SUPPORTING INFORMATION

Table S1. IDP dataset.

IDP	N	R_h [a]	Q [b]	$f_{PPII,chain}$ [c]	sequence
p53(1-93) WT	93	32.3	15	0.489	MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDL MLSPDDIEQWFTEDPGPDEAPRMPEAAPPVAPAPAAPTPAAPA PAPSWPL
p53(1-93) ALA^-	93	30.4	15	0.458	MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQGMDDL MLSPDDIEQWFTEDPGPDEGPRMPEGGPPVGPGPGGPTPGGPG PGPSWPL
p53(1-93) PRO^-	93	27.5	15	0.283	MEEGQSDGSVEGGLSQETFSDLWKLLGENNVLSGLGSQAMDD LMLSGDDIEQWFTEDGGGDEAGRMGEAAGGVAGAGAAGTG AAGAGAGSWGL
p53(1-93) ALA^-PRO^-	93	27.4	15	0.252	MEEGQSDGSVEGGLSQETFSDLWKLLGENNVLSGLGSQGMDD LMLSGDDIEQWFTEDGGGDEGGRMGEGGGGVGGGGGGGTG GGGGGGGSWGL
p53 TAD	73	23.8	14	0.450	MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDL MLSPDDIEQWFTEDPGPDEAPRMPEAAPRV
Vmw65	89	28	19	0.328	GSAGHTRRLSTAPPTDVSLGDELHLDGEDVAMAHADALDDFD LDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFTDALGI DEYGG
Hdm2-ABD	97	31.7	29	0.335	ERSSSSESTGTPSNPDLDAGVSEHSGDWLDQDSVSDQFSVEFEV ESLDSEDYSLSEEGQELSDEDDEVYQVTVYQAGESDTDSFEED PEISLADYWK
prothymosin-α	110	33.6	43	0.363	MSDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNANEE NGEQEADNEVDEEEEEGGEEEEEEEEGDGEEEDGDEDEEAESA TGKRAAEDDEDDDVDTKKQKTDEDD
HIF1-α-403	202	44.3	29	0.402	PAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQ NINLAMSPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSF TMPQIQDQTPSPSDGSTRQSSPEPNSPSEYCFYVDSDMVNEFKL ELVEKLFAEDTEAKNPFSTQDTDLDLEMLAPYIPMDDDFQLRS FDQLSPLESSSASPESASPQSTVTVFQ
Fos-AD	168	35	16	0.378	GSHMSVASLDLTGGLPEVATPESEEAFTLPLLNDPEPKPSVEPV KSISSMELKTEPFDDFLFPASSRPSGSETARSVPDMDLSGSFYAA DWEPLHSGSLGMGPMATELEPLCTPVVTCTPSCTAYTSSFVFT YPEADSFPSCAAAHRKGSSSNEPSSDSLSSPTLLAL
Mlph(147-240)	97	28	15	0.353	RLQGGGGSEPSLEEGNGDSEQTDEDGDLDTEARDQPLNSKKK KRLLSFRDVDFEEDSDHLVQPCSQTLGLSSVPESAHSLQSLSGE PYSEDTTSLEP
tau-K45	198	45	19	0.399	MSSPGSPGTPGSRSRTPSLPTPPTREPKKVAVVRTPPKSPSSAKS RLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDL SNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGN IHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKK IETHKLTFRENAKAKTDHGAEIVY
Mlph(147-403)	260	49	28	0.370	RLQGGGGSEPSLEEGNGDSEQTDEDGDLDTEARDQPLNSKKK KRLLSFRDVDFEEDSDHLVQPCSQTLGLSSVPESAHSLQSLSGE PYSEDTTSLEPEGLEETGARALGCRPSPEVQPCSPLPSGEDAHA ELDSPAASCKSAFGTTAMPGTDDVRGKHLPSQYLADVDTSDE DSIQGPRAASQHSKRRARTVPETQILELNKRMSAVEHLLVHLE NTVLPPSAQEPTVETHPSADTEEETLRRRLEELTSNISGSSTSSE
p57-ID	73	24	6	0.364	VRTSACRSLFGPVDHEELSRELQARLAELNAEDQNRWDYDFQ QDMPLRGPGRLQWTEVDSDSVPAFYRETVQV
PDE-γ	87	24.8	4	0.412	MNLEPPKAEIRSATRVMGGPVTPRKGPPKFKQRQTRQFKSKPP KKGVQGFGDDIPGMEGLGTDITVICPWEAFNHLELHELAQYGII
LJIDP1	94	24.52	4	0.356	MARSFTNIKAISALVAEEFSNSLARRGYAATAQSAGRVGASMS GKMGSTKSGEEKAAAREKVSWVPDPVTGYYKPENIKEIDVAE LRSAVLGKN
cad136	136	28.1	9	0.403	RLEQYTSAVVGNKAAKPAKPAASDLPVPAEGVRNIKSMWEKG NVFSSPGGTGTPNKETAGLKVGVSSRINEWLTKTPEGNKSPAP KPSDLRPGDVSGKRNLWEKQSVEKPAASSSKVTATGKKSETN GLRQFEKEP
α-synuclein	140	28.2	9	0.374	MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYV GSKTKEGVVHGVATVAEKTKEQVTNVGGAVVTGVTAVAQKT VEGAGSIAAATGFVKKDQLGKNEEGAPQEGILEDMPVDPDNE AYEMPSEEGYQDYEPEA

CFTR-R-region	189	32	5	0.364	GAMESAERRNSILTETLHRFSLEGDAPVSWTETKKQSFKQTGE FGEKRKNSILNPINSIRKFSIVQKTPLQMNGIEEDSDEPLERRLSL VPDSEQGEAILPRISVISTGPTLQARRRQSVLNLMTHSVNQGQN IHRKTTASTRKVSLAPQANLTELDIYSRRLSQETGLEISEEINEE DLKECLFDDME
SNAP25	206	39.7	14	0.351	MAEDADMRNELEEMQRRADQLADESLESTRRMLQLVEESKD AGIRTLVMLDEQGEQLERIEEGMDQINKDMKEAEKNLTDLGK FCGLCVCPCNKLKSSDAYKKAWGNNQDGVVASQPARVVDER EQMAISGGFIRRVTNDARENEMDENLEQVSGIIGNLRHMALDM GNEIDTQNRQIDRIMEKADSNKTRIDEANQRATKMLGSG
ShB-C	146	32.9	4	0.376	MTLGQHMKKSSLSESSSDMMDLDDGVESTPGLTETHPGRSAV APFLGAQQQQQQPVASSLSMSIDKQLQHPLQQLTQTQLYQQQ QQQQQQQQNGFKQQQQQTQQQLQQQQSHTINASAAAATSGS GSSSGLTMRHNNALAVSIETDV
HIF1-α-530	170	38.3	10	0.390	NEFKLELVEKLFAEDTEAKNPFSTQDTDLDLEMLAPYIPMDDD FQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATT TTATTDELKTVTKDRMEDIKILIASPSPTHIHKETTSATSSPYRD TQSRTASPNRAGKGVIEQTEKSHPRSPNVLSVALSQR
securin	202	39.7	1	0.413	MATLIYVDKENGEPGTRVVAKDGLKLGSGPSIKALDGRSQVST PRFGKTFDAPPALPKATRKALGTVNRATEKSVKTKGPLKQKQP SFSAKKMTEKTVKAKSSVPASDDAYPEIEKFFPFNPLDFESFDL PEEHQIAHLPLSGVPLMILDEERELEKLFQLGPPSPVKMPSPPW ESNLLQSPSSILSTLDVELPPVCCDIDI
Abeta(1-40)	40	14.36	3	0.308	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
sml1	104	23.4	5	0.363	MQNSQDYFYAQNRCQQQQAPSTLRTVTMAEFRRVPLPPMAEV PMLSTQNSMGSSASASASSLEMWEKDLEERLNSIDHDMNNNK FGSGELKSMFNQGKVEEMDF
PGR	135	37.7	7	0.535	AEPGKPAEPGKPAEPGKPAEPGTPAEPGKPAEPGTPAEPGKPAE PGKPAEPGKPAEPGKPAEPGTPAEPGTPAEPGKPAEPGTPAEPG KPAEPGTPAEPGKPAESGKPVEPGTPAQSGAPEQPNRSMHSTD NKNQ

[a] from literature reports[21-39] with the exception of the WT, PRO^-, ALA^-PRO^- variants of p53(1-93), Hdm2-ADB, prothymosin-α, and PGR which were measured in the current study; in Å

[b] net charge estimated from sequence as the absolute value for the number of ASP and GLU residues minus the number of LYS and ARG residue

[c] calculated from sequence using PP_{II} propensities determined experimentally for each amino acid type[79]

Table S2. R_h from simulated ensembles and sequence properties of IDP fragments.

IDP	25 residue fragment	R_h [a]	$f_{PPII,chain}$ [b]	ncd [c]	$f_{PPII,chain}$ [d]	ncd [e]
p53(1-93)	N-term: MEEPQSDPSVEPPLSQETFSDLWKL	12.04	0.443	1.00	0.489	1.555
	C-term: AAPPVAPAPAAPTPAAPAPAPSWPL	11.56	0.606	0.00		
	Center: LPSQAMDDLMLSPDDIEQWFTEDPG	11.83	0.398	1.40		
p53(1-93) PRO^-	N-term: MEEGQSDGSVEGGLSQETFSDLWKL	11.01	0.304	1.00	0.283	1.555
	C-term: AAGGVAGAGAAGTGAAGAGAGSWGL	10.50	0.258	0.00		
	Center: LGSQAMDDLMLSGDDIEQWFTEDGG	10.73	0.294	1.40		
p53(1-93) ALA^-PRO^-	N-term: MEEGQSDGSVEGGLSQETFSDLWKL	11.01	0.304	1.00	0.252	1.555
	C-term: GGGGVGGGGGGGTGGGGGGGGSWGL	9.89	0.162	0.00		
	Center: LGSQGMDDLMLSGDDIEQWFTEDGG	11.21	0.284	1.40		
PGR	N-term: AEPGKPAEPGKPAEPGKPAEPGTPA	11.19	0.562	0.20	0.535	0.602
	C-term: PGTPAQSGAPEQPNRSMHSTDNKNQ	11.36	0.439	0.00		
	Center: EPGKPAEPGTPAEPGTPAEPGKPAE	11.11	0.554	0.60		
Hdm2-ABD	N-term: ERSSSSESTGTPSNPDLDAGVSEHS	11.55	0.341	0.80	0.335	2.945
	C-term: VYQAGESDTDSFEEDPEISLADYWK	11.55	0.353	1.40		
	Center: QFSVEFEVESLDSEDYSLSEEGQEL	11.70	0.321	1.80		
prothymosin-α	N-term: MSDAAVDTSSEITTKDLKEKKEVVE	11.79	0.376	0.60	0.363	4.100
	C-term: TGKRAAEDDEDDDVDTKKQKTDEDD	11.46	0.373	1.40		
	Center: NGEQEADNEVDEEEEEGGEEEEEEE	11.46	0.365	3.40		
ShB-C	N-term: MTLGQHMKKSSLSESSSDMMDLDDG	11.51	0.306	0.60	0.376	0.331
	C-term: TSGSGSSGLTMRHNNALAVSIETDV	11.36	0.286	0.20		
	Center: MSIDKQLQHPLQQLTQTQLYQQQQQ	12.10	0.429	0.00		
securin	N-term: MATLIYVDKENGEPGTRVVAKDGLK	11.84	0.368	0.00	0.413	0.070
	C-term: LLQSPSSILSTLDVELPPVCCDIDI	11.71	0.391	0.80		
	Center: AKKMTEKTVKAKSSVPASDDAYPEI	11.63	0.434	0.20		
Fos-AD	N-term: GSHMSVASLDLTGGLPEVATPESEE	11.47	0.351	1.00	0.378	1.234
	C-term: AAHRKGSSSNEPSSDSLSSPTLLAL	11.04	0.343	0.00		
	Center: TARSVPDMDLSGSFYAADWEPLHSG	11.80	0.338	0.60		
HIF1- α -530	N-term: NEFKLELVEKLFAEDTEAKNPFSTQ	11.16	0.376	0.60	0.390	0.767
	C-term: KGVIEQTEKSHPRSPNVLSVALSQR	11.79	0.403	0.40		
	Center: FQQTQIQEPTANATTTTATTDELKT	11.28	0.395	0.40		
tau-K45	N-term: MSSPGSPGTPGSRSRTPSLPTPPTR	11.80	0.474	0.60	0.399	1.350
	C-term: KIETHKLTFRENAKAKTDHGAEIVY	11.99	0.355	0.20		
	Center: LSNVQSKCGSKDNIKHVPGGGSVQI	11.60	0.348	0.40		

[a] calculated from simulated ensemble generated for fragment sequence; in Å

[b] calculated from fragment sequence using PP_{II} propensities determined experimentally for each amino acid type[79]

[c] net charge density (*ncd*) calculated for the fragment using equation 1

[d] calculated for parent sequence using PP_{II} propensities determined experimentally for each amino acid type [79]

[e] net charge density (*ncd*) calculated for parent sequence using equation 1

Table S3. Charge-based properties calculated from sequence for each dataset IDP.

IDP	*ncd* [a]	*ncpr* [b]	κ[c]
p53(1-93) WT	1.555	0.161	0.202
p53(1-93) ALA^-	1.555	0.161	0.202
p53(1-93) PRO^-	1.555	0.161	0.202
p53(1-93) ALA^-PRO^-	1.555	0.161	0.202
p53 TAD	1.639	0.191	0.171
Vmw65	2.014	0.213	0.203
Hdm2-ABD	2.945	0.299	0.209
prothymosin-α	4.100	0.391	0.423
HIF1-α-403	2.040	0.144	0.171
Fos-AD	1.234	0.095	0.211
Mlph(147-240)	1.523	0.155	0.388
tau-K45	1.350	0.096	0.113
Mlph(147-403)	1.736	0.108	0.317
p57-ID	0.702	0.082	0.207
PDE-γ	0.429	0.046	0.327
LJIDP1	0.413	0.043	0.145
cad136	0.772	0.066	0.096
α-synuclein	0.761	0.064	0.172
CFTR-R-region	0.367	0.026	0.284
SNAP25	0.975	0.068	0.160
ShB-C	0.331	0.027	0.320
HIF1-α-530	0.767	0.059	0.158
securin	0.070	0.005	0.203
Abeta(1-40)	0.474	0.075	0.211
sml1	0.490	0.048	0.143
PGR	0.602	0.052	0.058

[a] net charge density (*ncd*) calculated using equation 1

[b] net charge per residue (*ncpr*) calculated as the net charge divided by *N*

[c] κ calculated from sequence[49] using the program localCIDER downloaded from pappulab.github.io/localCIDER/

Table S4. Sequence properties and amino acid scales with best correlation (R^2) to equation 6 error.

R^2	scale
0.32	frequency of α helix in all-α class[104]
0.28	aperiodic indices for α proteins[105]
0.27	α helix indices for α proteins[105]
0.23	ASN fractional composition
0.23	frequency of occurrence in beta-bends[101]
0.22	weights for coil at the window position of -5[106]
0.21	normalized frequency of α helix[107]
0.21	information measure for loop[108]
0.21	normalized frequency of α helix from LG[104]
0.21	information measure for turn[108]
0.21	normalized frequency of α helix[109]
0.21	normalized frequency of α helix[110]
0.20	normalized frequency of α helix[111]
0.20	normalized frequency of α helix, unwieghted[112]
0.20	relative preference at C3, specific locations at ends of α helices[113]
0.20	normalized frequency of α helix[99]
0.20	information measure for α helix[108]
0.19	average probability of a helix[114]
0.19	helix-coil equilibrium constant[115]
0.18	positional frequency at helix termini N5[96]

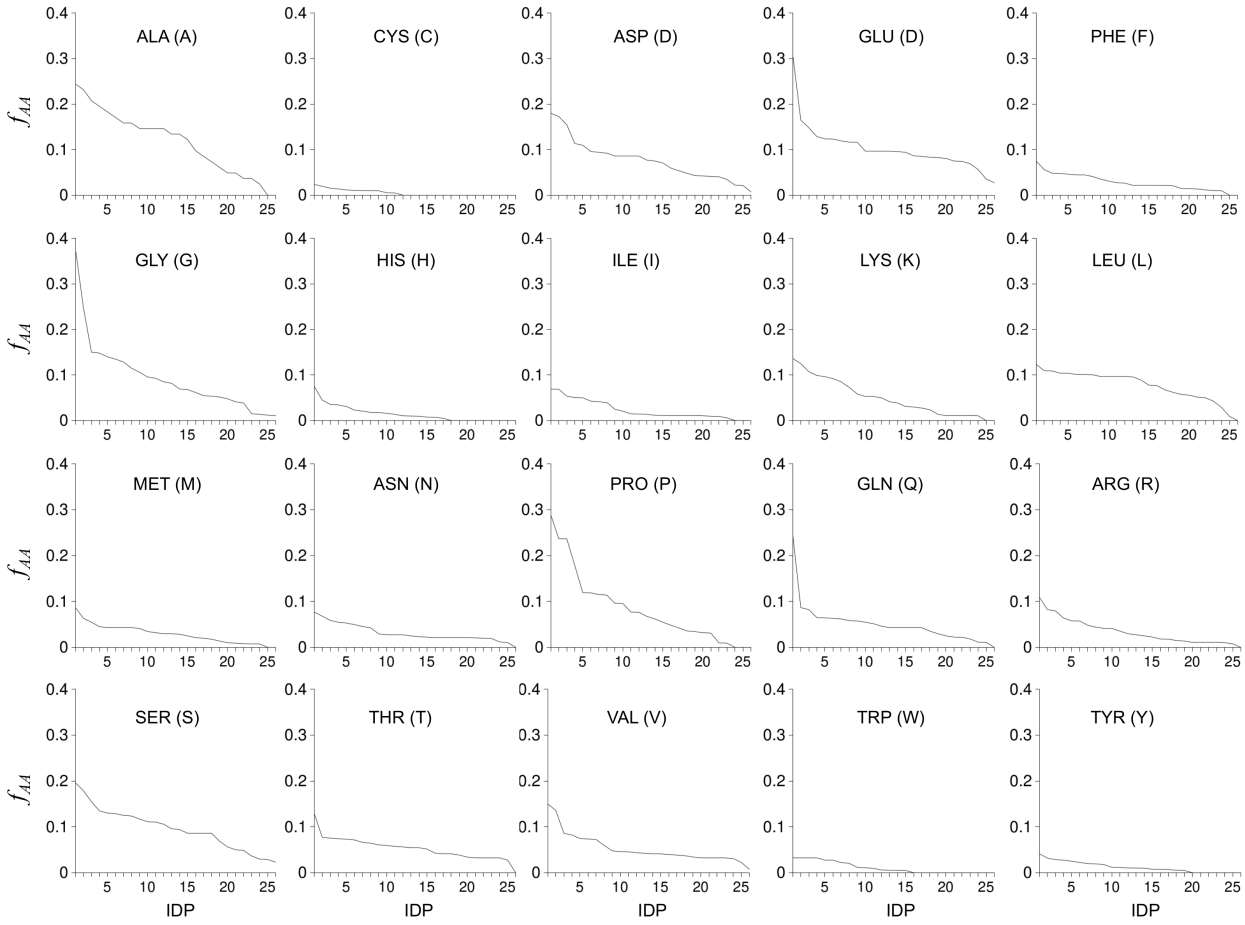

Figure S1. Fractional amino acid composition (f_{AA}) in the IDP dataset. Fractional composition was determined by summing the number of amino acids of a specific type in a sequence and dividing by N, shown in rank order from left to right for each of the 26 IDPs. Amino acid type is indicated in the panel figures.

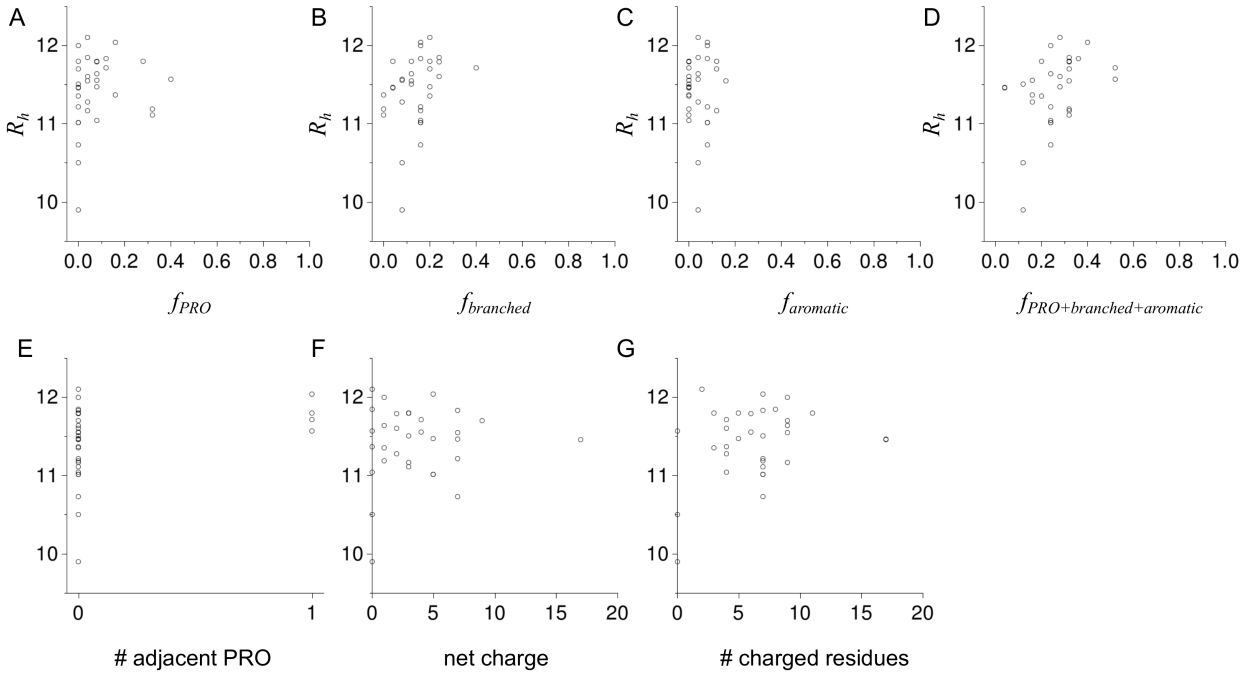

Figure S2. Open circles show R_h calculated from ensembles that were simulated for each 25-residue IDP fragment and the corresponding **A**) fractional PRO composition (f_{PRO}), **B**) fractional composition of branched residues ($f_{branched}$), **C**) fractional composition of aromatic residues ($f_{aromatic}$), **D**) fractional composition of PRO + branched + aromatic residues ($f_{PRO+branched+aromatic}$), **E**) number of sequentially adjacent PRO residues, **F**) net charge, and **G**) number of charged residues in the fragment sequence.

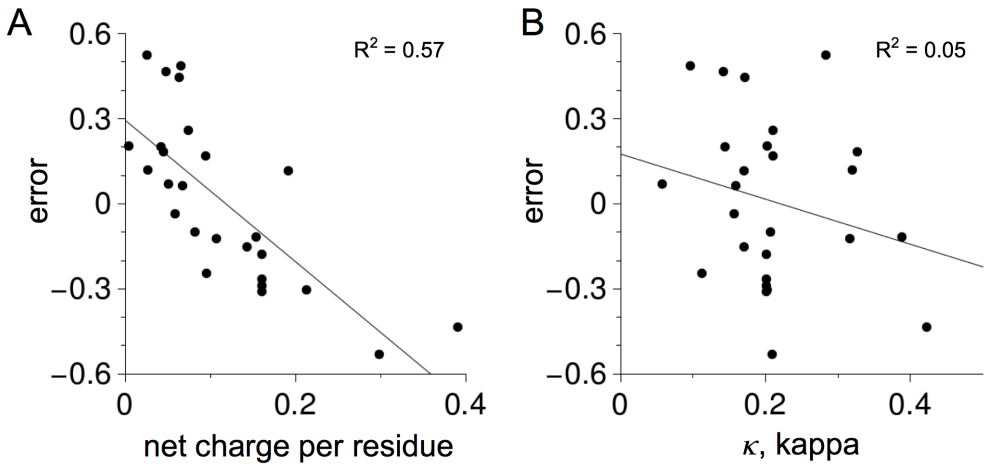

Figure S3. Comparison of the error from predicting R_h from sequence using equation 4, normalized for IDP size (equation 5), to net charge per residue (panel **A**) and κ (panel **B**). Net charge per residue and κ from sequence for each IDP is given in Table S3.

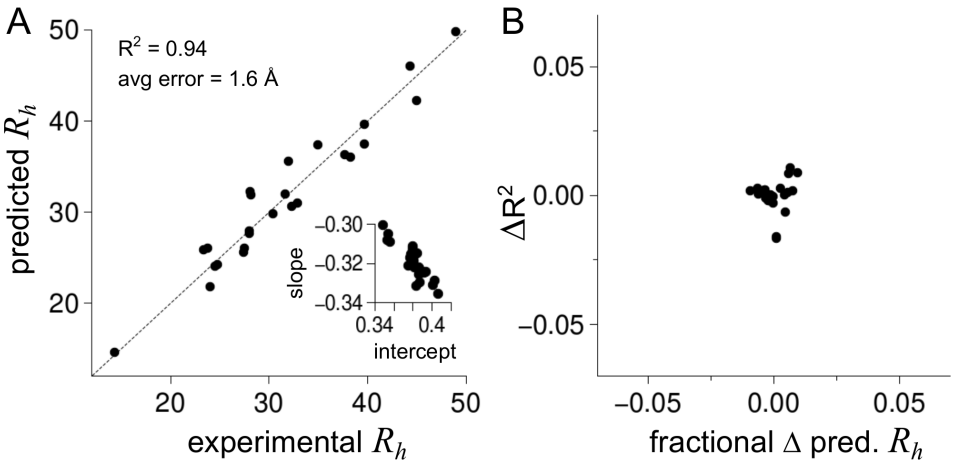

Figure S4. The effect of removing prothymosin-α from the IDP dataset on predicting R_h from intrinsic PP_{II} propensities and net charge. **A**) Removing prothymosin-α caused the linear trend in Fig. 4C to change from error = -0.25·(net charge density) + 0.31, with R^2 = 0.58, to error = -0.32·(net charge density) + 0.38, with R^2 = 0.61. Applying these trend changes to equation 6 yields $R_h = 2.16 \cdot N^{0.503-0.11 \cdot \ln(1-fPPII)} + 0.32 \cdot Q - 0.38 \cdot N^{0.5}$. Using this new equation to predict R_h gives the panel A data. The stippled line is the identity line. Inset: slope and intercept variations in the error trend owing to removal of individual IDPs (in addition to prothymosin-α) from the training set. **B**) Each circle represents the removal of a singular IDP, in addition to prothymosin-α, from the training set. The concomitant change in correlation for predicted and experimental R_h (ΔR^2) is compared to the fractional change in predicted R_h for the removed IDP.

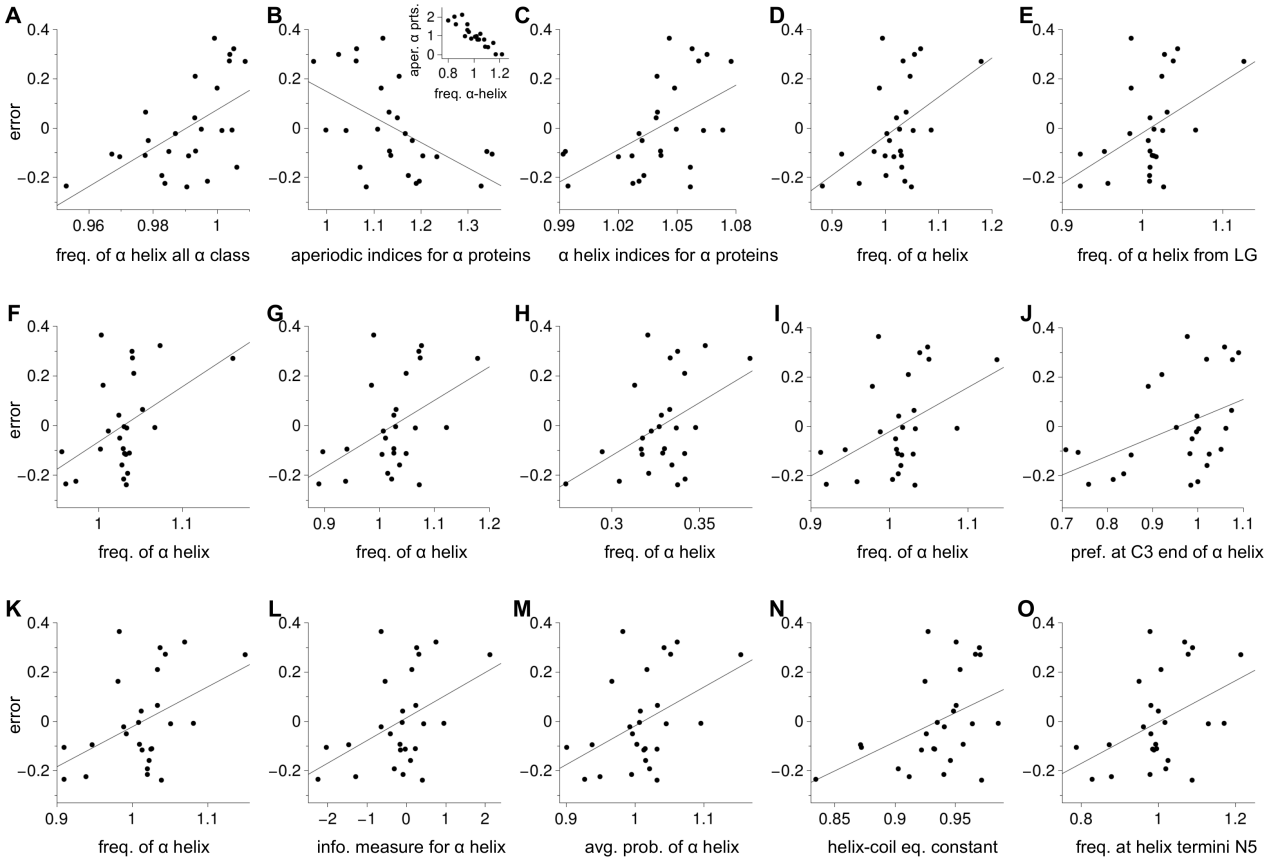

Figure S5. Correlation of equation 6 error to each α helix scale that is listed in Table S4. Panel order from **A** through **O** follows Table S4 order from top to bottom. The aperiodic indices for α proteins[105] in panel **B** represents the preference for non-regular (i.e., aperiodic) structure in α proteins and shows a strong anti-correlation to α helix propensity scales. For example, the inset compares the normalized frequency of α helix scale[104] (x-axis) from panel **A** to the aperiodic indices for α proteins scale (y-axis), which correlates by $R^2 = 0.826$. Thus, decreases in aperiodic indices represent increases in α helix propensity. This anti-correlation is why the trend in **B** is opposite to trends in panels **A** and **C** through **O**.