Title: Hold the fold: how delayed folding aids protein secretion

Nicholas McCaul[1], Ineke Braakman[2]

[1] Department of Biological and Geographical Sciences, School of Applied Sciences, University of Huddersfield, Huddersfield, UK

[2] Cellular Protein Chemistry, Bijvoet Centre for Biomolecular Research, Science4Life, Faculty of Science, Utrecht University, Utrecht, The Netherlands

Standfirst:

In bacteria, N-terminal signal peptides mark proteins for transport across the plasma membrane. A recent study by *Smets et al.* (2022) followed the folding of a pair of structural twins to shed light on how evolution has optimised the secretory process.

Article:

Signal sequences target roughly one third of the *E. coli* proteome (the secretome) for transport across the plasma membrane into the periplasm. A third of the secretome proteins are co-translationally translocated and pass directly from the ribosome into the SecYEG translocon. The other 65% are post-translationally targeted to SecYEG after chain termination and remain largely unfolded before translocation (Cranford-Smith & Huber, 2018). The cytosol is a protein-rich environment, packed with molecular chaperones such as Trigger Factor, SecB and Hsp70 (DnaK), which may slow down or prevent folding. To study the evolutionary adaptations that drive translocation, *Smets et al (2022)* compared the folding and export of PpiA and PpiB: a pair of very similar proteins that differ in their cellular localisation. Crucially, both paralogs are structurally similar and fulfil the same enzymatic function, meaning that sequence differences are likely to be relevant to their folding and export.

Previous *in-vitro* refolding studies have revealed that secreted proteins fold more slowly than their non-secreted counterparts. *Smets et al. (2022)* extend these findings by identifying the key changes in secreted proteins that modulate their folding behaviour. Hydrogen-deuterium-exchange mass spectrometry (HDX-MS) is used for this, in two approaches that profile both “global” and “local” (~3-residue resolution) changes in protein folding and dynamics. Both proteins showed similar global folding dynamics, yet periplasmic PpiA had a larger number of ‘dynamic islands’: small flexible regions among more rigid secondary-structure elements.

Studying structural twins in this way gave access to the question of whether periplasmic PpiA and cytoplasmic PpiB follow the same folding pathway, or whether their different folding speeds reflect distinct pathways. Both proteins adopted a 3-state folding path (unfolded–folding intermediate–folded), and contained similar foldons, 15-35 amino acid stretches that quickly form native-like structures that stabilise their surrounding regions. A key observation was that the order in which the two Ppis acquired the foldons differed. In fast-folding cytoplasmic PpiB, an initial foldon formed at the C-terminus, followed quickly by N-terminal foldons. Secreted PpiA folded more vectorially with N-terminal foldons forming first, before the C-terminal foldons. While N-terminal folding may seem problematic for translocation, the pore of the SecYEG translocon is wide enough to allow translocation of some secondary structure. Secretory PpiA has a larger number of sub-optimal or “frustrated” (energetically

unfavourable) amino acids, which reside mostly in β-hairpins and loops. In this way, folding is initially limited to the formation of basic secondary-structural elements, with bulkier tertiary structures (that could plug the SecYEG pore) forming later.

In an elegant experiment, key frustrated amino acids were grafted from PpiA to cytoplasmic PpiB. This slowed PpiB folding to the same rate as PpiA, with the reverse substitutions being similarly effective. *In vivo*, in a strain that bypasses the need for signal peptides, these changes supported secretion of PpiB and decreased that of PpiA. Folding rate thus appears to decrease with increasing secretion efficiency, suggesting that cellular machinery is tuned to identify proteins that are incompletely folded and direct them for export. Perhaps it is sufficient to populate numerous folding intermediates to provide sufficient time and opportunity for recognition of the signal peptide and for starting translocation. Evolution therefore must strike a balance: if folding is made too slow or difficult, proteins may aggregate in the cytosol or find themselves unable to fold in the periplasm, and fast folders may lose translocation competence before transit is initiated.

Positioned at the N-terminus of secreted proteins, signal peptides in *E. coli* must be recognised by either SRP or SecA if they are to be targeted to the translocon. The flexibility of the C-terminal end of the signal-peptide, often rich in structure-disrupting residues such as proline and glycine, was thought to facilitate access by signal peptidases I and II (De Bona *et al*, 2012). However, *Smets et al (2022)* demonstrate a novel function for this region in disrupting the folding of downstream regions. This effect relies on the folding pathway of the whole protein, however. If folding is mainly initiated at the C-terminus and/or when extensive hydrophobicity in the side chains cause rapid collapse, the signal peptide alone cannot maintain translocation competence. This explains why simply attaching a signal peptide is not always sufficient to induce secretion (Huber *et al*, 2005). Notably, signal peptides and the attached 'mature' protein portion accumulate polymorphisms at comparable rates, underscoring the co-evolution of secreted proteins and their signal peptides (Williams *et al*, 2000).

The importance of this co-evolution becomes clear when we consider that significant folding can (and does) already occur at the ribosome (Nilsson *et al*, 2015). Chain lengths reach at least 50 residues (and often many more) before signal peptides or other sequences are exposed enough to bind folding and export factors (Yang *et al*, 2019). During this time, folding is driven both by intrinsic properties of the sequence and interactions with the ribosome surface. Indeed, already at this early stage, the range and compactness of nascent chain folding intermediates is likely to influence both chaperone interactions and cellular localisation.

Post-translationally translocated substrates arrive at the SecYEG translocon in complex with SecA. Highly diverse signal peptides bind SecA in a flexible, elongated groove (Gelis *et al*, 2007). Next to SecA, the chaperones Trigger Factor and SecB may play a role in maintaining export competence. Strikingly, however, the confirmed pool of Trigger-Factor and SecB interactors only represents a small subset of the secretome; the same can be said for secretory proteins that bind to SecA co-translationally (Loos *et al*, 2019). This may suggest that only the most recalcitrant of proteins require help in remaining translocation competent. As roughly 30% of the *E. coli* proteome is capable of chaperone-independent folding, a key question is whether the secretome is overrepresented in this population (Ciryam *et al*, 2013).

Smets et al. (2022) therefore show us that secretory protein folding is delayed and translocation improved by the effects of amino acids at key points in the mature protein in combination with the N-

terminal signal sequence. Their data explain how a secreted protein has evolved to fold more slowly than its cytosolic paralog, and could be extrapolated to a wide range of secretory proteins, both co- and post-translationally translocated. Similar careful investigations will be required to unpick the complicated interactions between signal peptide, protein sequence, chaperones, and export machinery, but implications for protein quality control in eukaryotic mitochondria and endoplasmic reticulum are already becoming clear.

References

Ciryam P, Morimoto RI, Vendruscolo M, Dobson CM, O'Brien EP (2013) In vivo translation rates can substantially delay the cotranslational folding of the Escherichia coli cytosolic proteome. *Proc Natl Acad Sci U S A* 110: E132-140

Cranford-Smith T, Huber D (2018) The way is the goal: how SecA transports proteins across the cytoplasmic membrane in bacteria. *FEMS Microbiol Lett* 365

De Bona P, Deshmukh L, Gorbatyuk V, Vinogradova O, Kendall DA (2012) Structural studies of a signal peptide in complex with signal peptidase I cytoplasmic domain: the stabilizing effect of membrane-mimetics on the acquired fold. *Proteins* 80: 807-817

Derman AI, Puziss JW, Bassford PJ, Jr., Beckwith J (1993) A signal sequence is not required for protein export in prlA mutants of Escherichia coli. *EMBO J* 12: 879-888

Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Economou A, Kalodimos CG (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* 131: 756-769

Huber D, Cha MI, Debarbieux L, Planson AG, Cruz N, Lopez G, Tasayco ML, Chaffotte A, Beckwith J (2005) A selection for mutants that interfere with folding of Escherichia coli thioredoxin-1 in vivo. *Proc Natl Acad Sci U S A* 102: 18872-18877

Loos MS, Ramakrishnan R, Vranken W, Tsirigotaki A, Tsare EP, Zorzini V, Geyter J, Yuan B, Tsamardinos I, Klappa M *et al* (2019) Structural Basis of the Subcellular Topology Landscape of Escherichia coli. *Front Microbiol* 10: 1670

Nilsson OB, Hedman R, Marino J, Wickles S, Bischoff L, Johansson M, Muller-Lucks A, Trovato F, Puglisi JD, O'Brien EP *et al* (2015) Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep* 12: 1533-1540

Smets D, Tsirigotaki A, Smit JH, Krishnamurthy S, Portaliou AG, Vorobieva A, Vranken W, Karamanou S, Economou A (2022) Evolutionary adaptation of the protein folding pathway for secretability. *EMBO J*: e111344

Williams EJ, Pal C, Hurst LD (2000) The molecular evolution of signal peptides. *Gene* 253: 313-322

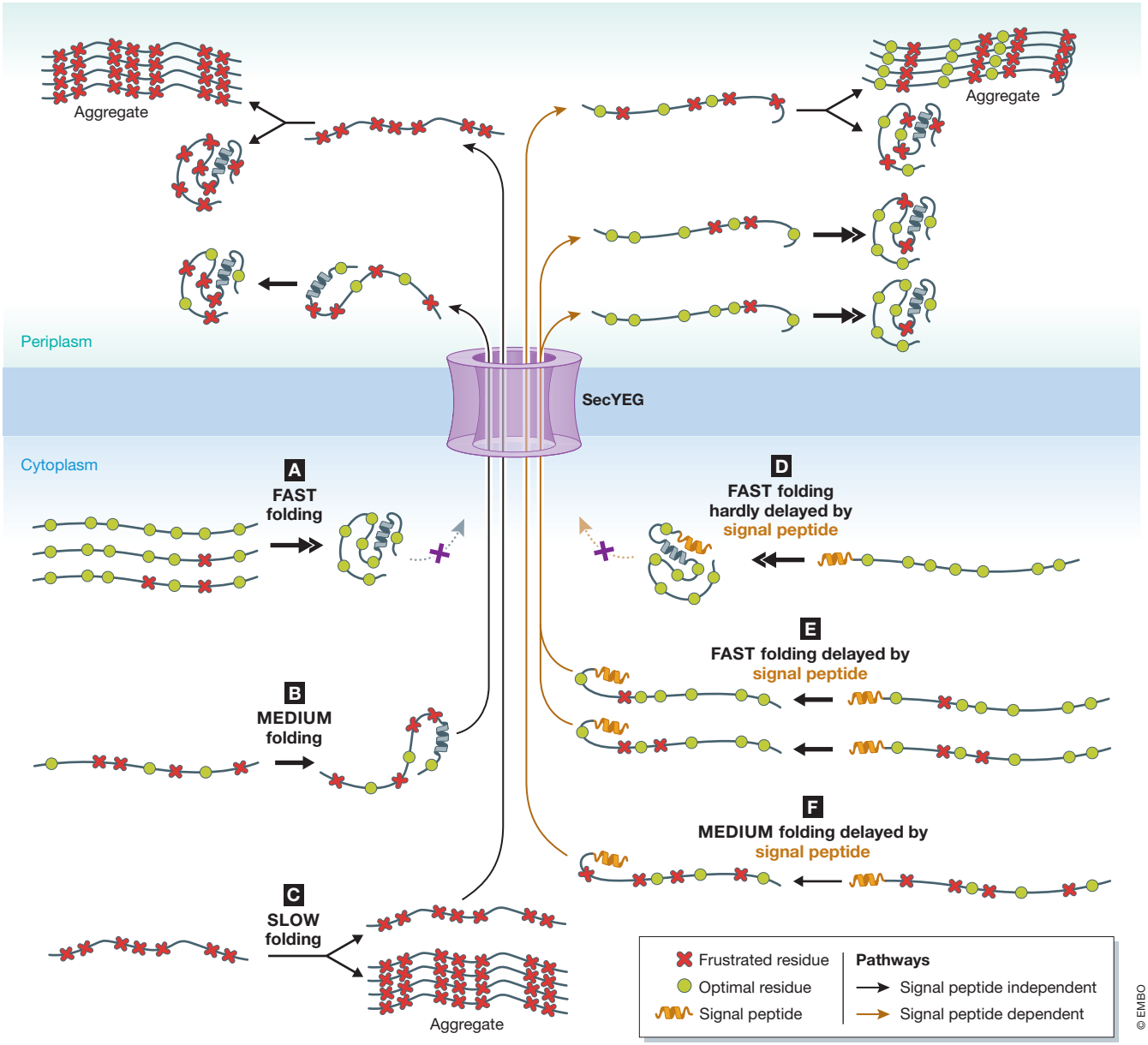
Yang CI, Hsieh HH, Shan SO (2019) Timing and specificity of cotranslational nascent protein modification in bacteria. *Proc Natl Acad Sci U S A* 116: 23050-23060

Legend

Figure 1 – Folding-dependent protein translocation across the plasma membrane.

Protein transport through the SecYEG translocon in the bacterial plasma membrane to the periplasm and beyond requires proteins to maintain export competence in incompletely folded form. Suboptimal,

also called frustrated, residues slow down folding. When a protein contains a very high ratio of optimal to frustrated residues, it will likely fold quickly, preventing translocation from taking place (A). By increasing the relative number of frustrated residues, folding is slowed, providing a window for translocation to occur. Once in the periplasm, the protein will finish folding (B). With too many frustrated residues, the folding rate may become too low, leading to increased risk of aggregation. Involvement of molecular chaperones may be necessary to maintain solubility and translocation competence and promote folding in the periplasm (C). A SecYEG mutant (prlA4) can translocate proteins lacking a signal peptide (black pathway (Derman *et al*, 1993)) but wild-type SecYEG cannot. The required N-terminal signal peptide then provides an additional layer of control by delaying folding of the protein's N-terminus (orange pathway). This delay alone is not sufficient to sustain translocation competence of extremely fast folding proteins, which remain cytosolic (D). Proteins that would otherwise have a folding rate too high for translocation however, are delayed enough to enable translocation (E). During translocation across the plasma membrane, the signal peptide is cleaved off, enabling folding into the functional, native form once in the periplasm, which also here requires a reasonable rate to have any yield. Molecular chaperones add a layer of control and may affect folding rates and translocation. Slowing folding down too much, however, may lead to difficulty in folding in the periplasm, resulting in aggregation (F).


Aggregate
Aggregate
Periplasm
SecYEG
Cytoplasm
A
FAST
folding
B
MEDIUM
folding
C
SLOW
folding
Aggregate
D
FAST folding
hardly delayed by
signal peptide
E
FAST folding delayed by
signal peptide
F
MEDIUM folding delayed by
signal peptide
Frustrated residue
Optimal residue
Signal peptide
Pathways
Signal peptide independent
Signal peptide dependent

© EMBO