Intrinsic Structural Disorder of DF31, a Drosophila Protein of Chromatin Decondensation and Remodeling Activities

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · JULY 2008

Impact Factor: 4.25 · DOI: 10.1021/pr700720c · Source: PubMed

8 AUTHORS, INCLUDING:

Mónika Zsuzsanna Bokor
Magyar Tudományos Akadémia Wigner Fizi…
55 PUBLICATIONS **501** CITATIONS
SEE PROFILE

András Perczel
Eötvös Loránd University
233 PUBLICATIONS **4,979** CITATIONS
SEE PROFILE

Peter Tompa
Vlaams Instituut voor Biotechnologie
132 PUBLICATIONS **8,545** CITATIONS
SEE PROFILE

research articles Journal of proteome research

Intrinsic Structural Disorder of DF31, a *Drosophila* Protein of Chromatin Decondensation and Remodeling Activities

Edit Szőllősi,[†] Monika Bokor,[‡] Andrea Bodor,[§] Andras Perczel,[§] Eva Klement,[||] Katalin F. Medzihradszky,[||,⊥] Kalman Tompa,[‡] and Peter Tompa*[,†]

Institute of Enzymology, Hungarian Academy of Sciences, Budapest, Hungary, Research Institute for Solid State Physics and Optics of Hungarian Academy of Sciences, Budapest, Hungary, Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Budapest, Hungary, Proteomics Research Group, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary, and Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California 94143

Received November 8, 2007

Protein disorder is predicted to be widespread in eukaryotic proteomes, although direct experimental evidence is rather limited so far. To fill this gap and to unveil the identity of novel intrinsically disordered proteins (IDPs), proteomic methods that combine 2D electrophoresis with mass spectrometry have been developed. Here, we applied the method developed in our laboratory [Csizmók et al., *Mol. Cell. Proteomics* 2006, *5*, 265−273] to the proteome of *Drosophila melanogaster*. Protein Df31, earlier described as a histone chaperone involved in chromatin decondensation and stabilization, was among the IDPs identified. Despite some hints at the unusual structural behavior of Df31, this protein has not yet been structurally characterized. Here, we provide evidence by a variety of techniques such as CD, NMR, gel-filtration, limited proteolyzsis and bioinformatics that Df31 is intrinsically disordered along its entire length. Further, by chemical cross-linking, we provide evidence that it is a monomeric protein, and suggest that its function(s) may benefit from having an extended and highly flexible structural state. The potential functional advantages and the generality of protein disorder among chromatin organizing proteins are discussed in detail. Finally, we also would like to point out the utility of our 2DE/MS technique for discovering—or, as a matter of fact, rediscovering—IDPs even from the complicated proteome of an advanced eukaryote.

Keywords: intrinsically unstructured protein • natively unfolded protein • 2D electrophoresis • chemical cross-linking • differential scanning calorimetry

Introduction

Intrinsically disordered proteins (IDPs) carry out their functions despite their lack of well-defined 3D structures.[1−4] Often, these proteins function by molecular recognition, in which structural disorder confers specific advantages, such as rapid and specific binding, the capacity of binding multiple partners, or binding promiscuity, that is, the ability to bind different partners with distinct functional outcomes.[5] Currently, we have solid experimental evidence for the disorder of about 500 proteins,[6] but bioinformatics predictors developed to recognize protein disorder from amino acid sequence suggest that several thousand IDPs may exist in the human proteome alone.[7,8] The extension of the structure−function paradigm to encompass IDPs requires the detailed structural−functional characterization of many novel IDPs, in order to arrive at useful generalizations in terms of how these proteins carry out their functions.

To achieve these goals, proteomic approaches for the *en masse* identification of IDPs have been developed.[9,10] These almost invariably rely on an initial enrichment of IDPs by heat-induced or chemical denaturation of globular proteins, separation of the of remaining proteins by 2D electrophoresis, and their subsequent identification by mass spectrometry (MS). A variant of these 2D-based approaches, described in our laboratory, combines an initial native electrophoresis with a denaturing second one utilizing [8 M] urea.[11] MS identification of proteins along the diagonal line of the second gel provided evidence that this method not only separates IDPs from globular proteins, but also provides direct evidence for their structural status. Although this proteomic technique principally enables the identification of a large number of IDPs, we only used it to identify a handful of IDPs; thus, the concept has not been really substantiated. Further, in our initial attempt, the resolving power of the technique had not been really put to the test, because IDPs were identified from the proteomes of

* To whom correspondence should be addressed at: Institute of Enzymology, Hungarian Academy of Sciences, 1113 Budapest, Karolina út 29; tel, +361-279-3143; fax, +361-466-5465; e-mail, tompa@enzim.hu.

[†] Institute of Enzymology, Hungarian Academy of Sciences.

[‡] Research Institute for Solid State Physics and Optics of Hungarian Academy of Sciences.

[§] Eötvös Loránd University.

[||] Proteomics Research Group, Biological Research Center, Hungarian Academy of Sciences.

[⊥] University of California San Francisco.

10.1021/pr700720c CCC: $40.75

Published on Web 05/17/2008

Escherichia coli and *Saccharomyces cerevisiae*, predicted not to feature protein disorder at very high proportions.[7,8,12]

To this end, this 2D/MS technique was applied to the proteome of *Drosophila melanogaster*, predicted to contain as much as 17% fully disordered proteins.[7] Several potential IDPs have been identified; among others, a protein of functional and structural interest was further characterized. This protein, Df31, has been previously described featuring multiple activities, such as removal of sperm basic chromosomal proteins, interaction with histones and facilitation of their loading onto DNA, and also association to poly nucleosomes and participation in the higher-order folding of chromatin.[13,14] In prior studies, this protein has already been noted for its odd behavior, such as unusual SDS/PAGE mobility, heat-stability and high apparent MW by gel-filtration. These features, often witnessed in the case of IDPs, recall that intrinsic disorder is prevalent in DNA-binding proteins. For example, a high level of disorder has been described in transcription factors (e.g., p53,[15] androgen receptor [16]), DNA-repair proteins (e.g., BRCA1[17] and XPA[18]), core- and linker histones,[19] and general chromatin-organizing factors (e.g., HMGA,[20] MeCP2[21] and Sir3p[22]).

Because of this apparently compelling generality, we have undertaken the detailed structural characterization of Df31. Since disorder can only be dependably established by a combination of independent techniques,[1] here we used a variety of techniques, such as bioinformatics analysis, limited proteolysis, CD and NMR spectroscopy, gel-filtration chromatography, chemical cross-linking and differential scanning calorimetry, to establish the structural status of Df31. Our results unequivocally support that Df31 is largely disordered and is probably a monomeric protein of extended polypeptide chain. This finding provides credibility to our newly developed 2D/MS technique for identifying IDPs, and also provides basic insight into the structure−function relationship of chromatin-organizing proteins in general, and Df31 in particular.

Materials and Methods

Methods. 1. Extract of *D. melanogaster*. *D. melanogaster* was grown on standard cornmeal medium at room temperature. A total of 2.5 mL of lysis buffer was added (50 mM Tris/HCl, pH 7.5, including Complete Mini, EDTA-free protease inhibitor cocktail tablet from Roche Applied Science, Germany) to 5 mL of imago and the cells were disrupted with a Potter-Elvehjem tissue grinder. The homogenate was centrifuged (100 000*g* × 30 min, 4 °C), and the supernatant was heated to 100 °C for 10 min. Heat-treated supernatant was centrifuged (100 000*g* × 30 min, 4 °C) again. The final supernatant usually contained 2 mg/mL protein and was kept at −80 °C in 200 μL aliquots for 2DE analysis.

2. 2D Electrophoresis. The method for separating IDPs from globular proteins has been described in detail elsewhere.[11] In short, our 2DE fractionation method combines a native gel and an 8 M urea gel, both prepared according to Laemmli,[23] but without SDS. To the supernatant of heat-treated *D. melanogaster* extracts (typically 100−200 μg proteins in 50−100 μL), 1/4 vol of native sample buffer was added, and the sample was run on a 0.7-mm thick native gel of 20 × 16 cm (Protean II XL, Bio-Rad). The band containing separated *D. malanogaster* proteins was cut out, soaked into the upper gel buffer containing 8 M urea for 45 min and placed on top of a second gel prepared with 8 M urea. This second dimension had no upper (stacking) gel, and it was cast between 1 mm spacers to accommodate the strip swollen in the urea solution. The second gel was run 2.5 times longer than the first at the same voltage and it was stained by colloidal Coomassie Brilliant Blue[24] overnight.

3. Mass Spectrometry. Potential IDPs to be identified were cut out of the 2D gel and analyzed by MS. Briefly, spots were in-gel digested by side-chain protected porcine trypsin as described in http://msf.ucsf.edu/ingel.html. The resulting tryptic digest was analyzed by LC-MS/MS on an Agilent 1100 nanoHPLC system online coupled with an XCT Plus Ion Trap mass spectrometer in information-dependent acquisition mode. HPLC conditions: sample was injected onto a precolumn (C18, 0.3 mm × 5 mm) at a flow rate of 10 μL/min for 5 min in solvent A; the separation column (C18, 75 μm × 150 mm) was equilibrated at 5% B for 15 min; gradient, 5−45% B in 16 min then up to 90% B for 3 min; flow rate 300 nL/min; solvent A was 0.1% formic acid in water; solvent B was 0.1% formic acid in acetonitrile. Data were processed by Data Analysis (version 5.2) and searched against the NCBI 20070926 protein database (5 519 594 sequences) using in-house Mascot search engine (version 2.1.04). Monoisotopic masses with peptide mass tolerance of ±1.2 Da and fragment mass tolerance of ±0.6 Da were considered. No species restriction was used, and 2 missed cleavage sites were allowed. Fixed modification, Cys carbamidomethylation; variable modifications, acetylation of protein N-termini, methionine oxidation and pyroglutamic acid formation from N-terminal Gln residues. Protein identification was accepted when the search engine indicated peptide score above significance threshold ($p < 0.05$).

4. Cloning of Df31. Genomic DNA from *D. melanogaster* was isolated by a standard Trizol procedure.[25] The full-length *D. melanogaster* Df31-encoding gene flanked by NdeI and XhoI sites was amplified by PCR from genomic DNA by the following two primers: sense, 5′- GGAATTCCATATGTCGGCTGCTACG-GAACAACAG-3′; and antisense, 5′-GGCTCGGACGATCAG-GCGCTCGAGCGG-3′. The amplified DNA fragments were digested with *NdeI* and *XhoI* and then ligated into *NdeI*/*XhoI*-digested pET-22(b). The product was transformed into *E. coli* Xl1-Blue cells, positive clones were selected by carbenicillin, and selected clones were confirmed by sequencing (MwG Biotech AG, Germany). For the expression of recombinant Df31, *E. coli* Bl21 cells were transformed with the pET-22(b)-Df31 plasmid, and induced by 0.5 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) at 30 °C for 3 h.

5. Purification of Df31. IPTG-induced *E. coli* cells were collected by centrifugation at 4000 rpm at 4 °C for 20 min and the pellet was resuspended and lysed in a buffer of 50 mM Na_2HPO_4, 150 mM NaCl, and 10 mM imidazole, pH 7.5, containing 5 mM BA, 0.5 mM PMSF and 2 mM dithioerythritol. After lysis for 2−3 min, cells were disrupted by sonication (10 × 10 s 24 μm at 0 °C, MSE sonicator, with 1 min breaks between runs). The cell lysate was centrifuged (36 000 rpm × 25 min, 4 °C) and the supernatant was boiled for 10 min. Heat-treated lysate was centrifuged again (36 000 rpm × 25 min, 4 °C) and the supernatant was purified on a Ni-NTA affinity column, eluting proteins with 500 mM imidazole. Elution was monitored by SDS-PAGE and fractions containing Df31 were pooled. Pooled Df31 was dialyzed against distilled water and lyophilized. Protein concentration was determined by measuring its mass, because the traditional assays for determining protein quantity are generally not applicable to IDPs.[26]

6. Bioinformatics Characterization of Protein Disorder. Sequence-specific assessment of protein disorder was obtained by the IUPred algorithm available at http://iupred.enzim.hu.[27,28]

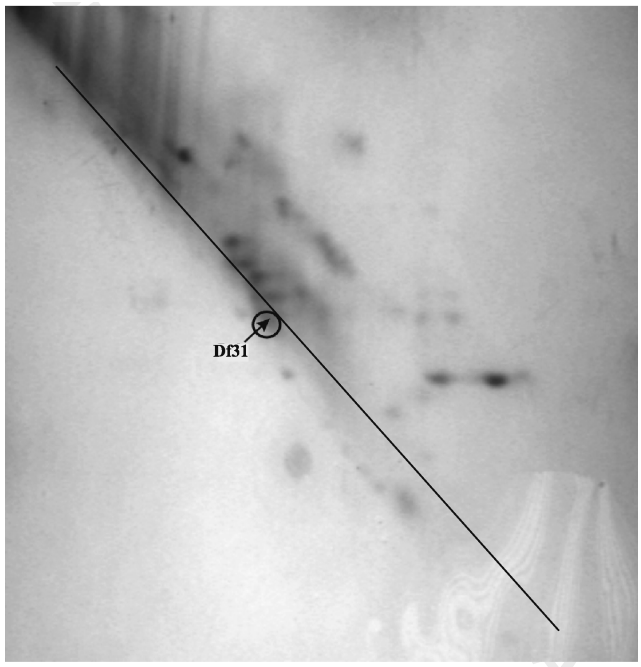

Figure 1. Identification of Df31 from *D. melanogaster* extracts. Heat stable extract was prepared from adult flies, and fractionated on a native/8 M urea 2D gel, as described previously (in both dimensions, a 7.5−15% gradient gel was used). The second gel was visualized by colloidal Coomassie Brilliant Blue staining. Potential IDPs were identified by mass spectrometry. The dot marked was identified as Df31, and further characterized for disorder.

The percentage of disorder is calculated as the ratio of residues with an IUPred score above 0.5.

The global tendency of the protein for disorder was also assessed by calculating its mean net charge <R> and mean hydrophobicity <H>. <R> was calculated as (Glu + Asp) − (Lys + Arg) divided by the length of the protein. <H> was calculated with the ProtScale server (us.expasy.org/tools/protscale.html), with settings as defined in ref 29. IDPs and globular proteins in the charge-hydrophobicity space thus defined (the CH plot) are usually separated by a line defined by the equation $<R> = 2785<H> - 1151$.[29] IDPs tend to be located above this line, whereas globular proteins tend to be located under this line.

7. Heat Stability. IDPs are known for their resistance to chemical and heat-induced denaturation. Thus, heat-treatment was also utilized to probe the structural status of Df31. The samples (BSA and Df31) were incubated in 50 mM Tris and 150 mM NaCl, pH 7,5 at 100 °C for 10 min, centrifuged (54 000 rpm × 20 min at 4 °C), and the supernatant was analyzed on SDS-PAGE.

8. Proteolysis. The extended conformational character of IDPs makes them very sensitive to proteolysis, which serves as a diagnostic marker for the lack of a stable structure. Here we used both wide- (subtilisin and proteinase NOVO) and narrow- (trypsin) specificity enzymes to digest Df31. Because of the extreme sensitivity of Df31, enzymes had to be used at low concentrations. The actual enzyme/Df31 ratios were 1:1000 (trypsin, subtilisin) and 1:2000 (proteinase NOVO). For the digestion of the globular BSA, much higher enzyme concentrations and longer incubation times had to be used. The experiments were typically performed at enzyme/BSA ratios of 1:50 (trypsin), 1:500 (subtilisin) and 1:100 (proteinase NOVO). Stock solutions (2 mg/mL) of all proteins (Df31, BSA, and enzymes) were prepared in a buffer of 50 mM Tris, 150 mM NaCl, and 1 mM EDTA, pH 7.5, and were diluted for the experiments. Digestion was started by the addition of the enzyme, and stopped after an incubation time of 1 h (BSA) or 5 min (Df31) by the addition of the SDS-PAGE sample buffer and 3-min boiling.

9. Far-UV Circular Dichroism (CD). CD allows the estimation of the secondary structure content of proteins in solution, and a largely negative ellipticity near 200 nm is typical of IDPs. The CD spectrum of Df31 was obtained on a Jasco-J-720 spectropolarimeter, at a path-length of 0.1 cm and at a protein concentration of 0.15 mg/mL in a buffer of 10 mM NaH_2PO_4 and 150 mM NaCl, pH 7.5. Measurements were performed at 20 °C.

10. Wide-Line NMR Measrurements. Mobile water molecules within the hydrate layer of proteins can be qualitatively and quantitatively characterized by proton free-induction decays (FIDs), as demonstrated earlier.[11,30] IDPs are characterized by a larger hydrate layer. The measurements were accomplished on a Bruker (Rheinstetten, Germany) SXP 4-100 pulse NMR spectrometer of $\sim 10^{-6}$ resolution at 44.14 and 82.57 MHz frequency. Determination of tightly bound (hydrate) water fraction is based on the comparison of the FID signal intensity extrapolated to $t = 0$ with that measured at a temperature where the whole sample is in liquid state. The zero-time FID signal intensity is proportional to the number of resonant nuclear spins in the sample. The measurements were made by a variable-temperature probe. The temperature was controlled to ±0.1 °C.

11. ^{1}H NMR Spectroscopy. Because of the disordered character of IDPs, their ^{1}H NMR spectra are usually characterized by line broadening and poor resonance dispersion. The spectrum of Df31 was recorded as follows: 0.8 mM solution of Df31 was prepared ($H_2O/D_2O = 9:1$, with a total volume of 550 μL), and placed in a high-quality NMR glass tube. ^{1}H NMR spectra were recorded at 500 MHz (Bruker DRX instrument), 16k complex data points were acquired in the direct dimension at 300 K using a spectral width of 12 ppm. Data were zero-filled and processed with a shifted quadratic sinbell plus exponential window functions. For water suppression the 3−9−19 pulse sequence with gradients was used.[31]

12. Chemical Cross-Linking. Cross-linking was performed in a buffer of 10 mM NaH_2PO_4 and 150 mM NaCl, pH 9.0, at 100 mM dimethyl-suberimidate-dihydrochloride concentration, at a protein concentration of 2.5 mg/mL of Df31, and the controls α-synuclein (NACP) and phosphoglycerate-mutase (PGM). Cross-linking was subsequently analyzed by SDS gel electrophoresis.

13. Gel Filtration Chromatography. A total of 100 μL (2 mg/mL) of protein was analyzed on an Amersham Biosciences Superdex 200 (1 × 30 cm) column at a flow rate of 0.5 mL/min in 10 mM NaH_2PO_4 and 150 mM NaCl, pH 7.5, buffer on an Amersham Biosciences FPLC system. The column was calibrated by the following globular proteins (molecular masses in parentheses): ribonuclease A (13.7 kDa), chymotrypsinogen A (25.0 kDa), ovalbumin (43.0 kDa), BSA (67.0 kDa) and alcohol dehydrogenase (146.8 kDa). Apparent MW of Df31 was determined from the linear fit of elution volumes against logarithms of MW of globular controls.

14. Differential Scanning Calorimetry (DSC). DSC measurements were carried out on a VP-DSC MicroCalorimeter

Table 1. Mascot Search Results of the Mass Spectrometric Identification of Df31 (gi|17737433 NCBI accession NP_523614, gi|24585692 NP_724350, gi|78707166 NP_001027388)[a]

gi|17737433 **Decondensation factor 31 CG2207-PA, isoform A** [***Drosophila melanogaster***]							
Mass: 18812 **Score:** 745 **Queries matched:** 10							
observed	M_r (expt)	M_r (calc)	delta	score	expect	rank	peptide
515.8600	1029.7054	1029.4978	0.2076	60	0.043	1	K.ADEAVATPEK.K
515.9100	1029.8054	1029.5342	0.2713	84	0.00012	1	K.VAAEEVDAVK.K
579.8800	1157.7454	1157.6291	0.1163	70	0.0035	1	K.VAAEEVDAVKK.D
601.9100	1201.8054	1201.5826	0.2229	110	3.7×10^{-7}	1	K.DAVAAEEVAAEK.A
665.9300	1329.8454	1329.6775	0.1679	59	0.043	1	K.KDAVAAEEVAAEK.A
796.4900	1590.9654	1590.7372	0.2282	67	0.0065	1	K.ASITENGGAEEESVAK.E
798.0700	1594.1254	1593.7522	0.3733	62	0.016	1	K.ASEPTVSFAADKDEK.K
850.1200	1698.2254	1697.8471	0.3783	79	0.00027	1	M.ADVAEQKNETPVVEK.V + Acetyl (Protein N-term)
982.4700	1962.9254	1962.8290	0.0965	91	2.7×10^{-5}	1	K.ENGAADSSATEPTDAVDGEK.A
781.6200	2341.8382	2341.2012	0.6370	69	0.0027	1	K.VAAEEVDAVKKDAVAAEEVAAEK.A

[a] Peptide matches were accepted above significance threshold ($p < 0.05$). Sequence coverage 54%. [b] Other entries containing this sequence from NCBI Entrez: gi|24585692, gi|78707166.

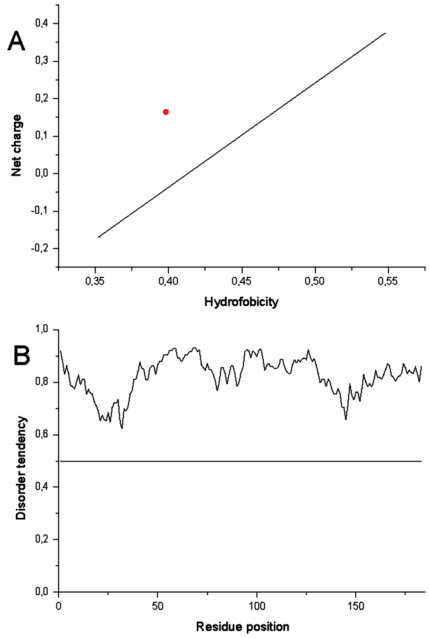

Figure 2. Disorder prediction of Df31. Mean charge of Df31 was plotted against average hydrophobicity of the protein (A). The line (<R> = 2785<H> − 1151) separates IDPs from globular proteins. Df31 is located in the disorder half of the 2D space thus defined. Disorder of Df31 is also predicted by the IUPred algorithm (B). For this method, score values above 0.5 indicate disorder, i.e., Df31 appears to be fully disordered.

with a cell volume of 500 μL, at a heating rate of 1 °C/min, keeping excess pressure at 3.1 atm. Protein concentration was 0.3 mg/mL, in a buffer of 10 mM NaH_2PO_4 and 150 mM NaCl, pH 7.5. Calorimetric scans were corrected for buffer contribution.

Materials

Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. The cDNA was prepared from total RNA of *D. melanogaster*, isolated with Trizol.[25] The pET22(b) plasmid was purchased from Novagen (Madison, WI). NucleoSpin mini-kit (Qiagen) was used for plasmid and DNA purification. Dimethyl-suberimidate was from Fluka Chemical Co., phosphoglycerate-mutase was from Boehringer Mannheim GmbH Co., side-chain protected porcine trypsin was from Promega, and all other chemicals were purchased from Sigma Chemical Co.

Results

Identification of *D. melanogaster* IDPs by 2DE/MS. We have reported that the combination of a native/8 M urea gel-electrophoresis allows the identification of intrinsically disordered proteins and provides direct evidence for their disordered character.[11] In the original paper, we analyzed *E. coli* and *S. cerevisiae* proteomes, but since protein disorder shows a sharp increase with evolution,[7,8] we wanted to demonstrate that the technique can also be applied to a higher eukaryote featuring significantly more IDPs. To this end, we fractionated extracts of *D. melanogaster* on the above-described 2D gel electrophoresis (Figure 1), and identified spots at, and above, the diagonal line by mass spectrometry.

Several proteins were identified (data not shown), one of them deemed particularly interesting for its functional implications, worthy of further studies. This protein, marked in Figure 1, was earlier named Df31. Details of the mass spectrometric identification are summarized in Table 1. This protein was functionally characterized some years ago.[13,14] In these studies, some observations might have actually been interpreted in terms of the disorder of this protein, but the concept of protein disorder was not yet generally accepted at that time. Thus, we decided to characterize the disorder of this protein in detail.

Bioinformatics Predictions. The first line of evidence for disorder is provided by the charge-hydrophobicity plot (Figure 2A). Although it has not been elaborated if the distance from the line separating ordered from disorder proteins is proportional to the disorder of a given protein,[11] it seems plausible that the observed picture is characteristic of a largely disordered protein. This is corroborated by sequence-specific prediction of disorder by the IUPred algorithm, which suggests that 100% of residues of Df31 fall into locally disordered regions (Figure 2B).

Heat-Stability and Proteolytic Sensitivity. A simple diagnostic test of mostly disordered proteins is their heat-stability that results from their generally highly hydrophilic character.[1,10] To test if Df31 falls into this category, we have subjected Df31 and BSA to boiling temperatures and analyzed their supernatants on SDS-PAGE afterward (Figure 3A). As seen, Df31 resists boiling, whereas BSA precipitates, underscoring the mostly disordered character of Df31.

Another indirect indication of disorder is proteolytic sensitivity, signaling accessibility of the polypeptide chain. To explore this opportunity, we have tested the effect of wide-specificity

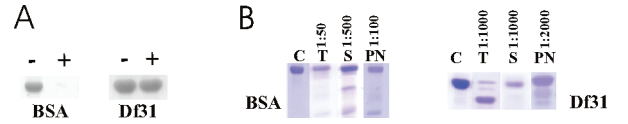

Figure 3. Heat-stability and proteolytic sensitivity of Df31. Df31 and BSA control were analyzed on SDS-PAGE before (-) and after (+) 10 min of boiling (A). Heat-stability of Df31 indicates that it is mostly disordered. The two proteins were also subjected to limited proteolysis (B), with a variety of enzymes (C, control; T, trypsin; S, subtilisin; PN, proteinase Novo) for 1 h (BSA) or for 5 min (Df31). Df31 was degraded in less time at much lower enzyme concentrations, which indicates steric accessibility, i.e., structural disorder, of its polypeptide chain.

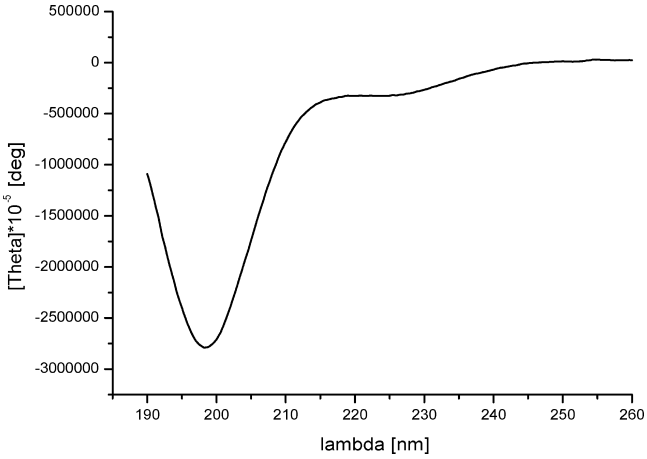

Figure 4. Far-UV CD spectrum of Df31. The CD spectrum of Df31 in the range of 190−250 nm is the average of 9 measurements. Distinctive feature of the spectrum is a large negative peak at 200 nm, typical of IDPs. Further, the lack of a shoulder in the range 210−230 nm indicates that the protein has practically no secondary structural elements.

proteases (subtilisin and proteinase NOVO) and also a narrow-specificity protease (trypsin) on Df31. As seen (Figure 3B), each enzyme digested Df31 within 5 min, even though the enzymes were used at very low enzyme/substrate ratios, typically in the order of 1:1000 to 1:2000 (w/w). The significance of this sensitivity is underlined by the fact that the globular control protein BSA required 1 h at much higher enzyme concentrations (1:50 to 1:500). Although this effect could be accounted for by the local accessibility of the polypeptide chain where the actual effective cleavage occurred, in our experience, uniform sensitivity to a panel of proteases is related to the largely disordered character of a protein.[32]

Far-UV Circular Dichroism (CD). The CD spectum of an IDP is dominated by a largely negative ellipticity near 200 nm, and decomposition of the spectrum in fact can be used for the quantitative characterization of the secondary-structural composition of a protein. Figure 4 shows the CD spectrum of Df31 in the far UV region in the native state (pH 7.5). The spectrum is typical of a completely unfolded state.[1,33]

Wide-Line NMR Measurements. The temperature dependence of the extrapolated FID and CPMG-echo train amplitude of mobile water protons of Df31 and the globular control BSA was also determined (Figure 5). We cooled the protein solutions to −70 °C, to freeze the whole water content. Upon subsequent heating, a melting point is detected for bound (hydration) water; bulk water melts above −5 °C.[11,30] (Hydration water freezes at the same temperature without hysteresis on cooling.) The bound (hydrate) water freezes without hysteresis between

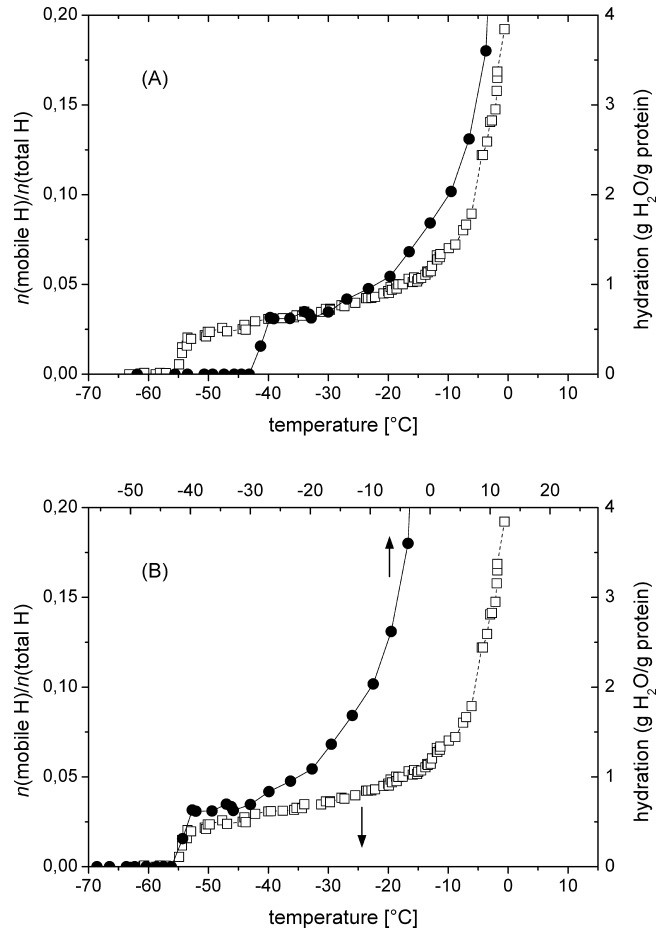

Figure 5. Temperature dependence of FID intensity of Df31 and BSA. Temperature dependence of the concentration of mobile of water protons in Df31 (●) was measured at 50 mg/mL protein concentration in water by wide-line NMR (extrapolated FID and CPMG echo-train amplitude of mobile water protons). Because the NMR amplitude is proportional to the number of protons, the intensity values yield the quantity of hydrate water below ∼−5 °C where the bulk water is frozen, indicating a much larger hydrate layer of the latter. Graph (B) compares the hydration curves at temperature scales shifted to have common hydration melting points. This way, the differences between the two kinds of proteins become strikingly obvious.

about −73 to −53 °C (200−220 K) as suggested by Bizzarri et al. for globular proteins.[34] There is 10−15 °C difference in the melting temperature for globular and intrinsically disordered proteins: hydration water melts at a lower temperature for the globular BSA (−54 °C) than for the IDP Df31 (−41 °C) (Figure 5A). The temperature trend of the mobile (hydration) water concentration is characteristically different for the two kinds of proteins. The globular BSA has unfrozen hydration water well below the melting point of the hydration water for the ID Df31. The mobile water concentrations coincide between −40 and −30 °C. The ID Df31 has much higher hydration above −30 °C than the globular BSA (see Figure 5A and Table 2), under the given protein concentrations. The hydration is 11% higher at −25 °C and 47% higher at −5 °C for Df31 than for BSA. If we compare the hydration curves at temperature scales shifted to have common hydration melting points, the differences between the two kinds of proteins become strikingly obvious (Figure 5B). In our experience, these differences are indicative of the open and largely exposed character of Df31

Table 2. Hydration of BSA and Df31 Solutions (50 mg/cm^3 protein in Tris buffer) Measured by Wide-Line NMR

		BSA	Df31
$T = -25.0 \pm 0.1$ °C	n (mobile H)/n (total H)	0.0403 ± 0.0008	0.0448 ± 0.0009
	hydration	0.81 ± 0.02	0.90 ± 0.02
$T = -5.0 \pm 0.1$ °C	n (mobile H)/n (total H)	0.109 ± 0.002	0.159 ± 0.003
	hydration	2.17 ± 0.04	3.18 ± 0.06

ensured by its IDP character, which ensures binding of much more water in the hydration shell.

^{1}H NMR Measurements. ^{1}H NMR spectrum of Df31 is shown in Figure 6. Considering the molecular mass of Df31, the one-dimensional ^{1}H NMR spectrum has unexpectedly sharp resonance lines. In addition, the lack of signal dispersion typical of folded proteins suggests that Df31 is either unfolded or may have very little stable 3D structure. Note that neither low-field (e.g., NH resonance) nor high-field proton resonances (e.g., methyl resonances of Val, Ile or Leu) can be detected as isolated NMR signals.

Gel Filtration Chromatography. Aberrant behavior was reported in the original publication on the characterization of Df31, although the possible reason was not elucidated and the conclusion of structural disorder was missed.[13] An MW of ~250 kDa for the 18-kDa protein was concluded. Here we also performed gel filtration to check on the apparent MW of Df31. In our case, on a Superdex 200 gel-filtration column, Df31 elutes at an apparent MW of about 67 kDa (Figure 7). This value reflects the behavior of full-length native Df31, as confirmed by SDS-PAGE analysis of the protein eluting from the column. On the SDS gel, the protein migrates as a single band with an apparent MW of 31 kDa (data not shown), in accord with its apparent MW reported originally.[13,14] This confirms that the MW of 67 kDa corresponds to native Df31 chain, indicating a hydrodynamic behavior characteristic of IDPs. The difference between our value and the MW reported earlier might come from post-translational modification of Df31 originally isolated from *D. melanogaster*,[13] as opposed to the recombinant protein used in this study. Such a modification may cause a significant change in protein conformation, as reported for the IDP tau protein,[35] or cause a change in the oligomeric state of the protein, as suggested for Df31.[13] Either of these might result in a high apparent MW, but it should be stressed that both values argue strongly for the extended state of Df31 characteristic of structural disorder.

Cross-Linking Studies. In accord with the previous paragraph, high apparent MW observed by gel filtration chromatography may in principle be explained by an oligomeric structure of the protein that falls apart in the presence of

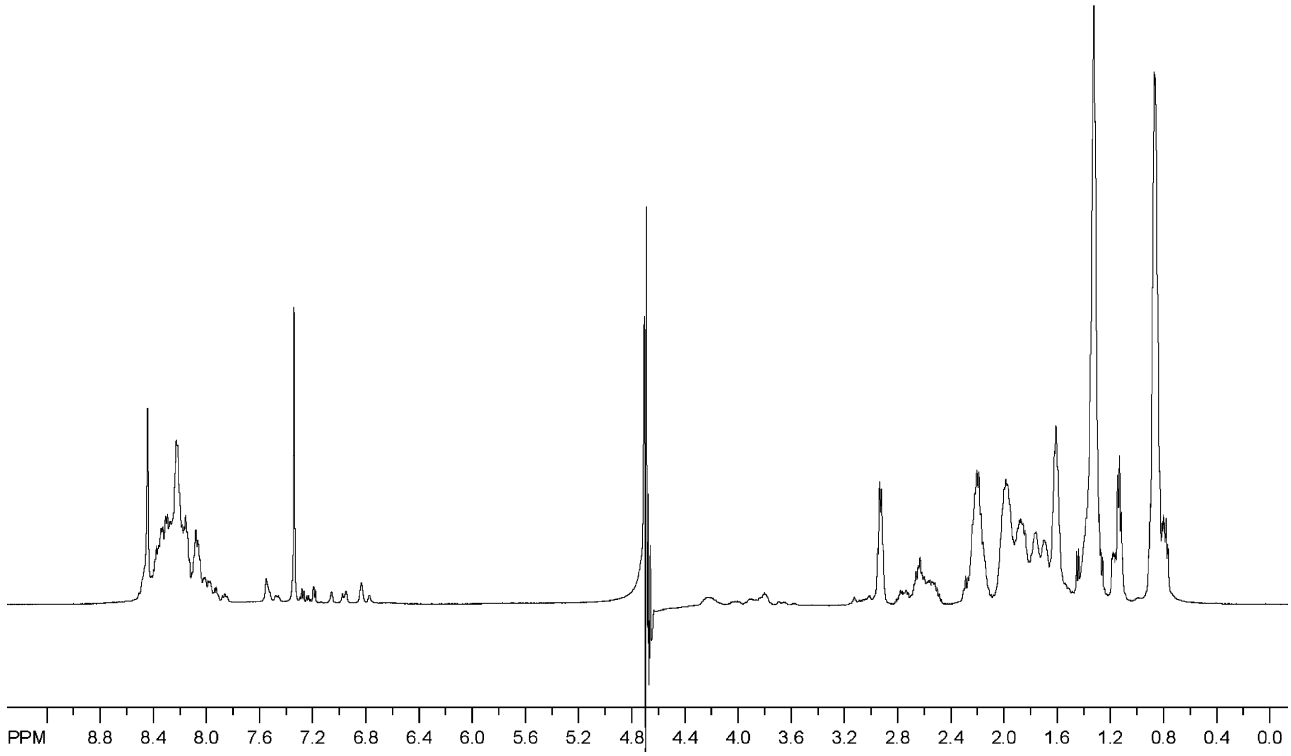

Figure 6. Proton NMR spectrum of Df31, ^{1}H spectrum of Df31 was recorded at 500 MHz: 16k complex data points were acquired in the direct dimension at 300K using a spectral width of 12 ppm. For water suppression, the 3−9−19 pulse sequence with gradients was used. Considering the molecular mass of Df31, the spectrum has unexpectedly sharp resonance lines, and a lack of signal dispersion in both the low-field and high-field regions.

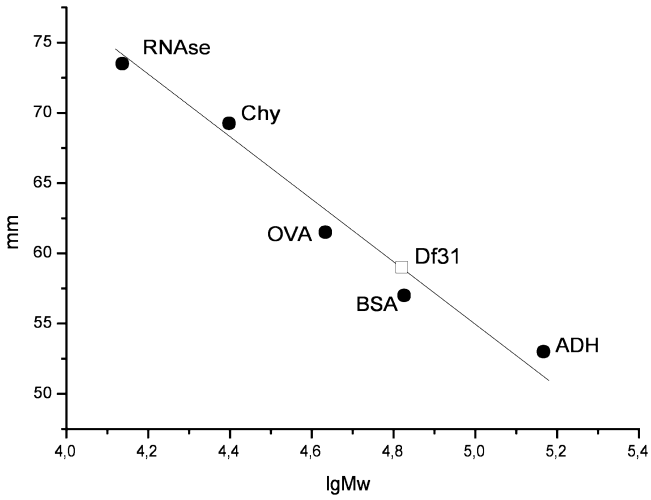

Figure 7. Hydrodynamic behavior of Df31. Hydrodynamic properties of Df31 were assessed by gel filtration on Superdex 200 column. Elution volume against the logarithm of apparent MW is plotted for a range of globular control proteins (●). The position of Df31 (□) suggests a large apparent MW, about 67 kDa. Provided Df31 is monomeric (cf. Figure 8), this large apparent MW may suggest it has an extended conformational ensemble pertaining to a disordered protein.

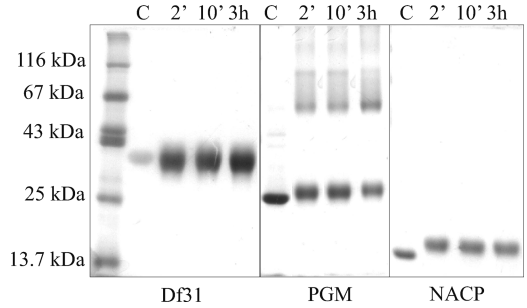

Figure 8. Chemical cross-linking for studying the quaternary structure of Df31. Intrinsically disordered α-synuclein (NACP) and homodimeric globular PGM, along with Df31, were incubated with 100 mM dimethyl-suberimidate for 3 h. In Df31 and α-synuclein, apparently no cross-linking occurred, whereas in PGM, a new band at about 2.5 times the original MW appeared. It is known that α-synuclein is a monomeric protein, whereas PGM is a dimeric protein. Thus, the lack of cross-linking indicates that Df31 does not form oligomers in solution.

denaturing SDS. To probe this possibility, that is, to explore the quaternary organization of the protein, systematic cross-linking studies were carried out. Such analyses have been routinely performed with oligomeric globular proteins but, to our knowledge, have not yet been attempted with IDPs. Conditions (100 mM dimethyl suberimidate, pH 9.0, incubation 1−60 min) were optimized for proper cross-linking of the homodimeric glycolytic enzyme, phosphoglycerate-mutase (PGM) (Figure 8). The experiment was then repeated with a known monomeric IDP, α-synuclein (NACP), as well as with Df31. As seen (Figure 8), α-synuclein shows no sign of cross-linking even under the longest incubation time, 60 min, which indicates a monomeric structure. Df31 displays exactly the same behavior; that is, this protein also must be monomeric. Thus, its apparent hydrodynamic behavior on gel filtration chromatography must be associated with its extended, disordered structural state. It should be added that the cross-linker dimethyl-suberimidate reacts with Lys residues, and the two

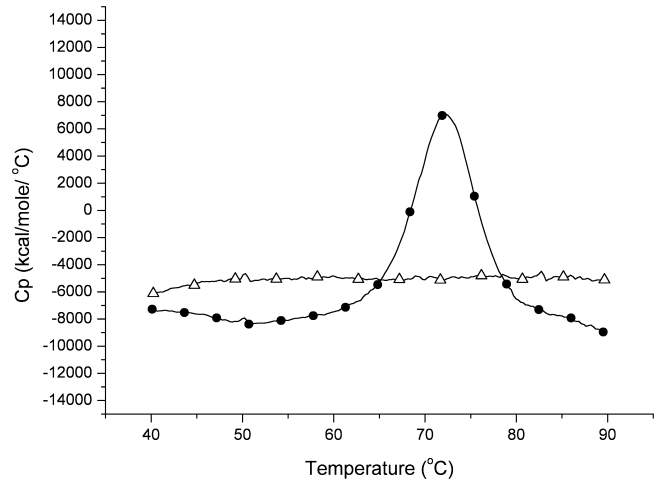

Figure 9. Differential scanning calorimetry of Df31 and lysozyme. The DSC curve of Df31 (Δ) and globular lysozyme (●) was recorded by continuous 1 °C/min heating and the change in heat capacity was calculated from the observed flow of heat. Curves, corrected for buffer contribution, show a distinct behavior of the two proteins: lysozyme undergoes a cooperative structural transition (melting) with a Tm of 72 °C, whereas Df31 lacks such transition. This latter behavior suggests a disordered structural state.

IDPs have very high Lys content (Df31, 21/183 = 11.5%; α-synuclein, 15/140 = 10.7%). Although the application of cross-linking to IDPs is novel, we might be assured that Df31 and α-synuclein, given their very high Lys content, would have been cross-linked if they had been in an oligomeric state.

Differential Scanning Microcalorimetry. DSC curves are characteristic of the lack of a cooperative structural transition in IDPs.[36] To test if Df31 is disordered by this criterion, the temperature dependence of the heat capacity of Df31, and of a control globular protein, lysozyme, was recorded (Figure 9). Lysozyme undergoes a critical exothermic transition with a peak temperature of 72 °C. Such transitions are characteristic of the melting of globular structures, and in fact, the transition seen for lysozyme corresponds to the melting temperature reported in the literature for this protein, 73.8 °C.[37] As opposed to this behavior, the calorimetric curve of Df31 is entirely flat, showing no signs of a major cooperative structural transition. Thus, Df31 lacks a folded structure; that is, it is disordered by this criterion.

Discussion

Our results provide unequivocal evidence that Df31 is intrinsically disordered. In previous studies,[13,14,38] the structural status of Df31 was not directly addressed, but several of the reported observations pointed toward this conclusion. Contradictions among MW values determined by mass spectrometry (18 kDa), SDS-PAGE (31 kDa) and gel-filtration (250 kDa) are typical of IDPs, due to their low SDS binding, and an extended conformational state.[1] Heat stability and proteolytic sensitivity reported are also prime characteristics of IDPs. Results of functional studies, such as unusual, distributed binding surfaces, rapid and transient, still specific binding function, large interaction surfaces enabling bridging distinct nucleosomes,[13,14,38] also fit perfectly into the concept of protein disorder.[39] Although the studies presented here have not provided sequence-specific data, they lend support to, and extend, these previous observations. In combination, results

of the different studies suggest that Df31 is fully disordered, with maybe some local transient structural order. In this sense, this protein belongs to the coil-type or premolten globule type of disordered proteins.[40]

These conclusions have two important implications. First, our results provide evidence that our novel 2DE fractionation strategy to identify IDPs[11] works in a proteome as complex as that of *D. melanogaster*, which potentially contains variants of about 14 000 proteins. Our study resulted in the rediscovery of an IDP, first described more than 10 years ago. As detailed above, many of its observed physicochemical[13] and functional[38] characteristics are in line with intrinsic disorder, although this feature has never been explicitly stated. This leads to the other, maybe more important, implication of our finding that chromatin organizing proteins in general show a high level of intrinsic disorder that may shed new light on their mechanisms of action, and in general on chromatin organization. Df31 has been first described as a chromatin decondensation factor,[13] later refined as a histone chaperone and general chromatin organization protein.[14] With respect to its ability to decondense chromatin and chaperone nucleosome assembly, Df31 is functionally related to nucleoplasmin.[7]

Its chaperone activity may be interpreted in terms of our previous suggestion that the chaperone function is compatible with structural disorder.[41] Both protein and RNA chaperones have a very high level of predicted and experimentally characterized disorder. These observations and site-directed mutagenesis studies on the correlation of the flexibility of the chain and its chaperone function could be conceptually united in an "entropy transfer" model of chaperone activity by IDPs.[41] This activity could explain how Df31 and its kin histone chaperones help the folding of histones and DNA in a nucleosome-assembly competent state. Further, facilitating the nucleosome and chromatin assembly process, that is, delivering histone to DNA in an assembly competent state, may be related to an old observation that disordered, charged binding proteins significantly accelerate association reactions due to an initial nonspecific binding (stickiness), which increases the lifetime of the encounter complex and, thus, the probability of productive interaction.[42] In this sense, disorder of Df31 may be instrumental in its ability to assist nucleosome assembly.

A further important implication of our observation is that the disorder of Df31 is probably also intimately related to its function in higher-level chromatin organization, as suggested by the generality of disorder among such proteins (HMGA, linker histones, Sir2p, MeCP2, cf Introduction). Disorder has been suggested to enable proteins to interact with more partners than a globular protein of similar size,[39] which is exploited in the functioning of scaffolding proteins.[1,2,43] Further, disorder also enables proteins to adapt to the structure of distinct binding partners, termed binding promiscuity or "moonlighting".[5] These points clearly pertain to the observed functions of Df31 and the other chromatin organizing proteins, which can interact with a variety of partners possibly including both DNA and other proteins. Their intimate flexibility may be instrumental in binding to DNA, because this partner has an elongated shape with binding sites facing various directions along the molecule. Also, unfolding transitions of DNA creates structural variants of the same molecule with binding motifs of varied spatial orientations, which can only be handled by the extreme flexibility endowed by disorder of the molecule. The extended conformation and distributed binding surfaces may also enable binding distinct nucleosomes, which is fundamental in their involvement in higher-order chromatin organization. A final interesting point is that Df31 interacts with the disordered tail region of H3,[38] but probably not with DNA, which sets it apart from the other factors. Since the tail region of H3 is also intrinsically disordered,[19] the two disordered regions might undergo mutual induced folding upon interaction. In this sense, this interaction is an example of mutual synergistic folding (or cofolding) of two IDPs, for which only a handful of examples have been reported.[1]

In conclusion, the results presented in this paper shed new light on the structural background of how a protein involved in regulating chromatin organization may carry out its multiple functions. The disorder of this protein appears to be of general relevance that may help to clarify basic issues with respect to the regulation of the dynamics of organization. This, in turn, will get us closer to deciphering how gene transcription is actually regulated.

Acknowledgment. This research was supported by grants OTKA K60694 from the Hungarian Scientific Research Fund, ETT 245/2006 from the Hungarian Ministry of Health, and the International Senior Research Fellowship ISRF 067595 from the Wellcome Trust. Mr. Péter Bánki is thanked for his help with the wide-line NMR measurements. Kromat Ltd. is acknowledged for providing the use of an Agilent 1100 nanoLC-XCT Plus IonTrap system.

References

(1) Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **2002**, *27*, 527–533.
(2) Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **2005**, *579*, 3346–3354.
(3) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **2005**, *18*, 343–384.
(4) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
(5) Tompa, P.; Szasz, C.; Buday, L. Structural disorder throws new light on moonlighting. *Trends Biochem. Sci.* **2005**, *30*, 484–489.
(6) Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N.; Obradovic, Z.; Dunker, A. K. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* **2007**, *35*, D786–793.
(7) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. Intrinsic protein disorder in complete genomes. *Genome Inf. Ser.* **2000**, *11*, 161–171.
(8) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **2004**, *337*, 635–645.
(9) Cortese, M. S.; Baird, J. P.; Uversky, V. N.; Dunker, A. K. Uncovering the unfoldome: enriching cell extracts for unstructured proteins by acid treatment. *J. Proteome Res.* **2005**, *4*, 1610–1618.
(10) Galea, C. A.; Pagala, V. R.; Obenauer, J. C.; Park, C. G.; Slaughter, C. A.; Kriwacki, R. W. Proteomic studies of the intrinsically unstructured mammalian proteome. *J. Proteome Res.* **2006**, *5*, 2839–2848.
(11) Csizmok, V.; Szollosi, E.; Friedrich, P.; Tompa, P. A novel two-dimensional electrophoresis technique for the identification of intrinsically unstructured proteins. *Mol. Cell. Proteomics* **2006**, *5*, 265–273.
(12) Tompa, P.; Dosztanyi, Z.; Simon, I. Prevalent structural disorder in E. coli and S. cerevisiae proteomes. *J. Proteome Res.* **2006**, *5*, 1996–2000.
(13) Crevel, G.; Cotterill, S. DF 31, a sperm decondensation factor from Drosophila melanogaster: purification and characterization. *EMBO J.* **1995**, *14*, 1711–1717.
(14) Crevel, G.; Huikeshoven, H.; Cotterill, S. Df31 is a novel nuclear protein involved in chromatin structure in Drosophila melanogaster. *J. Cell Sci.* **2001**, *114*, 37–47.
(15) Bell, S.; Klein, C.; Muller, L.; Hansen, S.; Buchner, J. p53 contains large unstructured regions in its native state. *J. Mol. Biol.* **2002**, *322*, 917–927.

(16) McEwan, I. J.; Lavery, D.; Fischer, K.; Watt, K. Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. *Nucl. Recept. Signaling* **2007**, *5*, e001.
(17) Mark, W. Y.; Liao, J. C.; Lu, Y.; Ayed, A.; Laister, R.; Szymczyna, B.; Chakrabartty, A.; Arrowsmith, C. H. Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions. *J. Mol. Biol.* **2005**, *345*, 275–287.
(18) Iakoucheva, L. M.; Kimzey, A. L.; Masselon, C. D.; Bruce, J. E.; Garner, E. C.; Brown, C. J.; Dunker, A. K.; Smith, R. D.; Ackerman, E. J. Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Sci.* **2001**, *10*, 560–571.
(19) Hansen, J. C.; Lu, X.; Ross, E. D.; Woody, R. W. Intrinsic protein disorder, amino Acid composition, and histone terminal domains. *J. Biol. Chem.* **2006**, *281*, 1853–1856.
(20) Reeves, R. Molecular biology of HMGA proteins: hubs of nuclear function. *Gene* **2001**, *277*, 63–81.
(21) Adams, V. H.; McBryant, S. J.; Wade, P. A.; Woodcock, C. L.; Hansen, J. C. Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J. Biol. Chem.* **2007**, *282*, 15057–15064.
(22) McBryant, S. J.; Krause, C.; Hansen, J. C. Domain organization and quaternary structure of the *Saccharomyces cerevisiae* silent information regulator 3 protein, Sir3p. *Biochemistry* **2006**, *45*, 15941–15948.
(23) Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227*, 680–685.
(24) Rosenfeld, J.; Capdevielle, J.; Guillemot, J. C.; Ferrara, P. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal. Biochem.* **1992**, *203*, 173–179.
(25) Chomczynski, P.; Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **1987**, *162*, 156–159.
(26) Szollosi, E.; Hazy, E.; Szasz, C.; Tompa, P. Large systematic errors compromise quantitation of intrinsically unstructured proteins. *Anal. Biochem.* **2007**, *360*, 321–323.
(27) Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433–3434.
(28) Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* **2005**, *347*, 827–839.
(29) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions. *Proteins* **2000**, *41*, 415–427.
(30) Bokor, M.; Csizmok, V.; Kovacs, D.; Banki, P.; Friedrich, P.; Tompa, P.; Tompa, K. NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. *Biophys. J.* **2005**, *88*, 2030–2037.
(31) Piotto, M.; Saudek, V.; Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR.* **1992**, *2*, 661–665.
(32) Csizmok, V.; Bokor, M.; Banki, P.; Klement, E.; Medzihradszky, K. F.; Friedrich, P.; Tompa, K.; Tompa, P. Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. *Biochemistry* **2005**, *44*, 3955–3964.
(33) Receveur-Brechot, V.; Bourhis, J. M.; Uversky, V. N.; Canard, B.; Longhi, S. Assessing protein disorder and induced folding. *Proteins* **2005**, *62*, 24–45.
(34) Bizzarri, A. R.; Cannistraro, S. Molecular dynamics of water at the protein-solvent interface. *J. Phys. Chem. B* **2002**, *106*, 6617–6633.
(35) Hagestedt, T.; Lichtenberg, B.; Wille, H.; Mandelkow, E. M.; Mandelkow, E. Tau protein becomes long and stiff upon phosphorylation: correlation between paracrystalline structure and degree of phosphorylation. *J. Cell Biol.* **1989**, *109*, 1643–1651.
(36) Uversky, V. N.; Kirkitadze, M. D.; Narizhneva, N. V.; Potekhin, S. A.; Tomashevski, A. Structural properties of alpha-fetoprotein from human cord serum: the protein molecule at low pH possesses all the properties of the molten globule. *FEBS Lett.* **1995**, *364*, 165–167.
(37) Maeda, Y.; Yamada, H.; Ueda, T.; Imoto, T. Effect of additives on the renaturation of reduced lysozyme in the presence of 4 M urea. *Protein Eng.* **1996**, *9*, 461–465.
(38) Guillebault, D.; Cotterill, S. The Drosophila Df31 protein interacts with histone H3 tails and promotes chromatin bridging in vitro. *J. Mol. Biol.* **2007**, *373*, 903–912.
(39) Gunasekaran, K.; Tsai, C. J.; Kumar, S.; Zanuy, D.; Nussinov, R. Extended disordered proteins: targeting function with less scaffold. *Trends Biochem. Sci.* **2003**, *28*, 81–85.
(40) Uversky, V. N. What does it mean to be natively unfolded. *Eur. J. Biochem.* **2002**, *269*, 2–12.
(41) Tompa, P.; Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **2004**, *18*, 1169–1175.
(42) Pontius, B. W. Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem. Sci.* **1993**, *18*, 181–186.
(43) Bhattacharyya, R. P.; Remenyi, A.; Good, M. C.; Bashor, C. J.; Falick, A. M.; Lim, W. A. The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science* **2006**, *311*, 822–826.

PR700720C