PROTEINS: Structure, Function, and Genetics 41:415–427 (2000)

Why Are "Natively Unfolded" Proteins Unstructured Under Physiologic Conditions?

Vladimir N. Uversky,[1,2]* Joel R. Gillespie,[1] and Anthony L. Fink[1]
[1] *Department of Chemistry and Biochemistry, University of California, Santa Cruz, California*
[2] *Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region, Russia*

ABSTRACT **"Natively unfolded" proteins occupy a unique niche within the protein kingdom in that they lack ordered structure under conditions of neutral pH in vitro. Analysis of amino acid sequences, based on the normalized net charge and mean hydrophobicity, has been applied to two sets of proteins: small globular folded proteins and "natively unfolded" ones. The results show that "natively unfolded" proteins are specifically localized within a unique region of charge-hydrophobicity phase space and indicate that a combination of low overall hydrophobicity and large net charge represent a unique structural feature of "natively unfolded" proteins. Proteins 2000;41:415–427.**

Key words: natively unfolded protein; unstructured protein; disordered protein; α-synuclein

INTRODUCTION

The number of proteins and protein domains that have been shown to have little or no ordered structure under physiologic conditions has increased exponentially over the past 10 years, starting from 1 study in 1989 and ending with 22 in the past year. The special term "natively unfolded" has been introduced as a descriptive term for such proteins.[1] Since the classic experiments of Anfinsen,[2] it is generally accepted that all the necessary information for the correct folding of a protein is included in its amino acid sequence. From this point of view, the absence of regular structure at physiologic conditions raises the general question, "are there particular features of amino acid sequence that are responsible for the lack of ordered structure in "natively unfolded" proteins?"

One distinction of the amino acid sequences of "natively unfolded" proteins has been suggested in the literature, namely, the presence of numerous uncompensated charged groups (often negative), i.e., a large net charge at neutral pH, arising from the extreme pI values in such proteins.[3–5] A low content of hydrophobic amino acid residues has been also noted for several "natively unfolded" proteins.[3,4] The purpose of this study is to show that it is possible to predict whether a given amino acid sequence encodes a native (folded) or a "natively unfolded" protein on the basis of sequence alone.

MATERIALS AND METHODS

An initial set of 275 protein sequences was chosen from the SWISS-PROT protein sequence data bank and its supplement TrEMBL.[6] Selection criteria included that the candidate sequence coded for a small globular monomeric protein of 50 to 200 amino acid residues, with no disulfide bonds and with no known interaction either with natural ligands or with membranes. Such proteins have been combined into the first set, native proteins. A second group of 91 proteins (or their isolated domains), which at physiologic conditions have been reported to have the NMR chemical shifts of a random-coil, and/or lack significant ordered secondary structure (as determined by CD or FTIR), and/or show hydrodynamic dimensions close to those typical of an unfolded polypeptide chain, have been combined into the set of "natively unfolded" proteins.

By using the Swiss Institute of Bioinformatics (SIB) server ExPASy[7] the following information was extracted for each individual protein: (1) number of amino acid residues, (2) molecular mass, (3) total number of negatively (Asp + Glu) and positively charged (Arg + Lys) residues, and (4) theoretical pI value. The hydrophobicity of each amino acid sequence was calculated by the Kyte and Doolittle approximation,[8] by using a window size of 5 amino acids. The hydrophobicity of individual residues was normalized to a scale of 0 to 1 in these calculations. The mean hydrophobicity is defined as the sum of the normalized hydrophobicities of all residues divided by the number of residues in the polypeptide. The mean net charge is defined as the net charge at pH 7.0, divided by the total number of residues.

RESULTS

The current list of 91 known "natively unfolded" proteins and their major characteristics are presented in Table I. Figure 1 represents the length distribution of "natively unfolded" proteins. The largest number (32) of "natively unfolded" proteins have between 50 and 100 residues. However, the histogram clearly shows that there is no consistency in the size of the "natively unfodled" proteins: they range from ~50 to 1,827 residues. Table I also shows that the net charge at pH 7.0 may be as high as +59, or as low as −117, or close to zero. The distribution of "natively unfolded" proteins within the scale of pI values is presented in Figure 2.

Grant sponsor: NATO Collaborative Linkage Grant; Grant sponsor: National Institutes of Health; Grant sponsor: National Parkinson Foundation.

*Correspondence to: Vladimir Uversky, Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064. E-mail: uversky@hydrogen.ucsc.edu

Received 11 April 2000; Accepted 21 July 2000

TABLE I. Major Characteristics of the "Natively Unfolded" Proteins

Protein	Evidence for "unfoldedness"	Length (a. a. r.)	Net charge	Hydrophobicity Total	Hydrophobicity Mean	Ligand/substrate	Ref.
a) Proteins for which "unfoldedness" has been established by NMR or multiparametric approach							
Heat shock transcription factor, N-terminal activation domain (*K. lactis*)	Heteronuclear NMR, far UV CD	194	−11	80.707	0.4248	DNA	34
Heat shock transcription factor, N-terminal activation domain (*S. cerevisiae*)	Heteronuclear NMR, far UV CD	167	−3	66.577	0.4083	DNA	34
4E-binding protein I	Heteronuclear NMR, far UV CD	118	−2	47.554	0.4171	eIF4E	35
4E-binding protein II	Heteronuclear NMR, far UV CD	120	−2	48.766	0.4204	eIF4E	35
Cyclin-dependent kinase inhibitor $p21^{Waf1/Cip1/Sdi1}$	Heteronuclear NMR, Proteolytic mapping, far UV CD	164	+3	60.369	0.4086	Cdk2, Zn^{2+}	36
Cyclin-dependent kinase inhibitor $p21^{Waf1/Cip1/Sdi1}$, N-terminal domain	Heteronuclear NMR, Proteolytic mapping, far UV CD	94	0	36.342	0.4038	Cdk2, Zn^{2+}	36
SNase, Δ131Δ fragment	Heteronuclear NMR, far UV CD	131	+16	52.676	0.4148	DNA, Ca^{2+}	21
Dessication-related protein	Heteronuclear NMR, proteolytic mapping, far UV CD	155	+6	56.023	0.3710		37
eIF4G1, functional domain (393–490)	Heteronuclear NMR, proteolytic mapping	98	−7	40.241	0.4281	eIF4E	38
Anti-sigma factor FlgM	Heteronuclear NMR	97	−1	38.288	0.4117	σ^{28}	39
Synaptobrevin, cytoplasmic domain (1–96)	Heteronuclear NMR	96	+1	37.867	0.4116	Syntaxin	40
Prion protein, N-terminal part	Heteronuclear NMR	96	+6	33.704	0.3663	Prion protein	41
LEF-1, C-terminal HMG domain	Heteronuclear NMR	101	+12	37.834	0.3900	DNA	42
TAF_{II}-$230_{11–77}$, N-terminal region	Heteronuclear NMR	67	−13	30.273	0.4805	TBP	43
Antitermination protein N	Heteronuclear NMR	107	+14	40.868	0.3968	RNA	44
Snc1, cytoplasmic domain	Heteronuclear NMR	93	0	34.796	0.3910	Sso1, Sec9	45
Prothymosin α	^{1}H-NMR, far UV CD, gel filtration, SAXS	110	−44	29.442	0.2778	Nucleic acids, Zn^{2+}	4,27
Nonhistone chromosomal protein HMG-17	^{1}H-NMR, far UV CD, FTIR, SAXS	89	+12	26.139	0.3075	DNA	15
Fibronectin-binding domain, D-type, *S. aureus*	^{1}H-NMR, far UV CD	134	−22	45.403	0.3493	Fibronectin	46
GCN4, DNA-binding domain	^{1}H-NMR, far UV CD	57	+7	19.739	0.3724	DNA	47
β-Tubulin, 394–445 fragment	^{1}H-NMR, far UV CD	52	−19	16.443	0.3426	GTP	48

TABLE I. (Continued)

Protein	Evidence for "unfoldedness"	Length (a. a. r.)	Net charge	Hydrophobicity Total	Hydrophobicity Mean	Ligand/substrate	Ref.
Nonhistone chromosomal protein HMG-14	^{1}H-NMR, far UV CD	100	+6	30.760	0.3204	DNA	14
Nonhistone chromosomal protein HMG-T	^{1}H-NMR, far UV CD	204	−1	67.772	0.3389	DNA	49
Nonhistone chromosomal protein HMG-H6	^{1}H-NMR, far UV CD	69	+14	24.848	0.3823	DNA	49
EMB-1 protein	^{1}H-NMR	92	+1	30.680	0.3478	Unknown	50
Negative factor, NEF protein	^{1}H-NMR	206	−4	87.605	0.4337	GTP	51
Osteocalcin	^{1}H-NMR	49	−6	18.846	0.4188	Ca^{2+}	16
Neutral zinc finger factor 1, two-domain fragment (487–606)	^{1}H-NMR	120	+8	44.987	0.3878	DNA, Zn^{2+}	52
Microtubule-associated protein tau	Far UV CD, FTIR, SAXS	441	+2	175.806	0.4023	Tubulin	1
Caldesmon	Electron microscopy, far UV CD, gel-filtration	771	−14	219.899	0.2867	Myosin, calmodulin, actin, tropomyosin	53
Microtubule-associated protein MAP2	Far UV CD, sedimentation, gel filtration	1,827	−117	738.133	0.4049	Microtubules	54
Secretogranin (chromatogranin B)	Far UV CD, gel filtration, *bis*-ANS	657	−53	207.051	0.3171	Ca^{2+}	20
α-Synuclein	Far UV CD, FTIR, sedimentation, gel filtration	140	−10	61.761	0.4541	Unknown	5
α_s-Casein	Far UV CD, FTIR, SAXS	199	−12	82.529	0.4232	Ca^{2+}	55
Vmw65, carboxyl-terminal transactivation domain	Far UV CD, gel filtration	87	−19	37.785	0.4552	DNA	56
DNA topoisomerase I, N-terminal domain	Far UV CD, gel filtration, sedimentation	197	+15	43.468	0.2252	DNA	57
Chromogranin A	Far UV CD, *bis*-ANS fluorescence	439	−51	157.216	0.3614	Ca^{2+}	19
Glucocorticoid receptor, 77-262 fragment	Far UV CD, Trp fluorescence	186	−16	77.863	0.4278	DNA, Zn^{2+}	25
Extracellular domain of the low affinity NGF receptor	Far UV CD, Trp fluorescence	222	−27	96.727	0.4437	NGF, membrane	58
cAMP-dependent protein kinase inhibitor (PKI-alpha)	Far UV CD, FTIR	75	−7	29.837	0.4202	Kinase	59
SdrD protein, B1–B5 fragment	Far UV CD, Trp fluorescence	562	−48	219.645	0.3936	Ca^{2+}	18
Caldesmon, 636–771 fragment	Far UV CD, Trp fluorescence, DSC	136	+9	53.399	0.4045	Calmodulin	Uversky, personal communication
Soluble transducer HtrX	Far UV CD, Trp fluorescence	489	−82	222.974	0.4597	Transducers	60

TABLE I. (Continued)

Protein	Evidence for "unfoldedness"	Length (a. a. r.)	Net charge	Hydrophobicity Total	Hydrophobicity Mean	Ligand/substrate	Ref.
Soluble transducer HtrXI	Far UV CD, Trp fluorescence	451	−80	194.144	0.4343	Transducers	60
Salivary proline-rich glycoprotein	Far UV CD, dansyl fluorescence	293	+12	82.658	0.2860		61
MAX-protein	Far UV CD, proteolitic mapping, mass spectrometry	163	+4	55.792	0.3509	DNA	13
EM-protein	Far UV CD, gel filtration	93	−1	30.496	0.3427		62
Apo-cytochrome c	Far UV CD, gel filtration	104	+9	40.266	0.4027	Heme	23
Cardiac muscle troponin I	Far UV CD, epitope mapping, Trp fluorescence	209	+14	80.978	0.3950	Actin, troponin C	63
Dopamine- and cAMP-regulated neuronal phosphoprotein, DARPP-32	Gel filtration, sedimentation	202	−23	67.594	0.3414	CAMP, dopamine	3
b) Proteins for which "unfoldedness" has been shown by one direct physicochemical method							
Fibronectin-binding domain PAQ8, *S. dysgalactiae*	Far UV CD	157	−32	62.076	0.4057	Fibronectin	64
Fibronectin-binding domain A-type, *S. dysgalactiae*	Far UV CD	124	−25	49.414	0.4118	Fibronectin	64
Fibronectin-binding domain B-type, *S. dysgalactiae*	Far UV CD	112	−25	43.234	0.4003	Fibronectin	64
Fibronectin-binding domain P-type, *S. pyogenes*	Far UV CD	143	−22	53.993	0.3884	Fibronectin	64
Sec9, SNAP-25-like domain	Far UV CD	251	−4	88.982	0.3603	Sso1, Snc1	65
Thyroid transcription factor-1, N-terminal domain	Far UV CD	156	+15	66.941	0.4404	TATA-binding protein	66
GAGA factor, glutamine-rich domain	Far UV CD	383	+6	152.889	0.4031	DNA	67
GAGA factor, central domain	Far UV CD	307	+7	125.169	0.4131	DNA	67
Small proline-rich protein 2	Far UV CD	72	+6	24.560	0.3612	Transglutaminase	68
CREB, truncated form (ACT265)	Far UV CD	265	−9	109.593	0.4199	DNA	69
HPV16 E7 protein	Far UV CD	98	−14	41.425	0.4407	Zn^{2+}	70
Histidine-rich Protein II	Far UV CD	327	−30	132.964	0.4117	Hem	26
Calsequestrin	Far UV CD	367	−80	158.289	0.4361	Ca^{2+}	71
Reduced RNase T1	Far UV CD	104	−10	43.789	0.4389	RNA	72
Manganese stabilizing protein of photosystem II, 1–244 fragment	Far UV CD	244	−5	101.626	0.4165	Mn^{2+}	73
HIV-1 integrase, N-terminal domain	Far UV CD	55	−4	33.698	0.4260	Zn^{2+}	74

TABLE I. (Continued)

Protein	Evidence for "unfoldedness"	Length (a. a. r.)	Net charge	Hydrophobicity Total	Hydrophobicity Mean	Ligand/substrate	Ref.
Bob-1, N-terminal domain	Far UV CD	65	+6	26.557	0.4354	Oct-1 POU, DNA	75
SPARC, BM-40, osteonectin	Far UV CD	285	−28	122.691	0.4266	Ca^{2+}	17
Human protamine 2	Far UV CD	102	+26	26.600	0.2612	DNA	12
Histone H1	Far UV CD	224	+59	94.062	0.4276	DNA, ATP, Ca^{2+}	22
Hirudin	Far UV CD	66	−7	24.761	0.3991	α-Thrombin	76
Myelin-binding protein	Far UV CD	196	+24	75.208	0.3917	Lipids	77
RNase HI, C-terminal domain (88–155)	Far UV CD	68	−1	25.137	0.3928	N-terminal domain	78
Parathyroid hormone-related protein	Far UV CD	141	+10	49.395	0.3606	Ca^{2+}	24
50S ribosomal protein L2	Far UV CD	272	+33	112.739	0.4207	Ribosome	79
50S ribosomal protein L27	Far UV CD	84	+11	34.329	0.4291	Ribosome	79
50S ribosomal protein L31	Far UV CD	70	+6	28.752	0.4356	Ribosome	79
50S ribosomal protein L32	Far UV CD	56	+9	19.349	0.3721	Ribosome	79
50S ribosomal protein L33	Far UV CD	54	+10	20.081	0.4016	Ribosome	79
30S ribosomal protein S12	Far UV CD	123	+21	50.761	0.4216	Ribosome	79
30S ribosomal protein S18	Far UV CD	74	+12	29.271	0.4182	Ribosome	79
30S ribosomal protein S19	Far UV CD	91	+12	37.558	0.4317	Ribosome	79
30S ribosomal protein S21	Far UV CD	70	+14	24.017	0.3639	Ribosome	79
Sperm protamine P3	Far UV CD	103	−20	32.252	0.3291	DNA, RNA	11
Galline	Far UV CD	61	+36	10.253	0.1799	DNA	80
Helix destabilizing protein (Pf1 gene 5 protein)	Gel filtration	144	0	62.566	0.4469	DNA	81
β-Synuclein, phosphoneuroprotein 14, 14 kDa brain-specific protein	Gel filtration	134	−15	56.967	0.4382	Unknown	82
Protein phosphatase inhibitor-1	Gel filtration	171	−3	63.150	0.3781	Phosphatase	83
Stathmin	Gel filtration	148	−4	50.463	0.3504	Tubulin	84
cGMP-dependent protein kinase inhibitor	Gel filtration, thermal stability	155	+6	57.926	0.3836	Kinase	85
c) Proteins for which "unfoldedness" has been shown by indirect method							
Gelsolin, 173–245 peptide	Proteolysis	73	−4	27.701	0.4015	Actin	86

Although this distribution has two definite maxima at pH ~4 and ~10.5, the isoelectric points cover a very wide range (between pH 3 and 13). Thus, parameters such as length of polypeptide, net charge, or pI cannot be used as a signature of "natively unfolded" proteins. For instance, Figure 3a shows that there is no essential difference between native and "natively unfolded" proteins in terms of mean net charge as a function of sequence length. Statistical analysis shows that small globular native proteins and "natively unfolded" proteins are characterized by mean net charges of 0.04 ± 0.04 and 0.12 ± 0.09, respectively. This finding suggests that a high value of net charge may represent a necessary, but not sufficient, condition for the given protein to be "natively unfolded."

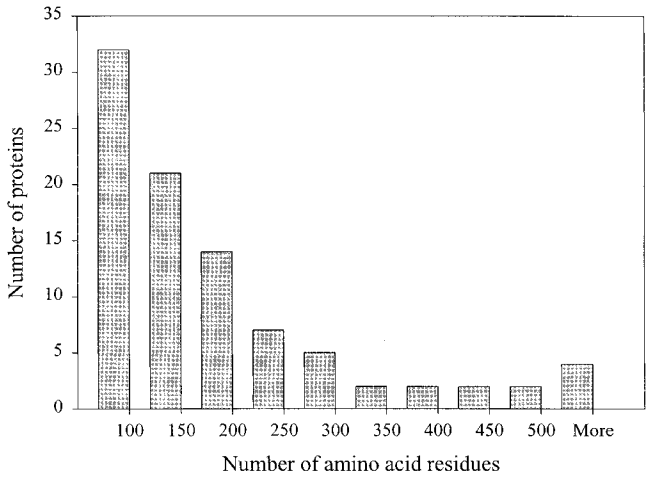

Fig. 1. Histogram represents the distribution of "natively unfolded" proteins as a function of length of amino acid sequences. The range of polypeptide length covered by each bar (with the exception for last one) is equal to 50 residues.

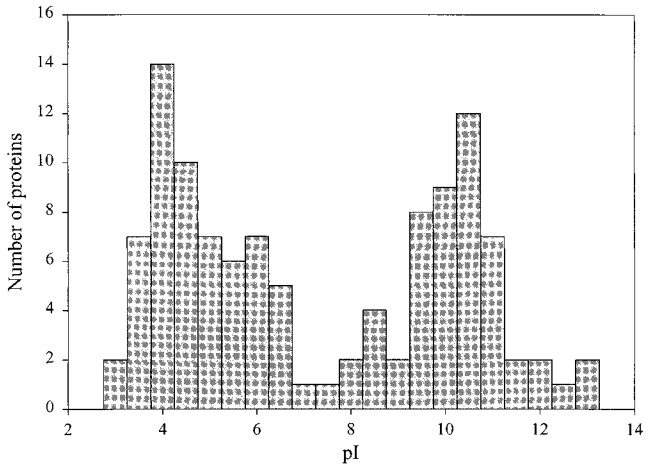

Fig. 2. Histogram, representing the distribution of "natively unfolded" proteins as a function of isoelectric point. Centers of bars are located with a step size of 0.5 pH unit, the range of pI covered by each bar is equal to ±0.25 pH unit from the central position.

A different situation is observed when the hydrophobicity of the polypeptide chain is taken into account. Figure 3b shows that, in this case, the set of "natively unfolded" proteins is relatively well separated from the set of native proteins. Statistical analysis of these data gives a mean hydrophobicity of 0.48 ± 0.03 and 0.39 ± 0.05 for native and "natively unfolded" proteins, respectively. This indicates that the mean hydrophobicity is a significant contributing factor in determining whether a protein will be folded or unfolded.

Figure 3c illustrates that consideration of both factors (mean hydrophobicity and mean net charge) allows us to reliably separate the "natively unfolded" proteins from native ones. Indeed, Figure 3c shows that these two classes of proteins occupy different areas within the charge-hydrophobicity phase space. It also can be seen that three "natively unfolded" proteins, α-synuclein, negative factor (NEF) and helix destabilizing protein (in Fig. 3c marked as green, yellow, and white circles, respectively) do not fit the general trend. Amino acid sequence analysis of these proteins has established that their N- and C- terminal regions are very distinct in overall hydrophobicity and possess charges of opposite sign (see Table II). In particular, the C-terminal 44 residues of α-synuclein, the N-terminal 56 residues of NEF and the C-terminal 35 residues of helix destabilizing protein have parameters typical of "natively unfolded" proteins, whereas the hydrophobicity and charge of the major part (~70%) of all three sequences are typical of native proteins. In Figure 3c, the data points corresponding to the C-terminal part of α-synuclein, the N-terminal part of NEF and the C-terminal part of helix destabilizing protein are marked as green, yellow, and white triangles, respectively. It is conceivable that the disordered regions of these molecules prevent the remainder of the protein from normal folding, perhaps through extensive electrostatic attractions.

The set of 91 proteins described in the literature as "natively unfolded," have at least 242 homologues, that are also expected to be natively unfolded, based on their relative net mean charge and mean hydrophobicity (light blue circles in Fig. 3d). Unfortunately, at present there are limited experimental data reported to confirm these predictions. In addition, by analysis of the Swiss protein database, we were able to find 130 different, nonhomologous proteins with sequences sharing low mean hydrophobicity and relatively high net charge (green symbols in Fig. 3d; Table III). We predict that these proteins will be natively unfolded. Because many of the proteins shown by green symbols in Figure 3d also have numerous homologues, the actual number of predicted natively unfolded proteins is much larger.

DISCUSSION

Our data are consistent with the conclusion that the combination of low mean hydrophobicity and relatively high net charge represent an important prerequisite for the absence of regular structure in proteins under physiologic conditions, thus leading to "natively unfolded" proteins. Many globular proteins are unfolded by extremes of pH,[9] and substantial evidence indicates that this is caused by charge-charge repulsion.[10] However, some globular proteins do not unfold under conditions of extreme pH.[9] It is likely that the outcome is determined by the balance in the competition between the charge repulsion driving unfolding and hydrophobic interactions driving folding. Thus, the situation is analogous to that with natively unfolded proteins. This hypothesis is supported by preliminary analysis of mean net charge against mean hydrophobicity at pH 2 for a limited number of proteins (data not shown). Analogously, a limited sampling of small globular proteins containing disulfide bonds, that are known to be unfolded on reduction of the disulfide, were shown to belong to the natively unfolded charge-hydrophobicity phase space when analyzed in their reduced state.

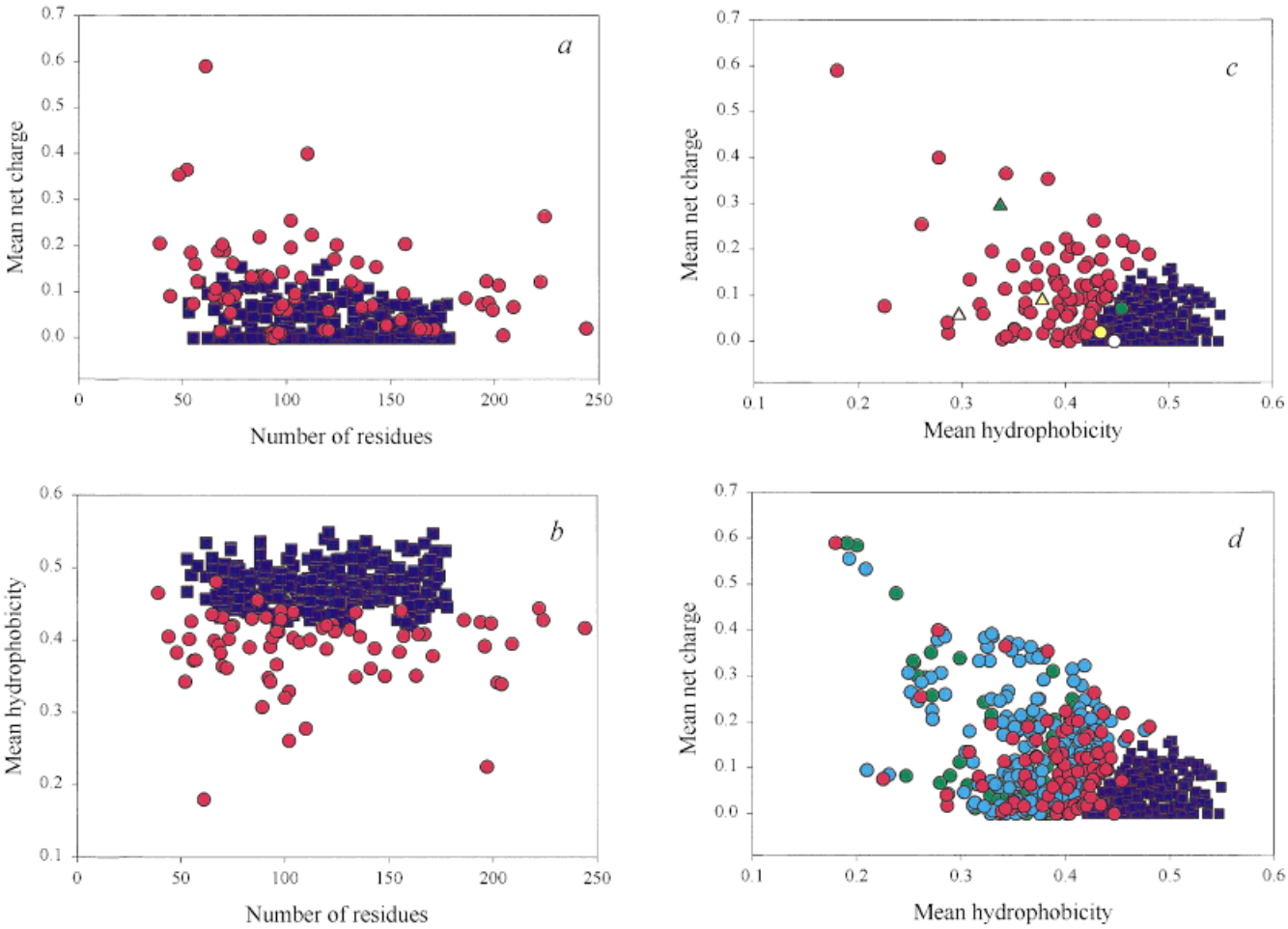

Fig. 3. Comparison of the mean net charge and the mean hydrophobicity for the set of 275 folded (blue squares) and 91 natively unfolded proteins (red circles). Data are presented as **a** dependence of the mean net charge on the length of polypeptide chain, **b** length-dependence of the mean hydrophobicity, **c** mean net charge vs. mean hydrophobicity (data for α-synuclein, negative factor and helix destabilizing protein are shown as green, yellow, and white circles, respectively; data for "natively unfolded" fragments of these proteins are presented by triangles of corresponding color, see text for explanation). In (**d**), the 242 homologues of the natively unfolded proteins are shown as cyan circles, and the 130 predicted natively unfolded proteins are shown as green circles.

TABLE II. Major Characteristics of α-Synuclein, Negative Factor, and Helix Destabilizing Protein and Their N- and C-Terminal Domains

					Hydrophobicity	
Protein	Length (a. a.)	M (Da)	pI	Net charge	Total	Mean
α-Sunuclein	140	14,460.1	4.44	−10	61.761	0.4541
α-Synuclein, N-terminal part	96	9,519.9	9.44	+4	46.893	0.5097
α-Synuclein, C-terminal part	44	4,958.2	3.40	−13	13.471	0.3368
Negative factor, F-protein, NEF protein, 3′ORF	206	23,366.3	6.14	−4	87.605	0.4337
Negative factor, N-terminal part	56	6,220.0	11.14	+5	19.614	0.3772
Negative factor, C-terminal part	150	17,628.9	5.58	−7	62.473	0.4497
Helix destabilizing protein (Pf1 gene 5 protein)	144	15,428.4	6.25	0	63.961	0.4569
Pf1 gene 5 protein, N-terminal part	109	11,765.4	4.99	−2	52.783	0.5027
Pf1 gene 5 protein, C-terminal part	35	3,681.0	9.69	+2	9.209	0.2971

It is known that unfolded proteins normally have very short lifetimes in the cell, thus, it is most probable that "natively unfolded" proteins are significantly folded in their normal cellular milieu. Table I shows that "natively unfolded" proteins in vivo are likely to be stabilized by binding of specific targets, ligands (such as a variety of

TABLE III. Proteins Predicted to be "Natively Unfolded"

Protein	Length (a. a.)	M (Da)	pI	Net charge	Hydrophobicity Total	Hydrophobicity Mean
Early E1A 11-kDa protein (*mouse adenovirus type 1*)	96	11,133.2	4.21	−14	35.322	0.3839
Fruit protein PKIWI501 (*Actinidia chinensis, kiwi*)	184	18,920.8	3.55	−46	73.130	0.4063
Male-enhanced antigen-1 (*Bos taurus*)	174	18,751.3	3.79	−27	70.770	0.4163
Male-enhanced antigen-1 (*Homo sapiens*)	172	18,544.0	3.79	−27	69.748	0.4152
Male-enhanced antigen-1 (*Mus musculus*)	174	18,584.0	3.78	−28	70.749	0.4162
Acidic protein MSYB (*Escherichia coli*)	124	14,259.1	3.40	−29	49.155	0.4096
Salivary acidic proline-rich phosphoprotein 1/2 (*Homo sapiens*)	150	15,372.4	4.40	−7	44.194	0.3027
Acidic proline-rich protein HP43A (*Mesocricetus auratus*)	169	18,298.1	4.16	−14	40.771	0.2471
DNA-directed RNA polymerase II 14.4-kDa polypeptide (*Caenorhabditis elegans*)	137	15,987.4	3.75	−23	51.563	0.3877
16.5-kDa Submandibular gland glycoprotein (*Mus musculus*)	119	12,893.9	3.60	−24	37.756	0.3283
Sulfated 50-kDa glycoprotein (*Mus musculus*)	132	14,154.3	4.23	−10	52.288	0.4080
Trophoblast-specific protein (*Mus musculus*)	106	12,117.3	3.95	−16	41.035	0.4023
Gene 3 protein (*spiroplasma virus 4*)	149	17,314.0	4.25	−14	54.725	0.3774
Head morphogenesis protein (*bacteriophage B103*)	101	11,704.0	4.25	−13	38.675	0.3997
Variant surface antigen A (*Mycoplasma hyorhinis*)	128	12,428.1	4.16	−5	40.648	0.3278
Variant surface antigen B (*Mycoplasma hyorhinis*)	145	13,099.7	4.03	−5	50.386	0.3573
Variant surface antigen D (*Mycoplasma hyorhinis*)	139	13,512.1	3.94	−8	45.217	0.3349
Variant surface antigen F (*Mycoplasma hyorhinis*)	145	14,040.9	3.47	−9	43.163	0.3061
Endocuticle Structural glycoprotein ABD-5 (*Locusta migratoria*)	82	8,828.5	4.19	−5	31.456	0.4053
C-terminal extension peptide (*Bombina orientalis*)	62	7,187.8	4.25	−5	22.904	0.3929
Procyclic form specific polypeptide B1-alpha (*Trypanosoma brucei brucei*)	94	10,045.3	3.53	−31	22.941	0.2549
Procyclic form specific polypeptide (*Trypanosoma brucei brucei*)	66	6,878.1	3.74	−17	16.865	0.2720
Brazzein (*Pentadiplandra brazzeana*)	54	6,498.3	7.91	+1	19.291	0.3858
Cation transport regulator CHAB (*E. coli*)	76	8,944.7	8.10	+1	24.190	0.3360
Disagregin (soft tick)	60	6,961.4	7.85	0	18.860	0.3368
EM-like protein GEA6 (*Arabidopsis thaliana*)	92	9,933.9	7.90	+1	29.897	0.3397
Late embryogenesis abandon protein (*Zea mays*)	91	9,683.6	6.61	−1	30.823	0.3543
Progonadoli-berin I (*Xenopus laevis*)	66	7,724.8	6.00	−1	25.626	0.3975
Matrix GLA-protein (MGP) (*Homo sapiens*) P08493	77	9,531.5	8.66	+2	26.743	0.3663
Late seed maturation protein P8B6 (*Raphanus sativus*)	83	8,9980.7	7.77	+1	24.748	0.3133
Tegument phosphoprotein (*Simian varicella virus*)	77	8,997.0	7.40	0	27.794	0.3807
Abscisic stress ripening protein 1 (*Lycopersicon esculentum*)	115	13,129.7	6.83	−1	40.207	0.3622
Abscisic stress ripening protein 2 (Tomato)	114	13,019.5	6.79	−2	39.209	0.3564
HMG1/2-like protein (*Triticum aestivum*)	161	17,213.8	7.28	0	56.921	0.3626
Pre-intermoult gene-1 protein (*Drosophila melanogaster*)	187	19,696.8	8.98	+2	69.529	0.3799
Putative RNA-binding protein (*Mus musculus*)	153	16,604.7	7.72	0	56.981	0.3824
Vacuolar ATP synthase subunit G (*Neurospora crassa*)	115	13,037.6	7.93	+1	42.385	0.3818
Dehydrin RAB18 (*Arabidopsis thaliana*)	186	18,463.8	7.93	+11	66.808	0.3671
50S ribosomal protein L29 (*Mycoplasma capricolum*)	138	15,637.7	7.60	+2	53.427	0.3987
Nucleic acid binding protein P15 (HTLV-1)	89	9,870.3	9.67	+7	33.664	0.3960
Water-stress inducible protein (rice)	163	16,543.2	9.83	+6	60.404	0.3799
Late embryogenesis abundant protein (upland cotton)	145	15,810.0	6.87	−1	49.219	0.3491
Cellular nucleic acid binding protein	177	19,462.2	9.65	+9	69.668	0.4027
ARS-binding factor 2 (*Saccharomyces cerevisiae*)	157	18,622.3	10.02	+12	57.119	0.3733
cAMP-regulated phosphoprotein 16 (*Bos taurus*)	96	10,665.2	10.09	+7	34.029	0.3699
BAD protein (*Homo sapiens*)	168	18,408.1	10.50	+5	59.406	0.3622
Prochole-cystokinin (*Homo sapiens*)	95	10,750.0	10.00	+3	35.354	0.3883
CCAAT/enhancer binding protein gamma (*Homo sapiens*)	150	16,408.4	10.22	+7	57.424	0.3933
Cold-inducible RNA-binding protein (glycine-rich RNA-binding proteinde cirp) (*Homo sapiens*)	172	18,648.0	9.74	+5	64.306	0.3828
Gastrula-specific protein 17 (*Xenopus laevis*)	147	16,901.7	10.58	+9	51.084	0.3572
GVPI protein, plasmid (*Halobacterium salinarium*)	144	16,258.7	10.44	+9	43.062	0.3076
Histone H2B.1 (*Homo sapiens*)	125	13,819.0	10.70	+18	50.651	0.4186
Histone H2B.2 (H2B/N) (*Homo sapiens*)	125	13,775.9	10.63	+16	50.892	0.4206
Histone H2B-III (*Volvox carteri*)	157	17,010.7	10.37	+18	63.204	0.4131
Histone H2B-IV (*Volvox carteri*)	155	16,841.5	10.32	+17	61.530	0.4075
Chromosomal protein MC1 (HMB) (*Methanosarcina barkeri*)	93	10,755.3	10.50	+8	33.152	0.3725

TABLE III. (Continued)

Protein	Length (a. a.)	M (Da)	pI	Net charge	Hydrophobicity Total	Hydrophobicity Mean
HMG-Y related protein A (SB16A protein) (*Glycine max*)	176	18,841.6	10.82	+18	65.627	0.3816
Mobility group protein 1B (*Chironomus tentans*)	110	12,149.6	10.15	+9	36.631	0.3456
High mobility group protein HMG-I (*Homo sapiens*)	106	11,544.8	10.82	+12	30.443	0.2985
Homeobox protein HOX-C6 (HOX-3C) (HHO.C8) (CP25) (*Homo sapiens*)	153	17,852.0	10.28	+8	54.394	0.3651
Thermonuclease (*Staphylococcus hyicus*)	143	16,670.0	10.06	+10	52.742	0.3794
Activated RNA polymerase II transcriptional coactivator P15 (*Homo sapiens*)	126	14,264.1	10.08	+7	45.906	0.3763
18-kDa seed maturation protein (*Glycine max*)	173	17,606.3	9.98	+6	65.894	0.3899
Serum amyloid A-3 protein (*Mus musculus*)	103	11,764.0	10.41	+8	36.469	0.3684
Retinoid X receptor α, DNA-binding domain (130–212) (*Homo sapiens*)	83	9,827.3	9.48	+11	30.812	0.3900
Retinal rod rhodopsin-sensitive cGMP 3′,5′-cyclic phosphodiesterase gamma-subunit (*Bos taurus*)	87	9,669.2	9.52	+4	34.819	0.4195
Small, acid-soluble spore protein gamma-type (*Thermoactinomyces thalpophilus*)	96	10,683.3	10.74	+3	30.861	0.3354
Signal recognition particle 14-kDa protein (SRP14) (*Arabidopsis thaliana*)	121	13,777.1	10.65	+17	46.629	0.3985
Transposase for transposon TN554 (*Staphylococcus aureus*)	125	14,804.3	10.51	+15	47.143	0.3896
Protein Z600 (*Drosophila melanogaster*)	90	10,507.0	10.31	+10	31.334	0.3643
Abscisic stress ripening protein 3 (*Lycopersicon esculentum*)	78	8,910.2	10.13	+6	25.346	0.3425
Coleoptericin (*Zophobas atratus*)	74	8,109.9	10.50	+5	24.938	0.3563
Hymenoptaecin (*Apis mellifera*)	93	10,286.5	10.15	+6	35.067	0.3940
Small, acid-soluble spore protein gamma-type (SASP) (*Bacillus megaterium*)	96	10,306.9	10.17	+5	31.378	0.3411
Alpha-inhibin-92 (*Homo sapiens*)	92	10,355.2	10.23	+6	27.944	0.3175
Sex-lethal protein, male-specific (*Drosophila melanogaster*)	48	5,601.0	10.01	+4	12.729	0.2893
Gene 0.6 protein (*bacteriophage t7*)	111	13,235.5	11.72	+20	37.885	0.3541
Dehydrin (*barley*)	139	14,235.6	9.60	+5	50.273	0.3838
Sperm-specific protein PHI-2B (*Mytilus californianus*)	148	15,688.4	13.14	+46	55.881	0.3881
Spermatid nuclear transition protein 1 (STP-1) (TP-1) (*Rattus norvegicus*)	54	6,264.3	12.47	+19	13.558	0.2712
Nuclear transition protein 2 (TP-2) (*Homo sapiens*)	78	8,912.3	11.82	+19	23.837	0.3221
Nonhistone chromosomal protein H6 (Histone T) (*Oncorhynchus mykiss*)	69	6,955.9	11.24	+14	24.848	0.3823
Nonhistone chromosomal protein HMG-14A (*Gallus gallus*)	104	11,225.3	10.00	+7	27.996	0.2797
Nonhistone chromosomal protein HMG-14B (*Gallus gallus*)	102	11,048.2	10.15	+7	32.145	0.3280
Nonhistone chromosomal protein HMG-14 (*Homo sapiens*)	99	10,527.6	10.12	+7	31.241	0.3289
Spermatid-specific protein T1 (*Sepia officinalis*)	78	10,631.8	13.03	+46	14.137	0.1910
Dehydrin RAB 15 (*Triticum aestivum*)	149	15,766.4	10.36	+8	52.504	0.3621
Dessication-related protein clone PCC6-19 (CDET6-19) (*Craterostigma plantagineum*)	155	15,568.0	10.07	+6	56.023	0.3710
Spermatid-specific protein T2 (*Sepia officinalis*)	77	10,485.6	13.02	+45	14.629	0.2004

small molecules, substrates, cofactors, other proteins, nucleic acids, membranes, etc.). Moreover, for the majority of proteins listed, the existence of pronounced ligand-induced folding has been established. Examples include DNA (or RNA) induced structure formation in protamines,[11,12] Max protein,[13] high mobility group proteins HMG-14[14] and HMG-17;[15] cation-induced folding of ostecalcine,[16] osteonectine,[17] SDRD protein,[18] chromatogranins A[19] and B,[20] Δ131Δ fragment of SNase,[21] histone H1,[22] and protamine;[12] folding of cytochrome *c* in the presence of heme;[23] membrane-induced secondary structure formation in parathyroid hormone related protein;[24] trimethylamine N-oxide induced structure formation in glucocorticoid receptor;[25] heme-induced folding of histidine-rich protein II;[26] zinc-mediated structure formation and compaction of prothymosin-α,[27] and many others (see Table I). Consequently, in contrast to in vitro experiments with purified protein, "natively unfolded" proteins probably have considerable structure in vivo as the result of their interaction with their natural "ligands."[28]

Our concept is that the combination of low mean hydrophobicity and high net charge leads to "natively unfolded" conformation. This suggests that any interaction of "natively unfolded" protein with natural ligand that will affect its mean net charge, mean hydrophobicity, or both, may change these parameters in such a way that they will approach those typical of folded native proteins. Unfortunately, the very attractive idea of calculating the joint

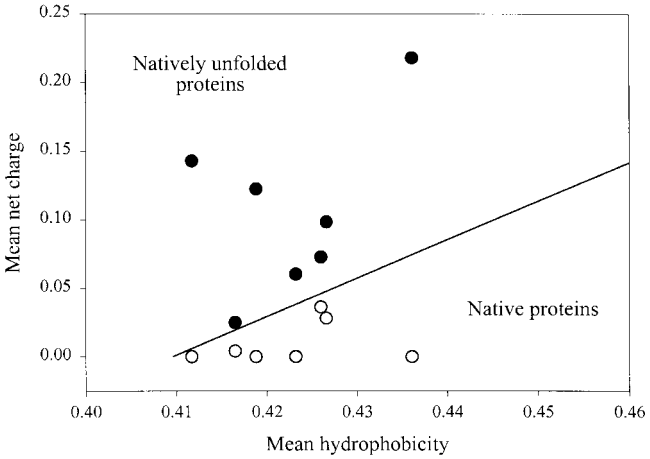

Fig. 4. Effect of cation binding on the mean net charge and the mean hydrophobicity of seven natively unfolded proteins. List includes ostecalcin, α-casein, HPV16 E7 protein, calsequestrin, manganase stabilizing protein, HIV-1 integrase, and ostenectin (SPARK). Filled and open symbols correspond to ligand-free and ligand-loaded forms, respectively. The solid black line represents the border between "natively unfolded" and native proteins calculated from the data presented in Figure 3d:

$$\langle R \rangle = 2.785 \langle H \rangle - 1.151,$$

Here, $\langle H \rangle$ and $\langle R \rangle$ are the mean hydrophobicity and the mean net charge of the protein, respectively.

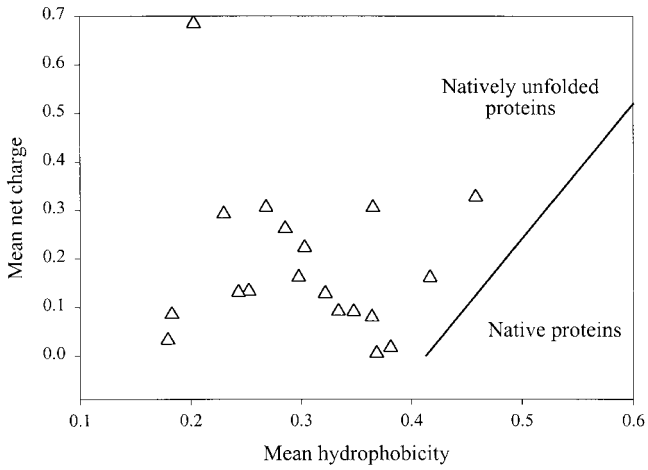

Fig. 5. Charge-hydrophobicity phase space for 19 proteins with the longest and/or strongest predictions of disorder published in Reference 32. List includes plasmodulin, 51-627 fragment; plasmodulin, 68-557 fragment; hypothetical gene 48 protein, 413-720 fragment; micronuclear linker histone polyprotein, 173-450 fragment; neurofilament triplet H protein, 523-761 fragment; slime mold RTOA protein, 77-303 fragment; human pre-mRNA splicing factor srp75, 186-410 fragment; DNA-directed RNA polymerase largest subunit, 1610-1830 fragment; lysostaphin precursor, 56-223 fragment; hypothetical 43.1-kDa protein in CYSG′ region, 57-210 fragment; transcription factor IIF, α-subunit, 246-398 fragment; pea histone H1, 100-252 fragment; 110-kDa antigen, 135-283 fragment; bifunctional endo-1,4β-xylanase precursor, 248-376 fragment; vitellogenin II precursor, 1142-1266 fragment; SNWA protein, 398-521 fragment; pre-RNA processing protein FHL1, 800-923 fragment; hyphally regulated protein, 621-741 fragment; ankyrin brain variant 1, 1778-1897 fragment. The solid line represents the border between "natively unfolded" and native proteins calculated from the data presented in Figure 3d (see legends to Fig. 4).

mean net charge and mean hydrophobicity of complexes of "natively unfolded" proteins with their natural ligands is rather difficult to implement. For the majority of proteins presented in Table I, there is no easy way to estimate the contribution of such natural partners as DNA, RNA, other proteins or peptides, membranes, etc. to the mean net charge and mean hydrophobicity of protein-partner complexes. In fact, there is only one class of proteins, namely metal binding proteins, for which such calculations can readily be done, as in this case only the mean net charge will be affected. Results of such calculations for seven proteins, ostecalcin, osteonectin, α-casein, HPV16 E7 protein, calsequestrin, manganese stabilizing protein and HIV-1 integrase, are presented in Figure 4. Black circles correspond to the ligand-free proteins, whereas open circles describe proteins complex with particular metal cations. The solid black line represents the border between "natively unfolded" and native proteins. This border satisfies the following relationship calculated from the data presented in Figure 3d:

$$\langle R \rangle = 2.785 \langle H \rangle - 1.151,$$

where $\langle H \rangle$ and $\langle R \rangle$ are the mean hydrophobicity and the mean net charge of the protein, respectively. Thus, the interaction of at least these seven proteins with their natural ligands results in a shift in their parameters to those characteristics of native proteins.

It has been suggested that the lack of rigid globular structure under physiologic conditions might represent a considerable functional advantage for "natively unfolded" proteins, as their large plasticity allows them to interact efficiently with several different targets.[29] Moreover, a disorder-order transition induced in "natively unfolded" proteins during the binding of specific targets in vivo might represent a simple mechanism for regulation of numerous cellular processes, including transcriptional and translational regulation and cell cycle control.[29] Evolutionary persistence of the "natively unfolded" proteins represents additional confirmation of their importance and raises intriguing questions on the role of protein disorders in biologic processes.

The results of the present analysis are intuitively reasonable. Amino acid sequences of proteins that have been shown to have little regular structure under physiologic conditions differ significantly from the those of "normal" globular proteins, due to the combination of low mean hydrophobicity and relatively high net charge. High net charge leads to charge-charge repulsion, and low hydrophobicity minimally means less driving force for a compact structure. It is clear that there are several ways in which such a specific sequence can lead to lack of a "normal" tightly packed globular structure. For example, in the case of α-synuclein the residues responsible are clustered mostly in the C-terminal region, and the isolated N-terminal region is predicted to fold. In other cases, the destabilizing residues are more uniformly distributed along the sequence. However, even in these cases, there may be local sequence effects: a survey of complexity of protein se-

quences suggested by Wootton,[30] and developed by Dunker and coauthors,[31–33] has shown that a large portion of the sequences of "native-unfolded" proteins contains segments of low complexity, and high predicted flexibility.[30–33] On the basis of such predictions by using a neural network algorithm, it has been forecast that more than 15,000 proteins in the Swiss Protein Database contain long (40 or more residues) disordered segments with more than 1,000 having especially high scores indicating disorder.[32] We have analyzed the sequences of 19 proteins with the longest and/or strongest predictions of disorder published in Reference 32). Figure 5 shows that all of these 19 proteins are located within the region of charge-hydrophobicity phase space characteristic of "natively unfolded" proteins.

Finally, fibrillar and coiled-coil proteins (which may also have low mean hydrophobicity and high net charge) were not included in our collection of "natively unfolded" proteins. However, its is well known that the majority of coiled-coil proteins (e.g., members of the collagen superfamily) possess rigid structure only due to the self-association and formation of quaternary complexes, being essentially unfolded in the monomeric state. This means that *monomeric* species of coiled-coil proteins can also be considered as the "natively unfolded" proteins.

ACKNOWLEDGMENT

V.N.U. was supported by a grant from National Parkinson Foundation.

REFERENCES

1. Schweers O, Schönbrunn Hanebeck E, Marx A, Mandelkow E. Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. J Biol Chem 1994;269:24290–24297.
2. Anfinsen CB, Haber E, Sela M, White FN. Kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci USA 1961;47:1309–1314.
3. Hemmings HG Jr, Nairin AC, Aswad DW, Greengard P. DARPP-32, a dopamine-and adenosine 3′:5′-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions: II. Purification and characterization of the phosphoprotein from bovine caudate nucleus. J Biol Chem 1984;4:99–110.
4. Gast K, Damaschun H, Eckert K, Schulze-Foster K, Maurer HR, Müller-Frohne M, Zirwer D, Czarnecki J, Damaschun G. Prothymosin α: A biologically active protein with random coil conformation. Biochemistry 1995;34:13211–13218.
5. Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT Jr. NACP, a protein implicated in Alzheimer's diseases and learning, is natively unfolded. Biochemistry 1996;35:13709–13715.
6. Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. Nucleic Acids Res 1999;27:49–54.
7. Appel RD, Bairoch A, Hochstrasser DF. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. Trends Biochem Sci 1994;19:258–260.
8. Kyte J, Doolittle BF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
9. Fink AL, Calciano LJ, Goto Y, Kurotsu T, Palleros DR. Classification of acid denaturation of proteins - intermediates and unfolded states. Biochemistry 1994;33:12504–12511.
10. Goto Y, Takahashi N, Fink AL. Mechanism of acid-induced folding of proteins. Biochemistry 29;1990:3480–3488.
11. Warrant RW, Kim SH. alpha-Helix-double helix interaction shown in the structure of a protamine-transfer RNA complex and a nucleoprotamine model. Nature 1978;271:130–135.
12. Gatewood JM, Schroth GP, Schmid CW, Bradbury EM. Zinc-induced secondary structure transitions in human sperm protamines. J Biol Chem 1990;265:20667–20672.
13. Horiuchi M, Kurihara Y, Katahira M, Maeda T, Saito T, Uesugi S. Dimerization and DNA binding facilitate alpha-helix formation of Max in solution. J Biochem (Tokyo) 1997;122:711–716.
14. Cary PD, King DS, Crane Robinson C, Bradbury EM, Rabbani A, Goodwin GH, Johns EW. Structural studies on two high-mobility-group proteins from calf thymus, HMG-14 and HMG-20 (ubiquitin), and their interaction with DNA. Eur J Biochem 1980;112: 577–580.
15. Abercrombie BD, Kneale GG, Crane Robinson C, Bradbury EM, Goodwin GH, Walker JM, Johns EW. Studies on the conformational properties of the high-mobility-group chromosomal protein HMG 17 and its interaction with DNA. Eur J Biochem 1978;84: 173–177.
16. Isbell DT, Du S, Schroering AG, Colombo G, Shelling JG. Metal ion binding to dog osteocalcin studied by 1H NMR spectroscopy. Biochemistry 1993;32:11352–11362.
17. Engel J, Taylor W, Paulsson M, Sage H, Hogan B. Calcium binding domains and calcium-induced conformational transition of SPARC/BM-40/osteonectin, an extracellular glycoprotein expressed in mineralized and nonmineralized tissues. Biochemistry 1987;26:6958–6965.
18. Josefsson E, O'Connell D, Foster TJ, Durussel I, Cox JA. The binding of calcium to the B-repeat segment of SdrD, a cell surface protein of *Staphylococcus aureus*. J Biol Chem 1998;273:31145–31152.
19. Yoo SH, Albanesi JP. Ca2(+)-induced conformational change and aggregation of chromogranin A. J Biol Chem 1990;265:14414–14421.
20. Yoo SH. pH- and Ca(2+)-induced conformational change and aggregation of chromogranin B. Comparison with chromogranin A and implication in secretory vesicle biogenesis. J Biol Chem 1995;270:12578–12583.
21. Alexandrescu AT, Abeygunawardana C, Shortle D. Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: a heteronuclear NMR study. Biochemistry 1994;33:1063–1072.
22. Tarkka T, Oikarinen J, Grundström T. Nucleotide and calcium-induced conformational changes in histone H1. FEBS Lett 1997; 406:56–60.
23. Stellwagen E, Rysary R, Babul G. The conformation of horse heart apocytochrome c. J Biol Chem 1972;247:8074–8077.
24. Willis KJ. Interaction with model membrane systems induces secondary structure in amino-terminal fragments of parathyroid hormone related protein. Int J Pept Protein Res 1994;43:23–28.
25. Baskakov IV, Kumar R, Srinivasan G, Ji YS, Bolen DW, Thompson EB. Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor. J Biol Chem 1999;274:10693–10696.
26. Lynn A, Chandra S, Malhotra P, Chauhan VS. Heme binding and polymerization by *Plasmodium falciparum* histidine rich protein: II. Influence of pH on activity and conformation. FEBS Lett 1999;459:267–271.
27. Uversky VN, Gillespie JR, Millett IS, Khodyakova AV, Vasilenko RN, Vasiliev AM, Rodionov IL, Kozlovskaya GD, Dolgikh DA, Fink AL, Doniach S, Permyakov EA, Abramov VM. Zn^{2+}-mediated structure formation and compaction of the "natively unfolded" human prothymosin α. Biochem Biophys Res Commun 2000;267:663–668.
28. Uversky VN, Narizhneva NV. Effect of natural ligands on structural properties and conformational stability of proteins. Biochemistry (Moscow) 1998;63:420–433.
29. Wright PE, Dyson HJ. Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. J Mol Biol 1999;293:321–331.
30. Wootton JC. Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput Chem 1994;18:269–285.
31. Dunker AK, Garner E, Guilliot S, Romero P, Albercht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput 1998;3:473–484.
32. Romero P, Obradovic Z, Kissinger C, Villafranca JE, Garner E, Guilliot S, Dunker AK. Thousands of proteins likely to have long disordered regions. Pac Symp Biocomput 1998;3:437–448.

33. Romero P, Obradovic Z, Dunker AK. Folding minimal sequences: the lower bound for sequence complexity of globular proteins. FEBS Lett 1999;462:363–367.
34. Cho HS, Liu CW, Damberger FF, Pelton JG, Nelson HCM, Wemmer DE. Yeast heat shock transcription factor N-terminal activation domains are unstructured as probed by heteronuclear NMR spectroscopy. Protein Sci 1996;5:262–269.
35. Fletcher CM, McGuire AM, Gingras AC, Li H, Matsuo H, Sonenberg N, Wagner G. 4E binding proteins inhibit the translation factor eIF4E without folded structure. Biochemistry 1998;37:9–15.
36. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21Waf1/ Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc Natl Acad Sci USA 1996;93:11504–11509.
37. Lisse T, Bartels D, Kalbitzer HR, Jaenicke R. The recombinant dehydrin-like desiccation stress protein from the resurrection plant *Craterostigma plantagineum* displays no defined three-dimensional structure in its native state. Biol Chem 1996;377:555–561.
38. Hershey PE, McWhirter SM, Gross JD, Wagner G, Alber T, Sachs AB. The cap-binding protein eIF4E promotes folding of a functional domain of yeast translation initiation factor eIF4G1. J Biol Chem 1999;274:21297–21304.
39. Daughdrill GW, Hanely LJ, Dahlquist FW. The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations. Biochemistry 1998; 37:1076–1082.
40. Hazzard J, Sudhol TC, Rizo J. NMR analysis of the structure of synaptobrevin and of its interaction with syntaxin. J Biomol NMR 1999;14:203–207.
41. Donne DG, Viles JH, Groth D, Mehlhorn I, James TL, Cohen FE, Prusiner SB, Wright PE, Dyson HJ. Structure of the recombinant full-length hamster prion protein PrP(29-231): the N terminus is highly flexible. Proc Natl Acad Sci USA 1997;94: 13452–13457.
42. Love JJ. Biophysical characterization of HMG-1 box domain of the lymphoid enhancer binding factor-1. PhD. thesis, University of California, San Diego, 1999.
43. Liu D, Ishima R, Tong KI, Bagby S, Kokubo T, Muhandiram DR, Kay LE, Nakatani Y, Ikura M. Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. Cell 1998;94:573–583.
44. Mogridge J, Legault P, Li J, Van Oene MD, Kay LE, Greenblatt J. Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. Mol Cell 1998;1:265–275.
45. Fiebig KM, Rice LM, Pollock E, Brunger AT. Folding intermediates of SNARE complex assembly. Nat Struct Biol 1999;6:117–123.
46. Penkett CJ, Redfield C, Dodd I, Hubbard J, McBay DL, Mossakowska DE, Smith RA, Dobson CM, Smith LJ. Structural and dynamical characterization of a biologically active unfolded fibronectin-binding protein from *Staphylococcus aureus*. J Mol Biol 1998;274:152–159.
47. Weiss MA, Ellenberger T, Wobbe CR, Lee JP, Harrison SC, Struhl K. Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. Nature 1990;347:575–578.
48. Jimenez MA, Evangelio JA, Aranda C, Lopez-Brauet A, Andreu D, Rico M, Lagos R, Andreu JM, Monasterio O. Helicity of alpha(404-451) and beta(394-445) tubulin C-terminal recombinant peptides. Protein Sci 1999;8:788–799.
49. Cary PD, Crane Robinson C, Bradbury EM, Dixon GH. Structural studies of the non-histone chromosomal proteins HMG-T and H6 from trout testis. Eur J Biochem 1981;119:545–551.
50. Eom JW, Baker WR, Kintanar A, Wurtele ES. The embryo-specific EMB-1 protein of *Daucus carota* is flexible and unstructured in solution. Plant Sci 1996;115:17–24.
51. Geyer M, Munte CE, Schorr J, Kellner R, Kalbitzer HR. Structure of the anchor-domain of myristoylated and non-myristoylated HIV-1 Nef protein. J Mol Biol 1999;289:123–138.
52. Berkovits HJ, Berg JM. Metal and DNA binding properties of a two-domain fragment of neural zinc finger factor 1, a CCHC-type zinc binding protein. Biochemistry 1999;38:16826–16830.
53. Lynch WP, Riseman VM, Bretscher A. Smooth muscle caldesmon is an extended flexible monomeric protein in solution that can readily undergo reversible intra- and intermolecular sulfhydryl cross-linking. A mechanism for caldesmon′s F-actin bundling activity. J Biol Chem 1987;262:7429–7437.
54. Hernández MA, Avila J, Andreu JM. Physicochemical characterization of the heat-stable microtubule-associated protein MAP2. Eur J Biochem 1986;154;41–48.
55. Bhattacharyya J, Das KP. Molecular chaperone-like properties of an unfolded protein, alpha(s)-casein. J Biol Chem 1999;274:15505–15509.
56. Donaldson L, Capone JP. Purification and characterization of the carboxyl-terminal transactivation domain of Vmw65 from herpes simplex virus type 1. J Biol Chem 1992;2671411–1414.
57. Stewart L, Ireton GC, Parker LH, Madden KR, Champoux JJ. Biochemical and biophysical analyses of recombinant forms of human topoisomerase I. J Biol Chem 1996;271:7593–7601.
58. Timm DE, Vissavajjhala P, Ross AH, Neet KE. Spectroscopic and chemical studies of the interaction between nerve growth factor (NGF) and the extracellular domain of the low affinity NGF receptor. Protein Sci 1992;1:1023–1031.
59. Thomas J, Van Patten SM, Howard P, Day KH, Mitchell RD, Sosnick T, Trewhella J, Walsh DA, Maurer RA. Expression in *Escherichia coli* and characterization of the heat-stable inhibitor of the cAMP-dependent protein kinase. J Biol Chem 1991;266: 10906–10911.
60. Larsen RW, Yang J, Hou S, Helms MK, Jameson DM, Alam M. Spectroscopic characterization of two soluble transducers from the Archaeon *Halobacterium salinarum*. J Protein Chem 1999;18:269–275.
61. Loomis RE, Bergey EJ, Levine MJ, Tabak LA. Circular dichroism and fluorescence spectroscopic analyses of a proline-rich glycoprotein from human parotid saliva. Int J Pept Protein Res 1985;26: 621–629.
62. McCubbin WD, Kay CM. Trypsin digestion of bovine cardiac troponin C in the presence and absence of calcium. Can J Biochem 1985;63:803–810.
63. Ferrieres G, Calzolari C, Mani JC, Laune D, Trinquier S, Laprade M, Larue C, Pau B, Granier C. Human cardiac troponin: I. Precise identification of antigenic epitopes and prediction of secondary structure. Clin Chem 1998;44:487–493.
64. House-Pompeo K, Xu Y, Joh D, Speziale P, Hook M. Conformational changes in the fibronectin binding MSCRAMMs are induced by ligand binding. J Biol Chem 1996;271:1379–1384.
65. Rice LM, Brennwald P, Brünger AT. Formation of a yeast SNARE complex is accompanied by significant structural changes. FEBS Lett 1997;415:49–55.
66. Tell G, Perrone L, Fabbro D, Pellizzari L, Pucillo C, De Felice M, Acquaviva R, Formisano S, Damante G. Structural and functional properties of the N transcriptional activation domain of thyroid transcription factor-1: similarities with the acidic activation domains. Biochem J 1998;329:395–403.
67. Agianian B, Leonard K, Bonte E, Van der Zandt H, Becker PB, Tucker PA. The glutamine-rich domain of the Drosophila GAGA factor is necessary for amyloid fibre formation in vitro, but not for chromatin remodelling. J Mol Biol 1999;285:527–544.
68. Tarcsa E, Candi E, Kartasova T, Idler WW, Marekov LN, Steinert PM. Structural and transglutaminase substrate properties of the small proline-rich 2 family of cornified cell envelope proteins. J Biol Chem 1998;273:23297–23303.
69. Richards JP, Bächinger HP, Goodman RH, Brennan RG. Analysis of the structural properties of cAMP-responsive element-binding protein (CREB) and phosphorylated CREB. J Biol Chem 1996;271: 13716–13723.
70. Pahel G, Aulabaugh A, Short SA, Barnes JA, Painter GR, Ray P, Phelps WC. Structural and functional characterization of the HPV16 E7 protein expressed in bacteria. J Biol Chem 1993;268: 26018–26025.
71. He Z, Dunker AK, Wesson CR, Trumble WR. Ca(2+)-induced folding and aggregation of skeletal muscle sarcoplasmic reticulum calsequestrin. The involvement of the trifluoperazine-binding site. J Biol Chem 1993;268:24635–24641.
72. Baskakov I, Bolen DW. Forcing thermodynamically unfolded proteins to fold. J Biol Chem 1998;273:4831–4834.
73. Lydakis-Simantiris N, Betts SD, Yocum CF. Leucine 245 is a critical residue for folding and function of the manganese stabilizing protein of photosystem II. Biochemistry 1999;38:15528–15535.
74. Zheng R, Jenkins TM, Craigie R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. Proc Natl Acad Sci USA 1996;93:13659–13664.

75. Chang JF, Phillips K, Lundbäck T, Gstaiger M, Ladbury JE, Luisi B. Oct-1 POU and octamer DNA co-operate to recognise the Bob-1 transcription co-activator via induced folding. J Mol Biol 1999;288: 941–952.
76. Otto A, Seckler R. Characterization, stability and refolding of recombinant hirudin. Eur J Biochem 1991;202:67–73.
77. Polverini E, Fasano A, Zito F, Riccio P, Cavatorta P. Conformation of bovine myelin basic protein purified with bound lipids. Eur Biophys J 1999;28:351–355.
78. Kanaya E, Kanaya S. Reconstitution of *Escherichia coli* RNase HI from the N-fragment with high helicity and the C-fragment with a disordered structure. J Biol Chem 1995;270:19853–19860.
79. Venyaminov SYu, Gudkov AT, Gogia ZV, Tumanova LG. Absorption and circular dichroism spectra of individual proteins from *Escherichia coli* ribosomes. Pushchino: AS USSR, Biological Research Center, Institute of Protein Research; 1981. 128 p.
80. Nakano M, Kasai K, Yoshida K, Tanimoto T, Tamaki Y, Tobita T. Conformation of the fowl protamine, galline, and its binding properties to DNA. J Biochem (Tokyo) 1989;105:133–137.
81. Bogdarina I, Fox DG, Kneale GG. Equilibrium and kinetic binding analysis of the N-terminal domain of the Pf1 gene 5 protein and its interaction with single-stranded DNA. J Mol Biol 1998;275:443–452.
82. Nakajo S, Omata K, Aiuchi T, Shibayama T, Okahashi I, Ochiai H, Nakai Y, Nakaya K, Nakamura Y. Purification and characterization of a novel brain-specific 14-kDa protein. J Neurochem 1990;55: 2031–2038.
83. Nimmo GA, Cohen P. The regulation of glycogen metabolism. Purification and characterisation of protein phosphatase inhibitor-1 from rabbit skeletal muscle. Eur J Biochem 1978;87:341–351.
84. Belmont LD, Mitchison TJ. Identification of a protein that interacts with tubulin dimers and increases the catastrophe rate of microtubules. Cell 1996;84:623–631.
85. Aswad DW, Greengard P. A specific substrate from rabbit cerebellum for guanosine 3′:5′-monophosphate–dependent protein kinase: I. Purification and characterization. J Biol Chem 256;1981: 3487–3493.
86. Ratnaswamy G, Koepf E, Bekele H, Yin H, Kelly JW. The amyloidogenicity of gelsolin is controlled by proteolysis and pH. Chem Biol 1999;6:293–304.