Supplemental Information for:

Properties of unfolded states of proteins suggest broad selection for expanded conformational ensembles

Micayla A. Bowman[1], Joshua A. Riback[2][†], Anabel Rodriguez[1], Hongyu Guo[3], Jun Li[3], Tobin R. Sosnick*[4,5] & Patricia L. Clark*[1]

[1]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556
[2]Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL 60637
[3]Department of Applied and Computational Modeling & Statistics, University of Notre Dame, Notre Dame, IN 46556
[4]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637
[5]Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637

*Corresponding authors:
pclark1@nd.edu
574-631-8353
trsosnic@uchicago.edu
773-218-5950

Supplemental Tables

Table S1. Raw and standardized hydropathy scales used in this work.

Amino Acid	MJ (35)		HW (11)		MJHW Mean Z-score
	Raw Score†	Z-score	Raw Score†	Z-score	
A	0.0200	-0.281	0.5	0.152	-0.0645
C	0.960	0.586	1.0	0.418	0.502
D	-0.720	-0.963	-3.0	-1.71	-1.34
E	-0.740	-0.982	-3.0	-1.71	-1.35
F	2.22	1.75	2.5	1.22	1.48
G	-0.380	-0.650	0.0	-0.115	-0.382
H	0.000	-0.299	0.5	0.152	-0.0737
I	1.89	1.44	1.8	0.845	1.14
K	-1.01	-1.23	-3.0	-1.71	-1.47
L	2.29	1.81	1.8	0.845	1.33
M	1.36	0.955	1.3	0.578	0.767
N	-0.630	-0.880	-0.2	-0.221	-0.551
P	-0.470	-0.733	0.0	-0.115	-0.424
Q	-0.560	-0.816	-0.2	-0.221	-0.518
R	-0.440	-0.705	-3.0	-1.71	-1.21
S	-0.550	-0.806	-0.3	-0.274	-0.540
T	-0.250	-0.530	0.4	0.099	-0.216
V	1.34	0.937	1.5	0.685	0.811
W	1.28	0.881	3.4	1.70	1.29
Y	0.880	0.512	2.3	1.11	0.812

† Raw scores have been multiplied by -1, such that hydrophobic clustering creates positive peaks in hydropathy plots (**Figure 2A**).

Table S2. Other IDPs with collapse properties measured by SAXS.

Protein	Citation	v[†]	HpC	norm HpC[‡]	κ	FCR[§]
PDomain	(36)	0.360	4.45	0.0449	0.330	0.056
NUL	(25)	0.523	0.68	0.0064	0.094	0.193
Il-1	(37, 38)	0.525	0.00	0.0000	0.097	0.182
N49	(25)	0.537	0.30	0.0100	0.418	0.053
IBB	(25)	0.538	0.06	0.0007	0.241	0.320
Ki-1/57	(37, 39)	0.539	0.76	0.0027	0.191	0.377
Met2	(40)	0.542	6.39	0.0306	0.157	0.216
FhuA	(3)	0.543	5.00	0.0368	0.176	0.177
RNasA	(3)	0.545	1.97	0.0170	0.140	0.140
Msh6 NTR	(41)	0.547	2.77	0.0091	0.356	0.335
ERM (1-122)	(37, 42)	0.548	2.46	0.0216	0.220	0.279
redAFP	(43)	0.550	0.00	0.0000	0.278	0.086
MeCP2	(37, 44)	0.550	5.78	0.0121	0.203	0.331
PIR	(37, 45)	0.553	0.72	0.0107	0.353	0.187
N98	(25)	0.556	2.73	0.0191	0.363	0.020
CSD1	(37, 46)	0.565	1.05	0.0075	0.179	0.311
p53	(37, 47)	0.569	1.41	0.0166	0.202	0.204
NUS	(25)	0.578	0.55	0.0076	N/A	0.000
ID3 of CBP	(48)	0.586	3.78	0.0095	0.196	0.108
Ash1	(15)	0.603	0.04	0.0005	0.194	0.181
NSP	(25)	0.603	1.69	0.0101	0.375	0.017
NLS	(25)	0.603	0.00	0.0000	0.391	0.500
HMPV (P1-60)	(49)	0.603	2.92	0.0449	0.217	0.247
HrpO	(37, 50)	0.603	2.84	0.0203	0.153	0.358

FEZ1	(37, 51)	0.603	0.55	0.0058	0.243	0.417

[†] See (4) and (5); [‡]Normalized HpC; [§]Fraction of charged residues

Supplemental Figures

Figure S1. (A) Sequence alignment of wild type PNt and sequence swap variants. **(B)** Amino acid sequences of wild type PNt, sequence swap variants and PCt. For Swaps1-4.1, swapped residues are shown in bold type. For Swap5 and Swap6, charged residues are shown in blue (acidic) and red (basic).

(A)

(B)

>PNt
DWNNQSIVKTGERQHGIHIQGSDPGGVRTASGTTIKVSGRQAQGILLENPAAELQFRNGSVTSS
GQLSDDGIRRFLGTVTVKAGKLVADHATLANVGDTWDDDGIALYVAGEQAQASIADSTLQGAGG
VQIERGANVTVQRSAIVDGGLHIGALQSLQPEDLPPSRVVLRDTNVTAVPASGAPAAVSVLGAS
ELTLDGGHITGGRAAGVAAMQGAVVHLQRATIRRGEALAGGAVPGGAVPGGAVPGGFGPGGFGP
VLDGWYGVDVSGSSVELAQSIVEAPELGAAIRVGRGARVTVPGGSLSAPHGNVIETGGARRFAP
QAAPLSITLQAGAH

>Swap1
DWNNQSIVKTGERQHGIHIQGSDPGGVRTASGTTIKVSGRQAQGILLENPAAELQFRNGSVTSS
GQ**K**SDDGIRRFLGTVTV**L**AGKLVADHATLANVGDTWDDDGIALYVAGEQAQASIADSTLQGAGG
VQIERGANVTVQRSAIV**L**GGLHIGALQSLQPED**D**PPSRVVLRDTNVTAVPASGAPAAVSVLGAS
LLTLDGGHITGGRAAGVAAMQGAVVH**E**QRATIRRGEALAGGAVPGGAVPGGAVPGGFGPGGFGP
VLDGWYGVDVSGSSVELAQSIVEAPELGAAIRVGRGARVTVPGGSLSAPHGNVIETGGARRFAP
QAAPLSITLQAGAH

>Swap2
DWNNQSIVKTGERQHGIHIQGSDPGGVRTASGTTIKVSGRQAQGILLENPAAELQFRNGSVTSS
GQ**K**SDDG**K**RR**G**LGTVTV**LAFI**LVADHATLANVGDTWDDDGIALYVAGEQAQASIADSTLQGAGG
VQIERGANVTVQR**LAIV**LL**GLHIGA**S**QS**G**QPED**D**PPSRVVLRDTNVTAVPASGAPAAVSVLGA**M**
LLTL**V**GGHITGGRAAGVAA**S**QGAV**D**H**E**QRATIRRGEALAGGAVPGGAVPGGAVPGGFGPGGFGP
VLDGWYGVDVSGSSVELAQSIVEAPELGAAIRVGRGARVTVPGGSLSAPHGNVIETGGARRFAP
QAAPLSITLQAGAH

>Swap3
DWNNQSIVKTGERQHGIHIQGSDPGGVRTASGTTIKVSGRQAQGILLENPAAELQFRNGSVTSS
GQ**KST**DG**T**RRFLG**DVI**VKAG**LL**VADHATLANVGDTWDDDGIALYVAGEQAQASIADSTLQGAGG
VQIERGANV**DVL**R**L**AIVDGGLHIGALQS**QQ**PE**TS**PPSRVVLRDTNVTAVPASGAPAAVSV**Q**GAS
E**Q**TLDGG**A**ITGGRAAGVAAM**LGH**VVHL**L**RATIRRGEALAGGAVPGGAVPGGAVPGGFGPGGFGP
VLDGWYGVDVSGSSVELAQSIVEAPELGAAIRVGRGARVTVPGGSLSAPHGNVIETGGARRFAP
QAAPLSITLQAGAH

>Swap4
DWNNQSIVKTGERQHGIHIQGSDPGGVRTASGTTIKVSGRQAQGILLENPAAELQFRNGSVTSS
GQLS**FV**GI**T**R**D**LG**RD**TVKAGKLVADHATLANVGDTWDDDGIALYVAGEQAQASIADSTLQGAGG
VQIERGA**DVR**VQR**E**AIVDGGLH**N**GALQSLQP**SI**LPPS**T**VVLRDTNVTAVPASGAPAAV**LVS**GAS
GLRLDGGHI**HE**GRAAGVAAMQGAVV**T**LQ**T**ATIRRGEALAGGAVPGGAVPGGAVPGGFGPGGFGP
VLDGWYGVDVSGSSVELAQSIVEAPELGAAIRVGRGARVTVPGGSLSAPHGNVIETGGARRFAP
QAAPLSITLQAGAH

>Swap4.1
DWNNQSIVKTGERQHGIHIQGSDPGGVRTASGTTIKVSGRQAQGILLENPAAELQFRNGSVTSS
GQLS**FV**GI**T**R**R**LG**DD**TVKAGKLVADHATLANVGDTWDDDGIALYVAGEQAQASIADSTLQGAGG
VQIERGA**DVE**VQR**R**AIVDGGLH**N**GALQSLQP**SI**LPPS**T**VVLRDTNVTAVPASGAPAAV**LVS**GAS
GLELDGGHI**HR**GRAAGVAAMQGAVV**T**LQ**T**ATIRRGEALAGGAVPGGAVPGGAVPGGFGPGGFGP

VLDGWYGVDVSGSSVELAQSIVEAPELGAAIRVGRGARVTVPGGSLSAPHGNVIETGGARRFAP
QAAPLSITLQAGAH

>Swap5
DWNNQSIV**K**TG**ER**QHGIHIQGS**D**PGGV**R**TASGTTI**K**VSG**R**QAQGILL**E**NPAA**E**LQF**R**NGSVTSS
GQLS**DD**GI**ED**FLGTVTV**D**AG**E**LVA**D**HATLANVG**D**TW**DDD**GIALYVAG**E**QAQASIA**D**STLQGAGG
VQI**ED**GANVTVQ**E**SAIV**D**GGLHIGALQSLQP**RR**LPPS**R**VVL**RK**TNVTAVPASGAPAAVSVLGAS
KLTL**R**GGHITGG**R**AAGVAAMQGAVVHLQ**R**ATI**RR**G**R**ALAGGAVPGGAVPGGAVPGGFGPGGFGP
VL**D**GWYGV**D**VSGSSV**E**LAQSIV**E**AP**E**LGAAI**R**VG**R**GA**R**VTVPGGSLSAPHGNVI**E**TGGA**RR**FAP
QAAPLSITLQAGAH

>Swap6
DWNNQSIV**K**TG**ER**QHGIHIQGS**D**PGGV**R**TASGTTI**K**VSG**R**QAQGILL**E**NPAA**E**LQF**R**NGSVTSS
GQLS**DR**GI**DR**FLGTVTV**E**AG**K**LVA**D**HATLANVG**D**TW**DKD**GIALYVAG**R**QAQASIA**D**STLQGAGG
VQI**RE**GANVTVQ**R**SAIV**D**GGLHIGALQSLQP**ER**LPPS**D**VVL**RD**TNVTAVPASGAPAAVSVLGAS
RLTL**D**GGHITGG**D**AAGVAAMQGAVVHLQ**R**ATI**ER**G**E**ALAGGAVPGGAVPGGAVPGGFGPGGFGP
VL**D**GWYGV**D**VSGSSV**E**LAQSIV**E**AP**E**LGAAI**R**VG**R**GA**R**VTVPGGSLSAPHGNVI**E**TGGA**RR**FAP
QAAPLSITLQAGAH

>PCt
AQGKALLYRVLPEPVKLTLTGGADAQGDIVATELPSIPGTSIGPLDVALASQARWTGATRAVDS
LSIDNATWVMTDNSNVGALRLASDGSVDFQQPAEAGRFKVLTVNTLAGSGLFRMNVFADLGLSD
KLVVMQDASGQHRLWVRNSGSEPASANTLLLVQTPLGSAATFTLANKDGKVDIGTYRYRLAANG
NGQWSLVGAKAPP
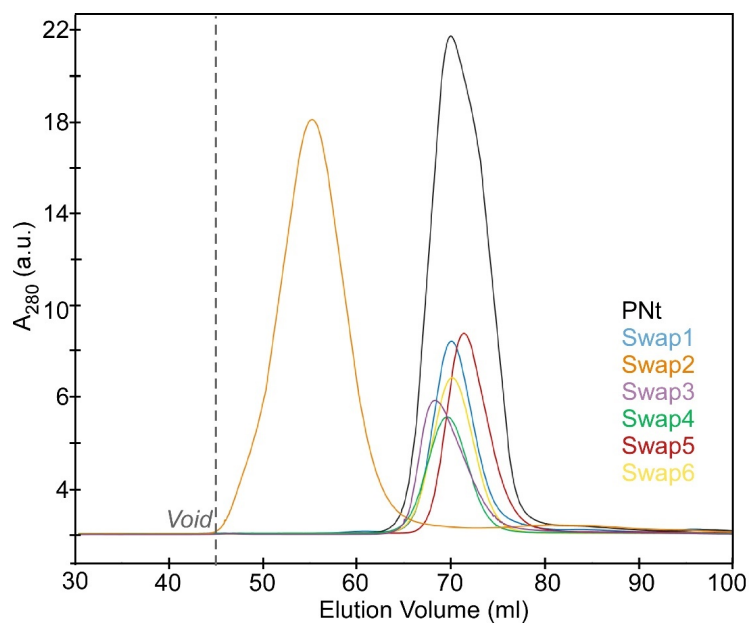
Figure S2. Size exclusion chromatography (SEC) of PNt and Swap variants. With the exception of Swap2, all variants elute as a monomer.
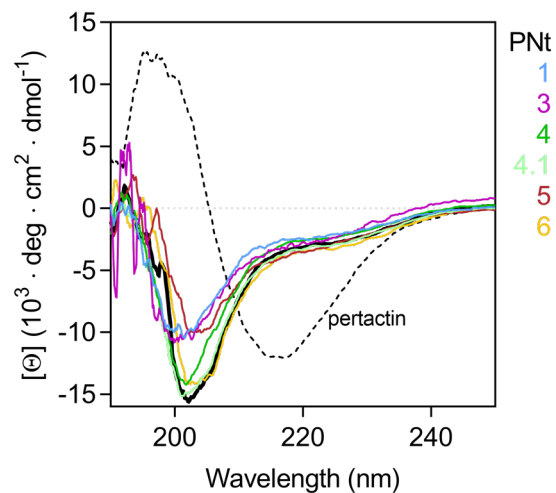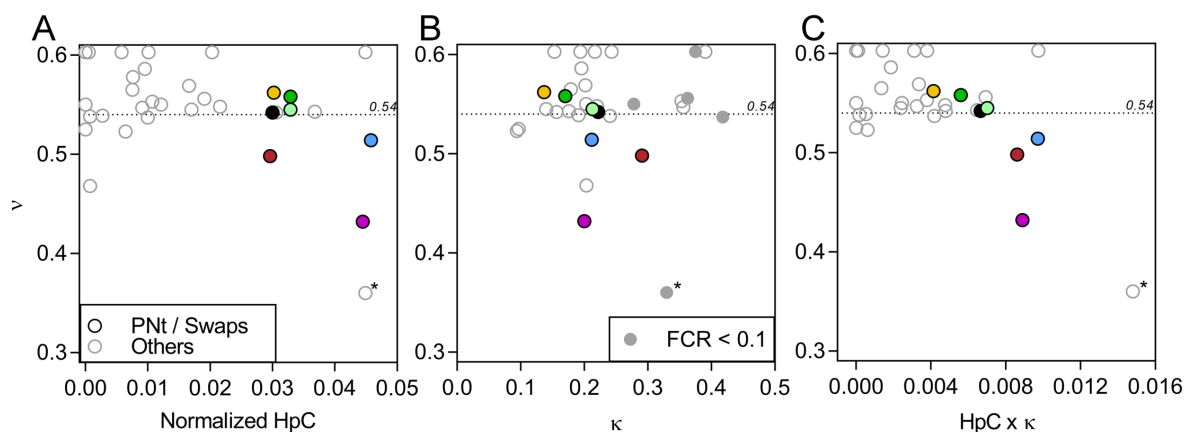
Figure S3. CD spectra of PNt and swap variants demonstrating random coil behavior. The spectrum of the folded β-helical pertactin passenger is included for comparison (dotted line).
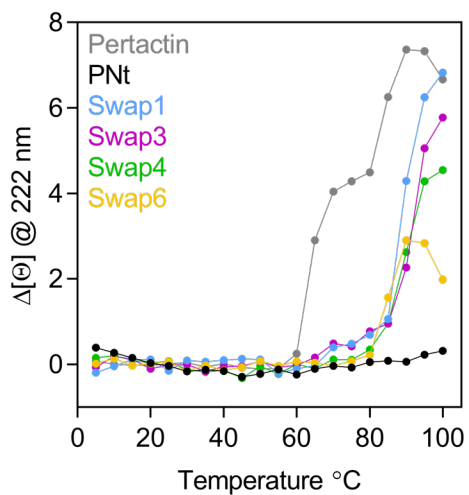
Figure S4. Normalized mean molar ellipticity (Θ) at 222 nm as a function of temperature for pertactin, PNt, or swap variants.

Figure S5. Comparison of the extent of chain collapse (quantified as the Flory exponent, ν, for wild type PNt and swap variants (filled points; colors as described in **Figure 1**) and other IDPs (gray) as a function of (**A**) normalized HpC, (**B**) κ, and (**C**) Normalized HpC x κ. The impact of charge patterns on polypeptide behavior is expected to be minimal for polypeptide chains with few charged residues (14); for this reason, chains with a fraction of charged residues (FCR) less than 0.1 are indicated by a solid gray point in panel **B**. The dotted line at $\nu = 0.54$ indicates the Flory exponent typical of an expanded polypeptide under native conditions (3). See **Table S2** for the list of other IDPs included here. Color scheme for swap variants as for other figures.
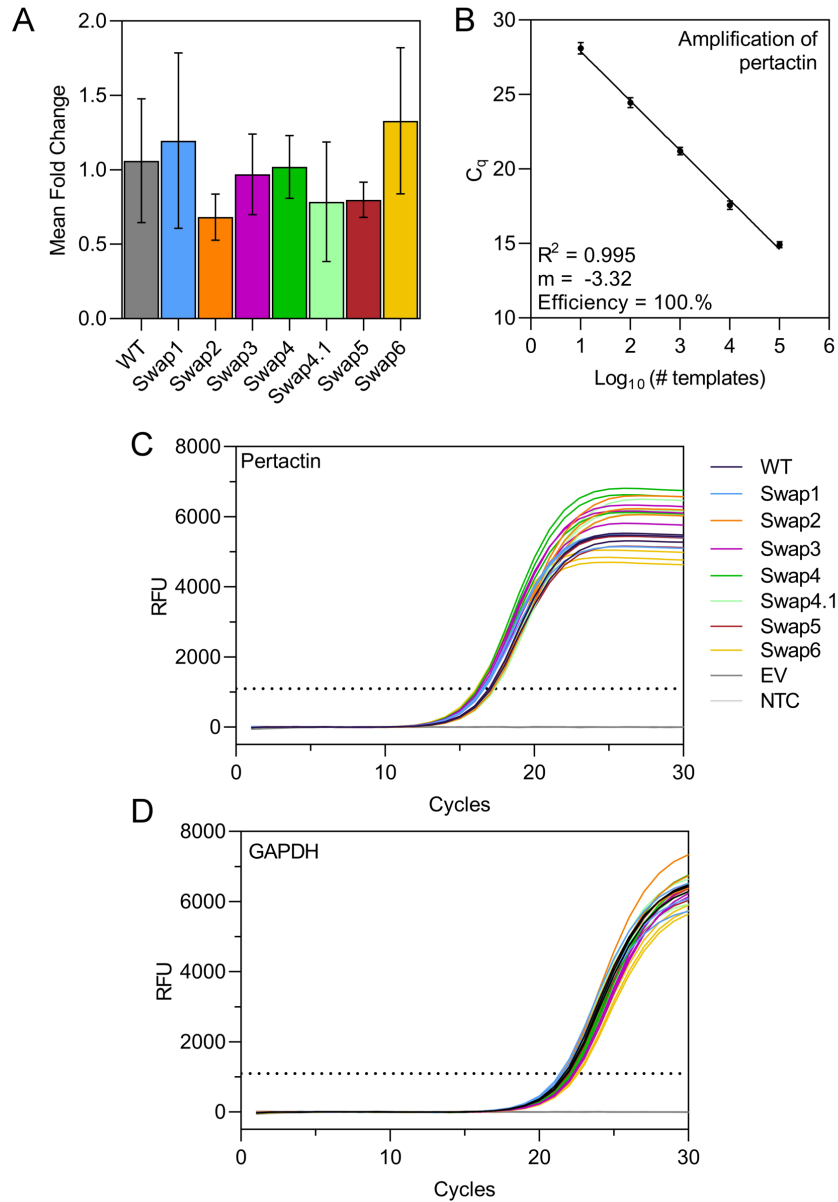
Figure S6. Quantification of RNA levels for full-length pertactin and swap variants chimeras. (**A**) Mean fold-change in mRNA level as measured by RT-qPCR, normalized to GAPDH, relative to wild type pertactin. Error bars represent the standard deviation from three biological replicates. (**B**) Primer amplification efficiency (100%). (**C-D**) Representative examples of raw amplification data.
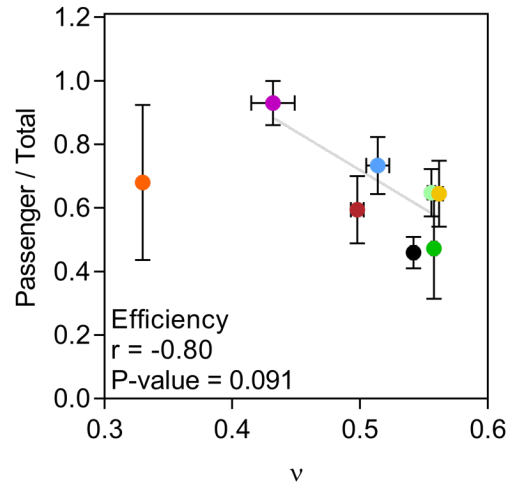
Figure S7. Secreted passenger fraction, calculated as the intensity of the cleaved passenger band divided by the sum of the passenger plus precursor intensities. Points for wild type PNt and swap variants are color-coded as in Figure 5 (also Figure S8, below). As explained in detail in the main text and Figure 5 legend, results for neither wild type PNt nor Swap2 were included in the correlation analysis. The correlation is not significant.
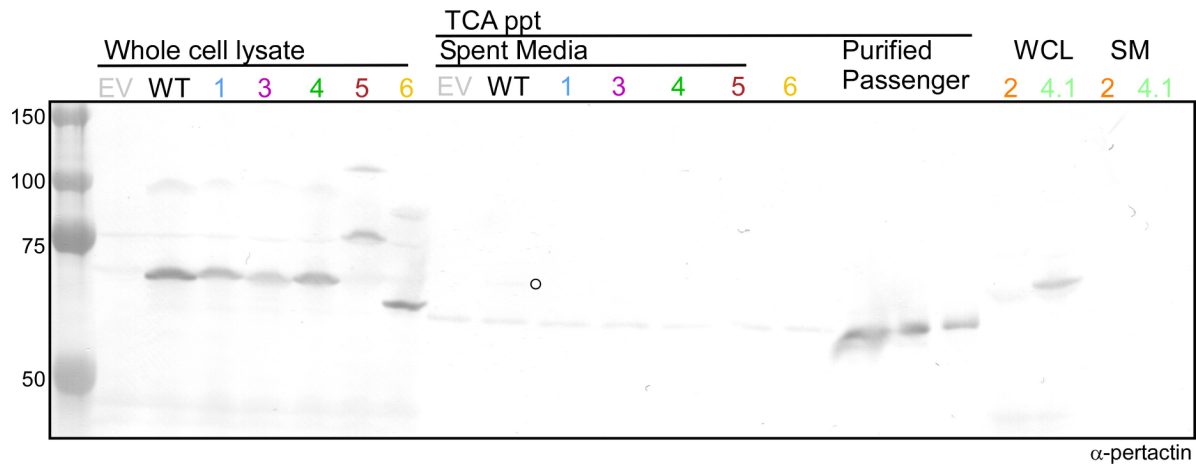
Figure S8. α-pertactin western blot of *E. coli* whole cell lysate (WCL) or trichloroacetic acid (TCA)-precipitated spent media from cultures expressing full length pertactin (WT or with a swap-mutated PNt region; numbering and color-coding as for other figures) or an empty vector (EV) control. TCA precipitated purified passenger included as a control for reproducibility of the TCA precipitation. Note that a very small amount of passenger was detected only in the spent media of the WT sample (open circle).

12
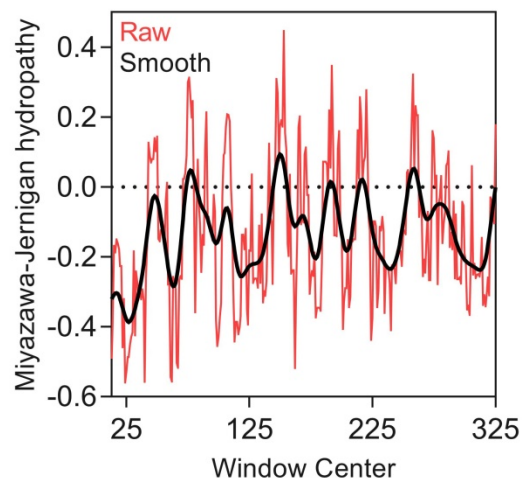
Figure S9. Example of averaging calculated hydropathy along sliding windows of nine amino acids (black curve) for PNt. Per-residue hydropathy values (red) were used to calculate HpC (Methods). Gaussian smoothed curves ($\sigma = 5$) were used for visual interpretation (**Figure 2A**).

Supplemental References

36. Riback JA*, et al.* (2017) Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell* 168(6):1028-1040 e1019.

37. Watson MC & Curtis JE (2014) Probing the average local structure of biomolecules using small-angle scattering and scaling laws. *Biophys J* 106(11):2474-2482.

38. Boze H*, et al.* (2010) Proline-rich salivary proteins have extended conformations. *Biophys J* 99(2):656-665.

39. Bressan GC*, et al.* (2008) Human regulatory protein Ki-1/57 has characteristics of an intrinsically unstructured protein. *J Proteome Res* 7(10):4465-4474.

40. Kolonko M*, et al.* (2016) Intrinsic disorder of the C-terminal domain of *Drosophila* methoprene-tolerant protein. *PLoS ONE* 11(9):e0162950.

41. Shell SS, Putnam CD, & Kolodner RD (2007) The N terminus of *Saccharomyces cerevisiae* Msh6 is an unstructured tether to PCNA. *Mol Cell* 26(4):565-578.

42. Lens Z*, et al.* (2010) Solution structure of the N-terminal transactivation domain of ERM modified by SUMO-1. *Biochem Biophys Res Commun* 399(1):104-110.

43. LaBreck CJ, May S, Viola MG, Conti J, & Camberg JL (2017) The Protein chaperone ClpX targets native and non-native aggregated substrates for remodeling, disassembly, and degradation with ClpP. *Front Mol Biosci* 4:26.

44. Yang C, van der Woerd MJ, Muthurajan UM, Hansen JC, & Luger K (2011) Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes. *Nucleic Acids Research* 39(10):4122-4135.

45. Moncoq K*, et al.* (2004) SAXS study of the PIR domain from the Grb14 molecular adaptor: a natively unfolded protein with a transient structure primer? *Biophys J* 87(6):4056-4064.

46. Konno T, Tanaka N, Kataoka M, Takano E, & Maki M (1997) A circular dichroism study of preferential hydration and alcohol effects on a denatured protein, pig calpastatin domain I. *Biochim Biophys Acta* 1342(1):73-82.

47. Wells M*, et al.* (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci USA* 105(15):5762-5767.

48. Contreras-Martos S*, et al.* (2017) Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. *Sci Rep* 7(1):4676.

49. Renner M, Paesen GC, Grison CM, & Reports G-S (2017) Structural dissection of human metapneumovirus phosphoprotein using small angle x-ray scattering. *Scientific Reports*.

50. Gazi AD, *et al.* (2008) Evidence for a coiled-coil interaction mode of disordered proteins from bacterial type III secretion systems. *J Biol Chem* 283(49):34062-34068.

51. Alborghetti MR, *et al.* (2010) Human FEZ1 protein forms a disulfide bond mediated dimer: implications for cargo transport. *J Proteome Res* 9(9):4595-4603.