Proofreading of substrate structure by the Twin-Arginine Translocase is highly dependent on substrate conformational flexibility but surprisingly tolerant of surface charge and hydrophobicity changes

Alexander S. Jones[1], James I. Austerberry[2], Rana Dajani[2], Jim Warwicker[2], Robin Curtis[3] Jeremy P. Derrick[2], Colin Robinson[1*]

[1]School of Biosciences, University of Kent, Canterbury CT2 7NJ, United Kingdom

[2]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, United Kingdom

[3]School of Chemical Engineering and Analytical Science, University of Manchester, Manchester M13 9PL, United Kingdom

* Corresponding author.

Email c.robinson-504@kent.ac.uk

Tel: +44 1227 823443

Abstract

The Tat system transports folded proteins across the bacterial plasma membrane, and in *Escherichia coli* preferentially transports correctly-folded proteins. Little is known of the mechanism by which Tat proofreads a substrate's conformational state, and in this study we have addressed this question using a heterologous single-chain variable fragment (scFv) with a defined structure. We introduced mutations to surface residues while leaving the folded structure intact, and also tested the importance of conformational flexibility. We show that while the scFv is stably folded and active in the reduced form, formation of the 2 intra-domain disulphide bonds enhances Tat-dependent export 10-fold, indicating Tat senses the conformational flexibility and preferentially exports the more rigid structure. We further show that a 26-residue unstructured tail at the C-terminus blocks export, suggesting that even this short sequence can be sensed by the proofreading system. In contrast, the Tat system can tolerate significant changes in charge or hydrophobicity on the scFv surface; substitution of uncharged residues by up to 3 Lys-Glu pairs has little effect, as has the introduction of up to 5 Lys or Glu residues in a confined domain, or the introduction of a patch of 4 to 6 Leu residues in a hydrophilic region. We propose that the proofreading system has evolved to sense conformational flexibility and detect even very transiently-exposed internal regions, or the presence of unfolded peptide sections. In contrast, it tolerates major changes in surface charge or hydrophobicity.

Introduction

The Tat pathway functions in the bacterial plasma membrane and plant thylakoid membrane to transport proteins to the periplasm and thylakoid lumen, respectively. The system is highly unusual among protein translocators because substrates are transported in a folded state. The Sec pathway, by contrast, transports proteins in an unfolded state using a very different mechanism (reviewed in [1-3]). Native Tat substrates cover a wide range of sizes (30 kDa – 150 kDa), they can be oligomeric (e.g. [4,5]) and it seems likely that the system has evolved to transport proteins that cannot be transported by Sec because they are too complex or have rapid folding kinetics. Tat plays a crucial targeting role for proteins involved in many important pathways, including energy metabolism, nutrient uptake and in some instances pathogenesis [6,7].

In Gram-negative bacteria such as *Escherichia coli*, the Tat machinery consists of three membrane embedded components – TatA, TatB and TatC (reviewed in [1,2, 8]). Substrates destined for the periplasm are targeted to the Tat apparatus via an N-terminal signal peptide containing a characteristic RR motif [9, 10]. Once at the membrane, substrates dock with a TatBC complex to which TatA protomers are then recruited and translocation occurs through an as-yet uncharacterised mechanism [1,2,10]. Once across the membrane, signal peptidase cleaves the signal peptide and the mature substrate is released into the periplasm.

The Tat system's ability to transport folded proteins does, however, present the problem that the substrates must be *correctly* folded, at least in most cases. For example, it would be futile to prematurely export proteins before they could bind their cognate redox cofactor in the cytoplasm, or to export globular proteins that rely on cytosolic chaperones for correct folding.

Some complex Tat substrates do have specific folding catalysts to assist in assembly prior to export (e.g. [11]) but a fascinating aspect of the Tat system is its inbuilt proofreading activity which enables preferential export of correctly folded proteins. This proofreading/quality control mechanism even applies to heterologous proteins that it never normally encounters, and is so strict that significant destabilisation of structure almost invariably results in 100% rejection of substrate [12-15]. However, very little is understood about the exact process the Tat machinery employs and what is 'sensed' in a substrate for this proofreading and quality control. Richter *et al.,* (2007) tested the Tat system with natively unfolded FG repeats from the yeast nuclear pore protein Nsp1p. These were exported with decreasing efficiency as a function of length and rejected upon insertion of a short hydrophobic sequence [16]. This did raise the possibility that Tat proofreading assesses whether a substrate should be transported based on its hydrophobicity, as such residues would usually be sequestered inside correctly folded substrates. Several studies have built on the work presented by Richter *et al*, utilising long sections of FG repeats [17] [18] or (Gly_4/Ser) linkers to separate two folded domains or inserted after the signal peptide, both of which intrinsically lack structure. In these studies, translocation was observed to varying degrees, attributed to the physical length of substrate rather than any proofreading mechanism. The earliest of these was carried out in plant thylakoids [19] where substrates were successfully transported into the thylakoid lumen, irrespective of folding state, until a certain physical size was exceeded. This would suggest Tat proofreading is not as infallible as previously thought, although it should be emphasised that the proofreading systems of thylakoid and bacterial Tat systems may well differ in significant respects [20,21].

Other studies have sought to probe Tat proofreading in the Tat components themselves. Rocco *et al*., [22] screened for mutations in Tat components and isolated several quality

control suppression (‘QCS’) mutants [22]. Their results pointed to the importance of residues in the Tat apparatus, since several mutations reduced the effectiveness of Tat-mediated proofreading, enabling export of misfolded proteins that are normally rejected. This clearly indicates proofreading is, in some part at least, carried out at the level of the Tat translocon.

While the Tat system is capable of quantitatively rejecting misfolded substrates, its proofreading system is not perfect and minor changes in structure can be tolerated. For example, it was shown that several disulphide-bonded proteins (including human growth hormone, a single-chain variable fragment (scFv) and interferon α2β) can be exported in the reduced state in wild type cells [23] and the native Tat substrate CueO is exported in its apo form [24]. However, these proteins are believed to adopt near-native structures even in the absence of disulphide bond formation or Cu ligation, and so the extent of misfolding is likely to be minimal. Other disulphide bond-containing proteins, such as PhoA and AppA, do not fold properly in the reduced form and these are rejected by Tat unless expressed in a strain with a more oxidising cytoplasm, such as 'CyDisCo' strains that permit disulphide bond formation in the cytoplasm [15, 25]. Human growth hormone, scFv, α2β and CueO are in a natively-folded, but more dynamic structural state, whereas PhoA and AppA are grossly misfolded in these scenarios.

A problem with some of the previous studies is that many of the substrates were severely misfolded, and under these conditions it is difficult to assess the features that Tat actually senses during its proofreading activities. In this study we have sought to ascertain the nuances of the Tat proofreading capacity and its biotechnological potential through the export of a single-chain variable fragment (scFv) and mutagenised variants thereof, where the wild-type has been shown to be folded and active as demonstrated in [26]. By altering the surface

properties without affecting the tertiary structure, we have been able to probe the requirements of the Tat proofreading system, while other mutations were designed to introduce unfolded regions and test their effects. The results show the Tat machinery is surprisingly tolerant of changes in substrate surface charge or hydrophobicity, allowing export of variants with additional charge-pairs or hydrophobic patches. On the other hand the Tat system displayed a notable preference for a more rigid substrate structure, suggesting that the substrate's conformational flexibility is a key factor that is assessed by the proofreading mechanism.

Materials and Methods

Plasmids used in this study

All constructs (see Table 1) were synthesised by GeneArt (Life Technologies, Thermo Fisher Scientific, UK) except 1SB, 5NLeu and 6NLeu which were generated through Quick Change PCR. PCR primers are shown in Supplementary Table 1. A 6xHis-tag (shown in bold) was added at the 3' end of the wild-type scFvM and all variants through successive rounds of amplification PCR. The first round of PCR used forward primer JASCFV1F (5'-GGTGGACCCCATATGCAGGAACAGCTGGTTGAATCTGGT-3') and appropriate reverse primer scFvMXRO. Second round PCR used the same forward primer as previously and reverse primer His_stop_BamHI_R (5'-GAATTCGGATCCGGATCCTTATTAGTG**ATGATGATGATGATG**-3'). These primers also added a 5' NdeI site (CATATG) and a 5' BamHI site (GGATCC), shown underlined in primers. PCR solutions were made using 1 µL template DNA (100 ng/µL), 1 µL dNTPs (200

μM final), 10 μL 5x GC buffer, 1.25 μL of each primer (0.5 μM final), 1.5 μL DMSO (3% final), 0.5 μL Phusion High-fidelity DNA polymerase (2 U/μL stock, New England Biolabs, UK) and made up to a final volume of 50 μL with milliQ H_2O. PCR was then carried out in Biometra T3 Thermocycler (Biometra Anachem, UK) as per Phusion polymerase instructions with an annealing temp of primer T_m minus 5°C.

PCR products were separated on 1% (w/v) agarose (Bio-Rad Laboratories Ltd, USA) gels by electrophoresis for 40 mins at 150V. Appropriate bands were visualised under UV transilluminescence and excised using a sharp scalpel blade before being purified with QIAprep Gel Extraction kit (Qiagen, Hilden, Germany) according to manufacturer's instructions.

Constructs were cloned into the vector pYU49 [25] through restriction cloning. Restriction digests were carried out with 16 μL of template DNA, 1 μL Buffer 3.1 (New England Biolabs, UK), 1.5 μL NdeI and incubated at 37°C for 30 min before adding 1 μL BamHI and incubating for a further 1hr at 37°C. Products were gel-purified in the same manner as described above and ligated into vector pYU49 or pHIA554, downstream of a TorA signal peptide and the first 4 amino acids of mature TorA using T4 DNA ligase (Roche). Insert and vector were mixed in a 3:1 ratio with 1 μL ligase buffer and 1 μL T4 ligase, before incubation overnight in Biometra T3 Thermocycler. The next day 10 μL ligation product was used to transform 100 μL *E. coli* DH5α and W3110 competent cells. Variants 1SB, 5N-Leu and 6N-Leu were generated through site-specific mutagenesis using the Quick-change protocol (Agligent), according to the manufacturer's instructions, using primers shown in Supplementary Table 1.

Cell culture and fractionation

5 mL Leuria Bertani (LB) medium (10g/L sodium chloride, 10g/L tryptone, 5g/L yeast extract) pre-cultures were inoculated from glycerol stocks and grown aerobically overnight at 30°C, 200 rpm with 1:1000 antibiotic (5 µL, 1M ampicillin). The next day cultures were diluted to OD_{600} = 0.05 in 50 mL fresh LB and antibiotic. Cultures were then grown at 30°C, 200 rpm in 250 mL Erlenmeyer flask until OD_{600} = 0.4 – 0.6 before induction of plasmid with 0.5 mM IPTG (25 µL, 1M IPTG). After 2, 3, 4 or 5hr induction, cells equivalent to a density of OD_{600} of 10 (~8 mL) were taken and fractioned into Cytoplasm (C), Membrane (M) and Periplasm (P) fractions by the EDTA/lysozyme/cold osmotic-shock method [23]. In brief, cells were pelleted by spinning at 3000 rpm, 4°C, 10 min. The supernatant was discarded and the cell pellet resuspended in 500 µL Buffer 1 (100 mM Tris-acetate pH 8.2, 500 mM sucrose, 5 mM EDTA pH 8.0) and 500 µL dH_2O before addition of 40 µL hen egg white lysozyme (1mg/mL) and incubation on ice for 5 minutes to digest the outer membrane. 20 µL $MgSO_4$ (1M) was added to stabilise the inner membrane before centrifugation at 14000 rpm, 4°C, 2 min. 500 µL of the supernatant was taken and frozen at -20°C as the Periplasmic fraction, P. The remaining supernatant was discarded and the pellet washed by resuspension in 750 µL Buffer 2 (50 mM Tris-acetate pH 8.2, 250 mM sucrose, 10 mM $MgSO_4$) and subsequently spun at 14000 rpm, 4°C, 5 min. The cell pellet was resuspended in 750 µL Buffer 3 (50 mM Tris-acetate pH 8.2, 2.5 mM EDTA pH 8.0) and sonicated for 6x 10 seconds, amplitude 8 µm to disrupt membranes (Soniprep 150plus, Sanyo Gallenkamp, Loughborough, UK). The resulting solution was then centrifuged at 70000 rpm, 4°C, 30 min to sediment the membrane fraction. 500 µL of the supernatant was then taken and frozen at -20°C and designated as the Cytoplasmic fraction, C. The remainder of the supernatant was discarded and the pellet resuspended in 500 µL Buffer 3 before freezing at -20°C, and designated as the Membrane fraction, M.

Western-blotting

Protein was transferred to PVDF-membrane (GE Healthcare, Buckinghamshire, UK) by wet-western-blotting, with electrophoresis for 1 hour at 80V, 300A. The PVDF membrane was then immersed in blocking solution (2.5% (w/v) skimmed milk powder in 50 mL 1x PBS – tween20 (0.1%)) and incubated at 4°C overnight. The following day the membrane was washed 3x for 5 min with 1x PBS-tween20 (0.1%) before incubation with primary antibody (3.5 µL anti-6x-His (Life Technologies, CA, USA) in 20 mL 1x PBS-tween20 (0.1%), 3% (w/v) BSA) for 1 hour at room temperature. Membranes were then washed 3x for 5 min with 1x PBS-tween20 (0.1%) before incubation with secondary antibody (4 µL anti-Mouse HRP conjugate (Promega, WI, USA) in 20 mL 1x PBS-tween20 (0.1%)) for 1 hour at room temperature. Finally membranes were washed for 8x for 5 min with 1x PBS-tween20 (0.1%). Immunoreactive bands were detected using enhanced chemiluminescence (ECL) kit (BioRad, Herts, UK) following manufacturer's instructions. Bands were visualised using BioRad Gel-doc chemiluminescence imager and associated software. Densitometry of band-intensity was also carried out on BioRad Gel-doc imager with appropriate software.

Protein A affinity chromatography

scFv variants were purified by expression of the mature-size proteins in *E. coli* BL21 DE3 cells, lysis of the cells and centrifugation at 100, 000 x g for 2 hr. The supernatant was passed through onto a 25 mL column packed with Protein A-Sepharose (GE Healthcare). Column binding and wash buffer was 50 mM Tris HCl (pH 8.5), 25 mM NaCl, and the protein eluted with 100 mM sodium citrate (pH 3.5). The protein samples were restored to pH 8.5 by passage through PD10 desalting column (GE Helathcare) equilibrated in 50 mM Tris HCl (pH 8.5), 25 mM NaCl.

Circular dichoism

Circular dichroism experiments were undertaken on the Applied Photophysics Chirascan. Protein solutions of 50 µL volume at 1 mg/mL were placed in a 1 mm pathlength quartz Hellma cell with Peltier temperature controlled stage at 25°C. Spectra were recorded between 190-260 nm at a scan rate of 10 s/nm and normalised by subtraction with a corresponding buffer blank.

Dynamic light scattering

Dynamic light scattering of each sample was determined using the Wyatt Dynapro plate reader system. The dynamic light scattering is determined using the laser wavelength of 830nm with a scattering angle of 158°. Hydrodynamic radii of the samples were determined after the samples (35 µL) were equilibrated at 25°C; results presented are an average of triplicate samples measured 5 times for 10s each.

Intrinsic fluorescence

Fluorescence spectra were recorded using the Avacta Optim 1000. 9 µL of each sample was loaded into the micro cuvette array and spectra recorded with laser excitation at 266 nm and emission spectra collected from 300-500 nm. Results presented are an average of duplicated samples with spectra collected at 25°C with default settings.

Results

scFvM is active in the reduced form but is only exported with high efficiency in the oxidised form

The aim of this study was to analyse the proofreading capacity of Tat in more detail, using a model substrate with a well-defined structure together with variants in which either (i) the surface features were changed while the folded state was unaffected or (ii) an unstructured domain was added. To this end we chose an scFv (termed scFvM, to distinguish it from the different scFv used by Alanen *et al*. [23]) and we then generated and tested a range of mutants, including some sourced from a biophysical study of the scFv and mutants in solution (manuscript in preparation). The predicted structure scFvM, based on homology modelling with PDB file 2GHW, is shown in Fig. 1A. The protein comprises two immunoglobulin domains that are connected by a loop region; each domain contains a single disulphide bond (circled in Fig. 1A). The remaining images show the charge and hydrophobicity characterostics from two different projections. For tests on the Tat proofreading system, scFvM was expressed with an N-terminal TorA signal peptide and a C-terminal 6x His-tag, as used in a previous study on a different scFv [23].

scFvM has been shown to be folded and active after expression in the reducing cytoplasm of wild type cells (manuscript in preparation, data shown below and [26]). We further confirmed that the disulphide bonds remain reduced during the purification process; Supplementary data Fig. 1 shows mass spectrometry analysis of the protein directly after purification or after addition of DTT to reduce any disulphide bonds. The data confirm that the protein has an identical mass of 26,806 Da (corresponding to the reduced form) in both cases. The biophysical/activity data experiments presented below were therefore carried out on scFvM in the reduced state.

We have previously shown that a different scFv could be exported in wild type or CyDisCo cells (*i.e*. in either the reduced or oxidised form; [23]); we therefore first tested whether the disulphide bonding status affects export of scFvM by Tat in these strains. The protein was expressed with a TorA signal peptide, and Fig. 1B shows time course analyses of export after induction of synthesis with IPTG. At the time points shown, cells were fractionated to yield cytoplasm, membrane and periplasm samples (C/M/P). The top panel of Fig. 1B shows that in wild type cells (*i.e*. in a strain with a reducing cytoplasm) there is a slow build-up of mature-size scFvM in the periplasmic fraction (P) over the 5 hr period, confirming that the protein is exported by Tat in the reduced state. However, export is very inefficient and only a small proportion of protein is present in the periplasm at the end of the incubation. A precursor band can be seen at a slightly larger molecular weight (31 kDa) in the membrane fraction (M) which corresponds to the unprocessed TorA-scFvM. A mature-size band can be seen in the cytoplasmic and membrane fractions, presumably due to clipping by cytoplasmic proteases. The lower panel shows expression of the precursor protein in CyDisCo cells. Here the periplasmic band is much stronger, with significant export observed from an early time point. Comparison of the periplasmic scFvM band intensities from several experiments (see below) reveals that export is ca. 10-fold more efficient in CyDisCo cells, indicating that the disulphide-bonded protein is far more readily transported by the Tat system than the reduced form (the overall signal intensity is also much higher, presumably reflecting degradation of the protein in the absnce of export). This is interesting because scFvM is clearly significantly folded after extraction from the cytoplasm of wild type cells in a reduced form, and we propose that the disulphide bonds generate a more rigid substrate structure that is far more acceptable to the proofreading system (discussed further below).

A range of mutated scFvM forms were analysed in this study and control experiments were carried out to verify that the fractionation protocol accurately demonstrates the levels of mature-size protein in the periplasm (Fig. 2). Panel A shows a Coomassie-stained gel of the cytoplasm, membrane and periplasm fractions, and a key point is that the protein band patterns are highly distinct, with no evidence for overlap between the cytoplasm and periplasm samples. Panel B shows scFvM export assays in CyDisCo cells and *tat* null mutant cells; the periplasmic mature band is clearly evident in the former but absent in the latter, confirming that there is minimal contamination of the periplasm by cytoplasm during fractionation. This is important because proteolytically-clipped mature-size scFvM is commonly found in the cytoplasm. Panel B also shows an assay for export of a mutated scFvM version, termed '26tail' (discussed below) in CyDisCo cells; the absence of a periplasmic band again shows that the fractionation protocol is effective. Finally, panel C shows periplasm samples from scFvM export assays in wild type or CyDisCo cells with the periplasm run adjacent to a sample of total cell extract that was not subjected to fractionation. This confirms that export efficiency is comparable to that observed as a proportion of the combined (C+M+P) samples.

Extensive alterations in substrate surface charge are largely tolerated by the Tat system

We considered it possible that Tat may have an inbuilt tendency to prefer certain surface characteristics in a substrate, with some characteristics perhaps being equated with an unfolded structure (for example, hydrophobic patches may be rejected as signifying incomplete folding). To test this point we changed the overall charge and hydrophobicity of the surface, while ensuring that the structure remained intact and folded. scFvM variants with various surface-residue alterations were designed and tested for export via the Tat system.

The initial surface alterations can be divided into three groups: 1) 2, 4 or 6 uncharged residues were substituted to form 1, 2 or 3 surface salt-bridges composed of Lys-Glu pairs; these variants are denoted 1SB/2SB/3SB, respectively (SB denotes salt bridge). 2) 5 uncharged surface residues were substituted with either Lys, Glu or Arg to create a more highly charged area (denoted 5Lys, 5Glu, and 5Arg). 3) A range of arginines were substituted with lysine on a large scale or, *vice versa*, to test whether the identity of the positively-charged side-chains affected export. All substitutions in groups 1 & 2 are located in a region that is uncharged in the wild-type, whereas group 3 substitutions are distributed globally across the scFvM surface.

Fig. 3 shows surface-charge images of the group 1 and 2 variants (those with increased surface charge). The region circled is relatively uncharged and it can be seen that variant 1SB contains a Lys-Glu salt bridge (side-chains highlighted in blue and red), 2SB contains a further bridge and 3SB contains a third. The lower images show the presence of 5 introduced Lys residues (in blue), 5 Glu (in red) or 5 Arg (blue); all are located in the region circled in the WT image.

In order to ensure that the mutations introduced minimal structural change, biophysical characterisation of the mutants was undertaken; the properties of the mutants are summarised in Table 2. It should be noted that all of these data were collected for proteins that were expressed in WT cells, so they are in the reduced state, whereas export assays were conducted in CyDisCo cells because export is much more efficient. However, since the reduced proteins are folded, it is reasonable to assume that disulphide bond formation by the CyDisCo components may cause the domains to be more rigid, but will not change the tertiary structure (particularly the surface topology) in any significant manner.

Secondary structural analysis of the wild type scFvM showed that the unmodified protein had a beta-sheet fraction of 0.34 regular β-sheet, 0.2 for disordered β-sheet, 0.12 present as turns and the remaining 0.34 as disordered (random coil). All of the scFvM variants showed a high degree of similarity to the wild type, with secondary structure fractions deviating by, at most, 0.03 from the wild type values. The fluorescence peak values, which provide an indication of overall compactness, illustrate that all variants share a similar globular density, as all lay within 0.6 nm of the fluorescence peak of the wild type at 334.6 nm. Additionally, the measured hydrodynamic size from dynamic light scattering correlated with the fluorescence measurements: the wild type had a radius of 2.4 nm in solution, and the variants lay within 0.2 nm of this value. These observations substantiate the rationale for creation of the mutants: *i.e*. changing charge and structural stability of the protein whilst leaving tertiary structure unaltered. These conclusions can be further substantiated by examination of the ability of each variant to bind Protein A. Each of the variants was expressed, extracted and subjected to Protein A affinity chromatography; the Coomassie-stained gel in Supplementary data Figure 2 shows samples of the eluates. All the variants are present at levels that are generally similar to that of the non-mutated protein, confirming that they have intact Protein A binding sites. Protein A affinity requires an intact V_H domain structure for binding and, combined with the further biophysical characterisation presented here, the variants can be considered to have very similar folded conformational states. Given these observations, it is reasonable to assume that any differences in export are due to changes in surface charge and not folding state.

All of the mutants were cloned into vector pYU49 and expression was induced with 0.5 mM IPTG for 3 hrs before fractionation into cytoplasm, membrane and periplasm samples. Panel

A of Figure 4 shows that introducing 1, 2 or 3 Lys-Glu salt bridges on the surface had little effect on export. A clear band is still visible in the periplasm (P) at the mature size (27 kDa), with precursor and mature bands apparent in the membrane and cytoplasm. If anything, the 2SB and 3SB variants appear to be exported more efficiently than the 1SB variant, but after several repeat experiments (see below) our overall conclusion is that all three SB variants are exported with broadly similar efficiency to the non-mutated scFv.

Fig. 4B shows that the 5Lys and 5Glu variants are exported with moderate efficiency (average export efficiency is discussed below), showing that significant changes to the overall surface charge can be tolerated. Creation of an ‘arginine patch’ in the 5Arg variant, however, greatly reduces export efficiency, as demonstrated by a comparatively weak band in the periplasmic fraction (P).

Fig 4C shows fractionation controls with the above mutants, where samples of periplasm were run adjacent to 'total cell' samples that were not fractionated. The export efficiencies appear similar to those observed in panels A and B. In addition, the control epxorts with TorA-scFvM on the left of panels A and B included total cell samples (lanes T) and these likewise show that any losses during fractionation are fairly minimal. The scFvM export assay in panel A also includes a sample of cells before induction (P), where the virtual absence of a band confirms that expression is tightly controlled by IPTG.

Because the 5Arg variant was poorly exported, we tested whether major changes in the Lys:Arg ratio affected export by changing a series of different surface residues in other places. 5 Arg residues were substituted in mutant 5R>K, while 7 lysines were changed to arginine in mutant 7K>R; Fig. 4D shows that both proteins are exported with reasonable

efficiency, suggesting that the low export observed in the 5Arg mutant was due to non-specific effects, rather than any 'aversion' to Arg on the part of the Tat system.

The above experiments were repeated several times and the bands on the blots were subjected to densitometry to compare export efficiencies to that of the wild-type. Fig 5 shows the export efficiencies calculated after 4 separate experiments. Here, it becomes evident that creation of a 5-Lys or 5-Glu charged patch reduces export efficiency only slightly, while introduction of the arginine patch in the 5Arg mutant greatly reduces export efficiency. Exchanging surface lysines with arginines does not drastically affect export, although it seems that surface arginines (7K>R) are slightly less well tolerated.

The Tat system tolerates significant changes in substrate surface hydrophobicity

We next wanted to test whether the Tat system assesses hydrophobicity in a substrate, as this could be a key method by which misfolding is recognised. In this set of substitutions, 4, 5 or 6 hydrophilic surface residues (including 3 charged residues) were substituted with leucine, with the initial substitutions made in the N-terminal domain. This generated 4N-Leu, 5N-Leu and 6N-Leu. Fig. 6 shows surface hydrophobicity profiles of these variants compared to the wild-type. The surface of the non-mutated protein (WT) is predicted to be relatively hydrophilic, with the hydrophobic character (in yellow) mainly contributed by a series of alanines, two valines and a single leucine in this projection. The 4 leucines introduced in variant 4N-Leu are arrowed, as is the additional leucine in 5N-Leu and the 6th Leu in 6N-Leu (this residue is partially hidden in this projection). It is, however, important to note that the structural characteristics of this group of variants were not assessed as they could not be prepared in sufficient quantities.

Expression of these variants results in efficient export of all three mutants, indicating that these changes are tolerated by the proofreading system. Export of 5N-Leu and 6N-Leu appears at first sight to be less efficient, in that the blot signals are slightly weaker for the periplasmic band. However, the key point is that none of these multiple changes blocks export by Tat, showing that significant increases in hydrophobicity are tolerated by the proofreading system.

Addition of an unfolded sequence to a folded structure, results in a complete loss of export

The above studies show that significant changes can be made to the surface of scFvM without blocking export by Tat. As a final part of this study we tested whether the addition of an unfolded section to a folded structure affected export of this protein. To this end we appended 26-residues (SNAIIIIITNKDPNSSSVDKLAAALE), chosen at random, to the C-terminus of scFvM, creating an unstructured 'tail', denoted scFvM-26tail. Secondary structure modelling using the Phyre^2 programme [27] indicates that the core scFvM is unaffected by addition of this tail while the tail itself is largely disordered (Supplementary Figure 3). An export assay is shown in Fig. 7, where it is clear that this variant is not exported at all by Tat throughout a 120-min induction period. Clearly, Tat detects the 26-residue C-terminal 'tail' as an unfolded domain and rejects the protein.

Discussion

In this study we have sought to characterise the factors that dictate whether a Tat substrate is transported or rejected. The primary aims were to test whether Tat senses conformational flexibility in a substrate, and whether specific surface features in a substrate are accepted or rejected by Tat. The actual proofreading mechanism is not understood in any sense, but in this study we assumed that the system may associate features such as hydrophobicity with an unfolded state, and we therefore manipulated these characteristics on the scFvM surface in order to test whether the changes provoked rejection by Tat. We targeted both the reduced and oxidised scFv to the Tat machinery and also appended a disordered 'tail' to the C-terminus to test whether the dynamic state of the substrate is sensed and rejected.

Whereas a different scFv has previously been shown to be efficiently exported by Tat in both the presence or absence of CyDisCo components [23], we have found export of scFvM to be almost totally dependent on the presence of CyDisCo. The protein is exported far more readily if the intra-domain disulphide bonds have formed. This is an unexpected and interesting finding because the protein is clearly correctly folded [26] in the reduced form. We interpret this to mean that the disulphide bonds help to 'lock' the scFv domains in a more rigid conformation whereas, without these anchors in place (*i.e*. in a reducing cytoplasm), scFvM is in a more dynamic state, perhaps similar to a molten globule. Somewhat surprisingly, reduced cysteine residues are more strongly hydrophobic than their cystine (S-S) equivalents [28], suggesting that the increased flexibility of the scFv in the oxidised state is attributable to a loss of these covalent linkages, rather than destabilization of the hydrophobic cores of the two domains. The entropic contribution of disulphide links to protein

stabilization has been well studied, and used as the basis for engineering stability (e.g. [29]). In this scenario, the 10% export we observe in WT cells may well occur with those molecules that happen to be in their most compact conformation when docking at the TatBC substrate-binding complex. These data strongly suggest that the folding state of the substrate is indeed monitored by the Tat system and that the system is sufficiently sensitive to be able to reject a protein even when it is not in any real sense 'unfolded' or even misfolded. Indeed, this scFv was specifically chosen and engineered for stable expression in the *E. coli* cytoplasm (manuscript in preparation and [26]). However, detailed analysis of this issue will require the use of additional test substrates and more sensitive experimental approaches, such as NMR, to link conformational flexibility to translocation-competence.

Among the scFvM variants constructed in this study, the first sets were designed to test the ability of the Tat components to sense the surface charge of substrates. In most of the cases the substitutions were made in a focused area, creating a substantial surface patch with very different properties. The main point to emerge is that multiple changes can be introduced without drastically affecting export. For example, the introduction of 3 Lys-Glu salt bridges represents a considerable change in surface charge with 6 non-polar residues replaced by charged residues, but this mutant is exported with high efficiency.

We further tested the importance of charge characteristics by substituting 5 uncharged residues by Glu, Lys or Arg and found that the Glu or Lys substitutions had only a modest effect on export. Clearly, the Tat system can tolerate significant increases in net negative or positive charge in the surface of a substrate, even when these changes are concentrated in a defined area. We did observe that the introduction of 5 Arg residues had an adverse effect on export, however replacement of 7 Lys residues by Arg in a different mutant (7K>R) had

relatively little effect, strongly suggesting that the Tat system has no preference for lysyl or arginyl side-chains on the substrate surface.

This study has also shown, perhaps surprisingly, that significant changes in surface hydrophobicity are tolerated in Tat-targeted substrates. Exposed hydrophobic regions are a hallmark of unfolded proteins and we considered it highly likely that the introduction of such a region would result in a reduction or block in export. Moreover, it was suggested that hydrophobic domains prevented export of a small unstructured protein in a previous study [16]. In this context it is interesting that the introduction of a substantial hydrophobic domain, containing 4, 5 or 6 Leu residues, did not markedly affect export. Our data show that these substitutions dramatically change the surface characteristics of the N-terminal domain of scFvM, and we conclude that if hydrophobicity is sensed by the Tat proofreading system, the level of hydrophobic character must be substantial for rejection to occur, or it must be sensed in conjunction with other factors (such as conformational flexibility).

In sharp contrast, a mutation designed to introduce an unfolded element to a folded substrate had very dramatic effects on export efficiency; the presence of a 26-residue unstructured tail at the C-terminus completely blocks export of scFvM. When taken in conjunction with the data showing the stimulatory effect of disulphide bond formation, these results indicate that Tat can proofread at different levels. The 26tail variant data show that it can identify a short, unfolded region, while the +/- CyDisCo results with unmodified protein strongly suggest that it can sense and respond to transiently-exposed regions in a globular protein that is active and relatively stable.

Acknowledgements

This work was supported by Biotechnology and Biological Sciences Research Council 'Bioprocessing Research Industry Club' grant BB/K011219/1 to C.R and BB/M006913/1 to JPD. We thank Kevin Howland for carrying out the mass spectrometry analysis.

References

[1] AR Osborne, TA Rapoport, B van der Berg, Protein translocation by the Sec61/SecY Channel, The Ann Rev Cell Dev Biol, 21(2005) 529-550.

[2] R Patel, SM Smith, C Robinson, Protein Transport by the bacterial Tat pathway, Biochim Biosphys Acta - Mol Cell Res 1843 (2014) 1620-1626.

[3] A Robson, I Collinson, The structure of the Sec complex and the problem of protein translocation, EMBO R, 7 (2006) 1099-1103.

[4] A Rodrigue, A Chanal, K Becks, M Muller, L Wu, Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial Tat pathway, J Biol Chem, 274 (1999) 13223-13228.

[5] L Wu, A Chanal, A Rodrigue, Membrane Targeting and translocation of bacterial hydrogenases, Arch Microbio, 173 (2000) 319-324.

[6] JA McDonough, JR McCann, EM Tekippe, JS Silverman, NW Rigel, M Braunstein, Identification of functional Tat signal sequences in Mycobacterium tuberculosis Proteins, J. Bact. 190 (2008) 6428-6438.

[7] UA Oschner, A Snyder, AI Vasil, ML Vasil, Effects of the Twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis, Pro. Nat. Ac. Sci. USA 99 (2002) 8312-8317.

[8] AM Chaddock, A Mant, I Karnauchov, S Brink, RG Hermann, RB Klosgen, C Robinson, A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the ΔpH-dependent thylakoidal protein translocase, EMBO J, 14 (1995) 2715-2722.

[9] NR Stanley, T Palmer, BC Berks, The Twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*, J Biol Chem, 275 (2000) 11591-11596.

[10] M Alami, I Luke, S Deitermann, G Eisner, HG Koch, J Brunner, M Muller, Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli, Mol Cell, 12 (2003) 937-946.

[11] MJ James, SJ Coulthurst, T Palmer, F Sargent, Signal peptide etiquette of a complex respiratory enzyme, Mol Microbiol, 90 (2013) 400-414.

[12] MP DeLisa, D Tullman, G Georgiou, Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway, PNAS, 100 (2003) 6115-6120.

[13] D Halbig, T Wiegert, N Blaudeck, R Freudi, GA Sprenger, The efficient export of NADP-containing gluscose-fructose oxidoreductase to the periplasm of Zymomonas mobilis depends both on an intact twin-arginine motif in the signal peptide and on the generation of a structural export signal induced by cofactor binding, Euro J Biochem, 263 (1999) 543-551.

[14] S Richter, T Bruser, Targeting of unfolded PhoA to the Tat translocon of *Escherichia coli*, J Biol Chem, 280 (2005) 42723-42730.

[15] CFRO Matos, C Robinson, A Di Cola, The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules, EMBO J, 27 (2008) 2055-2063.

[16] S Richter, U Lindenstrauss, C Luke, R Bayliss, T Bruser, Functional Tat transport of unstructured, small, hydrophilic proteins, J Biol Chem, 282 (2007) 33257-33264.

[17] U Lindenstraub, T Bruser, Tat transport of linker-containing proteins in *Escherichia coli*, FEMS Microl Let, 295 (2009) 135-140

[18] J Taubert and T Bruser, Twin-arginine translocation-arresting protein regions contact TatA and TatB, Biol Chem, 395 (2014) 827-836

[19] K Cline and M McCaffery, Evidence for a dynamic and transient pathway through the Tat protein transport machinery, EMBO J, 26 (2007), 3039-3049

[20] PJ Hynds, D Robinson, C Robinson, The Sec-independent Twin-arginine translocation system can transport both tightly folded and malfolded proteins across the thylakoid membrane, J Biol Chem, 273 (1998) 34868-34874

[21] RA Roffey and SM Theg, Analysis of the import of carboxyl-terminal truncations of the 23-kilodalton subunit of the Oxygen-Evolving complex suggests that its structure is an important determinant for thylakoid transport, Plant Physiol, 111 (1996) 1329-1338

[22] MA Rocco, D Waraho-Zhmayev, MP DeLisa, Twin-arginine translocase mutations that suppress folding quality control and permit export of misfolded substrate proteins, PNAS, 109 (2012) 13392-13397.

[23] HI Alanen, Kl Walker, MLV Suberbie, CFRO Matos, S Bonisch, RB Freedman, E Keshavarz-Moore, LW Ruddock, C Robinson, Efficient export of human growth hormone, interferon α2b and antibody fragments to the periplasm by the *Escherichia coli* Tat pathway in the absence of prior disulphide bond formation, Biochim Biophys Mol Cell res, 1853 (2015) 756-763.

[24] P Stolle, B Hou, T Bruser, The Tat substrate CueO is transported in an incomplete folding state, J Biol Chem, (2016) in press.

[25] CFRO Matos, HI Alanen, P Prus, Y Uchida, LW Ruddock, RB Freedman, E Keshavarz-Moore, C Robinson, Efficient export of prefolded, disulphide-bonded recombinant proteins to the periplasm by the Tat pathway in Escherichia coli CyDisCo strains, Amer Ins Chem Eng Biotech, 30 (2014) 281-290.

[26] S Edwardraja, R Neelamegam, V Ramadoss, S Venkatesan, S Lee, Redesigning anti-c-Met single chain Fv antibody for the cytoplasmic folding and its structural analysis, Biotech Bioeng, 106 (2010) 367-375.

[27]. LA Kelley, S Mezulis, CM Yates, MN Wass, MJE Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nat Prot 10 (2015) 845-858.

[28]. N Nagano, M Ota, K Nishikawa, Strongly hydrophobic nature of cysteine residues in proteins, FEBS Lett 458 (1999), 69-71.

[29]. M Zavodsky, CW Chan, JK Huang, M Zolkiewski, L Wen, P Krishnamorrthi, Disulfide bond effects on protein stability: designed variants of Curcurbita maxima trypsin inhibitor-V, Protein Sci 10 (2001), 149-160.

Table 1. Table showing plasmids used to express scFvM and variants for this study.

Plasmid	Function	Reference
pYU49	Expression of protein targeted to Tat via TorA signal peptide. Polycistronic Erv1p and human PDI expression.	[25]
pHIA554	Expression of protein targeted to Tat via TorA signal peptide.	[25]
pHAK13	TorA-scFvM wild-type	This study
pAJ5	TorA-3SB (L11Q, Q13K, A88E, L112K, T114E, S116K)	This study
pAJ6	TorA-5Lys (L11K, Q13K, A88K, L112K, S116K)	This study
pAJ7	TorA-5Glu (L11E, Q13E, A88E, L112E, S116E)	This study
pAJ8	TorA-1SB (L112K, T114E)	This study
pAJ9	TorA-5R>K (R19K, R87K, R150K, R203K, R240K)	This study
pAJ10	TorA-7K>R (K65R, K76R, K162R, K183R, K186R, K235R, K239R)	This study
pAJ22	TorA-C-term26 (Additional - SNAIIIIITNKDPNSSSVDKLAAALE ' 3' to C-terminus)	This study
pAJ24	TorA-2SB (A88E, L112K, T114E, S116K)	This study
pAJ31	TorA-4NLeu (Q13L, R19L, K65L, K76L)	This study
pAJ35	TorA-5Arg (L11R, Q13R, A88R, L112R, S116R)	This study
pAJ36	TorA-5NLeu (Q13L, R19L, K65L, K76L, A88L)	This study
pAJ39	TorA-6NLeu (Q13L, R19L, K65L, K76L, A88L, S116L)	This study

Table 2. Structural characteristics of scFv mutants characterised by circular dichroism, tryptophan fluorescence and dynamic light scattering. Secondary structural elements were obtained through fits of the spectra between 190-260 nm by CDNN secondary structural analysis programme. Samples were at a concentration of 1 mg/ml as determined by UV absorbance at 280nm, in 25 mM sodium acetate buffer, pH5. αR - regular alpha helix, αD - disordered alpha helix, βR - regular beta-sheet, βD - disordered beta sheet, T- beta turns, U - unordered. Secondary structural fractions are rounded.

Mutant	**pI**	**Secondary Structural Element**						**Fluoresence Peak [nm]**	**Hydrodynamic Radius [nm]**
		α_R	**α_D**	**β_R**	**β_D**	**T**	**U**		
WT	**7.8**	0	0	0.34	0.20	0.12	0.34	334.6 ± 0.1	2.41 ± 0.01
5Lys	**9**	0	0	0.35	0.20	0.13	0.33	334.6 ± 0.1	2.39 ± 0.01
5Glu	**5.3**	0	0	0.35	0.20	0.13	0.33	334.6 ± 0.1	2.60 ± 0.02
5Arg	**9**	0	0	0.36	0.20	0.11	0.33	334.6 ± 0.1	2.52 ± 0.04
3SB	**7.8**	0	0	0.35	0.20	0.12	0.33	334.1 ± 0.1	2.40 ± 0.01
2SB	**7.8**	0	0	0.34	0.20	0.13	0.34	334.6 ± 0.1	2.39 ± 0.01
1SB	**7.8**	0	0	0.36	0.20	0.11	0.34	334.9 ± 0.2	2.40 ± 0.01
7K>R	**7.8**	0	0	0.35	0.21	0.09	0.35	334.5 ± 0.1	2.37 ± 0.01
5R>K	**7.8**	0	0	0.35	0.19	0.11	0.35	335.15 ±0.1	2.43 ± 0.01

Figure legends

Figure 1. Tat-dependent export of scFvM is highly dependent on prior disulphide bond formation. A: The images show (i), ribbon diagram detailing the overall structure of scFvM (β-sheets = green, α-helix = turquoise, pink = 3-turn-loop). The disulphide bonds are circled. (ii). Surface potential map; blue = positive charge, red = negative, grey = uncharged/neutral. (iii). Hydrophobicity surface map; purple = hydrophilic, yellow = hydrophobic. B: TorA-scFvM was expressed in wild type or CyDisCo cells and after 2, 3 or 5 h induction the cells were fractionated to yield cytoplasm, membrane or periplasm samples (C, M, P). The samples were subjected to immunoblotting using antibodies to the C-terminal 6-His-tag. Mobility of precursor and mature forms of scFvM are indicated on the right and the mobilities of molecular mass markers (in kDa) are shown on the left.

Figure 2. Fractionation controls to assess export efficiency. TorA-scFvM was expressed in wild type and CyDisCo cells as detailed in Fig. 1B, and in *tat* null mutant cells lacking *tatABCDE* genes. After 2 h induction, samples of 'total' cells were removed and analysed directly (T) while other samples were fractionated into cytoplasm/ membrane/ periplasm (C/M/P) as in Fig. 1B. A: Coomassie-stained gel of the C/M/P fractions, with mobilities of molecular mass markers (in kDa) shown on the left. B: immunoblot showing export of scFvM in CyDisCo cells ('scFvM panel) and in *tat* null mutant cells. The central lanes show the export of an scFvM variant, '26tail' containing a 26-residue extension. C: TorA-scFvM was expressed in the presence (+) or absence (-) of CyDisCo components and samples of total cells and periplasm were analysed.

Figure 3. Structures of scFvM variants with introduced surface salt bridges or charged regions. The Figure shows electrostatic surface charge images of scFvM mutants containing substitutions. *Top row (from left)*: scFvM and the 1SB, 2SB and 3SB variants containing 1, 2 or 3 Lys-Glu pairs. Arrows indicate the introduced Lys-Glu pairs on the surface. *Bottom row (from left)*: the 5Lys, 5Glu and 5Arg variants in which residues in an uncharged domain (circled) were substituted by 5 x Lys, 5x Glu, or 5 x Arg. Red = negatively charged at physiological pH; blue = positively charged at physiological pH. Images generated using CCP4 molecular graphics.

Figure 4. Introduction of surface salt bridges or Lys/Glu patches in scFvM is tolerated by Tat, but the presence of a 5 xArg patch greatly reduces export efficiency. A: export assays using scFvM and variants in which 2, 4 or 6 residues were substituted to generate 1, 2 or 3 Lys-Glu salt bridges (1SB, 2SB, 3SB). Cells were grown and expression of the TorA-scFvM was induced as in Figure 1B, after which samples were fractionated into cytoplasm, membrane or periplasm (C/M/P). Samples of total cells (T) were also analysed from the scFvM control assay, together with an equivalent numbers of uninduced cells before addition of IPTG (U). B: export assays using scFvM and variants in which 5 uncharged residues were substituted by Arg, Glu or Lys. Lane labelling as in A. C: export assays with non-mutated TorA-scFvM, and the indicated variants were carried out as in A/B, after which samples of total cells and periplasm (T, P) were analysed. D: export assays using scFvM and variants in which 5 Arg residues were substituted by Lys (5R>K), or 7 Lys were substituted by Arg (7K>R).

Figure 5. Export efficiencies calculated for the scFvM variants used in this study. 4 separate export assays were carried out using the scFvM variants showed in the graph. After

each assay the signal strength for the exported periplasmic band was calculated as a percentage of total protein in the (C + M + P) fractions, and the average export efficiency is shown relative to that of unmodified TorA-scFvM expressed in CyDisCo cells . wt+, wt-: export efficiencies of TorA-scFvM expressed in CyDisCo and wild type cells, respectively.

Figure 6. Significant increases in substrate surface hydrophobicity are tolerated by the Tat system. The top left image depicts non-mutated scFvM with surface hydrophobic residues shown in yellow, while the remaining images show the 4N-Leu, 5N-Leu and 6N-Leu variants containing 4, 5 or 6 introduced leucine residues (introduced leucines are shown in yellow and arrowed). The images for 5N-Leu and 6N-Leu show only the single additional substitution in each case. Purple = non-hydrophobic, yellow = hydrophobic residue. The lower panel shows export assays using scFvM and the 3 mutated forms, in which cytoplasm, membrane and periplasm samples were analysed. A total cell sample of the scFvM assay was also run as a fractionation control.

Figure 7. **Addition of a disordered 26-residue C-terminal 'tail' results in a complete block in scFvM export**. Top panel: a '26tail' scFvM variant bearing a 26-residue C-terminal extension was expressed in CyDisCo cells. Cells were grown and expression was induced as in Figure 4; samples were remoeved after 60, 75, 90 and 120 min induction and fractionated into cytoplasm, membrane or periplasm (C/M/P). Samples of total cells (T) were also analysed from the 120 min time point, and an equivalent numbers of uninduced cells were analysed before addition of IPTG (U). Lower panel: control export assay using scFvM; samples were fractionated after 120 min induction.

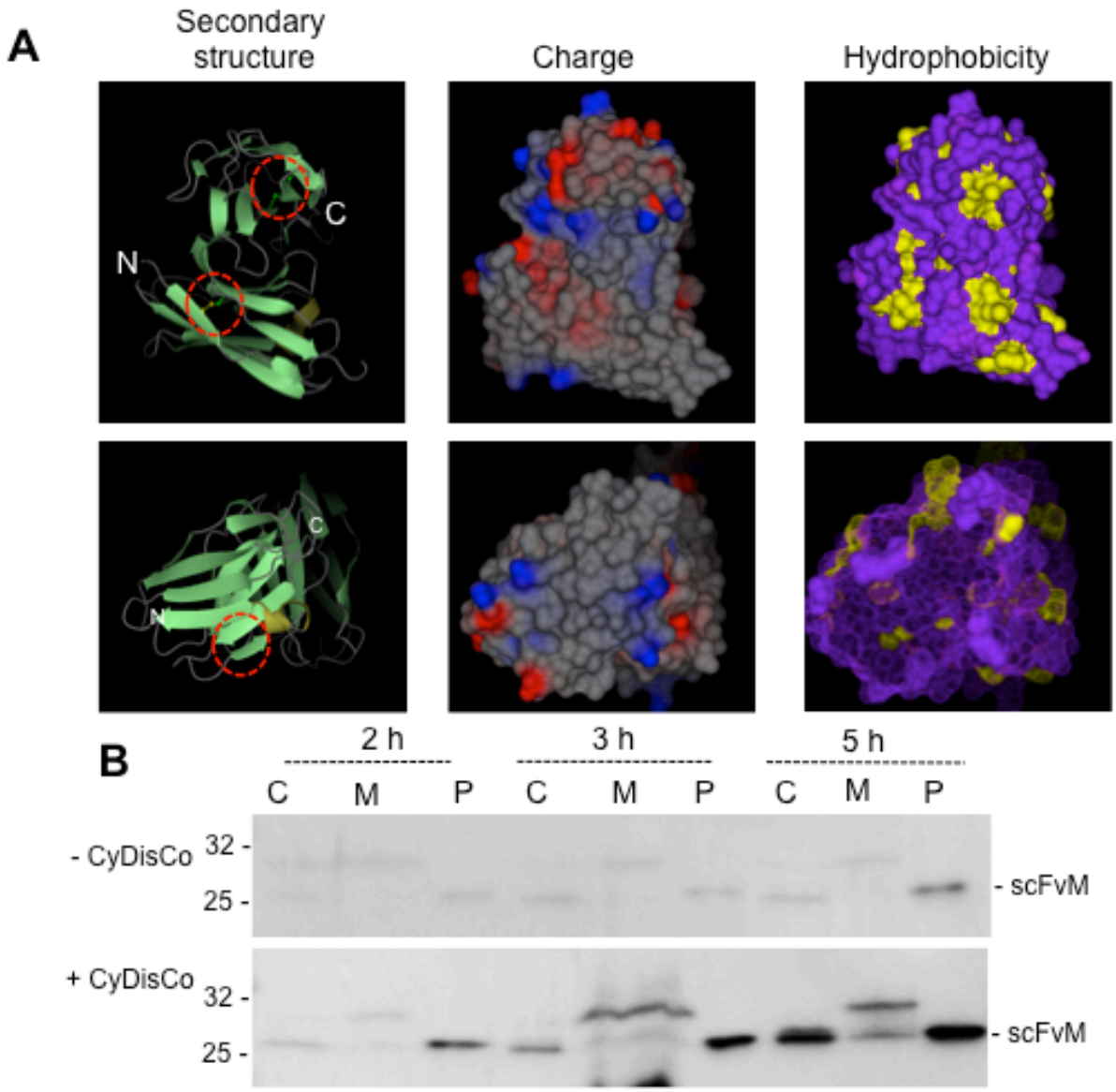
A
Secondary structure
Charge
Hydrophobicity
C
N
C
N
B
2 h
3 h
5 h
C M P C M P C M P
- CyDisCo
32 -
25 -
- scFvM
+ CyDisCo
32 -
25 -
- scFvM

Figure 1

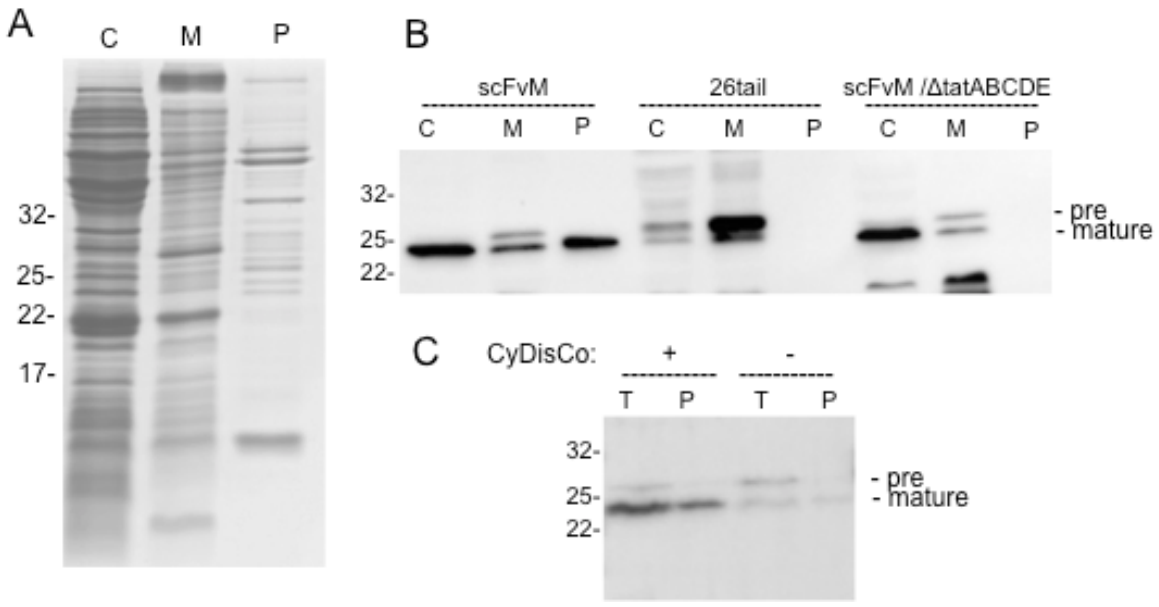
A
C
M
P
32-
25-
22-
17-
B
scFvM
26tail
scFvM /ΔtatABCDE
C
M
P
C
M
P
C
M
P
32-
25-
22-
- pre
- mature
C
CyDisCo:
+
-
T
P
T
P
32-
25-
22-
- pre
- mature

Figure 2

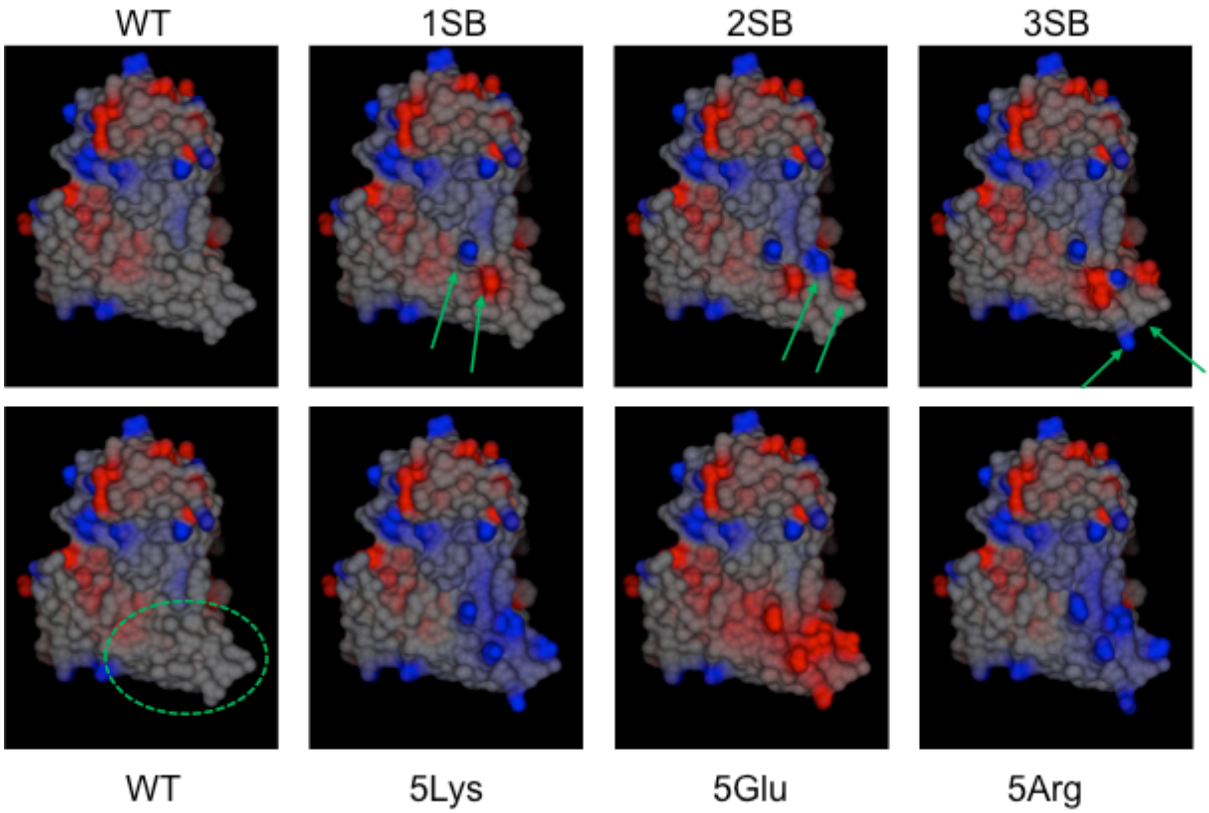

Figure 3

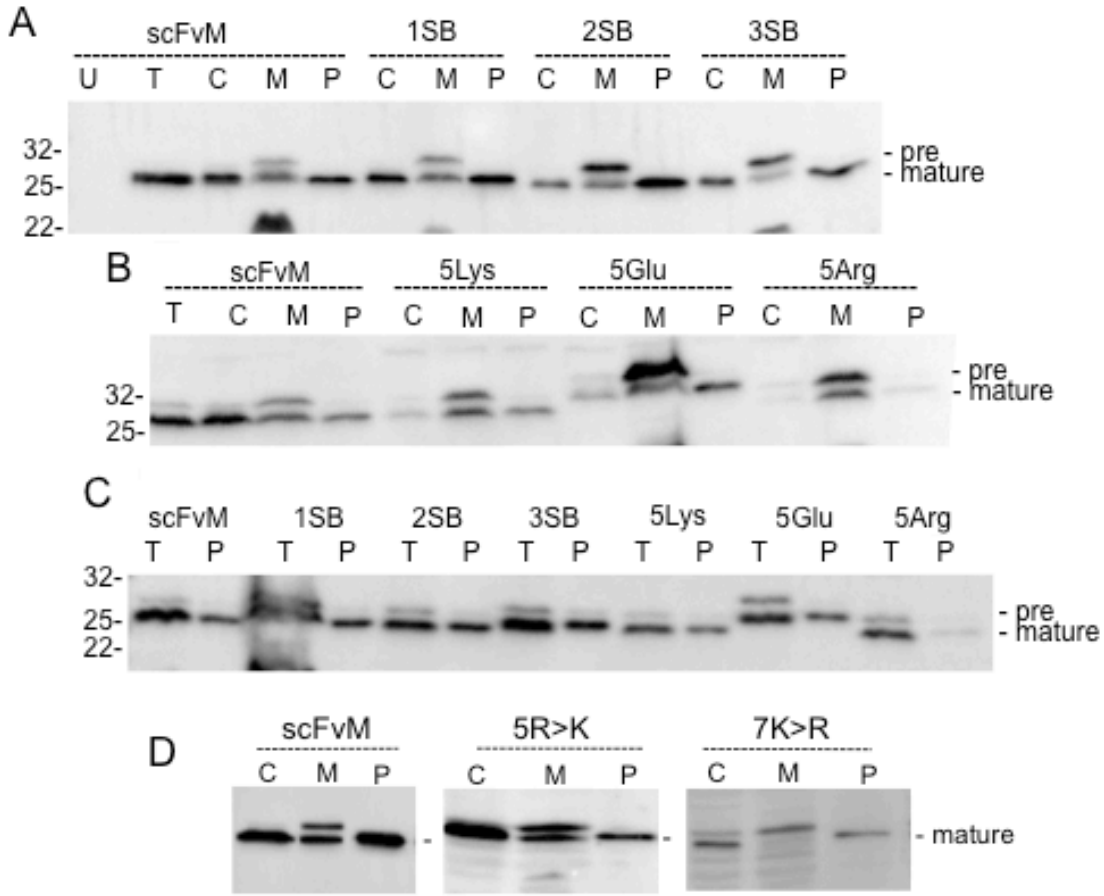
A
scFvM
1SB
2SB
3SB
U T C M P C M P C M P C M P
32-
25-
22-
- pre
- mature
B
scFvM
5Lys
5Glu
5Arg
T C M P C M P C M P C M P
32-
25-
- pre
- mature
C
scFvM
1SB
2SB
3SB
5Lys
5Glu
5Arg
T P T P T P T P T P T P T P
32-
25-
22-
- pre
- mature
D
scFvM
5R>K
7K>R
C M P C M P C M P
- mature

Figure 4

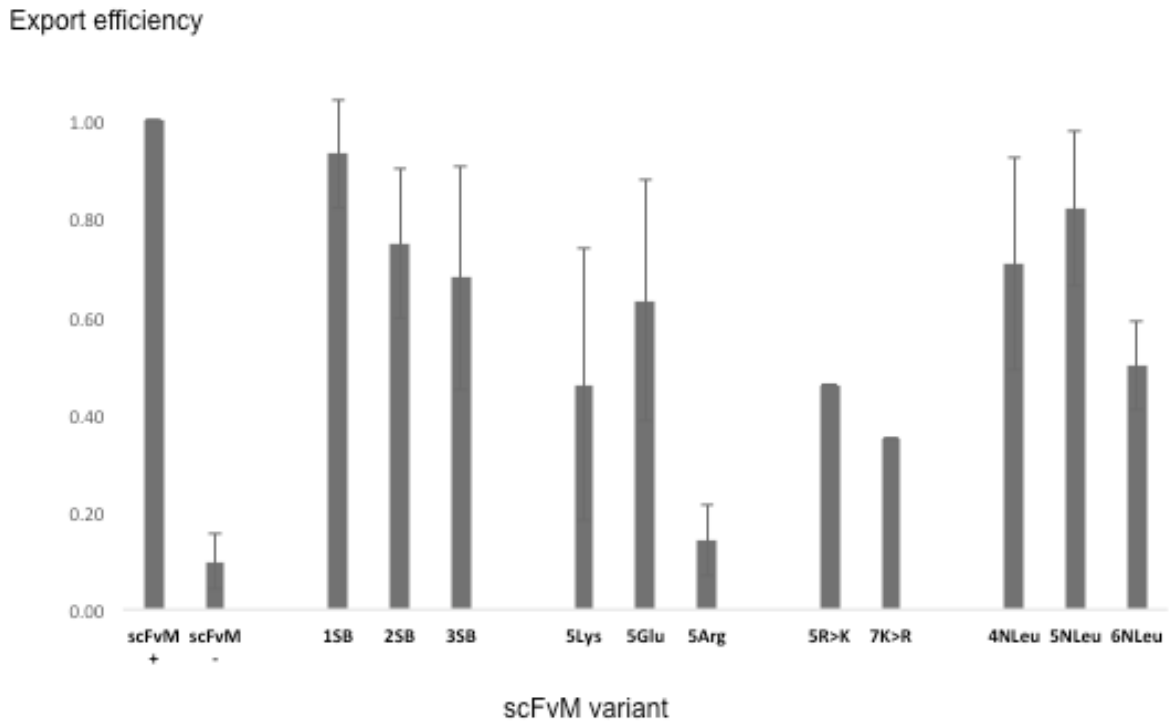
Export efficiency
1.00
0.80
0.60
0.40
0.20
0.00
scFvM +
scFvM -
1SB
2SB
3SB
5Lys
5Glu
5Arg
5R>K
7K>R
4NLeu
5NLeu
6NLeu
scFvM variant

Figure 5

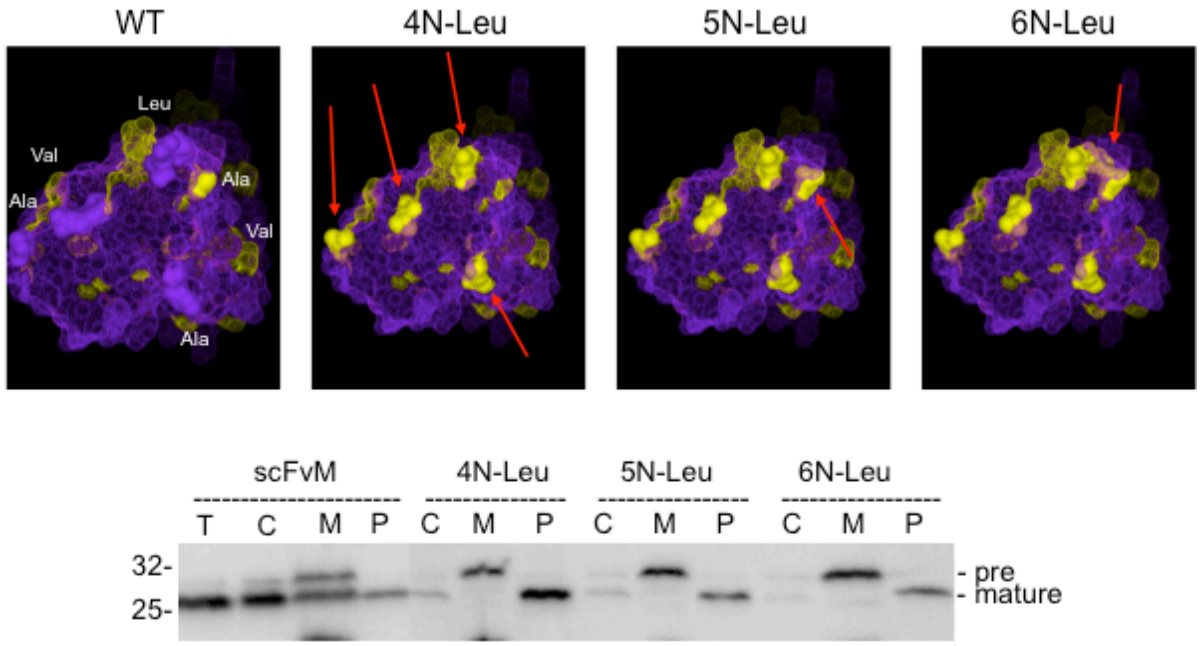
WT
4N-Leu
5N-Leu
6N-Leu
Leu
Val
Ala
Ala
Val
Ala
scFvM
4N-Leu
5N-Leu
6N-Leu
T C M P C M P C M P C M P
32-
25-
- pre
- mature

Figure 6

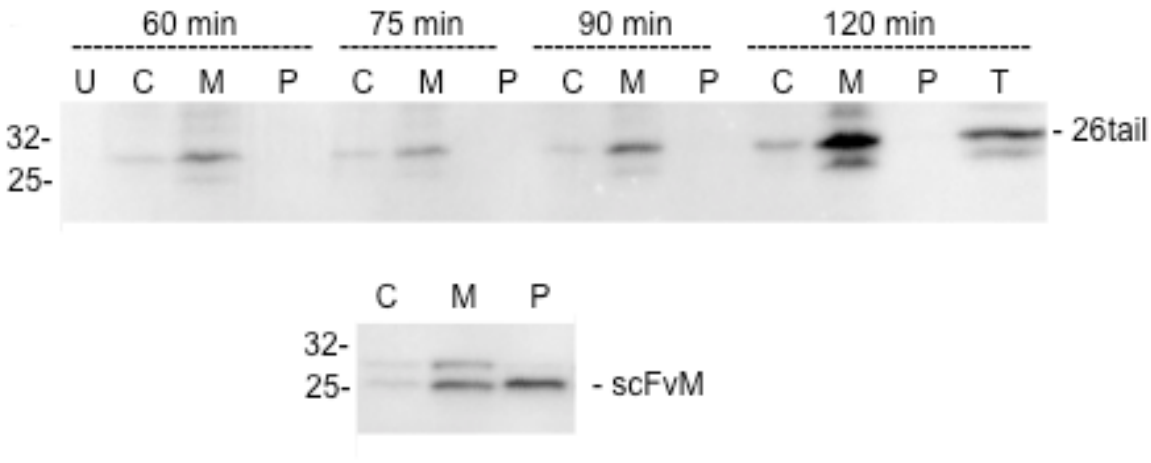
60 min
75 min
90 min
120 min
U C M P C M P C M P C M P T
32-
25-
- 26tail
C M P
32-
25-
- scFvM

Figure 7