A. et al., 2009. The yfiC gene of E. coli encodes an adenine-N6 methyltransferase that specifically modifies A37 of tRNA1Val(cmo5UAC). RNA (New York, N.Y.), 15(6), p.1134–41. Available at: http://www.rnajournal.org/cgi/pmidlookup?view=long&pmid=19383770.

Gouridis, G. et al., 2009. Signal peptides are allosteric activators of the protein translocase. *Nature*, 462(7271), p.363–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19924216.

Hobbs, E., Astarita, J. & Storz, G., 2010. Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in Escherichia coli: analysis of a bar-coded mutant collection. Journal of bacteriology, 192(1), pp.59–67. Available at: http://jb.asm.org/cgi/pmidlookup?view=long&pmid=19734312.

Huber, D. et al., 2011. SecA interacts with ribosomes in order to facilitate posttranslational translocation in bacteria.Molecular cell, 41(3), p.343–53. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097-2765(10)01013-

Huie, J. & Silhavy, T., 1995. Suppression of signal sequence defects and azide resistance in Escherichia coli commonly result from the same mutations in secA. Journal of bacteriology, 177(12), pp.3518–3526. Available at: http://jb.asm.org/content/177/12/3518.short

Kimura, E. et al., 1991. Determination of a region in SecA that interacts with presecretory proteins in Escherichia coli.The Journal of biological chemistry, 266(10), p.6600–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1826108.

Langridge, G. et al., 2009. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants.Genome research, 19(12), p.2308–16. Available at: http://genome.cshlp.org/cgi/pmidlookup?view=long&pmid=19826075.

Lederberg, J., 1950. The selection of genetic recombinations with bacterial growth inhibitors. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC385743/

Oliver, D. et al., 1990. Azide-resistant mutants of Escherichia coli alter the SecA protein, an azide-sensitive component of the protein export machinery. Proceedings of the National Academy of Sciences, 87(21), pp.8227–8231. Available at: http://www.pnas.org/content/87/21/8227.short.

Pogliano, K. & Beckwith, J., 1993. The Cs sec mutants of Escherichia coli reflect the cold sensitivity of protein export itself. Genetics, 133(4), p.763–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8462840.

van Stelten, J. et al., 2009. Effects of antibiotics and a proto-oncogene homolog on destruction of protein translocator SecY. Science (New York, N.Y.), 325(5941), p.753–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19661432.

Wolfe, A. et al., 2008. Signal integration by the two-component signal transduction response regulator CpxR.Journal of bacteriology, 190(7), p.2314–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18223085.