Supplementary Data

The Proline/Glycine-Rich Region Of The Biofilm Adhesion Protein Aap Forms An Extended Stalk That Resists Compaction

Alexander E. Yarawsky[1,2], Lance R. English[3], Steven T. Whitten[3], and Andrew B. Herr[2,4]

[1] Graduate Program in Molecular Genetics, Biochemistry & Microbiology, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0524, USA

[2] Division of Immunobiology and Center for Systems Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229-3026, USA

[3] Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas 78666, USA

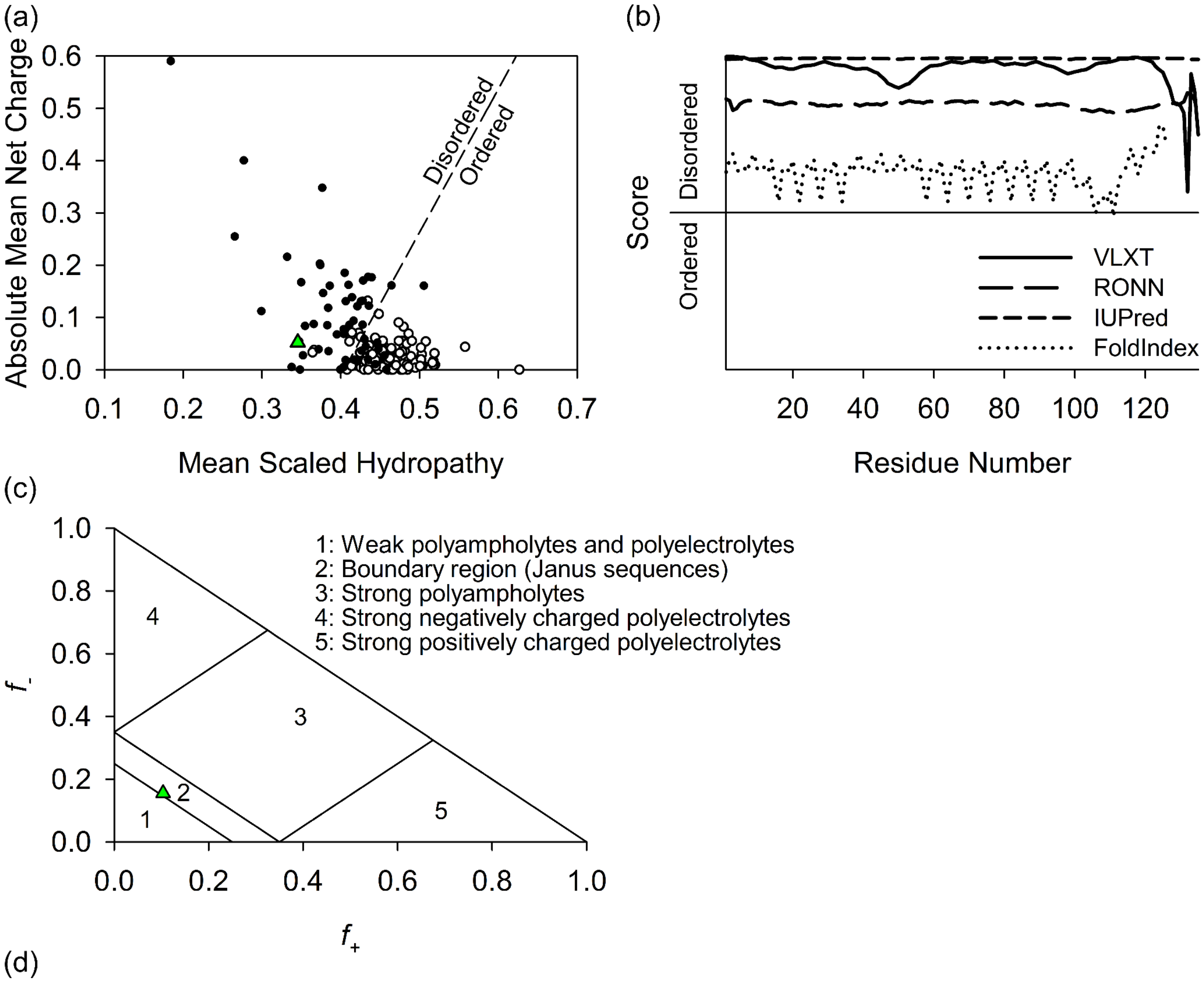
[4] Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229-3026, USA

Correspondence to Andrew B. Herr: Division of Immunobiology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave., Cincinnati, OH 45229, Telephone: (513) 803-7490; FAX: (513) 636-5355; E-mail: andrew.herr@cchmc.org

Supplementary data files present in this document include:

1. **Supplementary Figures S1—S2**
2. **Supplementary Tables S1—S5**

Supplementary Figures



(d)

```
        AEPGKPAEPGKPAEPGKPAEPGTPAEPGKPAEPGTPAEPGKPAEPGKPAEPGKPAEPGKPAEPGTP
SABLE   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SCRATCH cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
GOR     cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

        AEPGTPAEPGKPAEPGTPAEPGKPAEPGTPAEPGKPAESGKPVEPGTPAQSGAPEQPNRSMHSTDNKNQ
SABLE   ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SCRATCH ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccTcc
GOR     ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccEEcEEEc

                c: coil      T: turn      E: extended strand
```

Figure S1. PGR is predicted to be intrinsically disordered.
Panel (a) is the Uversky plot, showing PGR (*green triangle* – (0.3454, 0.0522)) lies on the portion of the plot where disordered proteins (*black circles*) tend to fall. Ordered proteins are shown in *white circles*. In panel (b), the results from several disorder predictions are plotted. The y-axis units should be considered arbitrary, as these algorithms have different ranges for their predictions; however, these results strongly support the hypothesis that PGR is an IDP. (c) shows PGR (*green triangle* – (0.10370, 0.15556)) in the boundary region (2) of the Das-Pappu phase plot separating weak polyampholytes/polyelectrolytes (1) and strong polyampholytes (3).

Panel (d) shows secondary structure predictions by SABLE, SCRATCH, and GOR servers, each predicting essentially complete random coil.

```
Q9L470_RP62A      EYGPTKAEPGKPAEPGKPAEPGKPAEPGTPAEPGKPAEPGTPAEP-----GKPAEPGKPA  55
Q8CQD9_ATCC 12228 EYGPTKAEPGKPAEPGKPAEPGKPAEPGKPAEPGTPAEPGKPAEP-----GKPAEPGKPA  55
Q6UV38_5179       EYGPTKAEPGKPAEPGKPAEPGKPAEPGTPAEPGKPAEPGTPAEP-----GKPAEPGKPA  55
Q6UV37_5179-R1    EYGPTKAEPGKPAEPGKPAEPGKPAEPGTPAEPGKPAEPGTPAEP-----GKPAEPGKPA  55
A0A075IHN3_1457   EYGPTKAEPGKPAEPGKPAEPGKPAEPGTPAEPGKPAEPGTPAEPGTSEPGKPAEPGKPA  60
                  ****************************.*****.*****.****     **********

Q9L470_RP62A      EPGKPAEPGKPAEPGTPAEPGTPAEPGKPAEPGTPAEPGKPAEPGTPAEPGKPAESG---  112
Q8CQD9_ATCC 12228 EPGKPAEPGKPAEPGTPAEPGKPAEPGKPAEPGKPAEPGTPAEPGKPAEPGTPAEPG---  112
Q6UV38_5179       EPGKPAEPGKPAEPGTPAEPGTPAEPGKPAEPGTPAEPGKPAEPGTPAEPGKPAESGKPV  115
Q6UV37_5179-R1    EPGKPAEPGKPAEPGTPAEPGTPAEPGKPAEPGTPAEPGKPAEPGTPAEPGKPAESGKPV  115
A0A075IHN3_1457   EPGKPAEPGKPAEPGKPAEPGTPAEPGKPAEPGKPAEPGTPAEPGTPAEPGKPAEPGKPA  120
                  ***************.*****.***********.*****.*****.*****.*** *

Q9L470_RP62A      ---KPVEPGTPAQSGAPEQPNRSMHSTDNKNQLPDTG  146
Q8CQD9_ATCC 12228 ---KPAEPGTPTQSGAPEQPNRSMHSTDNKNQLPDTG  146
Q6UV38_5179       EPGKPVEPGTPAQSGAPEQPNRSMHSTDNKNQSPDTG  152
Q6UV37_5179-R1    EPGKPVEPGTPAQSGAPEQPNRSMHSTDNKNQSPDTG  152
A0A075IHN3_1457   EPGKPAEPGTPAQSGAPEQPNRSMHSTDNKNQLPDTG  157
                     **.*****:******************* ****
```

Figure S2. The sequence of PGR is highly conserved among *S. epidermidis* strains. Identical residues are marked with an asterisk (*), highly conserved residues with a semicolon (:), and weakly conserved residues with a period (.). The UniProt accession number and the name of the strain identify each sequence in the alignment. The NCTC 11047 strain (UniProt accession no. **E0ACJ2**) is significantly longer than the strains shown above, having an additional 7 AEPGKP repeats compared to RP62A (UniProt accession no. **Q9L470**), and thus was omitted for clarity. The Aap from strain PM221 (GenBank accession no. **CDM15051**) contains a region between the last half B-repeat and the LPXTG motif which does not resemble the PGR of any of the above strains, neither in number of residues nor in amino acid content.

Supplementary Tables

Table S1. Concentration dependence of sedimentation velocity AUC data

PGR Concentration	$s_{20,w}$ [a]	f/f_0 [b]	M_{calc} [c]
25 µM	1.05	2.14	14.4 kDa
75 µM	1.05	2.11	14.0 kDa
150 µM	1.03	2.15	14.1 kDa
225 µM	1.02	2.18	14.1 kDa
300 µM	0.99	2.32	14.8 kDa

[a] The sedimentation coefficient standardized to 20° C and pure water.

[b] Frictional ratio – the experimental frictional coefficient divided by the frictional coefficient of an ideal, non-hydrated sphere

[c] The molecular weight calculated from the sedimentation coefficient and frictional ratio

Table S2. Temperature dependence of sedimentation velocity AUC data

Temperature (° C)	$s_{20,w}$ [a]	f/f_0 [b]	M_{calc} [c]
4° C	1.06	2.14	15.3 kDa
20° C	1.05	2.14	14.4 kDa
37° C	1.03	2.10	13.0 kDa

[a] The sedimentation coefficient standardized to 20° C and pure water.

[b] Frictional ratio – the experimental frictional coefficient divided by the frictional coefficient of an ideal, non-hydrated sphere

[c] The molecular weight calculated from the sedimentation coefficient and frictional ratio

Table S3. Salt dependence of sedimentation velocity AUC data

Temperature (° C)	NaCl Concentration	$s_{20,w}$ [a]	f/f_0 [b]	M_{calc} [c]
4° C	30 mM	1.04	2.05	13.8 kDa
	100 mM	1.06	2.02	13.9 kDa
	300 mM	1.06	2.00	13.8 kDa
	1 M	1.03	2.00	13.2 kDa
20° C	30 mM	1.04	2.09	13.7 kDa
	100 mM	1.05	2.08	13.8 kDa
	300 mM	1.04	2.08	13.7 kDa
	1 M	1.00	2.19	13.8 kDa
37° C	30 mM	1.04	2.04	12.6 kDa
	100 mM	1.05	2.05	12.9 kDa
	300 mM	1.03	2.09	13.0 kDa
	1 M	0.98	2.12	12.2 kDa

[a] The sedimentation coefficient standardized to 20° C and pure water.

[b] Frictional ratio – the experimental frictional coefficient divided by the frictional coefficient of an ideal, non-hydrated sphere

[c] The molecular weight calculated from the sedimentation coefficient and frictional ratio

Table S4. Comparison of hydrodynamic properties for PGR to a dataset of studied IDPs

Sequence	N[a]	Net Charge	R_h predicted[b]	f_{PPII}[c]	R_h Obs[d]	Reference
Aap-PGR	*135*	*7*	*38.50*	*0.5350*	*37.06*[e]	*This work*
p53(1-93)	93	15	29.51	0.4890	32.4	[1]
p53(1-93) ALA-	93	15	28.66	0.4581	30.4	[1]
p53 TAD	73	14	24.79	0.4500	23.8	[2]
Securin	202	1	42.57	0.4130	39.7	[3]
PDE-γ	87	4	26.51	0.4122	24.8	[4]
Cad136	136	9	33.77	0.4025	28.1	[5]
HIF1-α-403	202	29	42.13	0.4024	44.3	[6]
Tau-K45	198	19	41.52	0.3988	45.0	[7]
HIF1-α-530	170	10	37.81	0.3899	38.3	[6]
Fos-AD	168	16	37.17	0.3783	35.0	[8]
ShB-C	146	4	34.32	0.3764	32.9	[9]
α-synuclein	140	9	33.47	0.3744	28.2	[10]
Mlph(147-403)	260	28	47.00	0.3703	49.0	[11]
CFTR-R-region	189	5	39.18	0.3644	32.0	[12]
p57-ID	73	6	23.14	0.3636	24.0	[13]
prothymosin-α	110	43	29.02	0.3633	33.7	[14]
LJIDP1	94	4	26.46	0.3565	24.5	[15]
Mlph(147-240)	97	15	26.85	0.3528	28.0	[11]
SNAP25	206	14	40.60	0.3513	39.7	[16]
Hdm2-ABD	97	29	26.47	0.3345	25.7	[17]
Vmw65	89	19	25.13	0.3278	28.0	[18]
p53(1-93) PRO-	93	15	24.93	0.2832	27.4	[1]

[a] The number of amino acids in the sequence

[b] The predicted R_h from sequence and according to equation 6 in the main text

[c] The fractional number of PPII residues from sequence and according to intrinsic PPII propensities [19]

[d] The R_h of the IDP as measured experimentally in the reference listed in the final column

[e] As measured in this study, listed is the average of SEC and DLS measurements.

IDP dataset adapted from Tomasso et al. [19] and sorted by f_{PPII}

Table S5. Folded proteins and hydrodynamic measurements from literature

Sequence	N^a	$R_h{}^b$	Reference
staphylococcal nuclease	151‡	22.5	[20]
human recombinant lysozyme	132‡	21.8	[20]
bovine erythrocyte carbonic anhydrase	267‡	26.8	[20]
bovine pancreatic trypsin inhibitor	58	15.8	[21]
SH3 domain of PI3 kinase	90	18.6	[21]
horse heart cytochrome c	104	17.8	[21]
hen lysozyme	129	20.5	[21]
horse myoglobin	153	21.2	[21]
bovine alpha-lactalbumin	123	18.8	[21]
bovine pancreatic ribonuclease A	124	19.0	[21]
sperm whale apomyoglobin	153	20.9	[21]
ubiquitin	76	16.5	[21]
(apo)cytochrome C	104	18.5	[21]
α-lactalbumin	123	18.5	[21]
tumor supressor, p16	156	20.0	[21]
(apo)myoglobin	154	20.9	[21]
β-lactoglobulin	162	22.0	[21]
sarcoplasmic calcium binding	174	21.5	[21]
adenylate kinase	194	21.9	[21]
tryptophan synthase	268	24.2	[21]
β-lactamase	257	23.7	[21]
carbonic anhydrase B	260	23.3	[21]
RTEM β-lactamase	263	24.5	[21]

[a] The number of amino acids in the sequence, taken from [21] unless otherwise noted
[b] Hydrodynamic radius, in Å, reported by the reference listed in the final column
‡The number of residues was estimated from the *MW* using the average of 111.6 Da/residue
Folded protein data set adapted from [21], see [22-24] for additional details and individual references

Supplemental References

[1] R.B. Perez, A. Tischer, M. Auton, S.T. Whitten, Alanine and proline content modulate global sensitivity to discrete perturbations in disordered proteins, Proteins. 82 (2014) 3373-3384.

[2] D.F. Lowry, A. Stancik, R.M. Shrestha, G.W. Daughdrill, Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53, Proteins. 71 (2008) 587-598.

[3] N. Sanchez-Puig, D.B. Veprintsev, A.R. Fersht, Human full-length Securin is a natively unfolded protein, Protein science : a publication of the Protein Society. 14 (2005) 1410-1418.

[4] V.N. Uversky, S.E. Permyakov, V.E. Zagranichny, I.L. Rodionov, A.L. Fink, A.M. Cherskaya, et al., Effect of zinc and temperature on the conformation of the gamma subunit of retinal phosphodiesterase: a natively unfolded protein, Journal of proteome research. 1 (2002) 149-159.

[5] S.E. Permyakov, I.S. Millett, S. Doniach, E.A. Permyakov, V.N. Uversky, Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin, Proteins. 53 (2003) 855-862.
[6] N. Sanchez-Puig, D.B. Veprintsev, A.R. Fersht, Binding of natively unfolded HIF-1alpha ODD domain to p53, Molecular cell. 17 (2005) 11-21.
[7] A. Soragni, B. Zambelli, M.D. Mukrasch, J. Biernat, S. Jeganathan, C. Griesinger, et al., Structural characterization of binding of Cu(II) to tau protein, Biochemistry. 47 (2008) 10841-10851.
[8] K.M. Campbell, A.R. Terrell, P.J. Laybourn, K.J. Lumb, Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos, Biochemistry. 39 (2000) 2708-2713.
[9] E. Magidovich, I. Orr, D. Fass, U. Abdu, O. Yifrach, Intrinsic disorder in the C-terminal domain of the Shaker voltage-activated K+ channel modulates its interaction with scaffold proteins, Proceedings of the National Academy of Sciences. 104 (2007) 13022-13027.
[10] K.E. Paleologou, A.W. Schmid, C.C. Rospigliosi, H.Y. Kim, G.R. Lamberto, R.A. Fredenburg, et al., Phosphorylation at Ser-129 but not the phosphomimics S129E/D inhibits the fibrillation of alpha-synuclein, The Journal of biological chemistry. 283 (2008) 16895-16905.
[11] N.C. Geething, J.A. Spudich, Identification of a minimal myosin Va binding site within an intrinsically unstructured domain of melanophilin, The Journal of biological chemistry. 282 (2007) 21518-21528.
[12] J.M.R. Baker, Structural characterization and interactions of the CFTR regulatory region. (2009).
[13] J.N. Adkins, K.J. Lumb, Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2, Proteins. 46 (2002) 1-7.
[14] S. Yi, B.L. Boys, A. Brickenden, L. Konermann, W.Y. Choy, Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch, Biochemistry. 46 (2007) 13120-13130.
[15] S. Haaning, S. Radutoiu, S.V. Hoffmann, J. Dittmer, L. Giehm, D.E. Otzen, et al., An unusual intrinsically disordered protein from the model legume Lotus japonicus stabilizes proteins in vitro, The Journal of biological chemistry. 283 (2008) 31142-31152.
[16] U.B. Choi, J.J. McCann, K.R. Weninger, M.E. Bowen, Beyond the random coil: stochastic conformational switching in intrinsically disordered proteins, Structure (London, England : 1993). 19 (2011) 566-576.
[17] S.G. Sivakolundu, A. Nourse, S. Moshiach, B. Bothner, C. Ashley, J. Satumba, et al., Intrinsically unstructured domains of Arf and Hdm2 form bimolecular oligomeric structures in vitro and in vivo, Journal of molecular biology. 384 (2008) 240-254.
[18] L. Donaldson, J.P. Capone, Purification and characterization of the carboxyl-terminal transactivation domain of Vmw65 from herpes simplex virus type 1, The Journal of biological chemistry. 267 (1992) 1411-1414.
[19] M.E. Tomasso, M.J. Tarver, D. Devarajan, S.T. Whitten, Hydrodynamic Radii of Intrinsically Disordered Proteins Determined from Experimental Polyproline II Propensities, PLoS Comput Biol. 12 (2016) e1004686.
[20] T.D. Langridge, M.J. Tarver, S.T. Whitten, Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein, Proteins. 82 (2014) 668-678.
[21] J.A. Marsh, J.D. Forman-Kay, Sequence determinants of compaction in intrinsically disordered proteins, Biophys. J. 98 (2010) 2383-2390.
[22] D.K. Wilkins, S.B. Grimshaw, V. Receveur, C.M. Dobson, J.A. Jones, L.J. Smith, Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques, Biochemistry. 38 (1999) 16424-16431.
[23] O. Tcherkasskaya, E.A. Davidson, V.N. Uversky, Biophysical Constraints for Protein Structure Prediction, J. Proteome Res. 2 (2003) 37-42.

[24] O. Tcherkasskaya, V.N. Uversky, Denatured collapsed states in protein folding: Example of apomyoglobin, Proteins: Structure, Function, and Bioinformatics. 44 (2001) 244-254.