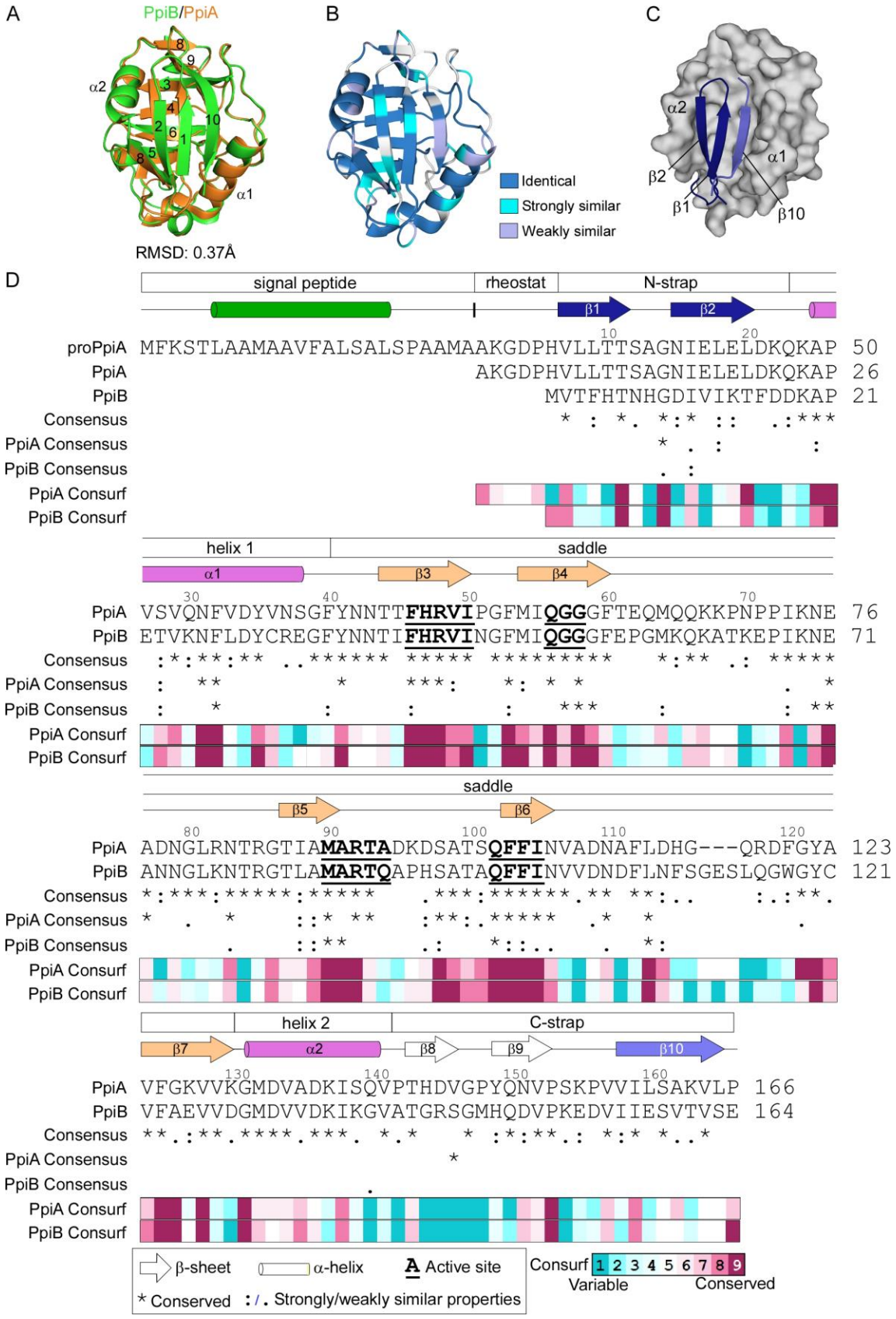
Appendix

Evolutionary adaptation of the protein folding pathway for secretability

Dries Smets, Alexandra Tsirigotaki, Jochem H. Smit, Srinath Krishnamurthy, Athina G. Portaliou, Anastassia Vorobieva, Wim Vranken, Spyridoula Karamanou and Anastassios Economou

Table of contents

Supplemental figures



Appendix Figure S1 Structure and sequence alignment of PpiA and PpiB with homology comparison across bacteria (related to Figure 1)

A. Structural alignment of periplasmic PpiA (PDB 1V9T: chainB 1.8Å, orange) and cytoplasmic PpiB (PDB 1LOP 1.7Å, green) using PyMOL, yielding an RMSD of 0.37Å. Both structures consist of an orthogonal β-barrel with the anti-parallel β-strands in the following sequence β1-10-3-4-6-5-7-2 (numbers indicated on the structure) with α-helices on either side of the barrel and two minor β-strands β8/9 located outside the β-barrel.

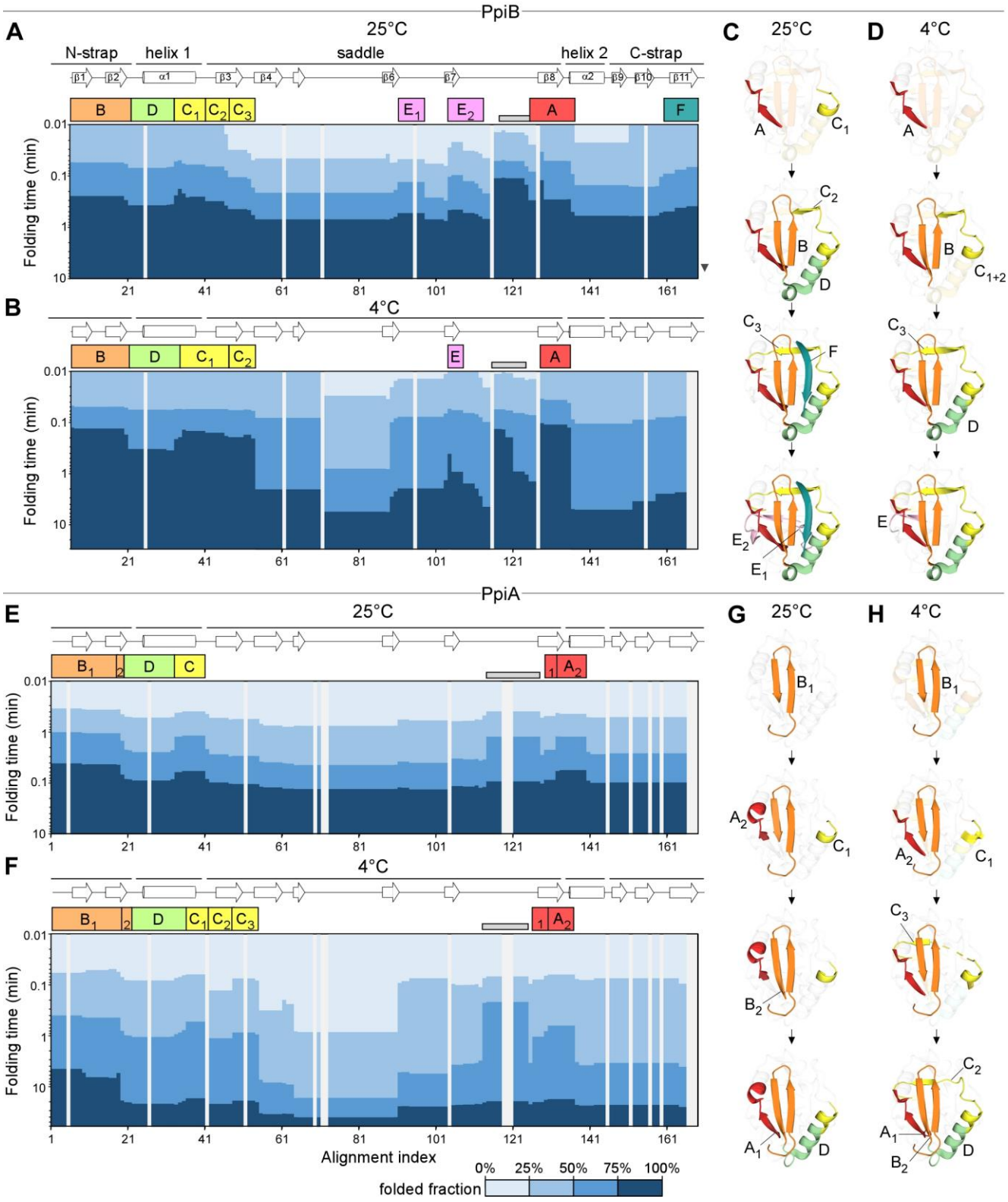
B. Similar residues between *E. coli* PpiA (P0AFL3) and PpiB (P23869). Identical, strongly similar physicochemical properties (scoring >0.5, Gonnet PAM 250 matrix) and weakly similar physico-chemical properties (scoring ≤ 0.5, Gonnet PAM 250 matrix) are depicted on the 3D structure of PpiA (PDB 1V9T: chainB).

C. Structural position of the 'front-facing' N- and C-strap (ribbons, dark blue and grey, respectively) within the cradle formed by the saddle in the back and the α-helices on either side (surface, light grey) using PpiA (PDB 1V9T: chainB).

D. Sequence alignment of *E. coli* (pro)PpiA with PpiB using Clustal Omega [1].

Top: the linear secondary structure is displayed with the different structural elements coloured and annotated (based on RCSB PDB). The residues of the active site are underlined [2]. '*': identical residues; ':' strongly similar and '.' weakly similar physicochemical properties.

Bottom: Consensus derived from 150 (pro)PpiA and PpiB sequences from across γ-proteobacteria that contain both twins (all sequences in Dataset EV1).

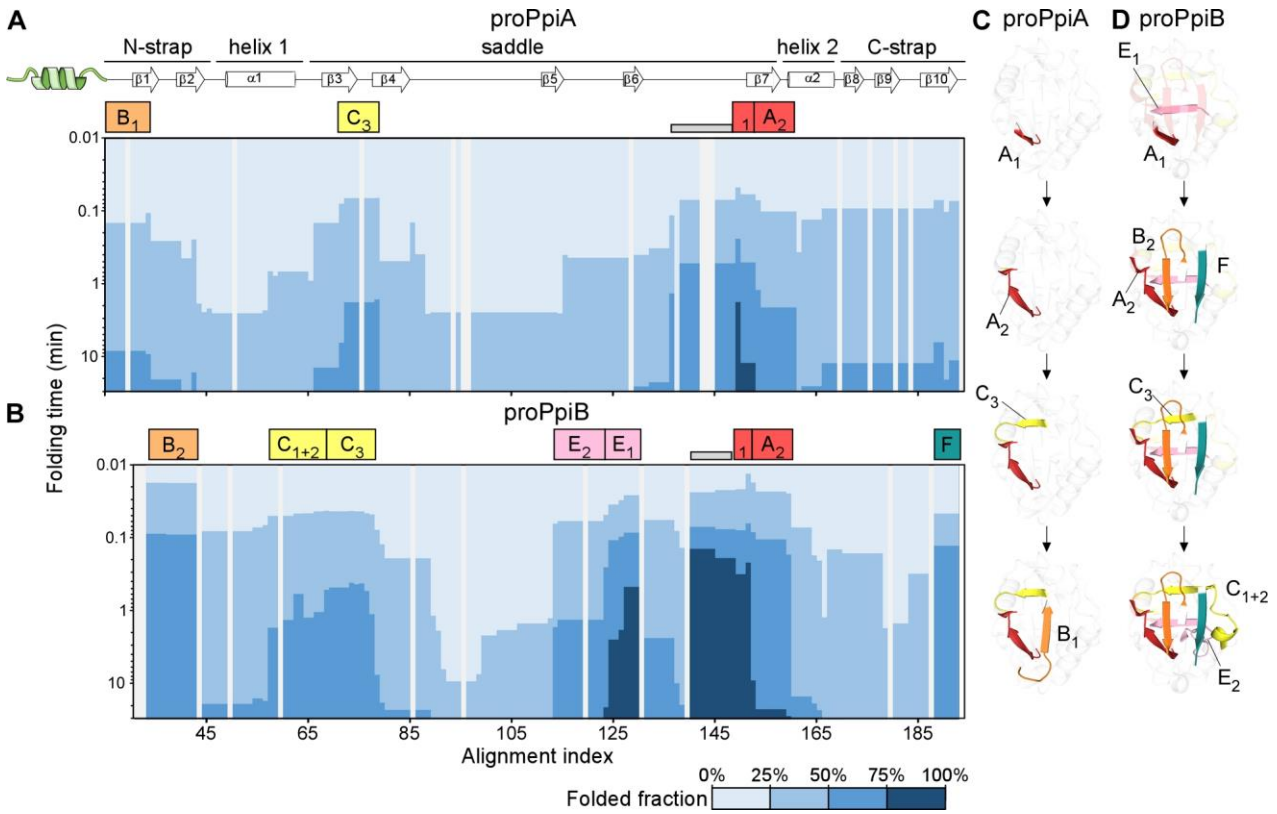


Appendix Figure S2 Folding kinetics of PpiA and PpiB at 25 and 4°C analyzed with local HDX-MS displayed as a colour map (related to Figure 3)

Folding kinetics of PpiB (**A-D**) and PpiA (**E-H**) at 25°C or 4°C (as indicated), monitored by local HDX-MS.

A, B, E, F. The HDX-MS refolding kinetics data for PpiA and PpiB at 25 and 4°C (Dataset EV4; *n=3* biological repeats) were further analyzed by PyHDX in order to determine the folded fractions per residue (Dataset EV5). The pipeline of analysis is shown in Fig.EV3B. Folding, displayed per residue (x-axis) over time (y-axis) in a colour map in steps of 25% folded fraction (as indicated at the bottom), is shown up to 10min for 25°C and 30-60min for 4°C (as indicated, complete data set in Dataset EV5). The alignment index (x-axis) is based on PpiA (extended N-tail; missing loop between β6-β7; Appendix Fig. S1D). For each peptide, 100% folding was set to the D-uptake of the final folded protein. Grey areas: residues absent in one of the twins, prolines or no experimental coverage. Colour-boxes below the linear secondary structure map (top) indicate foldons; named in alphabetical order and subscript numbers (if formed in gradual steps) following the order of formation sequence. Grey bar: unstructured regions that acquired final states fast (Fig. EV3) and were omitted from the analysis.

C, D, G, H. Foldons, colour-coded as in the left panels, are indicated relative to formation time and temperature on the PpiB (1LOP) and PpiA (1V9T) 3D-structures. The indicated time points were: for PpiB, 25°C ($t_{80\%}$ of 0.29-0.33-0.42-0.47 min); for PpiB, 4°C ($t_{80\%}$ of 0.09-0.29-0.90-1.75 min); for PpiA, 25°C ($t_{80\%}$ of 0.24-0.33-0.47-0.51 min); for PpiA, 4°C ($t_{50\%}$ of 0.34-0.55-0.79-0.99 min) (Dataset EV5).

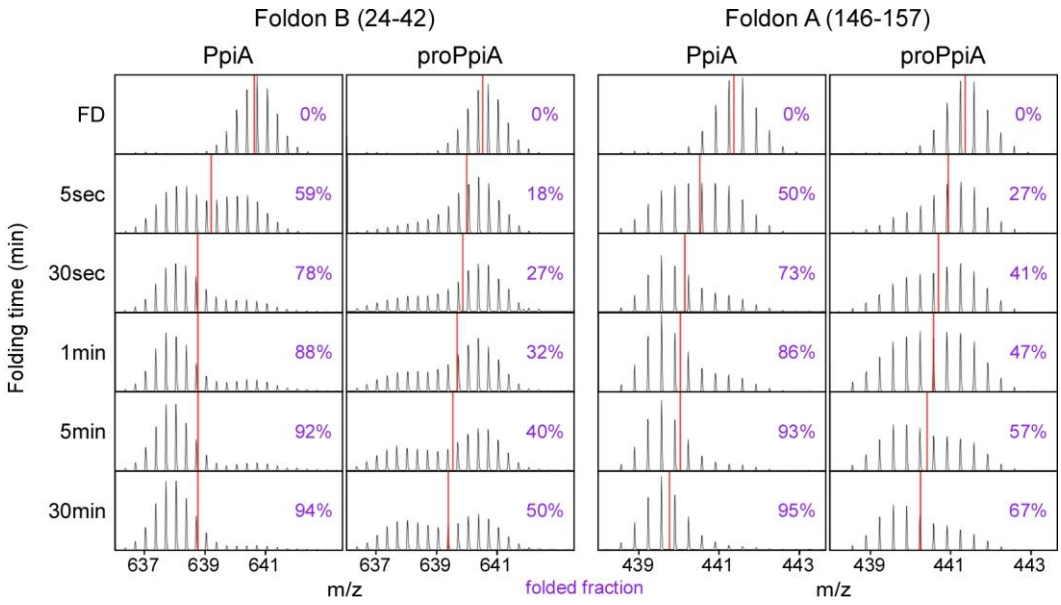


Appendix Figure S3 Refolding kinetics of (pro)PpiA and (pro)PpiB at 25°C analyzed with local HDX-MS (Related to Figure 5)

Folding kinetics of the mature domains of proPpiA (**A**, **C**) and proPpiB, carrying the proPpiA signal peptide, (**B**, **D**) at 25°C (as indicated), monitored by local HDX-MS.

A, B. The HDX-MS refolding kinetics data for (pro)PpiA and (pro)PpiB, at 25°C (Dataset EV4; *n=3* biological repeats) were further analyzed by PyHDX in order to determine the folded fractions per residue (Dataset EV5). The pipeline of analysis is shown in Fig.EV3B. Folding, displayed per residue (x-axis) over time (y-axis) in a colour map in steps of 25% folded fraction (as indicated at the bottom), is shown for up to 30 min (complete data set in Dataset EV5). The alignment index (x-axis) is based on the proPpiA sequence (signal peptide, extended N-tail; missing loop between β6-β7; Appendix Fig. S1D). For each peptide, 100% folding was set to the D-uptake of the corresponding native mature domain peptide. Grey areas: residues absent in one of the twins, prolines or indicating no experimental coverage. Colour-boxes below the linear secondary structure map (top) indicate foldons; named in alphabetical order and subscript numbers (if formed gradually, step-wise) following the order of formation sequence. Grey bar: unstructured fast folding regions (Fig. EV5D) omitted from analysis.

C, D. Foldons, colour-coded as in the left panels, are indicated relative to formation time and temperature on the PpiB (1LOP) and PpiA (1V9T) 3D-structures. The indicated time points are: for proPpiA ($t_{50\%}$ of 0.9-2.0-2.3-20.8 min) and for proPpiB ($t_{50\%}$ of 0.06-0.08-0.44-1.2 min), both at 25°C (Dataset EV5).



Appendix Figure S4 Effect of signal peptide on the initial foldons of PpiA (Related to Figure 5)

Peptides (proPpiA numbering followed) spanning two initial foldons, B (left) and A (right), were selected as examples to demonstrate the effect of the signal peptide on the folding of PpiA. Spectra at selected timepoints are displayed with the corresponding centroid indicated. The unfolded (0%) and native (100%) state (HDX data in Dataset EV4, *n=3* biological repeats) were used to calculate the folded fractions in PyHDX (purple; Dataset EV5). The pipeline of analysis is shown in Fig. EV3B.

Supplemental tables:

Appendix Table S1 Plasmids

Vector	Antibiotic resistance	promoter	Origin of replication	Reference/Source
pET22b	Ampicillin	T7(lac)	pBR322	Novagen (https://www.merckmillipore.com/)
pBAD501	Gentamycin	ara	p15A/pACYC	pBAD33proKLPhoA/Gem^R [9]
pET610	Ampicillin	Trc (Trp-lac)	pBR322	Driessen et al. [10]

Appendix Table S2 Primers

Primer	Forward/ Reverse	Gene	Restriction site or mutation inserted	Sequence (5'-3') (mutated codons are **bold,** restriction sites/mutations underlined)
X850	F	*ppiB*	NdeI	5'GGGAATTC**CATATG**GTTACTTTCCACACCAATCACGGC3'
X851	R	*ppiB*	XhoI	5'GACCCG**CTCGAG**CTCGCTAACGGTCACGCTTTCAATGAT3'
X743	F	*ppiA*	NdeI	5'GGGAATTC**CATATG**GCAGCGAAAGGGGACCCG3'
X1282	R	*ppiA*	HindIII	5′CCC**AAGCTT**CGGCAGGACTTTAGCGGAAAGGATAA3′
X1928	R	*ppiB*	HindIII	5'CCC**AAGCTT**CTCGCTAACGGTCACGCTTTCAATGAT3'
X2396	F	*ppiB*	I13L	5′CACGGCGATATTGTC**CTG**AAAACTTTTGACGAT3′
X2397	R	*ppiB*	I13L	5′ATCGTCAAAAGTTTT**CAG**GACAATATCGCCGTG3′
X2398	F	*ppiB*	L83I	5′AATACCCGTGGTACG**ATC**GCAATGGCACGTACT3′
X2399	R	*ppiB*	L83I	5′AGTACGTGCCATTGC**GAT**CGTACCACGGGTATT3′
X2400	F	*ppiB*	V160A	5′GTTATCATTGAAAGC**GCT**ACCGTTAGCGAGCTC3′
X2401	R	*ppiB*	V160A	5′GAGCTCGCTAACGGT**AGC**GCTTTCAATGATAAC3′
X2346	F	*ppiA*$_{>B,\ 6plet1}$	NdeI	5′CTTTAAGAAGGAGATATA**CATATG**GCGGCGAAAGGGGAC3′
X2428	R	*ppiA*$_{>B,\ 6plet1}$	HindIII	5′GAACAGGCATTTCTGGTGT**AAGCTT**CGGCAGGACTTTAGC3′
X2348	F	*ppiB*$_{>A,\ 6plet1}$	NdeI	5′CTTTAAGAAGGAGATATA**CATATG**GTTACTTTACACACCAATC3′
X2429	R	*ppiB*$_{>A,\ 6plet1}$	HindIII	5′GAACAGGCATTTCTGGTGT**AAGCTT**CTCGCTAACGGTAGC3′

Appendix Table S3 Strains

***E. coli* strain**	**Description (gene deleted)**	**Reference/source**
DH5a	*F– Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1*	Invitrogen
Lemo21(DE3)	T7 RNA polymerase gene under the control of the lacUV5 promoter.	New England BioLabs
BL21.19(DE3)	*secA13 (Am) supF (Ts) trp (Am) zch::Tn10 recA::cat clpA::kan)*	[5]
MC4100	*F-araD139 ¢(argF-lac)U169 rpsL150 (StrR) relA1 flbB5301 deoC1 pstF25 rbsR*	P. Genevaux [6-8]

Appendix Table S4 Cloned genes

Gene	**Uniprot accession number**	**Plasmid name**	**Vector**	**Description/reference**
proppiA	P0AFL3	pIMBB1042	pET22b	[11]
ppiA	P0AFL3	pIMBB1043	pET22b	[11]
ppiB	P23869	pIMBB1085	pET22b	[12]
proppiB		pLMB2094	pET22b	Addition of the PpiA signal peptide to the PpiB mature domain containing the N-terminal PpiA tail to avoid cleavage that is seen when the SP is directly attached to PpiB.
Grafted folding mutants (predicted from EFoldMine)				
Gene	**Construct**	**Plasmid name**	**Vector**	**Source**	**Description**
ppiA$_{>B,\ EFoldMine,4plet}$	sgLMB0075	pLMB2006	pET22b	Cloned synthetic gene (Genscript)	PpiA>B Quatdruplet EFoldMine-predicted mutant (E17V/A123C/G126A/A162V)
ppiA$_{>B,\ EFoldMine,\ Singlet}$	sgLMB0072	pLMB2003	pET22b	"	PpiB>A Singlet EFoldmine-predicted mutant (C121A)
ppiA$_{>B,\ EFoldMine,4plet}$	sgLMB0073	pLMB2004	pET22b	"	PpiB>A Quadruplet EFoldMine-predicted mutant (V12E/C121A/A124G/V160A)
ppiA$_{>B,\ EFoldMine,\ Multiplet}$	sgLMB0090	pLMB2021	pET22b	"	PpiB>A Multiplet EFoldMine-predicted mutant (H8A/V12E/D18Q/L28V/E33S/I40T/E66P/V103A/C121A/A124G/D128K/V133A)
Grafted folding mutants (derived from Native contacts)					
ppiA$_{>B,\ Singlet1}$	sgLMB0093	pLMB2083	pET22b	"	PpiA>B Singlet1 (L18I)
ppiA$_{>B,\ Doublet1}$	sgLMB0076	pLMB2007	pET22b	"	PpiA>B Doublet1 (E17V/L18I)

ppiA$_{>B,\ 3plet1}$	sgLMB0077	pLMB2008	pET22b	"	PpiA>B 3plet1 (E17V/L18I/G126A)
ppiA$_{>B,\ Singlet2}$	sgLMB0078	pLMB2009	pET22b	"	PpiA>B Singlet2 (L9F)
ppiA$_{>B,\ Singlet3}$	sgLMB0091	pLMB2022	pET22b	"	PpiA>B Singlet3 (V33L)
ppiA$_{>B,\ Singlet4}$	sgLMB0092	pLMB2023	pET22b	"	PpiA>B Singlet4 (A135V)
ppiA$_{>B,\ Doublet2}$	sgLMB0079	pLMB2010	pET22b	"	PpiA>B Doublet2 (V33L/A135V)
ppiA$_{>B,\ 3plet2}$	sgLMB0080	pLMB2011	pET22b	"	PpiA>B 3plet2 (L9F/V33L/A135V)
ppiA$_{>B,\ 3plet3}$	sgLMB0094	pLMB2084	pET22b	"	PpiA>B 3plet3 (L18I/I88L/A162V)
ppiA$_{>B,\ 6plet2}$	sgLMB0095	pLMB2085	pET22b	"	PpiA>B 6plet1 (L9F/L18I/V33L/I88L/A135V/A162V)
ppiA$_{>B,\ 6plet1}$	sgLMB0081	pLMB2012	pET22b	"	PpiA>B 6plet2 (L9F/E17V/L18I/V33L/G126A/A135V)
ppiA$_{>B,\ control}$	sgLMB0106	pLMB2096	pET22b	"	PpiA>B Negative Control (S28T/S101A/N151D)
ppiB$_{>A,\ 3plet1}$	PpiB>A (3plet 1)	pLMB2169	pET22b	Quick Change Mutagenesis	PpiB>A 3plet1 (I13L/L83I/V160A)
ppiB$_{>A,\ 3plet2}$	sgLMB0096	pLMB2086	pET22b	Cloned synthetic gene (Genscript)	PpiB>A 3plet2 (F4L/L28V/V133A)
ppiB$_{>A,\ 6plet1}$	sgLMB0097	pLMB2087	pET22b	"	PpiB>A 6plet1 (F4L/I13L/L28V/L83I/V133A/V160A)
ppiB$_{>A,\ 6plet2}$	sgLMB0089	pLMB2020	pET22b	"	PpiB>A 6plet2 (F4L/V12E/I13L/L28V/C121A/A124G/D128K/V133A)
ppiB$_{>A,\ Multiplet}$	sgLMB0098	pLMB2088	pET22b	"	PpiB>A Multiplet (T3L/F4L/H8A/I13L/T15L/L28V/C31V/L83I/C121A/V133A/V160A)
ppiB$_{>A,\ control}$	sgLMB0107	pLMB2097	pET22b	"	PpiB>A Negative control (T23S/A96S/D149N)
proppiB	sgLMB0104	pLMB2094	pET22b	"	Addition of the PpiA signal peptide to the PpiB mature domain containing the N-terminal PpiA tail (AKGDPH) to avoid cleavage that is seen when the signal peptide is directly attached to PpiB.
Constructs for *in vivo* secretion					
Gene	**Plasmid name**	**Vector**	**Description/reference**		
Secreted proteins					

ppiB phoA	pIMBB1571	pBAD501	The *ppiB* gene (495bp) was isolated by PCR from pIMBB1043 using primers X850 (Forw NdeI) and X1928 (Rev HindIII) and was cloned in the NdeI-HindIII sites of pIMBB1570 (pBAD501 pro(KL)PhoA), substituting the proPhoA signal peptide.
ppiA phoA	pIMBB1584	pBAD501	The *ppiA* gene (510bp) was isolated by PCR from pIMBB1085 using primers X743 (Forw NdeI) and X1282 (Rev HindIII) and was cloned in the NdeI-HindIII sites of pIMBB1570 (pBAD501 pro(KL)PhoA), substituting the proPhoA signal peptide.
ppiA$_{>B,\ 6plet1}$ *phoA*	pLMB2208	pBAD501	The *ppiA>B (S1)* gene (510 bp) was isolated by PCR from sgLMB0081 using primers X2346 (Forw NdeI) and X2428 (Rev HindIII) and was cloned in the NdeI-HindIII sites of pIMBB1570 (pBAD501 pro(KL)PhoA), substituting the proPhoA signal peptide.
ppiB$_{>A,\ 6plet1}$ *phoA*	pLMB2209	pBAD501	The *ppiB>A(S1)* gene (495 bp) was PCR isolated from DH5a using primers X2348 (Forw. NdeI) and X2429 (Rev. HindIII) and was cloned to the NdeI-HindIII site of pIMBB1570 (pBAD33proKLPhoAGemR), substituting the PhoA signal peptide.
Sec Translocase			
hissecY$_{prlA4(I408N/F286Y)}$*-EG*	pIMBB842	pET610	[11]

Appendix Table S5 Buffer list

Buffer S-A	50 mM Tris-HCl pH 8.0, 1 M NaCl, 5 mM Imidazole, 5% glycerol v/v
Buffer S-B	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 5 mM Imidazole, 5% glycerol v/v
Buffer S-C	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 5% glycerol v/v
Buffer S-D	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 50% glycerol v/v
Buffer U-A	50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 5 mM Imidazole, 5% glycerol v/v
Buffer U-B	50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 5 mM Imidazole, 5% glycerol v/v; 8M Urea
Buffer U-C	50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 5 mM Imidazole, 5% glycerol v/v; 6M Urea
Buffer U-D	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 5 mM Imidazole, 5% glycerol v/v; 6M Urea
Buffer U-E	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 100 mM Imidazole, 5% glycerol v/v; 6M Urea
Buffer U-F	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 5% glycerol v/v; 6M Urea
Buffer U-G	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 10% glycerol v/v; 6M Urea
Buffer A	5 mM MOPS pH 8.0; 5 mM NaCl
Buffer B	25 mM Tris-HCl pH 8.0, 25 mM KCl
Buffer C	25 mM Tris-HCl pH 8.0, 25 mM KCl, 8M Urea

References

1. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol Syst Biol, 2011. **7**: p. 539.
2. Kallen, J. and M.D. Walkinshaw, *The X-ray structure of a tetrapeptide bound to the active site of human cyclophilin A.* FEBS Lett, 1992. **300**(3): p. 286-90.
3. Ashkenazy, H., et al., *ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules.* Nucleic Acids Res, 2016. **44**(W1): p. W344-50.
4. Landau, M., et al., *ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W299-302.
5. Mitchell, C. and D. Oliver, *Two distinct ATP-binding domains are needed to promote protein export by Escherichia coli SecA ATPase.* Mol Microbiol, 1993. **10**(3): p. 483-97.
6. Casadaban, M.J., *Transposition and fusion of the lac genes to selected promoters in Escherichia coli using bacteriophage lambda and Mu.* J Mol Biol, 1976. **104**(3): p. 541-55.
7. Genevaux, P., et al., *Scanning mutagenesis identifies amino acid residues essential for the in vivo activity of the Escherichia coli DnaJ (Hsp40) J-domain.* Genetics, 2002. **162**(3): p. 1045-53.
8. Ullers, R.S., et al., *Trigger Factor can antagonize both SecB and DnaK/DnaJ chaperone functions in Escherichia coli.* Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3101-6.
9. Guzman, L.M., et al., *Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.* J Bacteriol, 1995. **177**(14): p. 4121-30.
10. van der Does, C., et al., *SecA is an intrinsic subunit of the Escherichia coli preprotein translocase and exposes its carboxyl terminus to the periplasm.* Mol Microbiol, 1996. **22**(4): p. 619-29.
11. Gouridis, G., et al., *Signal peptides are allosteric activators of the protein translocase.* Nature, 2009. **462**(7271): p. 363-7.
12. Tsirigotaki, A., et al., *Long-Lived Folding Intermediates Predominate the Targeting-Competent Secretome.* Structure, 2018.