Residue–dependent thermodynamic cost and barrel plasticity balances activity in the PhoPQ–activated enzyme PagP of *Salmonella typhimurium*

*Bharat Ramasubramanian Iyer and Radhakrishnan Mahalakshmi**

Molecular Biophysics Laboratory, Department of Biological Sciences, Indian Institute of Science Education and Research, Bhopal – 462023. India.

*E-mail: maha@iiserb.ac.in

Supporting Information

Supporting Information Figures

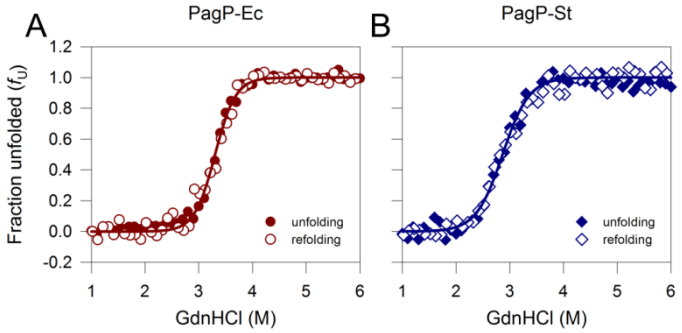

Fig. S1. Two-state (un)folding of PagP in DPC shows thermodynamic reversibility. Representative unfolding (filled symbols) and refolding (open symbols) profiles monitored at the λ_{em} of 342 nm, corresponding to the $\lambda_{em\text{-}max}$ of the folded protein, and normalized against pre- and post-transition Trp fluorescence intensities for (A) PagP-Ec (red circles) and (B) PagP-St (blue diamonds). Solid lines are the fits of the unfolding data.

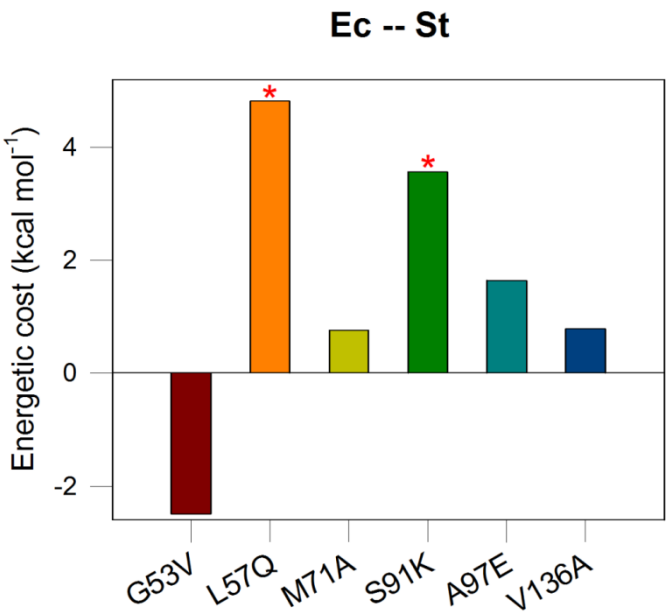

Fig. S2. Energetic cost of mutation. Energetic cost of substitution at the six positional differences in the transmembrane region between PagP-Ec and PagP-St was calculated from a previously reported scale for water-to-bilayer transfer free energy for each amino residue.[1] Labels for the abscissa follow: PagP-Ec residue, residue number, PagP-St residue. Energetic costs are greatest for mutations at sites 57 and 71 (indicated with *).

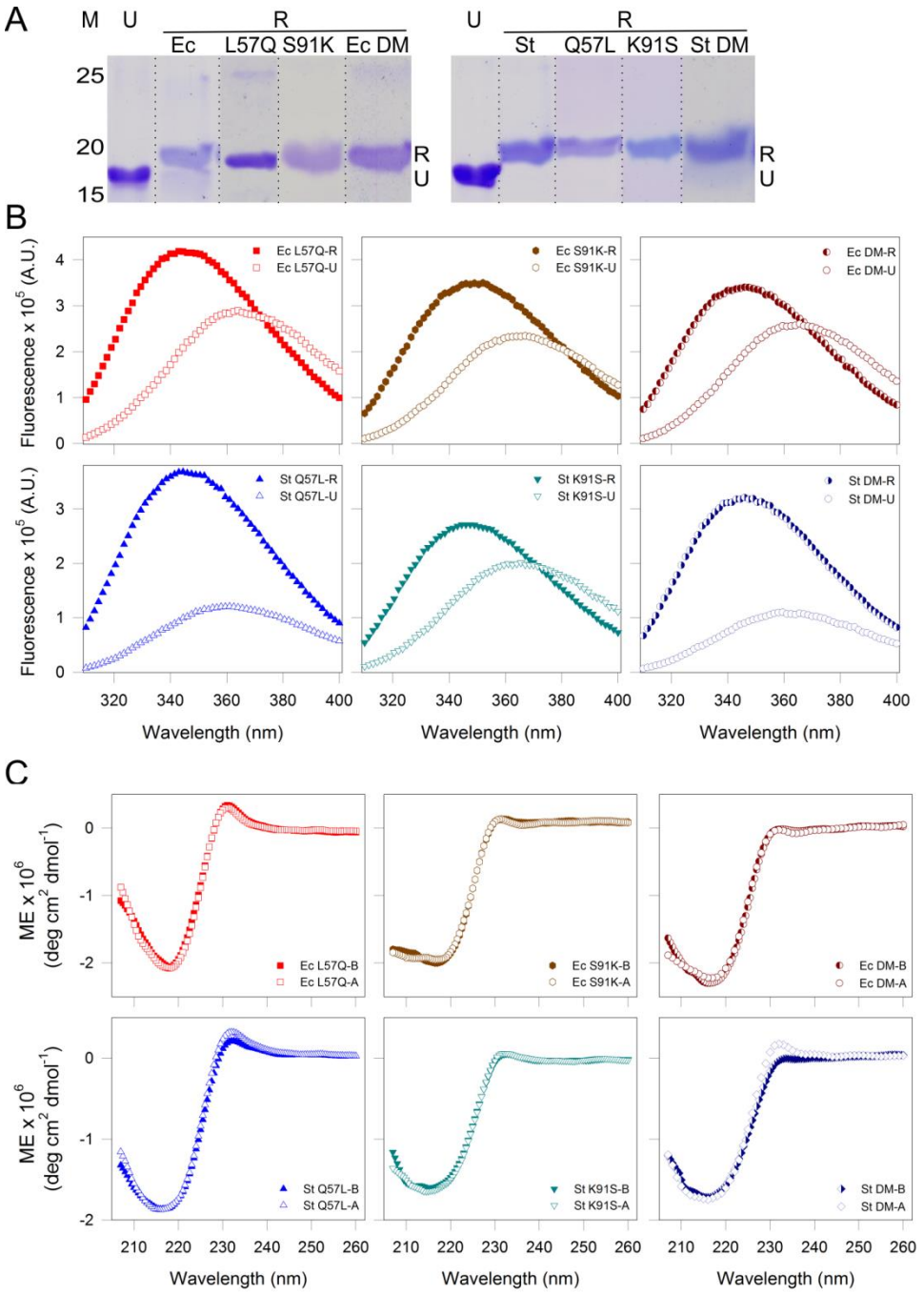

Fig. S3. Folding of PagP mutants monitored by PAGE and spectroscopy. (A) SDS-PAGE gels showing inverse mobility shift for all mutants characterized in this study. The samples were not boiled in gel loading dye, as reported earlier.[2] Gels on the left correspond to PagP-Ec and its mutants, while the gels on the right correspond to PagP-St and its mutants. M: molecular weight marker positions (in kDa) are indicated; U: PagP (PagP-Ec or PagP-St) unfolded using 8.0 M urea; R: PagP refolded in 10 mM DPC. Dotted lines separate different gel images that have been presented together. (B) Representative tryptophan emission spectra of PagP-Ec variants (top panels) and PagP-St variants (bottom panels) refolded (R, filled symbols) in 10 mM DPC or

unfolded (U, open symbols) using 8.0 M GdnHCl. (C) Representative far-UV CD spectra of PagP-Ec variants (top panels) and PagP-St variants (bottom panels) refolded in 10 mM DPC measured before (B, filled symbols) or after (A, open symbols) recovery from thermal denaturation. ME: molar ellipticity. The color and symbol scheme used in Fig. 4 of the main text is retained in panels B and C. PagP-Ec:L57Q (Ec L57Q): red, squares; PagP-St:Q57L (St Q57L): blue, upward triangles; PagP-Ec:S91K (Ec S91K): brown, hexagons; PagP-St:K91S (St K91S): dark cyan, downward triangles; PagP-Ec:DM (Ec DM): dark red, semi-filled circles; PagP-St:DM (St DM): dark blue, semi-filled diamonds.

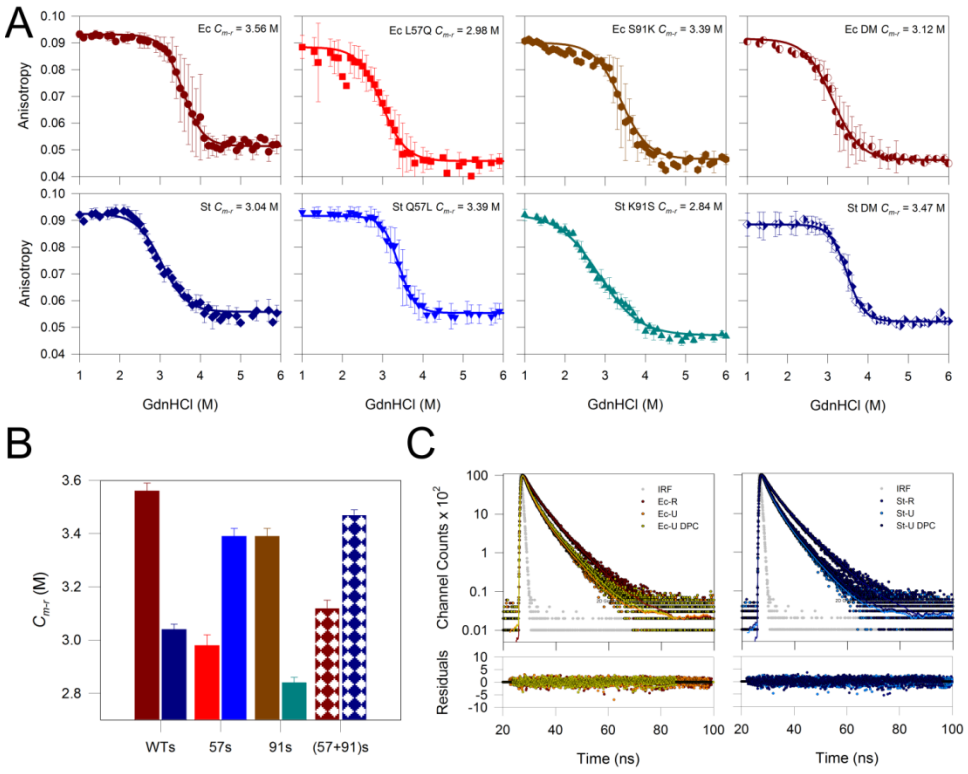

Fig. S4. Spectroscopic parameters compared across various PagP mutants. (A) Change in tryptophan fluorescence anisotropy measured across a GdnHCl gradient for the proteins described in this study. Data points have been averaged over a minimum of 3 independent experiments and fitted to a sigmoidal function (fits are shown as solid lines) to derive the chemical denaturation midpoint from changes in anisotropy ($C_{m\text{-}r}$). (B) Comparison of $C_{m\text{-}r}$ across all PagP variants shows a similar profile as the C_m summarized in Table 1. Color/symbol scheme: PagP-Ec (Ec): dark red, circles; PagP-St (St): dark blue, diamonds; PagP-Ec:L57Q (L57Q): red, squares; PagP-St:Q57L (Q57L): blue, upward triangles; PagP-Ec:S91K (S91K): brown, hexagons; PagP-St:K91S (K91S): dark cyan, downward triangles; PagP-Ec:DM (DM): dark red, semi-filled circles; PagP-St:DM (DM): dark blue, semi-filled diamonds. (C) Representative decay profiles for tryptophan fluorescence lifetime of PagP-Ec (left panel) and PagP-St (right panel) highlighting appreciable differences between the tryptophan behavior in the folded and the unfolded forms of both proteins.

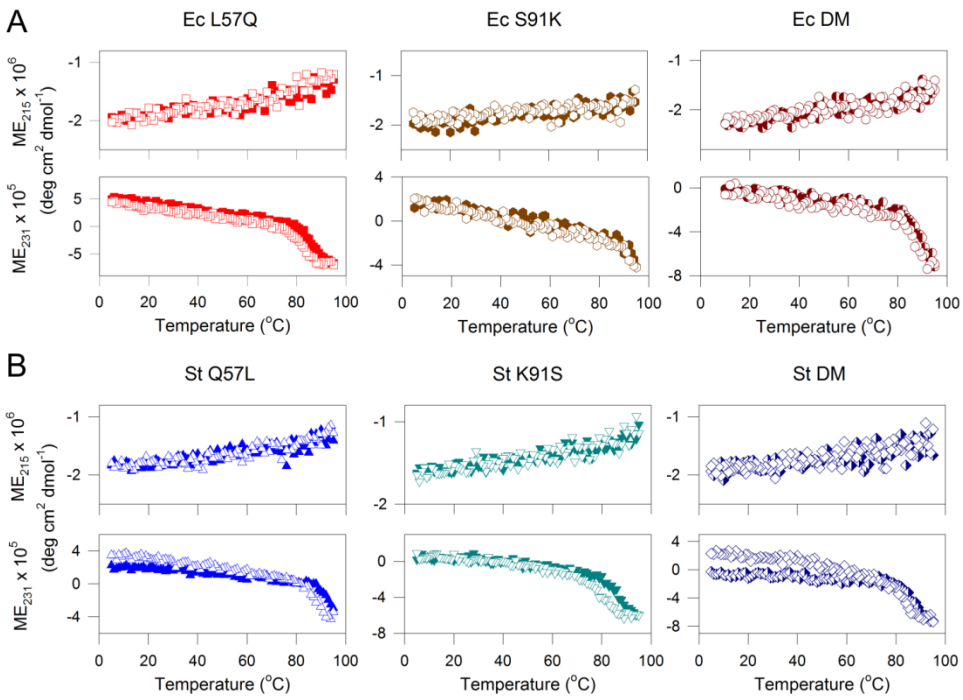

Fig. S5. Thermal denaturation of PagP mutants. (Continued from Fig. 6 of the main text). Representative thermal denaturation profiles for (A) PagP-Ec mutants and (B) PagP-St mutants monitored using far-UV CD from 5 °C to 95 °C (filled symbols) and 95 °C to 5 °C (open symbols). ME_{215} (change in molar ellipticity at 215 nm) profiles show ~30-40% loss in secondary structure, which is recovered at the end of the thermal denaturation and recovery cycles. ME_{231} profile, on the other hand, shows initiation of unfolding beyond ~80 °C for all mutants. There is also a clear incidence of hysteresis between the unfolding and recovery profiles at 231 nm. The color scheme used in Fig. S4 for the mutants has been retained in this figure.

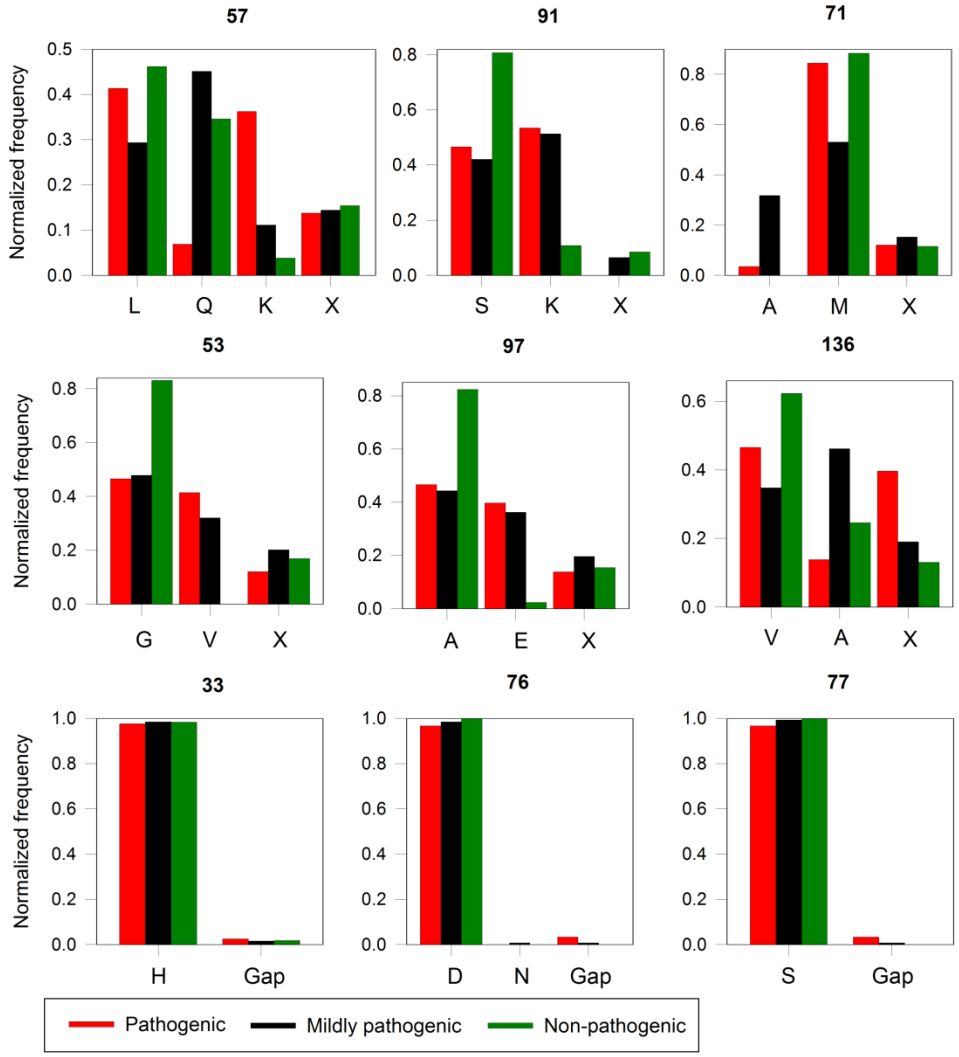

Fig. S6. Residue distribution for PagP from pathogenic and non-pathogenic bacteria. Frequency of occurrence of specific residues across various positions are normalized across bacteria that are pathogenic (red), mildly pathogenic (black) and non-pathogenic (green) for humans. Overall, our database analysis suggests that a preference for residue types at select positions may exist in pathogenic and mildly pathogenic *versus* non-pathogenic bacteria. However, it is not known whether these residue preferences have any bearing on the pathogenicity of the bacterium. Further, this analysis includes all known PagP sequences, irrespective of whether they are functional or inactive in the source organism, or whether the gene is dispensable for bacterial survival.

Top panel: At position 57 (left), Lys is generally observed in pathogenic bacteria, whereas Gln, which is observed primarily in non-pathogenic species, also occurs in *S. typhimurium*. At position 91 (middle), Lys is predominant in pathogens, including *S. typhimurium*. This Lys is present in PagP from other organisms including *Legionella pneumophila*, suggesting that it may have been retained by pathogenic bacteria by selection. Note, however, that there are several exceptions in each case. Residues with low frequency of occurrence have been grouped under "X" in both graphs. Position 71, one of the six specific differences noted in Fig. 1 of the main

text (and Fig. S2), does not show any variation across species. Met is preferred in this position and is equally abundant in pathogenic as well as non-pathogenic bacteria. Other residues, including Ala, are less preferred.

Middle panel: In the positions 53, 97 and 136, preponderance of selected residue(s) is observed only for the non-pathogenic variants, whereas mildly pathogenic and pathogenic bacteria do not exhibit an overwhelming preference for a particular residue class. For instance, at position 53, Gly is preferred over Val in non-pathogenic strains. But, the reverse trend is not observed, since, in pathogenic strains, there is a near-equal occurrence of Gly and Val at this position. A similar argument can be extended for residue preferences at position 97, with Ala being preferred among non-pathogens, but equal distribution of Ala and Glu is obtained among pathogens. The case of position 136 is similar to 71. Although Val is slightly preferred among non-pathogens, overall, this position does not show considerable variation across species, with residues exhibiting equal abundance in pathogenic and non-pathogenic strains.

Lower panel: The conserved H33 (right), D76 (middle) and S77 (right) belonging to the catalytic triad are retained across all bacteria or is missing (denoted by ‘Gap’).

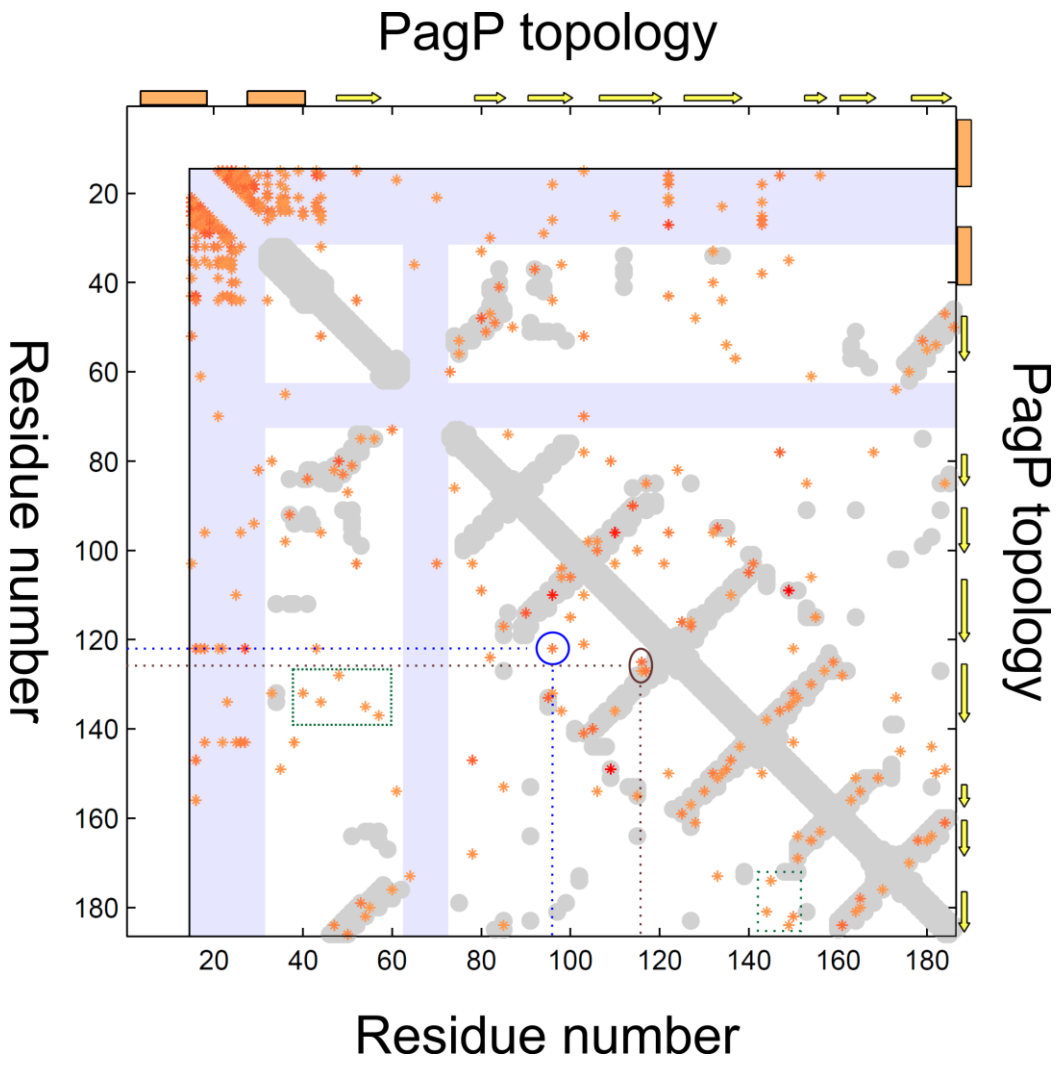

Fig. S7. Covariance analysis of PagP. Contact map representation of evolutionary couplings (ECs) derived from EVfold analysis[3, 4] of PagP Ec (Uniprot ID P37001), overlaid on known residue contacts (grey filled circles shown in the background) from the crystal structure (PDB ID 1THQ). The residue numbers are based on the primary sequence of PagP Ec (Uniprot ID P37001), which includes the signal sequence. Regions excluded from the plot or shaded in purple correspond to the segments that are absent in the 1THQ crystal structure; this includes the signal sequence, a part of the N-terminal helix and a loop region. The top 200 predicted ECs are plotted (✳) in shades of red and orange, to depict couplings with high and moderate EC-scores, respectively. The lines of ECs (in grey) running perpendicular to the diagonal, and overlapping with known residue contacts, are indicative of the beta-strand regions of the protein. We also observe certain scattered ECs (dotted green boxes), which may be of importance from the aspect of protein sequence coevolution. Several of the top-ranked evolutionarily constrained residues of PagP (Table S3) were present in the N-terminal helical region, the role of which has been extensively characterized previously.[5] We have looked at the role of the N-terminal helix in the stability and activity of both PagP-Ec and PagP-St (unpublished results), and we are able to

comment that despite having a definitive role in the stabilities of both proteins, the N-terminal helix does not significantly alter the catalytic properties of these two homologous barrels. The six positional differences in the transmembrane region that we have mapped between PagP-Ec and PagP-St are distributed across the top 40% of residue positions. Particularly, Ser91 (residue # 116 above), which shows a cumulative EC strength of 0.608 (Table S3), is predicted to be evolutionarily coupled with several other residues (brown circle and dotted drop lines). Our experimental findings on the 91st positional mutants of PagP-Ec and PagP-St, coupled with the bioinformatics analysis using sequences from the PFAM database (Fig. S6), point towards an important role for the chemical nature of the residue present in this position. Apart from this, Ala71 and Ala97 of PagP-Ec (residue # 96 and 122 in the above figure, respectively; blue circle and dotted drop lines) also show very high evolutionary coupling strengths, as well as a strong evolutionary correlation between them. Mutation of Ala71 on a background double mutant does not significantly affect the activity or the stability of PagP-Ec (unpublished results), whereas an important catalytic role for Ala97 is highly implausible, considering the fact that it is located on the face of the protein diametrically opposite to the active site as well as that it is positioned considerably away from the entry and exit points for the donor and the acceptor, respectively.

Supporting Information Tables

Table S1. Input molecules for the simulations.

System	Protein structure	No. of detergent molecules	No. of water molecules	No. of ions	Total no. of atoms
Ec-DPC	1THQ[6]	80	14302	3 K^+	50141
St-DPC	1THQ[#]	80	14480	3 K^+	50679

[#] PagP-St structure was generated by computationally mutating the positional differences of PagP-Ec (Ec) and PagP-St (St) using CHARMM-GUI input generator.[7-9]

Table S2. Spectroscopic properties of Trp in PagP.

Protein[#]	<τ> (ns)	χ^2	r	K_{SV} (M^{-1})	k_q (M^{-1} ns^{-1})
Ec-R	3.39	1.12	0.093	4.25	1.25
Ec-U	2.20	1.16	0.058	11.27	5.12
Ec-U DPC	2.39	1.09	0.052	-	-
St-R	3.62	1.13	0.092	4.61	1.27
St-U	2.23	1.18	0.060	12.20	5.47
St-U DPC	2.43	1.09	0.054	-	-

[#]Ec: PagP-Ec; St: PagP-St; R: Refolded; U: Unfolded.

Table S3. PagP evolutionary couplings (EC) analysis residue summary table.

Residue index[#]	Amino acid	Number of ECs	Cumulative strength[@]	EC strength	Conser-vation
2	A	11	2.528	12.3	16
11	E	12	2.205	10.8	33
97	A	10	2.16	10.5	29
1	N	10	1.997	9.7	19
3	D	9	1.856	9.1	19
10	R	10	1.826	8.9	26
4	E	7	1.764	8.6	31
19	Q	9	1.702	8.3	35
78	W	8	1.515	7.4	34
7	T	7	1.491	7.3	18
71	A	6	1.368	6.7	50
18	Q	6	1.329	6.5	26
118	N	7	1.326	6.5	50
85	A	5	1.163	5.7	38
5	W	6	1.146	5.6	51
124	V	4	1.107	5.4	47
125	L	5	0.923	4.5	38
73	A	5	0.899	4.4	80
159	F	4	0.811	4	36
84	I	2	0.775	3.8	74
122	L	3	0.761	3.7	38
100	N	3	0.717	3.5	45
55	F	3	0.712	3.5	45
14	A	4	0.66	3.2	40
27	I	3	0.653	3.2	33
108	T	3	0.648	3.2	86
53	G	3	0.647	3.2	38
13	I	3	0.618	3	46
91	S	3	0.608	3	62
140	M	3	0.603	2.9	47
8	T	3	0.597	2.9	42
15	Q	3	0.581	2.8	50
81	W	3	0.579	2.8	71
6	M	3	0.568	2.8	42
107	F	3	0.56	2.7	46
57	L	3	0.555	2.7	24
111	V	3	0.553	2.7	59

102	H	3	0.551	2.7	40
129	A	3	0.54	2.6	68
22	H	3	0.537	2.6	40
126	L	3	0.524	2.6	92
60	W	3	0.518	2.5	40
23	Y	2	0.493	2.4	53
109	A	3	0.492	2.4	69
136	V	2	0.471	2.3	36
92	T	2	0.441	2.1	48
12	N	2	0.426	2.1	63
75	K	2	0.418	2	74
113	A	2	0.394	1.9	75
28	P	2	0.393	1.9	83
25	L	2	0.383	1.9	78
45	R	2	0.378	1.8	40
35	R	2	0.376	1.8	93
157	M	2	0.365	1.8	36
90	E	2	0.359	1.8	70
110	G	2	0.354	1.7	73
103	L	2	0.342	1.7	62
139	Q	2	0.339	1.7	80
156	W	2	0.334	1.6	82
89	W	1	0.331	1.6	60
65	N	1	0.331	1.6	72
131	V	2	0.328	1.6	42
29	A	2	0.326	1.6	53
50	P	2	0.326	1.6	82
119	Y	2	0.32	1.6	86
148	N	2	0.32	1.6	82
70	Y	1	0.313	1.5	90
151	N	2	0.313	1.5	93
115	D	1	0.293	1.4	77
80	K	1	0.293	1.4	68
59	R	1	0.252	1.2	78
16	T	1	0.252	1.2	77
67	H	1	0.251	1.2	69
116	N	1	0.246	1.2	59
153	Y	1	0.244	1.2	45
154	F	1	0.23	1.1	89
58	S	1	0.227	1.1	77
24	D	1	0.227	1.1	78

48	E	1	0.22	1.1	93
134	G	1	0.217	1.1	48
56	G	1	0.212	1	95
26	Y	1	0.212	1	93
161	F	1	0.204	1	81
30	I	1	0.203	1	59
20	P	1	0.203	1	62
105	L	1	0.193	0.9	69
63	K	1	0.19	0.9	51
79	N	1	0.187	0.9	79
99	E	1	0.186	0.9	28
144	P	1	0.185	0.9	98
130	S	1	0.183	0.9	90
62	E	1	0.179	0.9	61
96	L	1	0.175	0.9	53
143	I	1	0.172	0.8	88
69	L	1	0.168	0.8	76
155	A	1	0.168	0.8	71
128	L	1	0.164	0.8	82
31	T	1	0.163	0.8	72
149	N	1	0.16	0.8	89
120	I	1	0.16	0.8	72
39	D	1	0.16	0.8	78
138	F	1	0.159	0.8	59
112	T	1	0.157	0.8	93
32	W	1	0.157	0.8	91
145	G	1	0.157	0.8	90
34	A	1	0.157	0.8	47
61	D	1	0.154	0.8	85
49	R	1	0.154	0.8	65
40	K	1	0.154	0.7	61
36	F	1	0.153	0.7	54
132	G	1	0.153	0.7	71
9	F	1	0.153	0.7	47

[#]The six positional differences in the transmembrane region, mapped between PagP-Ec and PagP-St in this study, are highlighted in green. Residues corresponding to the signal sequence have been removed. [@] Average cumulative strength = 0.205.

References:

[1] Moon, C. P., and Fleming, K. G. (2011) Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers, *Proc. Natl. Acad. Sci. U.S.A. 108*, 10174-10177.

[2] Maurya, S. R., Chaturvedi, D., and Mahalakshmi, R. (2013) Modulating lipid dynamics and membrane fluidity to drive rapid folding of a transmembrane barrel, *Sci. Rep. 3*, 1989.

[3] Marks, D. S., Hopf, T. A., and Sander, C. (2012) Protein structure prediction from sequence variation, *Nat. Biotechnol. 30*, 1072-1080.

[4] Sheridan, R., Fieldhouse, R. J., Hayat, S., Sun, Y., Antipin, Y., Yang, L., Hopf, T., Marks, D. S., and Sander, C. (2015) EVfold.org: Evolutionary Couplings and Protein 3D Structure Prediction, DOI: 10.1101/021022.

[5] Huysmans, G. H., Radford, S. E., Brockwell, D. J., and Baldwin, S. A. (2007) The N-terminal helix is a post-assembly clamp in the bacterial outer membrane protein PagP, *J. Mol. Biol. 373*, 529-540.

[6] Ahn, V. E., Lo, E. I., Engel, C. K., Chen, L., Hwang, P. M., Kay, L. E., Bishop, R. E., and Prive, G. G. (2004) A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin, *EMBO J. 23*, 2931-2941.

[7] Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM, *J. Comput. Chem. 29*, 1859-1865.

[8] Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program, *J. Comput. Chem. 30*, 1545-1614.

[9] Cheng, X., Jo, S., Lee, H. S., Klauda, J. B., and Im, W. (2013) CHARMM-GUI micelle builder for pure/mixed micelle and protein/micelle complex systems, *J. Chem. Inf. Model. 53*, 2171-2180.