The *Bacillus subtilis* TatAdCd system exhibits an extreme level of substrate selectivity

Kelly M. Frain[1], Alexander S. Jones[1], Ronald Schoner[2], Kelly L. Walker[1] and Colin Robinson[1*]

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

[2]Biopharmaceutical Development, MedImmune LLC, Gaithersburg, MD 20878, U.S.A.

*Corresponding author.

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

Tel: +44 1227 823443

Abstract

The Tat system preferentially transports correctly folded proteins across the bacterial membrane although little is known of the proofreading mechanism. Most research has focused on TatABC systems from Gram-negative bacteria, especially *Escherichia coli*, and much less is known of the TatAC-type systems from Gram-positive organisms. We have previously shown that the *Bacillus subtilis* TatAdCd system is functional in an *E. coli tat* null background and able to transport TorA-GFP and native TorA (TMAO reductase); here, we examined its ability to transport other proteins bearing a TorA signal sequence. We show that whereas *E. coli* TatABC transports a wide range of biotherapeutics including human growth hormone, interferon α2b, a VH domain protein and 2 different scFvs, TatAdCd transports the scFvs but completely rejects the other proteins. The system also rejects two native *E. coli* substrates, NrfC and FhuD. Moreover, we have shown that TatABC will transport a wide range of folded scFv variants with the surface altered to incorporate multiple salt bridges, charged residues (5 glutamate, lysine or arginine), or hydrophobic residues (up to 6 leucines). In contrast, TatAdCd completely rejects many of these variants including those with 5 or 6 added Leu residues. The combined data show that the TatABC and TatAdCd systems have very different substrate selectivities, with the TatAdCd system displaying an extreme level of selectivity when compared to the *E. coli* system. The data also provide a preliminary suggestion that TatAdCd may not tolerate surface domains with a level of hydrophobicity above a certain threshold.

Introduction

Bacteria use two main systems to transport proteins across the plasma membrane: the Sec and Tat pathways. The two systems use completely different mechanisms to achieve transmembrane translocation; the Sec pathway transports proteins in an obligatorily unfolded state, while the Tat pathway is specialised for the transport of fully folded proteins. The precise mechanism of the Tat system is still a matter of debate, but the available evidence indicates that in Gram-negative bacteria, substrates bearing a Tat signal peptide interact initially with a TatBC complex, which then assembles with a separate TatA complex to form the full translocation system [1,2].

In Gram-negative bacteria, a number of Tat substrates are redox proteins that bind any of a range of cofactors, such as FeS or NiFe centres. Since these cofactors are only inserted in the cytoplasm, there is a clear need for substrate 'proofreading' to avoid exporting incompletely assembled proteins, and a range of assembly chaperones have been identified, several being substrate-specific [3]. However, the Tat proofreading systems go much further than this, and numerous studies have shown that incorrectly folded proteins - even heterologous proteins that Tat does not encounter *in vivo* - are only transported in a correctly folded state [4-7]. In these cases, incorrectly folded substrates are quantitatively rejected.

Tat systems from Gram-positive organisms are usually different in structure due to the lack of a TatB subunit, with the TatA subunit being bifunctional and able to fulfil the functions of both TatA and TatB from Gram-negative organisms, probably as a result of an early gene duplication event [8-10]. *B. subtilis* has two Tat systems, TatAdCd and TatAyCy, each with different substrate specificities. The only native substrate of the TatAdCd system is the

phosphodiesterase PhoD, but the system has been expressed in an *E. coli tat* null mutant where it provides a partial complementation [9]. In this strain, TatAdCd was shown to export a Tat signal peptide- (TorA-) GFP construct, as well as the native TorA protein (TMAO reductase, an 86 kDa protein bearing a complex molybdenum-based cofactor). Virtually nothing is known about the substrate proofreading activities of Tat systems from Gram-positive organisms, and we considered it possible that some of these systems may not engage in such extensive substrate proofreading - the TatAdCd system, for example, has evolved to transport only a single protein and may not require a rigorous quality control system.

In this study we have tested the ability of TatAdCd to transport a range of heterologous proteins that have all been shown to be exported by the *E. coli* TatABC system. These proteins are all disulphide bond-containing proteins, which raises an additional level of complexity because disulphide bonds cannot form in the reducing cytoplasm of wild type *E. coli* strains [11]. However, several of these proteins have been shown to be exported by TatABC because they form near-native structures even without disulphide bonds [12]. One of the substrates does require prior disulphide bond formation to be exported by Tat, and this substrate was expressed in 'CyDisCo' (Cytoplasmic Disulphide formation in *E. coli*) strains that promote the catalysis of disulphide bonds via a yeast mitochondrial thiol oxidase (Erv1p) and human protein disulphide isomerase (PDI) [13].

This range of substrates, together with a series of mutated versions of one scFv, were tested for export by TatAdCd with surprising results: the system efficiently exported two different scFvs but rejected a series of other biotherapeutic proteins. Moreover, we also tested a range of scFv mutated forms, all of which are efficiently exported by TatABC, and found that a large proportion were totally rejected by TatAdCd. The combined data

demonstrate that the TatAdCd system exhibits an extreme level of substrate selectivity when compared with systems from Gram-negative bacteria.

Materials and Methods

Materials

All chemicals, unless specified otherwise, were obtained from Sigma Chemical Co. Ltd. (Poole, Dorset, UK) and were of analytical grade.

Plasmids and bacterial strains

scFvM constructs were as described in [14] except for the P172S mutant which was generated through Quick change PCR (see below), and subsequently cloned into the vector pYU49. Prior to cloning into pYU49, a 5' NdeI site, a 3' hexahistidine-tag and a 3' BamHI site were added through successive rounds of PCR. PCR primers were:

Primer Name	Sequence	Use
scFvMwtP172SF	T CAG CAG AAA TCG GGT CAG GCA CCG CGT TAC CT	Forward primer
scFvMwtP172SR	TC GCG TGG TAT CAG CAG AAA TCG GGT CAG GCA C	Reverse primer

Overproduction of pCM11: preFhuD was cloned as in [15] into pBAD24 with a 5' EcoR1 site, 3'hexahisdidine-tag and a 3' Pst1 site for the native protein export studies [16].

Primer Name	Sequence	Use
	GCATATGAATTCATGAGCGGCTTACCTC	Forward primer
	GGTACGCTGCAGTCACGCTTTACCTCCGATG	Reverse

		primer

All constructs were amplified using Phusion high fidelity DNA polymerase (New England Biolabs). Products for pYU49 were digested with Nde1 and BamH1 and inserted into pYU49 with a C-terminal hexahistinidine tag. TatABC and TatAdCd were digested with Nde1 and Xho1 and inserted into pLysSBAD, with a C-terminal Strep tag II tag. All plasmid purification was performed using the QIAprep spin miniprep kit (Qiagen) and purified from agarose gels using the gel extraction kit (Qiagen), according to the protocol. All plasmids were fully sequenced (see Table 1). Plasmids encoding Tor-hGH, TorA-IFN and TorA-VH domain are described in [13]. Table 1 lists the plasmids used in this study.

Growth conditions

Following cotransformation of pYU49 and pLysSBAD, a single colony was used to inoculate 5 ml of LB containing 1 mM chloramphenicol and 1 mM ampicillin and grown overnight at 37°C, 200 rpm. This culture was used to inoculate 50 mL of LB in 250 mL conical flasks to OD600= 0.05 and were grown until OD600= 0.5 at 37°C, 200 rpm with the same antibiotics as previously mentioned. Cells were then induced with 0.5 mM IPTG and 0.5 mM arabinose and left for 3 hours at 30°C (unless stated otherwise).

Fractionation

Cells were standardized to the equivalent of OD600=10 and then collected by centrifugation (3000rpm, 10 min). The periplasm (**P)** was collected using EDTA/lysozyme/cold osmotic shock and centrifugation [13]. The resulting pellet was washed and resuspended in 50mM Tris-acetate pH8.2, 2.5mM EDTA pH8.0 before sonication (4-6 X 10 sec) at 8μm amplitude (Soniprep 150plus, Sanyo Gallenkamp,

Loughborough, UK). The sonicate was centrifuged for 30 min at 70,000rpm, 4°C and the supernatant collected as 'Cytoplasmic fraction' **C.** The 'Membrane fraction' **M** is prepared by resuspending the pellet in buffer as before [13]

Detection

Following fractionation protein samples were analysed by SDS-PAGE. Protein was then electrotransferred onto PVDF membrane (Amersham, UK) to be immunoblotted with C-terminal his antibodies (blocked with PBS-T containing 5% (w/v) dried skimmed milk for 1 h). The membranes were washed with PBS-T before a 1 h incubation with the primary antibody, Anti-his C term (Life Technologies, CA, USA), Membrane was washed prior to a 1 hour incubation with the secondary antibody, Anti-mouse IgG (H+L) HRP conjugate (Promega, WI, USA). Membranes were washed and proteins were visualized using an ECL (enhanced chemiluminescence) detection system (BioRad, Herts, UK) subject to the manufacturer's instructions.

Results

Partial complementation of the ΔtatABCDE filamentous cell phenotype upon expression of the B. subtilis TatAdCd machinery

E. coli ΔtatABCDE (*Δtat*) cells exhibit a range of phenotypic effects, highlighting the important roles of the Tat system in this organism. These effects include decreased motility, increased sensitivity to detergents and impaired septation [17]. With the latter, light microscopic analysis showed that *Δtat* strains display defective cell division, forming chains up to 10 cells long [18,19]. This phenotype is complemented when the TatABC machinery is expressed from a plasmid. However, while the *B. subtilis* TatAdCd system is able to function in *Δtat* strains, its expression does not fully complement the phenotype

[20]; at most stages of lab-scale growth the filamentous phenotype is still apparent. It has been suggested that this partial rescue reflects the ability of TatAdCd to efficiently recognize GFP constructs bearing the *E. coli* Tat signal peptide, while at least some other *E. coli* signal peptides are not recognized [20]. This could explain why a number of *E. coli* Tat substrates are mislocalised in these TatAdCd-expressing cells [21].

In this study we analysed the export of a number of proteins, all bearing the TorA signal peptide, in *Δtat* cells expressing pTatABC (*E. coli*) or pTatAdCd (*B. subtilis).* The operons were expressed from a different plasmid to that used in previous studies (pLysS-BAD; see Materials and Methods) and we first confirmed that *Δtat* cells expressing pTatAdCd exhibited the previously-observed filamentous phenotype. Fig. 1 confirms that the *Δtat* strain expressing TatAdCd does indeed show the characteristic filamentous phenotype, whereas expression of TatABC complements this phenotype and single cells are generally observed.

TatAdCd fails to export a range of proteins that are exported by TatABC

Δtat cells expressing TatAdCd have been shown to export TorA-GFP but to date this is the only heterologous protein tested for export by TatAdCd when expressed in the *E. coli Δtat* strain. To further analyse the export capacity of TatAdCd we tested for export of a range of proteins, including human growth hormone (hGH), interferon α2b (IFN), a single chain antibody variable fragment (scFv) and a VH domain construct [13, 22]. All of these constructs have been successfully exported by TatABC in wild type *E. coli* cells, and it was further shown that none of them require prior disulphide bond formation for export by Tat. All proteins tested were expressed on a pET23-based plasmid system either with or without the CyDisCo components, Erv1p and PDI [12]. After expression, cells were fractionated to yield samples of cytoplasm, membrane and periplasm (C, M, P). Fig. 2

shows immunoblots using antibodies to the His-tagged target proteins in tests carried out in the absence of CyDisCo components (identical results were obtained in CyDisCo strains; data not shown).

TorA-hGH (23.6 kDa), TorA-IFN (20.7 kDa) and a TorA-VH domain construct (9.5 kDa) are detected in TatABC-expressing cells, with the vast majority of protein exported to the periplasm and present as mature, processed protein, while no export is observed in *Δtat* cells, as expected. Unexpectedly, TatAdCd exported none of the proteins, with essentially no mature protein detected in the periplasm samples. Only low levels of protein are present in the cytoplasm or membrane fractions, presumably because these proteins are subject to proteolysis within the cytoplasm, as reported previously [12].

We also tested twio native *E. coli* Tat substrates, namely NrfC and FhuD. Both have been shown to be exported by the Tat export pathway [15,16] and the left hand panels of Fig. 3 show that they are exported to the periplasm with very high efficiency in wild type cells (WT) under these expression conditions. In contrast, essentially no export is observed in *Δtat* cells expressing TatAdCd, and the protein is observed primarily in the membrane fraction (which includes insoluble material).

TatAdCd efficiently exports two different scFv proteins

We next tested the ability of TatAdCd to export two TorA-scFv constructs (Fig. 4). One scFv (28.9 kDa) is raised against the omega peptide of β-galactosidase [23] and the other (27 kDa) is described in [14,24]; these scFvs are termed scFvO and scFvM, respectively. Export of TorA-scFvO was reported in [12] and its export was shown to be CyDisCo-independent. In contrast, we recently carried out a study of the export of TorA-scFvM and showed that export is strongly dependent on prior disulphide bond formation [14].

The upper panel of Fig. 4 shows that both scFvs are exported by TatAdCd in CyDisCo-expressing cells, with export of scFvM indeed more efficient in TatAdCd-expressing cells than in TatABC-expressing cells. Two bands are present in the C and M fractions; the higher band is precursor protein while the lower band is believed to represent proteolytically clipped protein. The scFvs differ in their dependence on disulphide bond formation for export by TatAdCd. TorA-scFvO can be exported in the absence of CyDisCo by the *E. coli* TatABC [14]; the lower panel of Fig. 4 shows that scFvM is exported well in the presence of the CyDisCo components, whereas virtually no export is observed in wild type cells (-CyDisCo). These data resemble those obtained with TatABC-expressing cells in a previous study, where export of TorA-scFvM was shown to be far more efficient in CyDisCo strains [14], so in this case the TatABC and TatAdCd systems do exhibit similar preferences in terms of substrate structure.

Effects of scFvM surface mutations: TatAdCd is far more selective than TatABC

In order to probe the proofreading activity of the *E. coli* TatABC system, we recently tested a range of variants of scFvM. Most of the variants were shown to be properly folded, with the mutations designed to change the hydrophobicity or charge on the surface [14]. In the first group, uncharged residues were substituted to create 1, 2 or 3 Lys-Glu salt bridges (SB1, SB2, SB3). This increased the surface charge but in a net-neutral manner. In the 2nd group, 5 uncharged surface residues were substituted with either Lys, Glu or Arg to create a more positively- or negatively-charged area (denoted 5Lys, 5Glu, and 5Arg). In the third group, 4, 5 or 6 neutral, polar surface residues were substituted with leucine, with the substitutions made in the N-terminal domain (mutants 4N-Leu, 5N-Leu and 6N-Leu). These changes thus create a domain that is much more hydrophobic (see Supplementary Figure 1 for further details).

Fig. 5 shows that the 'SB' mutants in group 1 were all exported by TatABC as shown previously [14]. SB1, with one salt bridge, is exported with particularly high export efficiency although it is notable that this mutant is almost totally degraded in the *Δtat* cells. However, SB2 and SB3 are also exported with reasonable efficiency by TatABC. The data obtained with TatAdCd are slightly different: SB1 and SB2 are exported to a reasonable extent, but no export of SB3 is observed.

The second group of mutations (Fig. 6) which create a highly positively- or negatively-charged area, have interesting effects. 5Lys, 5Glu and 5Arg are all exported by TatABC, with the 5Glu variant exported more efficiently than 5Lys. Although 5Glu and 5Arg are exported by TatAdCd, no export of 5Lys is observed. Given this result, we sought to test whether the TatAdCd inherently discriminates against Lys residues, and we therefore constructed another mutant in which 5 different Arg residues were substituted by Lys. Data obtained with this protein (5R>K) are shown in the bottom panel of Fig. 6. The results show that the mutant is exported well by TatABC and, to a very low extent, by TatAdCd. On the basis of these data we believe that TatAdCd does not have an inherent aversion to Lys residues, and probably rejects the 5Lys mutant for other, undefined reasons.

Tests with the 'hydrophobic patch' group of mutant scFvM proteins, in which 4, 5 or 6 charged surface residues were substituted with leucine, are shown in Fig. 7. The results show that the substitutions had little effect on TatABC export, with all 3 substrates exported well. In contrast, mutant 4N-Leu is exported only very weakly by TatAdCd, while 5N-Leu and 6N-Leu are not exported at all. Clearly, these substitutions render the scFvM particularly unsuitable for transport by TatAdCd.

Overall, these data show that TatAdCd exhibits a far more stringent level of substrate selectivity when compared with TatABC, and the data therefore raise the interesting possibility that the system has an inherent tendency to reject substrates bearing exposed hydrophobic domains.

TatAdCd rejects misfolded variants of scFvM

While the data in Figs. 2-6 point to an unusual level of substrate selectivity by TatAdCd, these tests did not address the question of whether TatAdCd rejects *misfolded* proteins because virtually all of the substrates were shown to be highly folded. We therefore tested TatAdCd with another scFvM variant that contains an unfolded region that causes it to be quantitatively rejected by TatABC [14]. This mutant (‘26tail’) has 26 residues appended to the C-terminus of the scFvM, and this unfolded section is clearly identified as 'inappropriate' and rejected by TatABC [14]. Fig. 8 shows that this mutant is likewise completely rejected by TatAdCd. The other is a P172S substitution that is located between the two major scFvM domains (shown in Fig. 7). We expected this mutation to cause a destabilisation in the overall structure but the variant is transported with high efficiency by TatABC as shown. However, Fig. 8 also shows that this variant is not exported at all by TatAdCd, strongly suggesting that TatAdCd is more sensitive than TatABC in responding to changes in the substrate conformation. We conclude from these data that TatAdCd does have an inherent proofreading system that enables it to identify and reject mis-folded substrates.

Discussion

The primary aim of this study was to further investigate the capabilities of the *B. subtilis* TatAdCd system, and to compare its substrate selectivity and proofreading abilities to

those of the *E. coli* TatABC system. The expression of *B. subtilis* TatAdCd is co-regulated with its single substrate PhoD (a phosphodieasterase) from the *phoD* operon in phosphate starvation conditions [25]. A second system, TatAyCy, exports other Tat substrates [26]. However, when expressed in *E. coli ΔtatABCDE* cells, TatAdCd has been shown to partially complement the *E.coli Δtat* phenotype and to efficiently export TorA-GFP and the native TorA protein (TMAO reductase; an 86 kDa MGD-containing protein). TatAdCd is clearly capable of transporting diverse proteins and we set out to compare this Gram-positive TatAC-type system with the well-studied *E. coli* TatABC system.

In addition to probing any mechanistic differences, these systems also have potential for biotechnological exploitation, and a second aim was to understand the capabilities of TatAdCd in this context. Previous results during batch fermentation using a model Tat substrate (TorA-GFP) showed export to the periplasm and also release into the culture medium [20]. This platform could decrease downstream purification needs and shield the protein from proteases within the cell. A range of biopharmaceuticals, including hGH, IFN, scFvs and a VH domain have been shown to be exported by TatABC [12,13] but they have not been tested for export by TatAdCd.

This study has shown for the first time that a Tat system from a Gram-positive organism can exhibit a far more stringent level substrate specificity when compared with its *E. coli* counterpart. Of the biotherapeutics described above, only the scFvs were exported by TatAdCd even though all of the above proteins are exported with high efficiency by TatABC using the TorA signal peptide. No export of IFN, hGH or VH domain protein to the periplasm was observed via TatAdCd. It is unclear why TatAdCd is unable to transport these proteins. The proteins are all smaller than GFP and substantially smaller than PhoD (62 kDa) and TorA (86 kDa) so size is not the issue.

The export of two different scFvs has also been shown via TatABC previously [13] and [14] and this study has shown that TatAdCd can also export both of these proteins. Moreover, export one of the proteins (scFvM) by TatABC is largely CyDisCo-dependent, and this study has shown that the same applies for export by TatAdCd. We previously explored the phenomenon of Tat 'quality control' by analyzing the ability of TatABC to export different variants of scFvM, which has a well-characterized structure. We found that a variety of surface changes, leading to the addition of salt bridges, changes in charge distribution and the creation of hydrophobic regions, were all tolerated by TatABC with minimal effect on export. Using the same plasmids under the same growth conditions and methods, we observed complete rejection or significantly decreased export of most of these scFv variants when tested with TatAdCd.

Changing the surface charge on the scFv gives clear-cut data but it is difficult to discern the underlying reasons. The presence of 1 or 2 salt bridges (variants SB1 and SB2) is tolerated by TatAdCd, whereas the SB3 variant with 3 salt bridges is completely rejected; it is, however, unclear whether this mutant is rejected because of the position of the third salt bridge or because of structural changes The 5Glu variant (5 surface Glu residues added) is exported with reasonable efficiency whereas the 5Lys variant is completely rejected. Further tests would be required before any conclusions can be reached regarding TatAdCd's preferences for substrate surface charge.

Increases in surface hydrophobicity, however, appear to provide a clearer picture. Variants 4N-Leu, 5N-Leu and 6N-Leu, in which the surface of the N-terminal domain is rendered increasingly hydrophobic, are all exported well by TatABC but 4N-Leu is exported only very weakly by TatAdCd and the 5N-Leu and 6N-Leu proteins are wholly rejected. These

data suggest that the TatAdCd proofreading mechanism may reject substrates in which the hydrophobicity is above a certain threshold, with the corresponding threshold for TatABC perhaps being set at a significantly higher level. However, further studies are required to systematically explore these possibilities before a definitive conclusion can be drawn. Nevertheless, the present study does show that while the TatABC and TatAdCd systems can both transport certain proteins, they have very different substrate selectivity mechanisms, and will be of interest to probe the underlying reasons.

Acknowledgements

This was supported by a Biotechnology and Biological Sciences Research Council CASE studentship to K.M.F. We thank MedImmune for part-funding of the studentship, Dr Ray Field for co-supervision of the project and Dr. Soojin Han for helpful discussions throughout the project. This work was supported by Biotechnology and Biological Sciences Research Council 'Bioprocessing Research Industry Club' grant BB/K011219/1 to C.R.

References

[1] K Cline, Mechanistic Aspects of Folded Protein Transport by the Twin Arginine Translocase (Tat), J Biol Chem, 290 (2015) 16530-8.

[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] K Hatzixanthis, DJ Richardson, F Sargent, Chaperones involved in assembly and export of N-oxide reductases, Biochem Soc Trans, 33 (2005) 124-6.

[4] 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.

[5] CF 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) 2055-63.
[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.

[6] MP DeLisa, D Tullman, G Georgiou, Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway, Proc Natl Acad Sci U S A, 100 (2003) 6115-20.

[7] 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.

[8]. JDH Jongbloed, R van der Ploeg, J-M van Dijl, Bifunctional TatA subunits in minimal Tat protein translocases, Trends Microbiol.; 14 (2006) 2-4.

[9] JP Barnett, RT Eijlander, OP Kuipers, C Robinson, A minimal Tat system from a gram-positive organism: a bifunctional TatA subunit participates in discrete TatAC and TatA complexes, J. Biol. Chem, 283 (2008) 2534–42.

[10] MR Yen, YH Tseng, EH Nguyen, LF Wu, MH Saier, Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system, Arch. Microbiol, 177 (2002) 441–50.

[11] KJ Woycechowsky, RT Raines. Native Disulfide Bond Formation in Proteins Curr Opin Chem Biol, 4 (2000) 533.

[12] HI Alanen, KL Walker, ML Velez Suberbie, CFRO Matos, S Bönisch, 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 disulfide bond formation, BBA 1853 (2015) 756- 763.

[13] CFRO Matos, C Robinson, HI Alanen, P Prus, Y Uchida, LW Ruddock, RB Freedman, E Keshavarz-Moore, Efficient export of pre-folded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains, Biotechnol Prog, 30 (2014) 281-90.

[14] AS Jones, JI Austerberry, R Dajani', J Warwicker, R Curtis, JP Derrick, C Robinson. 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. Biochim Biophys Acta - Mol. Cell Res *(in press)*

[15] CFRO. Matos, C. Robinson, A. Di Cola. The Tat system proofreads FeS protein subsrates and directly initiates the disposal of rejected molecules, The EMBO Journal 27 (2008) 20155-2063.

[16] 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. 21 (2000) 11591-6.

[17] NR Stanley, K. Findlay, B.C. Berks, T. Palmer. *Escherichia coli* strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope. J Bacteriol (2001) 183 139–144.

[18] J. Harrison, H. Ceri, EA Badry, NJ Roper, KL Tomlin, RJ Turner. Effects of the twin-arginine translocase on the structure and antimicrobial susceptibility of *Escherichia coli* biofilms, Can. J. Microbiol 51(2005) 671-683.

[19] B Ize, NR Stanley, G Buchanan, T Palmer, Role of the *Escherichia coli* Tat pathway in outer membrane integrity, Mol Microbiol, 48 (2003) 1183–1193.

[20] AM Albiniak, CFRO Matos, SD Branston, RB Freedman, E, Keshavarz-Moore, C Robinson, High-level secretion of a recombinant protein to the culture medium with a *Bacillus subtilis* twin-arginine translocation system in Escherichia coli, FEBS J, 280 (2013) 3810-21.

[21] MM Reynolds, L Bogomolnaya, J Guo, L Aldrich, D Bokhari, CA Santiviago, M McClelland, H Andrews-Polymenis, Abrogation of the twin arginine transport system in *Salmonella enterica* serovar Typhimurium leads to colonization defects during infection, PLoS ONE, 6 (2011) e5800.

[22] K Dudgeon, K Famm, D Christ, Sequence determinants of protein aggregation in 564 human VH domains, Protein Eng. Des. Sel, 22 (2008) 217–220.

[23] P Martineau, P Jones, G Winter, Expression of an antibody fragment at high levels in the bacterial cytoplasm, J. Mol. Biol, 280 (1998) 117–127.

[24] S Edwardraja, R Neelamegam, V Ramadoss, S Venkatesan, SG Lee, Redesigning of Anti-c-Met Single Chain Fv Antibody for the Cytoplasmic Folding and Its Structural Analysis, Biotechnol. Bioeng, 106 (2010) 367-375.

[25] O Pop, U Martin, C Abel, JP Müller, The twin-arginine signal peptide of PhoD and the TatAd/Cd proteins of *Bacillus subtilis* form an autonomous Tat translocation system, J Biol Chem, 277 (2002) 3268-73.

[26] R van der Ploeg, U Mäder, G Homuth, M Schaffer, EL Denham, CG Monteferrante, M Miethke, MA Marahiel, CR Harwood, T Winter, M Hecker, H Antelmann, JM van Dijl, Environmental salinity determines the specificity and need for Tat-dependent secretion of the YwbN protein in Bacillus subtilis, PLoS One, 6 (2011) e18140.

Table 1. Plasmids used in this study

Plasmid	Function	Reference
pYU49	Expression of protein targeted to Tat via TorA signal peptide. Polycistronic Erv1p and human PDI expression.	Matos *et al* 2014
pHIA554	Expression of protein targeted to Tat via TorA signal peptide. No CyDisCo expression	Matos *et al* 2014
pHA17	TorA-scFvO	Alanen et al 2015
pHA15	TorA-IFN	Alanen et al 2015
pHA14	TorA-hGH	Alanen et al 2015
pHA23	TorA-VH	Alanen et al 2015
pHAK13	TorA-scFvM wild-type	Jones et al 2016
pAJ5	TorA-3SB (L11Q, Q13K, A88E, L112K, T114E, S116K)	Jones et al 2016
pAJ6	TorA-5Lys (L11K, Q13K, A88K, L112K, S116K)	Jones et al 2016
pAJ7	TorA-5Glu (L11E, Q13E, A88E, L112E, S116E)	Jones et al 2016
pAJ8	TorA-1SB (L112K, T114E)	Jones et al 2016
pAJ9	TorA-5R>K (R19K, R87K, R150K, R203K, R240K)	Jones et al 2016
pAJ10	TorA-7K>R (K65R, K76R, K162R, K183R, K186R, K235R, K239R)	Jones et al 2016
pAJ11	TorA-P172S (P172S)	Jones et al 2016
pAJ24	TorA-2SB (A88E, L112K, T114E, S116K)	Jones et al 2016
pAJ31	TorA-4NLeu (Q13L, R19L, K65L, K76L)	Jones et al 2016
pAJ35	TorA-5Arg (L11R, Q13R, A88R, L112R, S116R)	Jones et al 2016
pAJ36	TorA-5NLeu (Q13L, R19L, K65L, K76L, A88L)	Jones et al 2016
pAJ39	TorA-6NLeu (Q13L, R19L, K65L, K76L, A88L, S116L)	Jones et al 2016
pCM1	preNrfC	Matos et al 2008
pCM11	preFhuD	This study

Figure legends

Figure 1. **TatAdCd does not complement the filamentous phenotype of *Δtat* cells**. The figure shows microscope images of *ΔtatABCDE* cells (top panels) and the same cells expressing TatABC or TatAdCd from the pLysS-BAD plasmid, after 2 or 5 hours of induction with arabinose (left and right images, respectively). The images show that TatABC complements the filamentous phenotype of *Δtat* cells whereas TatAdCd does not.

Figure 2. **TatAdCd fails to export human growth hormone, interferon and VH domain protein constructs**. TorA-hGH, TorA-IFN and TorA-VH domain constructs were expressed in *Δtat* cells and in *Δtat* cells expressing TatABC or TatAdCd. After 3 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 His tag on the target proteins. Mobility of mature target proteins is indicated on the right; the mobilities of molecular mass markers (in kDa) are shown on the left.

Figure 3. TatAdCd fails to export two native *E. coli* Tat substrates

The precursor forms of *E. coli* NrfC and FhuD were expressed in wild type cells (WT) or *Δtat* cells expressing TatAdCd as indicated. After 3 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 His tag on the target proteins. Mobility of mature target proteins and the precursor forms is indicated on the right; the mobilities of molecular mass markers (in kDa) are shown on the left.

Figure 4. **TatAdCd efficiently exports two different scFvs**. Upper panels: 2 different scFvs, designated scFvO and scFvM, were expressed with an N-terminal TorA signal

peptide in *Δtat* cells and in *Δtat* cells expressing TatABC or TatAdCd together with the CyDisCo components Erv1p and PDI. After 3 h induction the cells were fractionated to yield cytoplasm, membrane or periplasm samples (C, M, P). Lower panel: TorA-scFvM was expressed in *Δtat* cells expressing TatAdCd in the presence or absence of the CyDisCo components Erv1p and PDI. After 3 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 His tag on the target proteins. Mobility of mature target proteins is indicated on the right; the mobilities of molecular mass markers (in kDa) are shown on the left.

Figure 5. **Differential export of surface salt-bridge-containing scFvM variants by TatABC and TatAdCd**. TorA-scFvM variants SB1, SB2 and SB3 were expressed in *Δtat,* TatABC- or TatAdCd-expressing CyDisCo cells, and samples were fractionated and immunoblotted as detailed in Figure 4. Mobilities of mature-size SB1, SB2 and SB3 are indicated.

Figure 6. **Differential export of scFvM variants containing introduced Glu-, Lys- or Arg-rich domains by TatABC and TatAdCd**. TorA-scFvM variants 5Lys, 5Glu or 5Arg, and a further variant containing 5 Arg to Glu substitutions, were expressed in *Δtat,* TatABC- or TatAdCD-expressing CyDisCo cells, and samples were fractionated and immunoblotted as detailed in Figure 4. Mobilities of mature-size proteins are indicated.

Figure 7. **scFvM variants with increased surface hydrophobicity are transported by TatABC but rejected by TatAdCd**. TorA-scFvM variants 4N-Leu, 5N-Leu and 6N-Leu were expressed in *Δtat,* TatABC- or TatAdCD-expressing CyDisCo cells, and samples were fractionated and immunoblotted as detailed in Figure 4. Mobilities of mature-size

proteins are indicated.

Figure 8. **TatAdCd rejects scFvM variants that are misfolded**. Top left: homology modelled structure of scFvM with Pro172 indicated. TorA-scFvM variants P172S and 26tail (containing a 26-residue C-terminal extension), were expressed in TatAdCD-expressing CyDisCo cells, and the P172S mutant was also expressed in TatABC-expressing CyDisCo cells (top right). Samples were fractionated and immunoblotted as detailed in Figure 4. Mobilities of mature-size proteins are indicated.

Supplementary Figure 1**. Surface charge model of scFvM highlighting key residues altered in this study.** (**A**) and (**B**) show the unmodified protein in two different projections. (**C**): hydrophilic residues were substituted by Leu (*yellow*) to create 'hydrophobic patch' mutants; the image shows the 5N-Leu mutant. (**D**): other changes involved substitution of surface residues to create Lys-Glu salt bridge pairs or patches of charge (*green*) or changes to the Arg/Lys content (*pink*). Full details are given in [14].

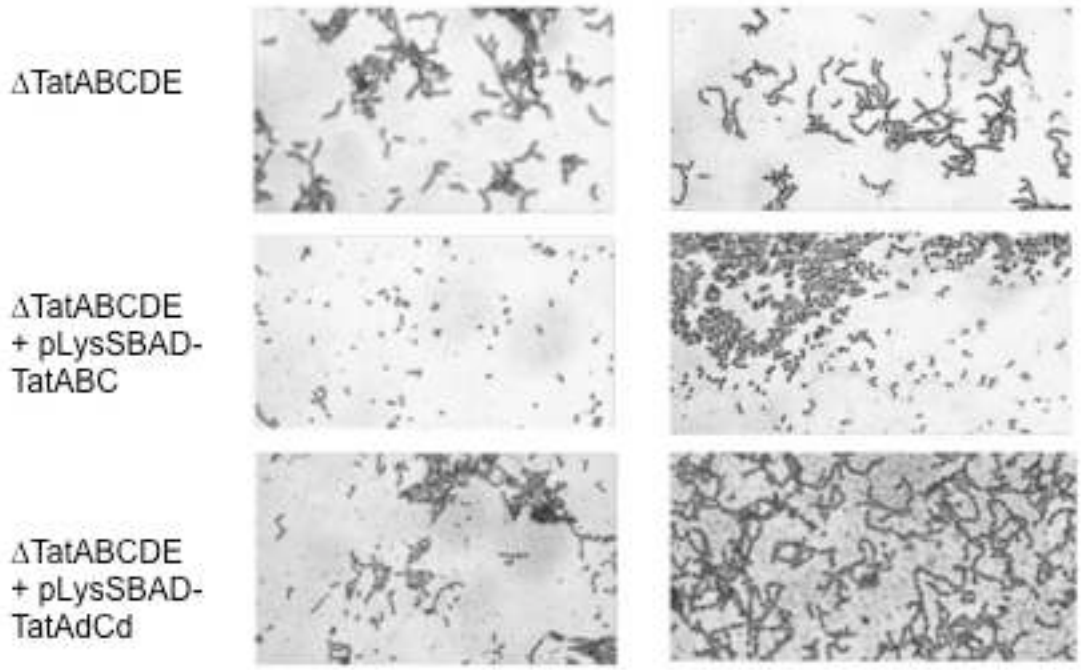
ΔTatABCDE
ΔTatABCDE + pLysSBAD-TatABC
ΔTatABCDE + pLysSBAD-TatAdCd

Figure 1

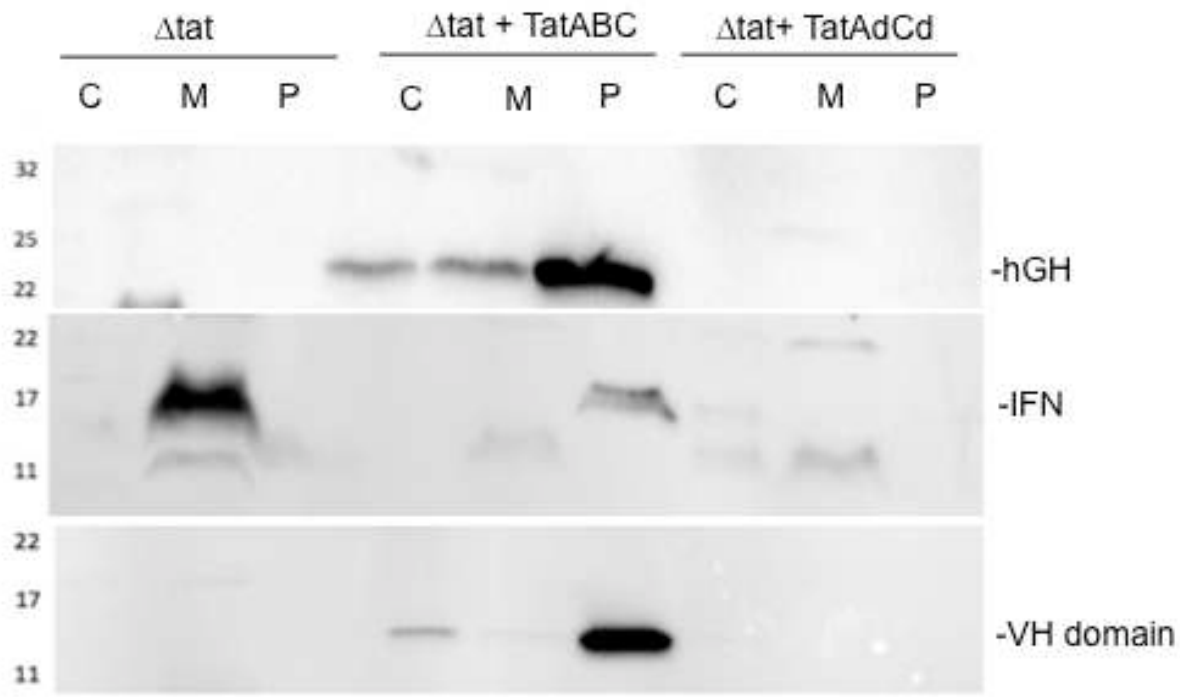
Δtat
Δtat + TatABC
Δtat+ TatAdCd
C
M
P
C
M
P
C
M
P
32
25
22
-hGH
22
17
11
-IFN
22
17
11
-VH domain

Figure 2

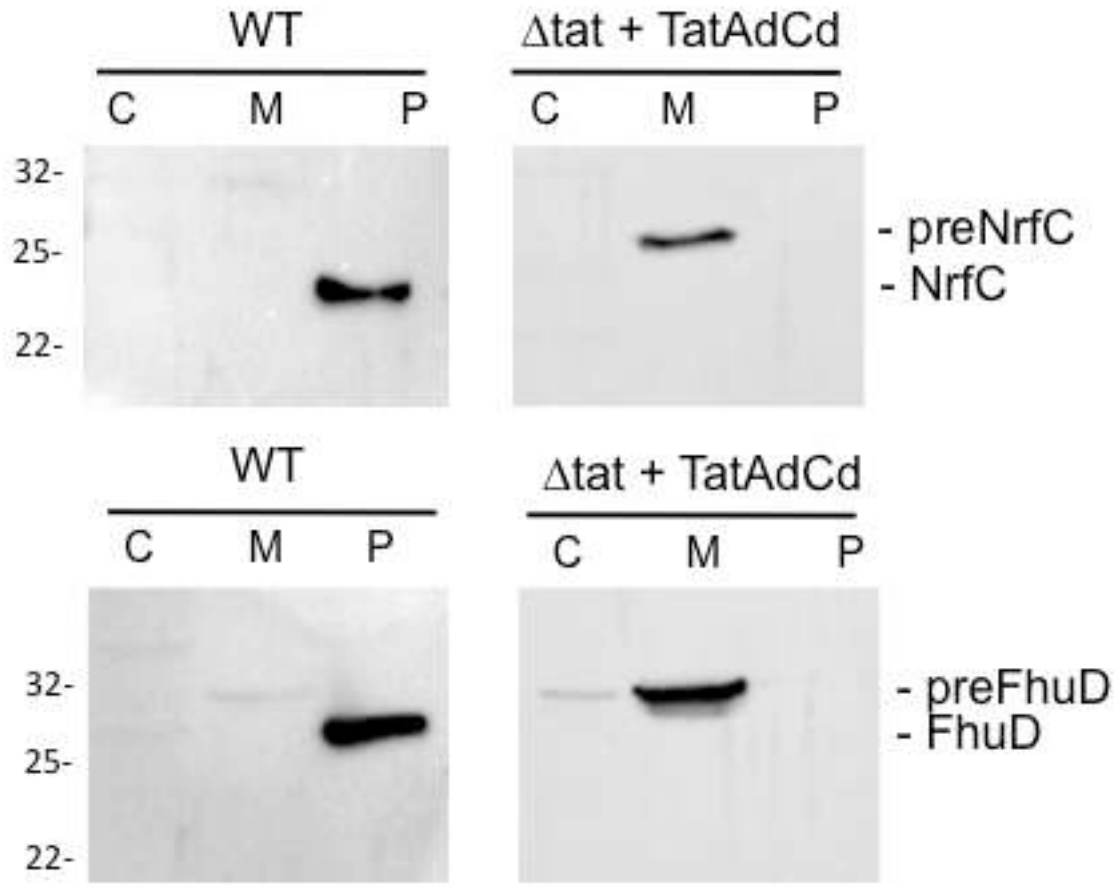
WT
Δtat + TatAdCd
C
M
P
C
M
P
32-
25-
22-
- preNrfC
- NrfC
WT
Δtat + TatAdCd
C
M
P
C
M
P
32-
25-
22-
- preFhuD
- FhuD

Figure 3

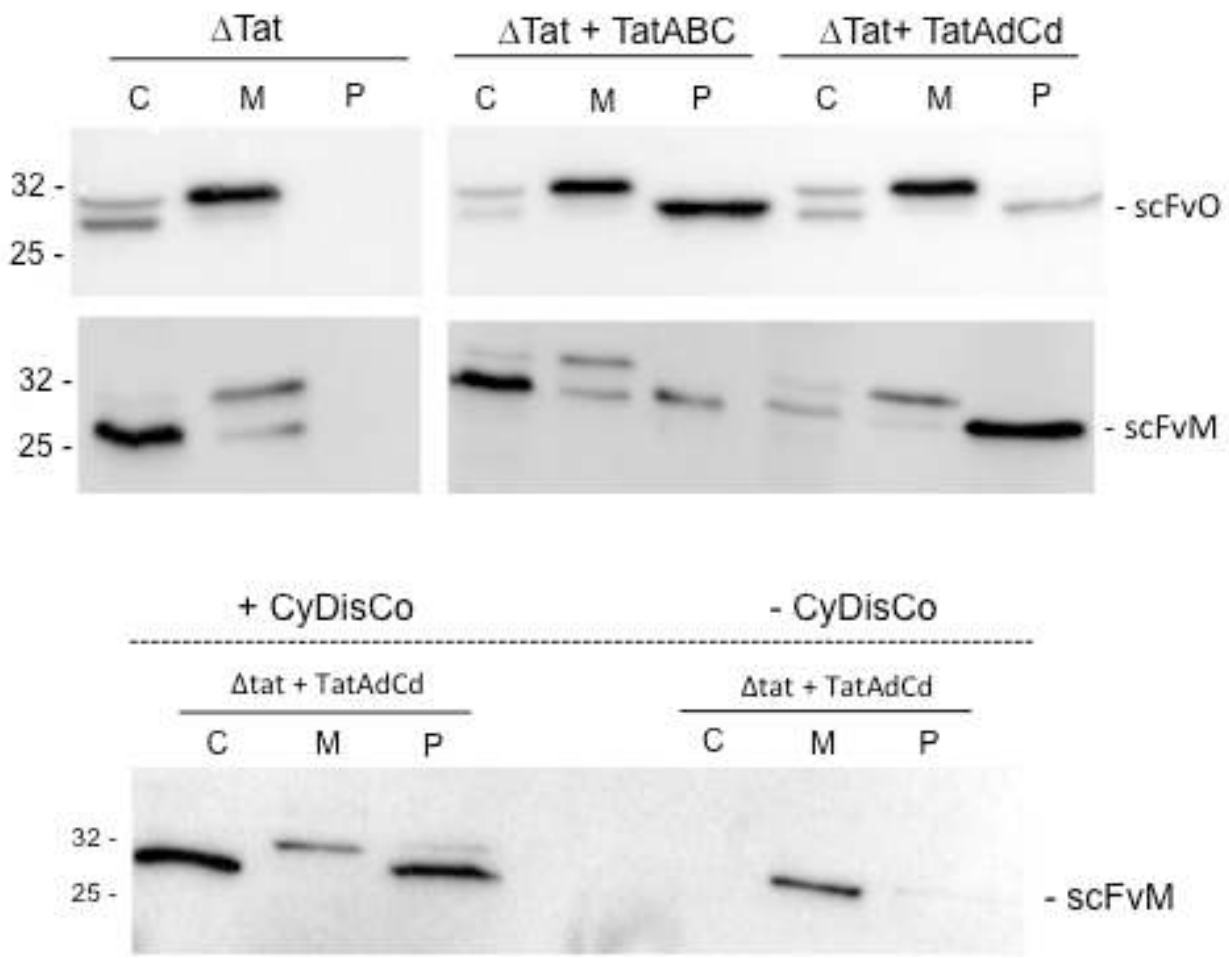
ΔTat
ΔTat + TatABC
ΔTat+ TatAdCd
C
M
P
32 -
25 -
- scFvO
- scFvM
+ CyDisCo
- CyDisCo
Δtat + TatAdCd

Figure 4

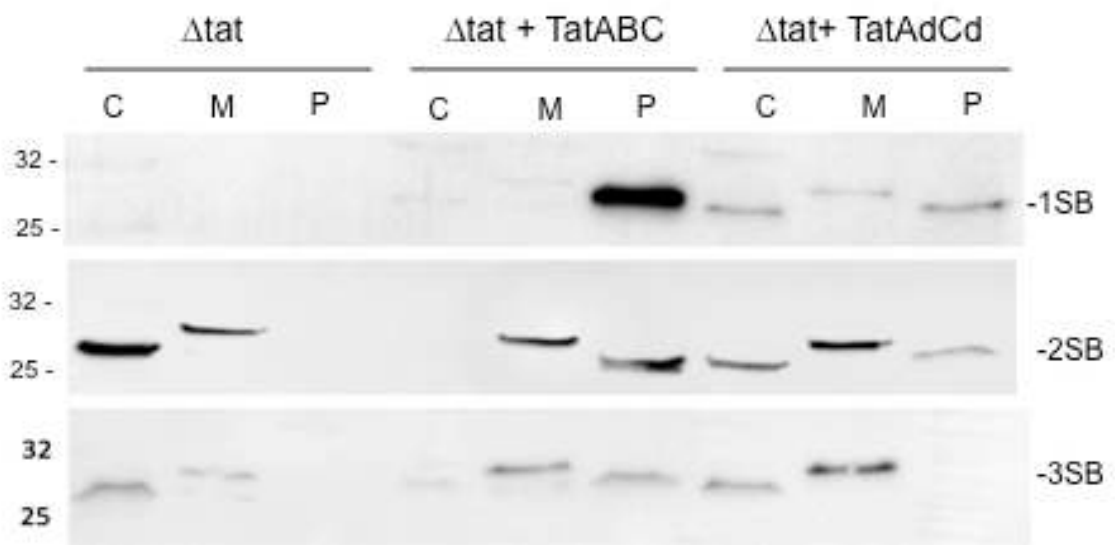
Δtat
Δtat + TatABC
Δtat+ TatAdCd
C
M
P
C
M
P
C
M
P
32 -
25 -
-1SB
32 -
25 -
-2SB
32
25
-3SB

Figure 5

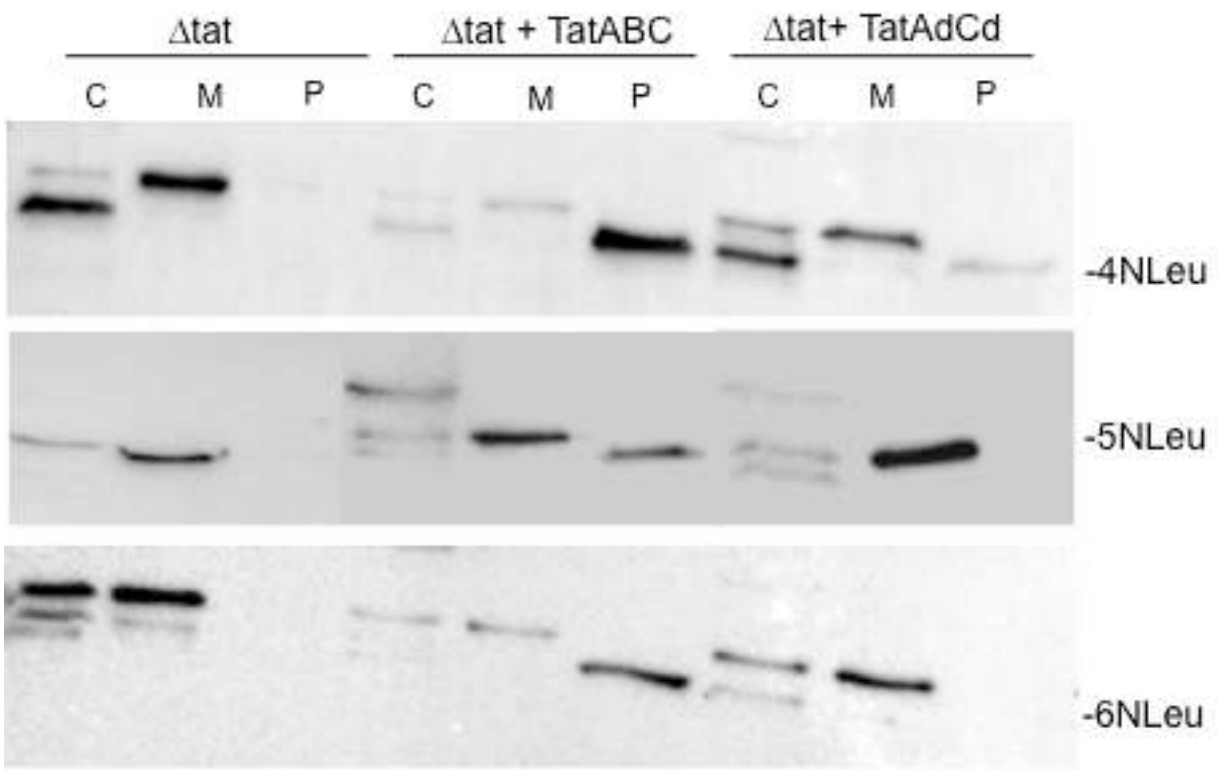

Figure 6

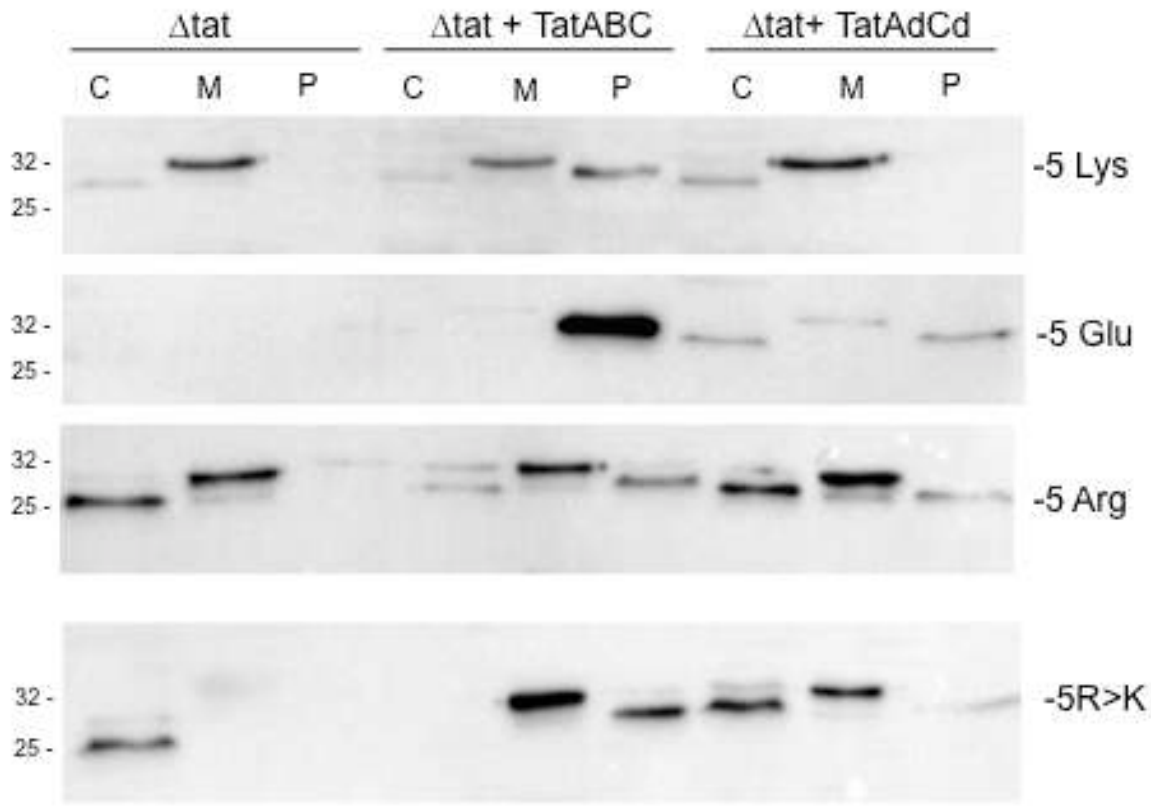
Δtat
Δtat + TatABC
Δtat+ TatAdCd
C
M
P
C
M
P
C
M
P
32 -
25 -
-5 Lys
32 -
25 -
-5 Glu
32 -
25 -
-5 Arg
32 -
25 -
-5R>K

Figure 7

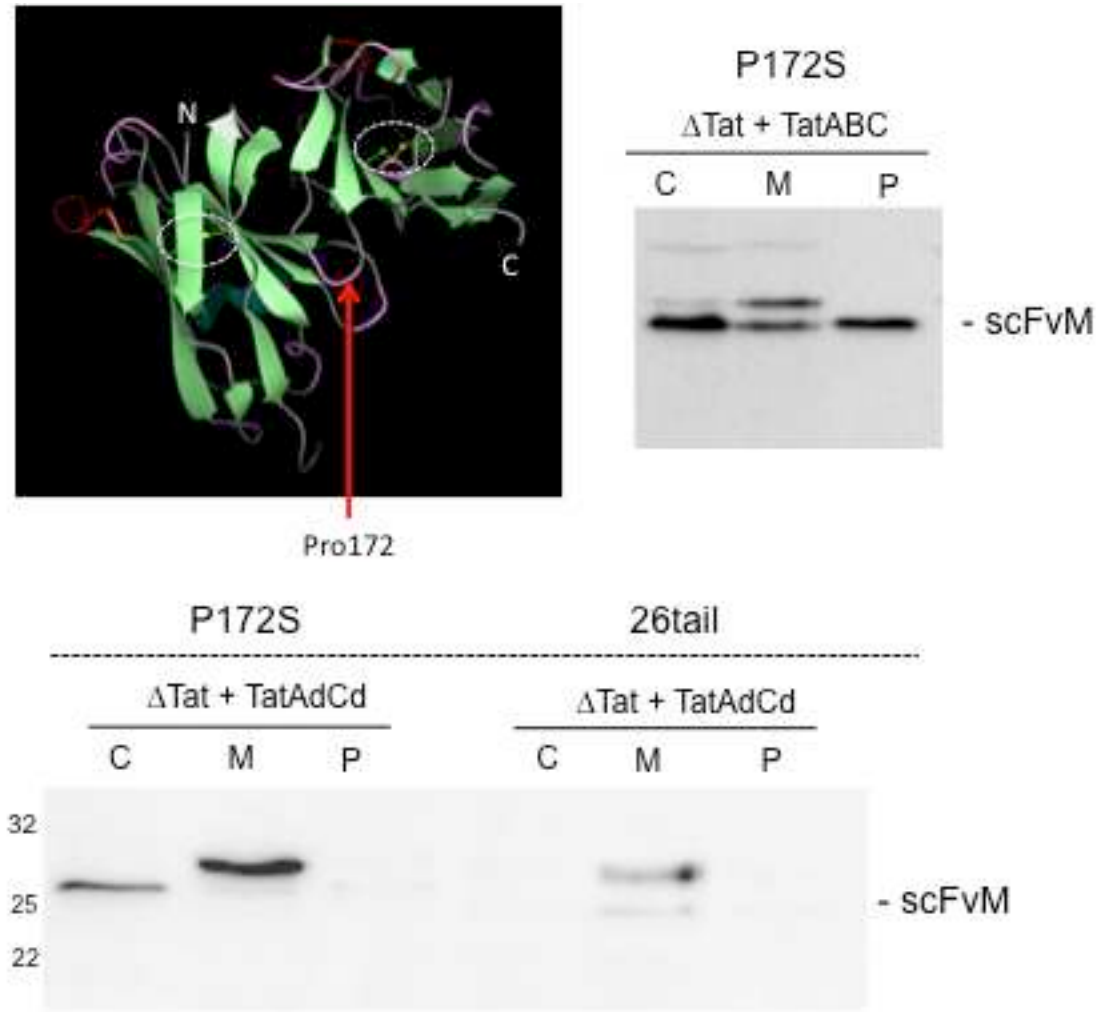
N
C
Pro172
P172S
ΔTat + TatABC
C
M
P
- scFvM
P172S
26tail
ΔTat + TatAdCd
ΔTat + TatAdCd
C
M
P
C
M
P
32
25
22
- scFvM

Figure 8