

Biophysical Journal, Volume 98

Supporting Material

Sequence determinants of compaction in intrinsically disordered proteins

Joseph A. Marsh and Julie D. Forman-Kay

Supporting Material for
'Sequence determinants of compaction in intrinsically disordered proteins'

Table S1. List of intrinsically disordered proteins with R_h measurements.

Name	# of residues	R_h	pH	Reference	Sequence
A. PFG-NMR measurements					
Abeta(1-40)	40	14.36	7.4	(1)	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
SBD	61	25.6	6	(2)	GSMMSASSQSPNPNNAEYECSTIPPLEYECSTIPPLQQAQASGALSSPPPTVMVPVGVVKHP
CTL9-I98A	92	21.7	6.6	(3)	AAEELANAKLKEQLEKLTVTIPAKAGEGGRFLFGSITSKQAAESLQAQHGLKLDKRKIELADAIRALGYTNVPVKLHPEVTATLKVHVTEQK
Hdm2-ABD	97	25.7	6	(4)	XXSSSESTGTSPNPDLDAGVSEHSGDWLDQDSVSDQFVSEFEVESLDSSEYSLSEEGQELSDDEDDEVYQVTVYQAGESDTDSFEEDPEISLADYWK
Sm1	105	23.4	7	(5)	MQNSQDYFYAQNRCQQQAPSTLRTVTMAEFRRVPLPMAEVPMLSTQNSMGSSASASASLEMWEKDLEERLNSIDHDMNNNKFGSGELKSMFNQGGKVEEMDF
prothymosin alpha	110	33.7	7	(6)	MSDAAVDTTSEITTKDLKEKKEVVEEAENGRDAPANGNANEENGEQEADNEVDEEEEGGEEEEEEEGDGEEEDGDEDEEAESATGKRAAEDDEDDDDVDTKKQKTDEDD
TC1	112	26.5	5	(7)	HHHHHHMKAKRSHQAIMSTSLRVSPSIHGYHFDTASRKKA VGNIFENTDQESLERLFRNSGDKKAEERAKIIFAIQDVVEEKTRALMALKKRRTKDKLFQFLKLRKYSIKVH
alpha-synuclein	140	28.2	7.4	(8)	MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLVYVGSKTKEGVVHG VATVAEKTKEQVTNVGGAVVTGVTAVAQKTVEGAGSIAAATGFVKDKQLGKNEEGAPQEGILEDMPVDPDNEAYEMPSEEGYQDYEPA
CFTR R region	189	32	6.8	(9)	GAMESAERRNSILTETLHRFSLEGDAPVSWTETKKQSFKQTGEFGEKRNKNSILNPINSIRKFSIVQKTPLQMNGIEEDSDEPLERRLSLVPDSEQGEAILPRISVISTGPTLQARRRQSVLNLMTHSVNVQGGNIHRKTTASTRKVSLAPQANLTELDIYSRRLSQTGLEISEEINEEDLKECLFDDME
Tau K45	198	45	6.5	(10)	MSSPGSPGTPGSRSRTPSLPTPPTREPCKVAVVRTPPKSPSSAKSRLQTA PVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNKIHVPGGGSVQIVYKPVDLKVTSKCGSLGNIHHKPGGGQVEVKSEKLDKFDKRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY
RYBP	234	39.5	7	(11)	HHHHHHMTMGDKKSPTRPKRQAKPAADEGFWDGCSVCTFRNSAEAFKCSICDVRKGTSTRKPRNSQLVAQQAQYATPPPPKKEKKEKVEKQDKEKPEKDKEISPSVTCKNTNKKTKPKSDILKDPPEANSIQSANATTKTSETNHTSRPRLKNVDRSTAQQLAVTVGNVTVIITDFKEKTRSSSTSSSTVTSSAGSEQNQSSSGSESTDKGSSRSSTPKGDMSAVNDESF
3D7-6H MSP2	237	34.3	7	(12)	MIKNESKYSNTFINNAYNMSIRRSMAESKPSTGAGGSAAGGSAGGSAGGSAGGSAGGSAGSGDNGADAEGSSSTPATTTTTTKTTTTTTTTNDAEASTSTSENPNHKNAETNPKGKGEVQEPNQANKETQNNNSNVQQDSQTKSNVPPTQDADTKSPTAQPEQAENSAPTAEQTESPELQSA PENKGTGQHGHMHGSRNNHPQNTSDSQKECTDGNKENC GAATSLLNSSNNHHHHHH
B. SEC measurements					
EHD-L16A	61	20.1	7	(13)	TNDEKRPRTA FSSSEQAARLKREFNENRYLTERRRQQLSELGLNEAQIKIWFQNKRAKIKK
p53-TAD	73	23.8	7.5	(14)	MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPPLSQA MDDLMLSPDDIEQWFTEDPGPDEAPRMP EAAAPRV
p57-ID	73	24	7	(15)	XXTSACRSLFGPVDHEELSRELQARLAELNAEDQNRWDYDFQQDMPLRGPGRQLQWTEVDSVSPAFYRETQV

PDE-gamma	87	24.8	7.5	(16)	MNLEPPKAEIRSATRVMGGPVTPRKGPPKFKQRQTRQF KSKPPKGVQFGDDIPGMEGLTDITVICPWEAFNHL ELHELAQYGI
Vmw65	89	28	8	(17)	GSAGHTRRLSTAPPTDVSGLDELHLDGEDVAMAHADA LDDFDLMDLGDGDSPPGFTPHDSAPYGALDMADFEF EQMFTDALGIDEYGG
LJIDP1	94	24.52	7	(18)	MARSFTNIKAISALVAEEFSNSLARRGYAATAQSAGRV GASMSGKMGSTKSGEKAAREKVSVPDPVTGYK PENIKEIDVAELRSVAVLGKN
Mlph(147- 240)	97	28	7.5	(19)	XXXGGGGSEPSLEEINGDSEQTDEDGDLDEARDQPL NSKKKRLLSFRDVFEEEDSHLVQPCSTLGLSSVPE SAHSLQSLSGEPYSEDTSLEP
TyrRS(D1)	107	21	7	(20)	MALFSGDIANLTAEEIEQGFKDVPFSVHEGGDVLVEL LVSAGISPSKRQAREDIQNGAIYNGERLQDVGAILTAE HRLGRFTVIRRGKKKYLYRYALEHHHHHH
ASR1	115	27.4	7	(21)	MEEKHHHHHHLFHHKDKAEEGPVDYKEIKHHKHLE QIGKLTVAAGAYALHEKHEAKKDPEHAHKHKIEEEI AAAAAVGAGGFAPHEHHEKDKAKKEKKLRGDTTIS SKLLF
Nup116	126	25.2	6.8	(22)	GSRRASVSGSALFGAKPASGGLFGQSAGSKAFGMNTN PTGTTGGLFGQTNQQSGGGLFGQQNSNAGGLFGQN NQSQNQSLFGQQNSNAFGQPQQGGLFGSKPAGGL FGQQGASTHHHHHH
Cad136	136	28.1	7	(23)	RLEQYTSAVVGNKAAPAKPAASDLVPAEGVRNIKS MWEKGNVFSSPGGTGTPNKETAGLKVGVSSRINEWLT KTPEGNKSPAPKPSDLRPGDVSGRNLWEKQSVKPA ASSKVATATGKKSETNGLRQFEKEP
NTAIL	139	27	7	(24)	MRGSHHHHHHXXXHTTEDKISRAVGRQAQVSFLHG DQSENELPRLGGKEDRRVKQSRGEARESYRETGPSRAS DARAAHLPTGTPLDIDTASESSQDPQDSRRSADALLRL QAMAGISEEQGSDTDTPIVYNDNRLLD
NL3-cyt	140	28.3	8.5	(25)	MGSSHHHHHSSGLVPRGSHMAYRKDKRRQEPLRQPS PQRGAGAPELGAAPPEELAAALQLGPTHHECEAGPPHDT LRLTALPDYTLTLRRSPDDIPLMTPNTITMIPNSLVGLQ TLHPYNTFAAGFNSTGLPHSHSTTRV
ShB-C	146	32.9	8	(26)	MXXGQHMKKSSLSESSDMMDLDDGVSTPGLTETHP GRSAVAPFLGAQQQQQPVASSLSMSIDKQLQHPLQQ LTQTQLYQQQQQQQQQNGFKQQQQTQQQLQQ QSHTINASAAAATSGSGSGLTMRHNNALAVSIETDV
Fos-AD	168	35	7	(27)	GSHMSVASLDLTGGLPEVATPESEEAFPLPLNDPEPKP SVEPVKSISSELMELKTEPFDDFLPASSRPSGSETARSVPD MDLSGSFYAADWEPLHSGSLGMGPMATELEPLCTPVV TCTPSCTAYTSSVFVTYPEADSFPSCAAHRKGSSSNEP SSDSLSSPTLLAL
HIF1-alpha 530	170	38.3	7.2	(28)	XEFKLELVEKLEAEDTEAKNPFSTQDLDLEMLAPYI PMDDDFQLRSFDQLSPLESSASPEASQSTVTVFQQT QIQEPTANATTTTATDELKTVTKDRMEDIKILIASPST HIHKETTSSPYRDTQSRTASPNRAGKGVIEQTEKSH PRSPNVLSVALSQR
HIF1-alpha 403	202	44.3	7.2	(28)	XAAGDTIISLDFGSNDTETDDQGLEEVPLYNDVMLPSP NEKLQININLAMSPLPTAETPKPLRSSADPALNQEVALK LEPNPESLELSFTMPQIQDQTPSPSDGSTRQSSPEPNSP EYCFYVDSDMVNEFKLELVEKLEAEDTEAKNPFSTQD TDLDEMLAPYIPMDDDFQLRSFDQLSPLESSASPEA SPQSTVTVFQ
Securin	204	39.7	7.2	(29)	XXMATLIYVDKENGEPGTRVVAKDGLKLGSGPSIKAL DGRSQVSTPRFGKTFDAPPALPKATRKALGTVNRAKTEK SVKTKGPKLKQKQPSFSAKKMTEKTVKAKSSVPASDDA YPEIEKFFPFNPLDFESFDLPEEHQIAHLPLSGVPLMILD EERELEKLFQLGPPSPVKMPSPPWESNLLQSPSSILSTLD VELPPVCCDIDI
Gliotactin	215	33.4	7.4	(30)	XXXXMWRNKRQSDRFYDEDVFINGEGLEPEQDTRG VDNAHMVNTNHHALRSRDNIEYRDSPTKTLASKAHT DTTSLRSPSLAMTQKSSSQASLKSGLKTNHGLVKQ SERAATPRSQQNGSTAKVASPPVEEKRLQLSSTPVT QLQAEPARKRVPTAASVSGSSRSTTPVPSARSTTTHTTA TLSSQPAQAQPRRTHLVEGVPHHHHHH
Mlph(147-	260	49	7.5	(19)	XXXGGGGSEPSLEEINGDSEQTDEDGDLDEARDQPL NSKKKRLLSFRDVFEEEDSHLVQPCSTLGLSSVPE

403)					SAHSLQSLSGEPYSEDTTSLPEGLEETGARALGCRPSP EVQPCSPPLPSGEDAHAELSPAASCKSAFGTTAMPGTD DVRGKHLPSQYLADVDTSEDSIQGPRAASQHSKRRA RTVPETQILELNKRMSAVEHLLVHLENTVLPPSAQEPT VETHPSADTEEETLRRRLEELTSNIGSSTSSE
------	--	--	--	--	--

Table S2. List of folded proteins with R_h measurements.

Name	# of residues	R_h	Reference
bovine pancreatic trypsin inhibitor	58	15.8	(31)
SH3 domain of PI3 kinase	90	18.6	(31)
horse heart cytochrome c	104	17.8	(31)
hen lysozyme	129	20.5	(31)
horse myoglobin	153	21.2	(31)
bovine alpha-lactalbumin	123	18.8	(31)
bovine pancreatic ribonuclease A	124	19	(31)
sperm whale apomyoglobin	153	20.9	(31)
ubiquitin	76	16.5	(32)
(apo)cytochrome C	104	18.5	(32)
alpha-lactalbumin	123	18.5	(32)
tumor suppressor, p16	156	20	(32)
(apo)myoglobin	154	20.9	(32)
beta-lactoglobulin	162	22	(32)
sarcoplasmic calcium binding	174	21.5	(32)
adenylate kinase	194	21.9	(32)
tryptophan synthase	268	24.2	(32)
beta-lactamase	257	23.7	(32)
carbonic anhydrase B	260	23.3	(32)
RTEM beta-lactamase	263	24.5	(32)

Table S3. List of chemically denatured proteins with R_h measurements.

Name	# of residues	R_h	Reference
lysozyme(49-64), 8 M urea, pH 2	16	10.4	(31)
D3(17-37), 8 M urea	22	13.3	(31)
D3(2-38), 8 M urea	32	15.5	(31)
PI3 SH3, 3.5 M GdnHCl	90	27.5	(31)
lysozyme, 8 M urea, pH 2, reduced	129	34.6	(31)
triosephosphate isomerase, 2.5 GdnHCl	247	49.7	(31)
reduced bovine pancreatic ribonuclease A, 6 M GdnHCl	124	31.4	(31)
ubiquitin, 8 M urea	76	24.6	(32)
ribonuclease A, 8 M urea	124	32.4	(32)
lysozyme, 8 M urea	129	33.1	(32)
myoglobin, 8 M urea	153	35.1	(32)
beta-lactoglobulin, 8 M urea	162	37.8	(32)
chymotrypsinogen, 8 M urea	245	45	(32)
carbonic anhydrase B, 8 M urea	260	47.8	(32)
beta-lactamase, 8 M urea	257	48.9	(32)
albetin, 6 M GdnHCl	73	24.7	(32)
beta-lactamase, 6 M GdnHCl	257	52	(32)
carbonic anhydrase B, 8 M urea	260	52	(32)
ubiquitin, 6 M GdnHCl	76	25.8	(32)
prothymosin alpha, 6 M GdnHCl	110	31.4	(32)
ribonuclease A, 6 M GdnHCl	124	32.8	(32)
alpha-lactalbumin, 6 M GdnHCl	123	31.8	(32)
alpha-synuclein, 6 M GdnHCl	140	34.3	(32)
tumor suppressor, p16, 6 M GdnHCl	156	37.1	(32)
dihydrofolate reductase, 6 M GdnHCl	158	37.6	(32)
beta-lactoglobulin, 6 M GdnHCl	162	37	(32)
adenyate kinase, 6 M GdnHCl	193	42.1	(32)

References

1. Danielsson, J., J. Jarvet, P. Damberg, and A. Graslund. 2002. Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer's Abeta(1-40) peptide. *Magn Reson Chem* 40:S89-97.
2. Chong, P. A., B. Ozdamar, J. L. Wrana, and J. D. Forman-Kay. 2004. Disorder in a target for the smad2 mad homology 2 domain and its implications for binding and specificity. *J Biol Chem* 279:40707-40714.
3. Li, Y., B. Shan, and D. P. Raleigh. 2007. The cold denatured state is compact but expands at low temperatures: hydrodynamic properties of the cold denatured state of the C-terminal domain of L9. *J Mol Biol* 368:256-262.
4. Sivakolundu, S. G., A. Nourse, S. Moshiach, B. Bothner, C. Ashley, J. Satumba, J. Lahti, and R. W. Kriwacki. 2008. Intrinsically unstructured domains of Arf and Hdm2 form bimolecular oligomeric structures in vitro and in vivo. *J Mol Biol* 384:240-254.
5. Danielsson, J., L. Liljedahl, E. Barany-Wallje, P. Sonderby, L. H. Kristensen, M. A. Martinez-Yamout, H. J. Dyson, P. E. Wright, F. M. Poulsen, L. Maler, A. Graslund, and B. B. Kragelund. 2008. The intrinsically disordered RNR inhibitor Sml1 is a dynamic dimer. *Biochemistry* 47:13428-13437.
6. Yi, S., B. L. Boys, A. Brickenden, L. Konermann, and W.-Y. Choy. 2007. Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch *Biochemistry* 46:13120-13130.
7. Gall, C., H. Xu, A. Brickenden, X. Ai, and W. Y. Choy. 2007. The intrinsically disordered TC-1 interacts with Chibby via regions with high helical propensity. *Protein Sci* 16:2510-2518.
8. Paleologou, K. E., A. W. Schmid, C. C. Rospigliosi, H.-Y. Kim, G. R. Lamberto, R. A. Fredenburg, P. T. Lansbury, C. O. Fernandez, D. Eliezer, M. Zweckstetter, and H. A. Lashuel. 2008. Phosphorylation at Ser-129 but not the phosphomimics S129E/D inhibits the fibrillation of alpha-synuclein. *J Biol Chem* 283:16895-16905.
9. Baker, J. M. R. 2009. Structural Characterization and Interactions of the CFTR Regulatory Region (PhD Thesis). Department of Biochemistry. University of Toronto, Toronto.
10. Soragni, A., B. Zambelli, M. D. Mukrasch, J. Biernat, S. Jeganathan, C. Griesinger, S. Ciurli, E. Mandelkow, and M. Zweckstetter. 2008. Structural characterization of binding of Cu(II) to tau protein. *Biochemistry* 47:10841-10851.
11. Neira, J. L., M. Roman-Trufero, L. M. Contreras, J. Prieto, G. Singh, F. N. Barrera, M. L. Renart, and M. Vidal. 2009. The transcriptional repressor RYBP is a natively unfolded protein which folds upon binding to DNA. *Biochemistry* 48:1348-1360.
12. Zhang, X., M. A. Perugini, S. Yao, C. G. Adda, V. J. Murphy, A. Low, R. F. Anders, and R. S. Norton. 2008. Solution conformation, backbone dynamics and lipid interactions of the intrinsically unstructured malaria surface protein MSP2. *J Mol Biol* 379:105-121.
13. Mayor, U., J. G. Grossmann, N. W. Foster, S. M. Freund, and A. R. Fersht. 2003. The denatured state of Engrailed Homeodomain under denaturing and native conditions. *J Mol Biol* 333:977-991.
14. Lowry, D. F., A. Stancik, R. M. Shrestha, and G. W. Daughdrill. 2008. Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53. *Proteins* 71:587-598.
15. Adkins, J. N., and K. J. Lumb. 2002. Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2. *Proteins* 46:1-7.
16. Uversky, V. N., S. E. Permyakov, V. E. Zagranichny, I. L. Rodionov, A. L. Fink, A. M. Cherskaya, and E. A. Permyakov. 2002. Effect of zinc and temperature on the conformation of the gamma subunit of retinal phosphodiesterase: a natively unfolded protein. *Journal of Proteome Research* 1:149-159.

17. Donaldson, L., and J. P. Capone. 1992. Purification and characterization of the carboxyl-terminal transactivation domain of Vmw65 from herpes simplex virus type 1. *J Biol Chem* 267:1411-1414.
18. Haaning, S., S. Radutoiu, S. Hoffmann, J. Dittmer, L. Giehm, D. E. Otzen, and J. Stougaard. 2008. An unusual intrinsically disordered protein from the model legume *Lotus japonicus* stabilizes proteins in vitro. *J Biol Chem* 283:31142-31152.
19. Geething, N. C., and J. A. Spudich. 2007. Identification of a minimal myosin Va binding site within an intrinsically unstructured domain of melanophilin. *J Biol Chem* 282:21518-21528.
20. Guez, V., S. Nair, A. Chaffotte, and H. Bedouelle. 2000. The anticodon-binding domain of tyrosyl-tRNA synthetase: state of folding and origin of the crystallographic disorder. *Biochemistry* 39:1739-1747.
21. Goldgur, Y., S. Rom, R. Ghirlando, D. Shkolnik, N. Shadrin, Z. Konrad, and D. Bar-Zvi. 2007. Desiccation and zinc binding induce transition of tomato abscisic acid stress ripening 1, a water stress- and salt stress-regulated plant-specific protein, from unfolded to folded state. *Plant Physiol* 143:617-628.
22. Krishnan, V. V., E. Y. Lau, J. Yamada, D. P. Denning, S. S. Patel, M. E. Colvin, and M. F. Rexach. 2008. Intramolecular cohesion of coils mediated by phenylalanine-glycine motifs in the natively unfolded domain of a nucleoporin. *PLoS Comput Biol* 4:e1000145.
23. Permyakov, S. E., I. S. Millett, S. Doniach, E. A. Permyakov, and V. N. Uversky. 2003. Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin. *Proteins* 53:855-864.
24. Longhi, S., V. Receveur-Brechot, D. Karlin, K. Johansson, H. Darbon, D. Bhella, R. Yeo, S. Finet, and B. Canard. 2003. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J Biol Chem* 278:18638-18648.
25. Paz, A., T. Zeev-Ben-Mordehai, M. Lundqvist, E. Sherman, E. Mylonas, L. Weiner, G. Haran, D. I. Svergun, F. A. Mulder, J. L. Sussman, and I. Silman. 2008. Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neuroligin 3. *Biophys J* 95:1928-1944.
26. Magidovich, E., I. Orr, D. Fass, U. Abdu, and O. Yifrach. 2007. Intrinsic disorder in the C-terminal domain of the Shaker voltage-activated K⁺ channel modulates its interaction with scaffold proteins. *Proc Natl Acad Sci U S A* 104:13022-13027.
27. Campbell, K. M., A. R. Terrell, P. J. Laybourn, and K. J. Lumb. 2000. Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos. *Biochemistry* 39:2708-2713.
28. Sanchez-Puig, N., D. B. Veprintsev, and A. R. Fersht. 2005. Binding of natively unfolded HIF-1 α ODD domain to p53. *Mol Cell* 17:11-21.
29. Sanchez-Puig, N., D. B. Veprintsev, and A. R. Fersht. 2005. Human full-length Securin is a natively unfolded protein. *Protein Sci* 14:1410-1418.
30. Zeev-Ben-Mordehai, T., E. H. Rydberg, A. Solomon, L. Toker, V. J. Auld, I. Silman, S. Botti, and J. L. Sussman. 2003. The intracellular domain of the *Drosophila* cholinesterase-like neural adhesion protein, gliotactin, is natively unfolded. *Proteins* 53:758-767.
31. Wilkins, D. K., S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, and L. J. Smith. 1999. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* 38:16424-16431.
32. Tcherkasskaya, O., E. A. Davidson, and V. N. Uversky. 2003. Biophysical constraints for protein structure prediction. *J Proteome Res* 2:37-42.