

# Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome

Peter Tsvetkov,<sup>1</sup> Gad Asher,<sup>1</sup> Aviv Paz,<sup>2,3</sup> Nina Reuven,<sup>1</sup> Joel L. Sussman,<sup>3</sup> Israel Silman,<sup>2</sup> and Yosef Shaul<sup>1\*</sup>

<sup>1</sup> Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

<sup>2</sup> Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel

<sup>3</sup> Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

## ABSTRACT

*Intrinsically unstructured proteins (IUPs), also known as natively unfolded proteins, lack well-defined secondary and tertiary structure under physiological conditions. In recent years, growing experimental and theoretical evidence has accumulated, indicating that many entire proteins and protein sequences are unstructured under physiological conditions, and that they play significant roles in diverse cellular processes. Bioinformatic algorithms have been developed to identify such sequences in proteins for which structural data are lacking, but still generate substantial numbers of false positives and negatives. We describe here a simple and reliable in vitro assay for identifying IUP sequences based on their susceptibility to 20S proteasomal degradation. We show that 20S proteasomes digest IUP sequences, under conditions in which native, and even molten globule states, are resistant. Furthermore, we show that protein–protein interactions can protect IUPs against 20S proteasomal action. Taken together, our results thus suggest that the 20S proteasome degradation assay provides a powerful system for operational definition of IUPs.*

Proteins 2008; 70:1357–1366.  
© 2007 Wiley-Liss, Inc.

**Key words:** IUP; proteasomal degradation; molten globule; natively unfolded; IDP; proteolysis.

## INTRODUCTION

The proteins that provide the machinery for growth and maintenance of the living cell turn-over by complex processes of synthesis and degradation. Thus, the extended polypeptide chain synthesized on the ribosome collapses to a compact molten globule (MG),<sup>1</sup> which subsequently acquires the precise three-dimensional structure of the functional “native” protein, with or without the aid of molecular chaperones.<sup>2</sup> Evidence has accumulated over the last two decades that many proteins contain extensive disordered regions, and that some proteins are even completely disordered under physiological conditions.<sup>3</sup> These proteins are termed natively unfolded<sup>4</sup> or intrinsically disordered proteins (IDPs),<sup>3</sup> and their involvement in many cellular processes, including transcriptional regulation and signal transduction, has been demonstrated.<sup>5,6</sup> Thus, as well formulated by Dyson and Wright,<sup>5</sup> functional proteins fall on a structural continuum, ranging from tightly packed proteins that are almost completely ordered, displaying well-defined tertiary structures, such as acetylcholinesterase,<sup>7</sup> through proteins that contain both large ordered and large disordered stretches, a good example being the p53 tumor suppressor,<sup>8,9</sup> on to whole proteins such as the ~440-residue tau protein, which appears to be intrinsically disordered from its NH<sub>2</sub>— to its COOH— terminus.<sup>10</sup> Furthermore, in multidomain proteins, tightly packed and ordered domains may be connected by intrinsically disordered and flexible disordered linker sequences, as is the case for a construct that contains the first three zinc fingers of transcription factor-III<sub>A</sub>.<sup>11</sup>

In general, intrinsic disorder appears to be associated with low complexity, and with a low content of hydrophobic amino-acids and/or a high net charge.<sup>4,12</sup> Since the recognition of the prevalence of intrinsically unstructured proteins (IUPs) by Dunker and coworkers,<sup>13</sup> a sizeable number of algorithms

Grant sponsor: The European Commission Sixth Framework Research and Technological Development Programme; Grant number: 031220; Grant sponsors: Bruce Rosen Foundation, The Jean and Julia Goldwurm Memorial Foundation, Kimmelman Center for Biomolecular Structure and Assembly, Benziyo Center for Neuroscience, Israel Ministry of Science, Culture, and Sport grant for the Israel Structural Proteomics Center (ISPC), The Divadol Foundation, The Neuman Foundation, The Israel Science Foundation, Israel Cancer Research Fund, Samuel Waxman Cancer Research Foundation, The Minerva Foundation with funding from the Federal German Ministry for Education and Research, Autism Speaks.

JLS is the Morton and Gladys Pickman Professor of Structural Biology, and YS is the Oscar and Emma Getz Professor.

Peter Tsvetkov, Gad Asher, and Aviv Paz contributed equally to the study.

\*Correspondence to: Yosef Shaul, Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: yosef.shaul@weizmann.ac.il

Received 26 March 2007; Revised 3 May 2007; Accepted 14 May 2007

Published online 18 September 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21614

have been developed aimed at predicting whether a given amino-acid sequence is ordered or disordered, starting with the PONDR predictor of the Dunker group.<sup>14</sup> A recent review compares several of the programs produced for this purpose.<sup>15</sup> Those available so far predict the ordered/disordered status of a given sequence based on the physicochemical properties/energetics of the protein sequence alone; thus not taking into account the complete folding of the mature protein.

Several methods are currently employed for identifying and characterizing IUPs *in vitro*. X-ray crystallography and NMR spectroscopy are the main tools in 3D structure determination of globular proteins at atomic resolution. Lack of backbone coordinates in X-ray structures, or diminished chemical shift dispersion in NMR experiments, can highlight disordered parts of a protein of which the bulk is fully folded. However, it is impossible to use X-ray crystallography to characterize IUPs, since they cannot be crystallized due to their inherent flexibility. NMR is a most suitable method for characterizing such flexible and dynamic structures; indeed, various NMR techniques have been employed in characterization of unfolded states of proteins in general, and of IUPs in particular.<sup>16</sup> Other available methods include far-UV circular dichroism spectroscopy,<sup>17</sup> Raman optical activity (ROA)<sup>18</sup> and 2D electrophoresis.<sup>19</sup> These latter methods can provide details about the average conformation of the polypeptide chain, but do not provide sequence-specific information. Additional methods, such as size-exclusion chromatography and small angle X-ray scattering, which provide information concerning the dimensions of a protein, have also been applied to characterization of IUPs. Most of these experimental methods demand substantial amounts of purified protein, special instrumentation and expertise, and are highly time consuming. Information obtained mainly by use of X-ray crystallography, CD, and NMR, permitted establishment of the DisProt database that contains proteins that possess at least one experimentally determined disordered region.<sup>20</sup> As already mentioned, a number of algorithms have been developed that can predict, based on the amino-acid sequence, whether a protein is intrinsically unstructured or contains large disordered segments. Nevertheless, there is a need of methods for simple and rapid experimental identification of IUPs and of disordered sequences.

The proteasome degrades the majority of cellular proteins in eukaryotes. It controls the levels of various regulatory proteins, and prevents the accumulation of misfolded mutant and damaged proteins.<sup>21</sup> The symmetrical core 20S proteasome consists of 28 subunits arranged in a cylindrical form of four rings. The two inner rings are catalytically active in protein degradation. Two 19S regulatory complexes associate with the 20S core particle to form the 26S proteasome. The 19S regulatory complex recognizes the polyubiquitin chains and catalyzes substrate deubiquitination, denaturation, and translocation of the substrate

into the 20S catalytic core for degradation, whereas the 20S proteasome possesses the ubiquitin-independent proteolytic functions.<sup>22,23</sup> The architecture of the proteasome suggests that correctly folded globular proteins in their native conformation should not be readily susceptible to 20S proteasomal attack unless denatured by the regulatory 19S subunit. It has been proposed that unfolded protein substrates are fed into the 20S proteasome via their termini as extended chains and are degraded progressively.<sup>24,25</sup> Liu *et al.*,<sup>26</sup> however, have provided evidence for susceptibility of internal disordered regions to endoproteolytic activity of the proteasome.

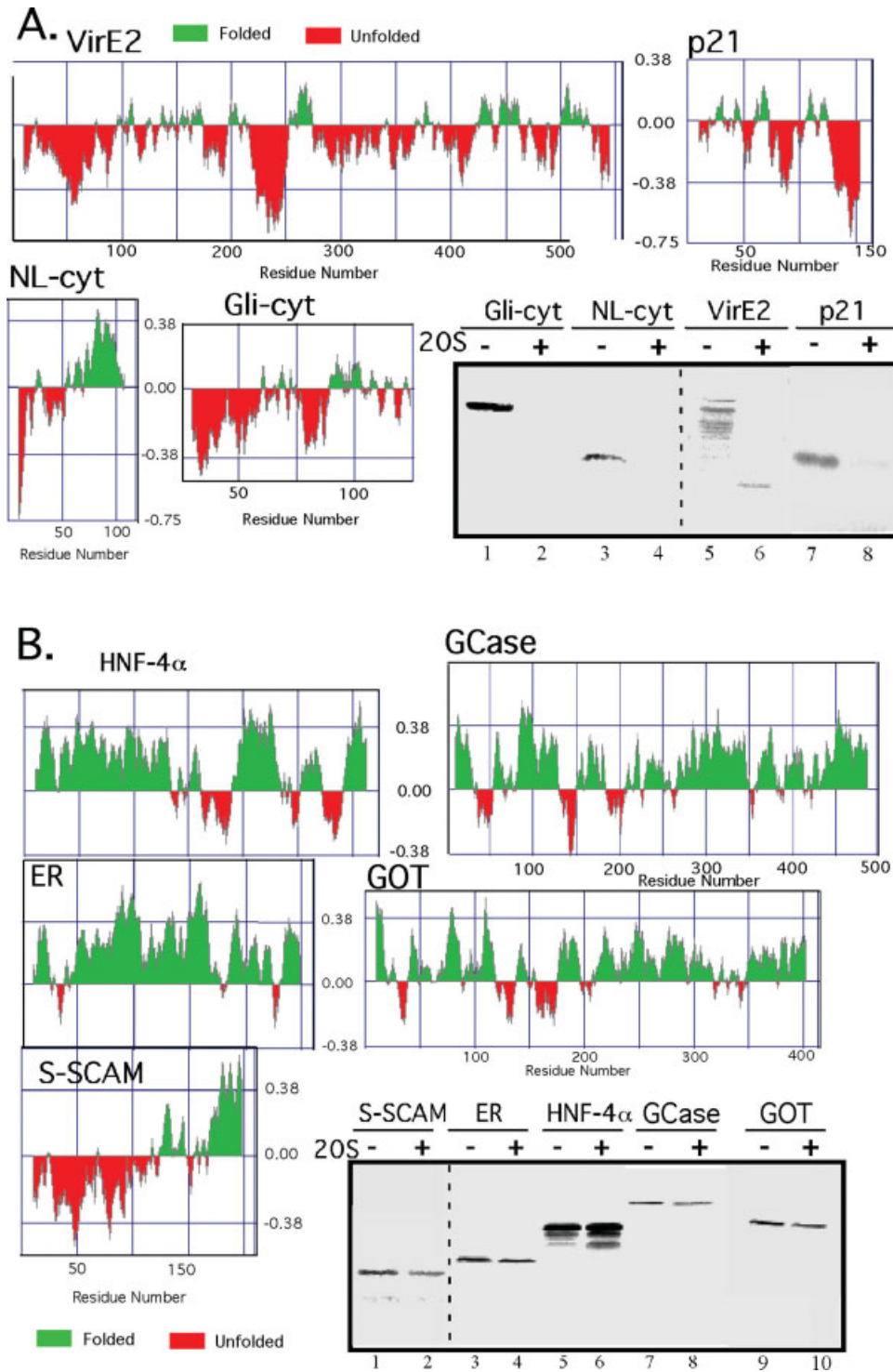
In the following, we examine the action of the 20S proteasome on a repertoire of folded and IUPs and protein domains. We show that susceptibility to proteasomal digestion is indeed correlated with intrinsic disorder, so that it may actually serve as a diagnostic measure. We further show that proteins in an MG state, although susceptible to digestion by the potent protease, proteinase K, are proteasome-resistant, despite their lack of tertiary structure, presumably due to their being relatively compact. Finally, we show that a protease-sensitive IUP becomes proteasome-resistant when complexed with a functional partner.

## RESULTS

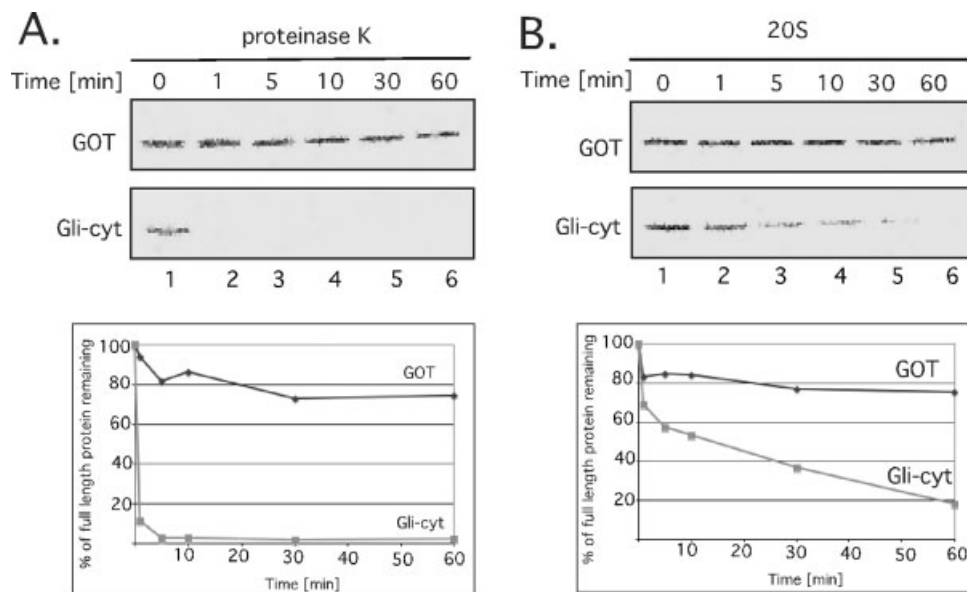
### Degradation of unstructured proteins by the 20S proteasome

The core 20S proteasome is a large, multicatalytic protease that degrades proteins to small peptides. The proteolytic properties of the 20S proteasome include “chymotrypsin-like,” “trypsin-like,” and “peptidylglutamyl-peptide hydrolyzing” or “caspase-like” activities.<sup>27</sup> In contrast to the 26S proteasome, the core 20S proteasome lacks the 19S regulatory subunits that are responsible for recognizing poly-ubiquitinated proteins and for unfolding protein substrates.<sup>27</sup> The core 20S proteasome is, therefore, considered to be capable of degrading unstructured proteins.<sup>23</sup>

To examine if the 20S proteasome is capable of distinguishing between folded and unstructured proteins, we examined its effect on a repertoire of folded and unstructured proteins and protein domains that were obtained as [<sup>35</sup>S]-labeled species by *in vitro* translation (Fig. 1). The unstructured protein domains that were tested include the cytoplasmic domains of gliotactin (Gli), and neuroligin-3 (NL3), VirE2 and p21. Gli and NL3 are members of the family of cholinesterase-like adhesion molecules (CLAMs), so named due to the sequence homology of their extracellular domains to the synaptic enzyme, acetylcholinesterase.<sup>28</sup> They are transmembrane proteins that contain both an extracellular ChE-like domain and an intracellular domain. The intracellular domains of both Gli and NL3 *viz.*, Gli-cyt and NL3-cyt

**Figure 1**

Degradation of folded and unstructured proteins by 20S proteasomes. [ $^{35}$ S]-labeled proteins, obtained by *in vitro* translation, were incubated without (–) or with (+) purified 20S proteasomes at 37°C for 1 h. The samples were then subjected to SDS-PAGE and [ $^{35}$ S]-labeled bands were detected by autoradiography. (A) Unstructured proteins: Gli-cyt, NL3-cyt, VirE2, p21; (B) folded proteins: S-SCAM, ER, HNF-4 $\alpha$ , human GCCase, GOT. In the Foldindex<sup>®</sup> profiles the green color-coding corresponds to sequences predicted to be folded, and red corresponds to those predicted to be unfolded.<sup>12</sup> The dashed line separates experiments that were run on different gels.

**Figure 2**

Degradation of GOT and Gli-cyt by proteinase K and by 20S proteasomes. [ $^{35}\text{S}$ ]-labeled GOT and Gli-cyt were incubated with proteinase K (A) and with 20S proteasomes (B), both at 37°C. Samples taken at indicated times were subjected to SDS-PAGE, and the [ $^{35}\text{S}$ ]-labeled bands were detected by autoradiography. Quantification was as described under Materials and Methods.

have been shown to be intrinsically disordered by a number of methods including NMR, circular dichroism, ROA, and size-exclusion chromatography<sup>29</sup> (Paz *et al.*, in preparation). The human Cdk inhibitor, p21<sup>cip1</sup>, is an IUP protein that plays a critical role in regulating cell division and DNA replication.<sup>30</sup> VirE2 is an *Agrobacterium tumefaciens* virulence protein that interacts with the specific molecular chaperone VirE1 to regulate its aggregation and its interaction with DNA. Uncomplexed VirE2 forms functional soluble filamentous aggregates that are capable of binding DNA.<sup>31</sup> The fact that various prediction algorithms indicate that VirE2 is unstructured, and the tendency of uncomplexed VirE2 to form aggregates, support the notion that it is indeed unstructured.

The folded proteins tested were a set of proteins that display a well-resolved 3D structure, and were also predicted to be predominantly folded using Foldindex, with the exception of S-SCAM (Fig. 1). These include human S-SCAM, a neural scaffolding protein (PDB ID: 1UEQ)<sup>32</sup>; human estrogen receptor (ER), a ligand-induced transcription factor nuclear receptor (PDB ID: 1A52/3ERT)<sup>33</sup>; human hepatocyte nuclear factor 4 alpha (HNF-4 $\alpha$ ), a major regulator of hepatocyte function (PDB ID: 1PZL)<sup>34</sup>; human acid- $\beta$ -glucosidase (GCase), a lysosomal enzyme that degrades glucosylceramide (PDB ID: 1OGS)<sup>35</sup>; and human glutamate-oxaloacetate transaminase (GOT), a pyridoxal-phosphate-dependent enzyme that plays a role in amino-acid metabolism and in the urea and tricarboxylic acid cycles (PDB ID: 1AJS).<sup>36</sup>

Incubation of *in vitro*-translated [ $^{35}\text{S}$ ]-labeled Gli-cyt, NL3-cyt, VirE2, and p21, all classified as completely disordered, in the presence of purified 20S proteasomes, resulted in their complete degradation [Fig. 1(A)]. In contrast, S-SCAM, ER, HNF4, Gcase, and GOT, all classified as completely folded, were resistant to degradation under the same conditions [Fig. 1(B)]. This strong correlation between susceptibility to 20S proteasomal degradation and the presence of unstructured sequences supports the possibility that our 20S proteasomal degradation assay can be used to distinguish between folded and unstructured proteins.

#### Time-course analysis of proteinase K digestion and 20S proteasomal degradation of folded and unstructured proteins

Limited proteolysis is commonly used as a tool to probe the conformational features of partially unstructured proteins or to cleave multidomain proteins at domain boundaries.<sup>37</sup> Proteases of broad substrate specificity, such as proteinase K, are frequently used for this purpose.<sup>38</sup> We compared the sensitivity of Gli-cyt, an essentially completely disordered protein domain, and of GOT, an almost completely folded globular protein, to proteolysis by proteinase K and to 20S-mediated proteasomal degradation. Whereas Gli-cyt was completely digested after 5 min in the presence of proteinase K [Fig. 2(A)], degradation by 20S proteasomes [Fig. 2(B)] was much slower under the experimental conditions employed. GOT was completely

resistant to both proteinase K and to 20S proteasomes under the same conditions. These results indicate that the 20S proteasomal degradation assay, like limited proteolysis with proteinase K, may be used to detect unstructured protein sequences.

### Relative susceptibility of proteins in their native and MG states to proteinase K and to 20S proteasomes

The concept of the “molten globule” (MG) was initially used by to describe an intermediate in protein folding en route from the extended polypeptide synthesized on the ribosome to the fully folded and functional native protein.<sup>1,39</sup> The term is now more generally used to refer to a collapsed state of a protein of slightly greater dimensions than those of the native protein. The MG state may also be obtained from the native state by employing mild denaturing procedures. The MG retains some native secondary and tertiary structure, but lacks well-packed side chains, and displays increased conformational flexibility relative to the native state, as evidenced by the ability of MG species that contain intrachain disulfide bonds to undergo spontaneous disulfide reshuffling.<sup>40</sup> We thought, therefore, that it would be of interest to examine whether the 20S proteasomal degradation assay is capable of distinguishing between the native and MG state of two proteins for which both states have been well characterized.

$\alpha$ -Lactalbumin ( $\alpha$ LA) is a calcium-binding globular protein present in mammalian milk, which functions as a specific modifier of the enzyme, galactosyltransferase.<sup>41</sup> The MG state of  $\alpha$ LA is one of the best-characterized folding intermediates, and has been studied intensively by a variety of physicochemical approaches.<sup>42</sup> Transition of  $\alpha$ LA to its MG state occurs at elevated temperatures (above 50°C), at intermediate denaturant concentrations (3–4M urea), and upon depletion of  $\text{Ca}^{+2}$  ions.<sup>1</sup>

Bovine  $\alpha$ LA was incubated in the presence or absence of 5 mM  $\text{Ca}^{+2}$ , and further analyzed either following incubation with proteinase K or 20S proteasomes. In the presence of 5 mM  $\text{Ca}^{+2}$ ,  $\alpha$ LA adopts a fully folded native conformation and, consequently, is resistant to degradation both by proteinase K and 20S proteasomes [Fig. 3(A,B)]. In contrast,  $\text{Ca}^{+2}$ -depleted  $\alpha$ LA is completely digested by proteinase K, as is to be expected, since it is known to be in an MG state [Fig. 3(A)]. However, this MG state of  $\alpha$ LA is not susceptible to degradation by the 20S proteasomes [Fig. 3(B)]. This is in agreement with an earlier report that even reduction of two of the three disulfide bridges in  $\alpha$ LA, yielding a species that is maintained in a compact configuration, does not render it susceptible to 20S proteasomal degradation.<sup>43</sup> To rule out any nonspecific effects of  $\text{Ca}^{+2}$  on the capacity of 20S proteasomes to degrade proteins, NL3-cyt was incubated in the presence of the 20S proteasomes either alone

or in the presence of 5 mM  $\text{Ca}^{+2}$ . The presence of the calcium ions had no effect on the ability of the 20S proteasome to degrade NL3-cyt, confirming their specific effect on the degradation of  $\alpha$ LA [Fig. 3(C)].

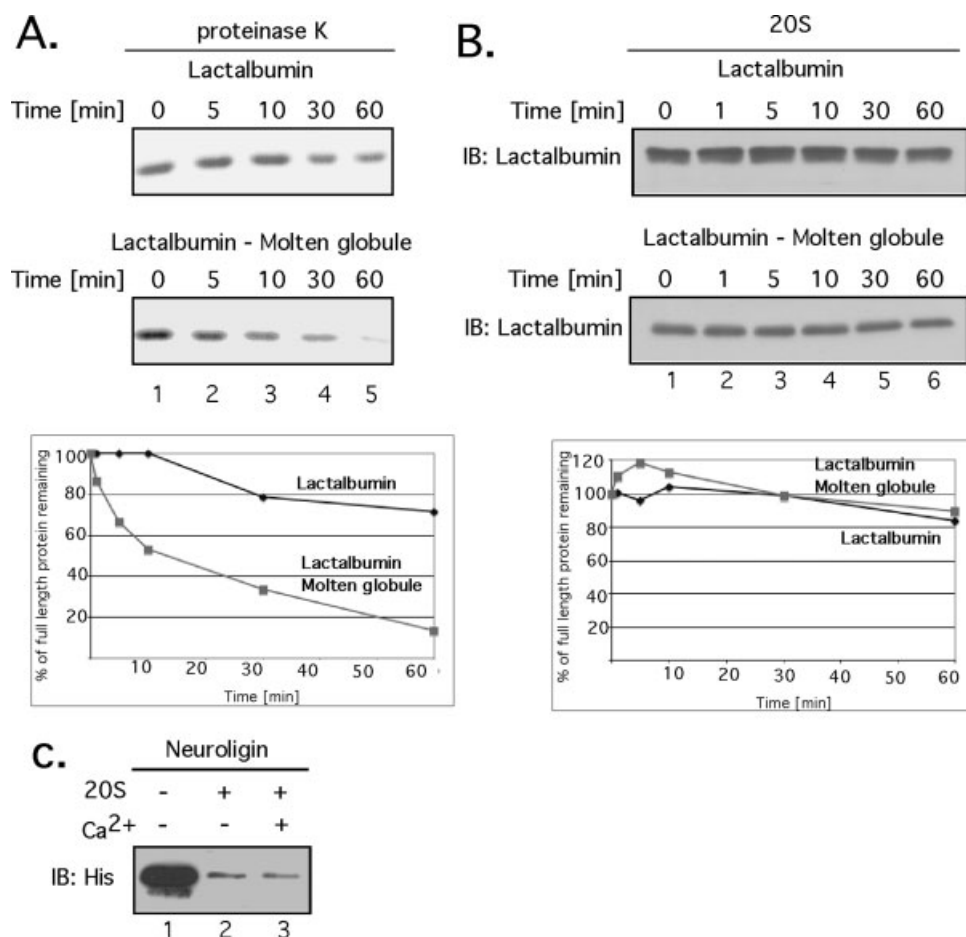
To further strengthen the contention that proteins in a MG state are resistant to 20S proteasomes, we performed similar experiments on another well-characterized long-lived MG species, that of *Torpedo californica* acetylcholinesterase (*TcAChE*). The crystal structure of *TcAChE* reveals the presence of a deeply buried free cysteine,<sup>7</sup> and it was subsequently demonstrated that chemical modification of its thiol group transforms native *TcAChE* to a MG.<sup>44</sup> Purified native *TcAChE* was resistant to degradation both by proteinase K and by 20S proteasomes [Fig. 4(A,B)]. The MG state of the enzyme was generated by treatment with *N*-ethylmaleimide, (NEM), as previously described<sup>44</sup> (see Materials and Methods). While the MG state of *TcAChE* was susceptible to proteinase K [Fig. 4(A)], it was resistant to 20S proteasomal degradation [Fig. 4(B)]. The intrinsically unstructured Gli-cyt served as a positive control to monitor for the degradative activity of the 20S proteasomal preparation [Fig. 4(B)]. Thus, not only are fully folded native proteins resistant to 20S proteasomes, but also partially unstructured but compact species.

### Complexing of an IUP can protect it from 20S proteasomal degradation

Many IUPs are believed to interact in situ with one or more partners to form functional complexes.<sup>5</sup> Such an interaction might potentially serve to protect the IUP from 20S proteasomal degradation. NL3-cyt is known to form a tertiary complex with S-SCAM and PSD95.<sup>32,45</sup> Susceptibility of [<sup>35</sup>S]-labeled NL3-cyt to 20S proteasomes was tested in the presence of [<sup>35</sup>S]-labeled S-SCAM, of [<sup>35</sup>S]-labeled PSD95, or of both. Incubation in the presence of [<sup>35</sup>S]-labeled HDMX, a human protein, which is a critical regulator of p53 activation,<sup>46</sup> and which is not expected to interact with NL3-cyt, served as a control. NL3-cyt was completely degraded by 20S proteasomes (Fig. 5, lanes 1,2), whereas the addition of either S-SCAM or PSD95 retarded degradation, and the addition of both was even more effective (Fig. 5, lanes 3–5). Thus complexation can indeed protect an IUP from 20S proteasomal degradation; consequently, susceptibility to 20S proteasomal degradation can provide not only an assay for identifying unstructured proteins, but also for screening for putative partners.

## DISCUSSION

The fact that many proteins and protein domains are intrinsically unstructured has only been widely recognized in recent years. It is now believed, based on various *in silico* predictors for protein disorder, that at least 25%

**Figure 3**

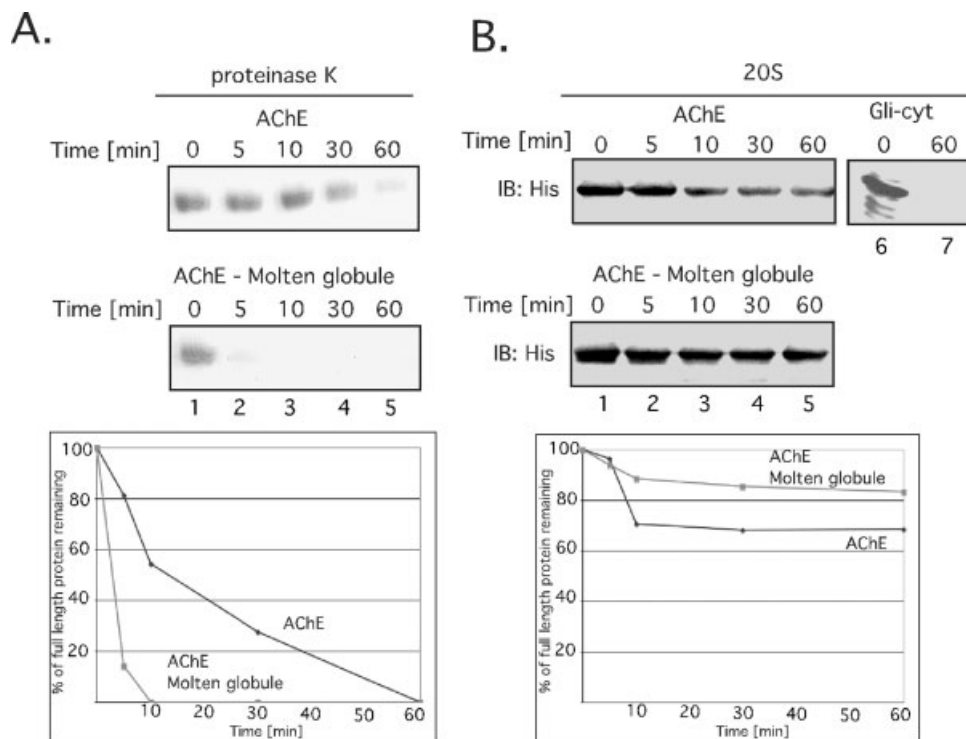
Degradation of  $\alpha$ LA in its native and MG states by proteinase K and by 20S proteasomes. (A)  $\text{Ca}^{2+}$ -depleted bovine  $\alpha$ LA was incubated with proteinase K at 37°C in the absence (–) or presence (+) of 5 mM  $\text{CaCl}_2$ ; (B)  $\text{Ca}^{2+}$ -depleted  $\alpha$ LA was incubated in the presence of purified 20S proteasomes at 37°C in the absence or presence of  $\text{Ca}^{2+}$ ; (C) NL3-cyt was incubated in the presence of purified 20S proteasomes at 37°C for 60 min in the absence or presence of  $\text{CaCl}_2$ . Samples were subjected to SDS-PAGE, and the protein bands were visualized by Gelcode staining (A) or by immunoblot analysis using rabbit anti- $\alpha$ LA antibodies (B) or mouse anti-His antibodies (C).

of the sequences in SwissProt contain long disordered regions, and that in eukaryotes this value may be as high as 63%.<sup>47</sup> The growing awareness of intrinsic disorder in proteins has, in turn, resulted in growing evidence that IUPs may play functional roles in such important processes as transcription and translation, as well as in signaling cascades, such as the complex network of the postsynaptic density in neurons.<sup>48</sup> In view of the growing interest in IUPs and their functional role, efforts have been made to develop both *in silico* and *in vitro* approaches for the rapid identification and characterization of IUPs, and for monitoring changes in their folding status and susceptibility to degradation upon interaction with functional partners.

Various *in silico* predictors of intrinsic disorder use different parameters and algorithms for their analysis, and by using a number of predictors one can get a more accurate representation of the protein's folding state.<sup>49</sup> For example, the construct of S-SCAM that we have uti-

lized herein is composed of two WW domains, linked to a PDZ domain. FoldIndex depicts both WW domains as disordered [Fig. 1(B)], in contrast to IUPred, which predicts that they are both mostly ordered (data not shown). This protein exemplifies the uncertainties involved in use of the current prediction algorithms. In the 20S degradation assay, the protein remained stable, indicating that it is not disordered, as is also suggested by the NMR structure of its PDZ domain (1UEQ). Thus, the 20S proteasomal degradation assay can give a more operational definition of IUPs in cases in which different prediction algorithms yield conflicting results.

We described earlier an assay for the identification of IUPs that is based on susceptibility to 20S proteasomal degradation. We have shown that the 20S proteasomes are capable of digesting only IUPs and unstructured domains, whereas globular proteins in their fully folded native conformation are resistant. Our technique is appli-

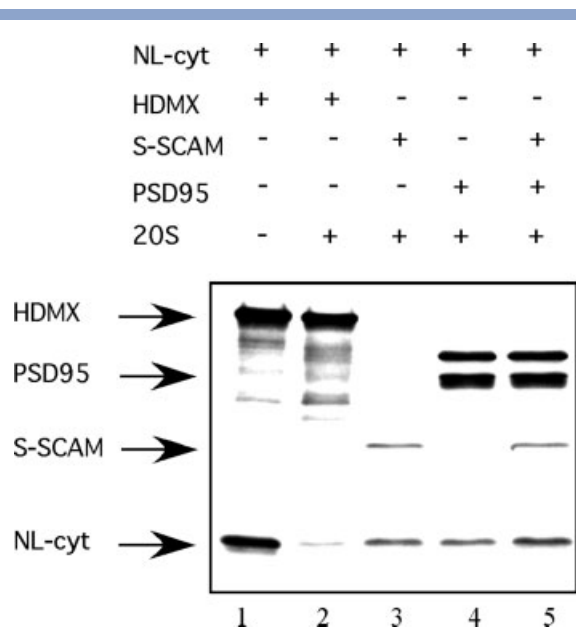
**Figure 4**

Degradation of TcAChE in its native and MG states by proteinase K and by 20S proteasomes. Native and MG preparations of TcAChE were incubated with proteinase K at 24°C (A), and with 20S proteasomes at 37°C, with Gli-cyt, incubated with 20S proteasomes at 37°C for 60 min, serving as a control (B). Samples were subjected to SDS-PAGE, and the protein bands were revealed using Gelcode staining (A), or immunoblot analysis using anti-TcAChE antibodies for TcAChE, and mouse anti-His antibodies for Gli-cyt (B).

cable to proteins that are only partially disordered, such as p53.<sup>8,50</sup> The question arises whether the structured portion escapes 20S degradation. In the present study, we used [<sup>35</sup>S]-methionine-labeled proteins that permit detection of residual fragments. With the exception of VirE2, we observed no smaller-labeled fragments after incubation with the 20S proteasomes, suggesting that the proteins were completely degraded to smaller fragments, rather than cleaved. However, certain motifs may be resistant to degradation as was reported for the polyQ chain.<sup>51</sup> We have further shown that well-defined preparations of proteins in a long-lived MG state, in which the protein is devoid of most of its tertiary structure, but retains its secondary structure, and is still quite compact,<sup>39</sup> are also resistant to the 20S proteasome. Taken together, our results suggest that the susceptibility to 20S proteasomal degradation is a simple, rapid, and reliable method of operationally defining IUPs and unstructured sequences. In addition, utilization of this method could lead to the characterization of motifs that sensitize IUPs to proteasomal degradation. An important issue is whether IUPs are present as such within the living cell. Susceptibility to the 20S proteasome *in vitro* may be extended in the future to address the issue of the status

of IUPs *in vivo*, either by inducing the 20S proteasomal degradation or by selectively blocking 26S proteasome activity, but not that of the 20S proteasome. Proteins found to undergo proteasomal degradation under such conditions are likely to have retained their IU status *in vivo*.

Both the native and MG conformations of TcAChE and bovine  $\alpha$ LA were resistant to 20S proteasomal degradation, whereas proteinase K digested the MG but not the native forms. These observations narrow the specificity of the 20S proteasomes to IUPs and other fully unstructured sequences. The MG species are significantly more flexible than the corresponding native species, based on their capacity to undergo disulfide reshuffling<sup>40</sup> as well as being highly susceptible to proteinase K. Nevertheless, they apparently do not present sufficiently extended unstructured sequences to the 20S proteasome to permit their penetration into its cavity so as to undergo peptide bond cleavage. Given our demonstration that folded proteins are resistant to both degradation methods, that IUPs are sensitive to both, and that proteins in the MG state are sensitive to proteinase K degradation, but resistant to 20S proteasomes, their utilization in tandem provides a convenient operational tool for discriminating between the three states.

**Figure 5**

Binding of S-SCAM and PSD95 protects NL3-cyt against 20S proteasomal degradation. *In vitro*-translated [<sup>35</sup>S]-NL3-cyt was incubated for 20 min at 37°C in the presence or absence of 20S proteasomes, with or without [<sup>35</sup>S]-PSD95, S-SCAM or both; incubation in the presence of [<sup>35</sup>S]-HDMX, which is not expected to bind to NL3-cyt, served as a control. Samples were subjected to SDS-PAGE, and [<sup>35</sup>S]-labeled proteins were detected by autoradiography.

As mentioned earlier, 36–63% of the proteins in eukaryotic proteomes contain long sequences that are intrinsically disordered *in vitro*.<sup>47</sup> This does not imply that these IUPs, or protein sequences, are necessarily disordered *in vivo*. On the contrary, in many cases convincing evidence has been presented that they form functional complexes within the cell, adapting a well-defined conformation upon interaction with their partner(s). We have, indeed, shown that binding of S-SCAM and PSD95 to NL3-cyt protects it from 20S proteasomal degradation. This example, taken together with several other reports (reviewed in Ref. 52), suggests that protein–protein interactions can protect IUPs from 20S proteasomal degradation. This may be achieved either by masking the unstructured domain or by folding it. It has further been argued that their intrinsically disordered character may enable IUPs to form relatively low-affinity complexes, thus permitting plasticity and rapid turnover, and, consequently, to serve as hub proteins interacting with multiple partners.<sup>53</sup>

Since the core 20S proteasome, as such, is present in eukaryotic cells,<sup>23</sup> it is plausible that it would rapidly degrade “free” IUPs. Thus, the 20S proteasome might have an important regulatory function in eliminating excess IUPs that had not yet been incorporated into functional complexes. We earlier described the *in vivo* degradation of p53, a protein containing substantial IU sequences,<sup>8</sup>

by a novel mechanism that we named “degradation by default.”<sup>52</sup> To demonstrate that degradation by default is executed by 20S proteasomes, we performed biochemical analyses, and revealed a tripartite protein complex composed of p53, the 20S proteasome and a third protein, NQO1.<sup>50</sup> In this complex, NQO1 plays the role of gatekeeper for the 20S proteasome, which, under appropriate conditions, protects p53 from 20S proteasomal degradation. Interestingly, it is possible to induce p53 degradation by default simply by inhibiting NQO1 with known drugs, such as dicoumarol.<sup>50</sup> The same phenomenon was demonstrated also with several other IUPs (unpublished observations). Thus, IUPs can serve as substrates for the 20S proteasome also *in situ*, and their intracellular level may be controlled by the novel mechanism that we have proposed.

Finally, we believe that the current terms, namely IUPs or IDPs, do not adequately describe the functional significance of this category of proteins. Both terms are inaccurate, inasmuch as these proteins may become structured/ordered at given points in time and space, both in the presence and absence of partners. Furthermore, the use of the terms IDP and IUP may also give the impression that these proteins are in some way “inferior” to structured proteins. This is completely wrong, given the fact that such proteins are mainly associated with higher organisms and functions, as can be seen, for example, by their prevalence at synapses in the nervous system. We would like to propose that they be called “4D proteins,” based on the fact that their structures are not fixed, as is generally the case for “3D proteins,” but rather defined by time and space.

## MATERIALS AND METHODS

### Reagents and proteins

Proteinase K and calcium-depleted bovine  $\alpha$ LA were obtained from Sigma (St Louis, MO). The G<sub>2</sub> dimeric form of TcAChE was purified from frozen *Torpedo* electric organ tissue (Aquatic Research Consultants, San Pedro, CA) by affinity chromatography, after solubilization with phosphatidylinositol-specific phospholipase C, as described previously.<sup>7</sup>

Gli-cyt was over-expressed in *Escherichia coli* BLR (DE3) cells, and subsequently purified as described.<sup>29</sup> NL3-cyt was similarly expressed, with a six-residue His-tag at its NH<sub>2</sub>-terminus, but was purified under denaturing conditions, so as to prevent proteolytic degradation. Thus, immediately after expression, the pellet was lysed in 8M urea/100 mM NaH<sub>2</sub>PO<sub>4</sub>/10 mM Tris, pH 8.0, and NL3-cyt was captured on an Ni-NTA column (GE Healthcare, NJ) from which it was eluted with the same buffer adjusted to pH 4.5. The eluted protein was further purified by absorption on an SP FF cation exchange column (GE Healthcare), from which it was eluted using a

0–1M NaCl gradient in 50 mM MES, pH 6.0. Finally, the pure fractions eluted from the cation-exchange column were pooled and concentrated, with concomitant replacement of the denaturing buffer by 1 mM EDTA/5 mM DTT/250 mM NaCl/100 mM Tris, pH 8.5, using a 3000 MWCO Vivaspin concentration device (Sartorius, Göttingen, Germany). The concentrated protein was further purified through a HiLoad 16/60 Superdex 75 pg column (GE Healthcare) pre-equilibrated with the same buffer (Paz *et al.*, in preparation). Anti-*TcAChE* rabbit polyclonal antibodies were prepared by the Biological Services Antibody Unit of the Weizmann Institute of Science. Mouse anti-His antibodies were obtained from Sigma, and rabbit antiovine  $\alpha$ LA antiserum was from Bethyl Laboratories (Montgomery, TX).

### Plasmids

*In vitro* translation was performed using the TNT<sup>®</sup> Quick Coupled Transcription/Translation System (Promega, WI). The plasmids employed were the following: pET28-Gli-cyt,<sup>29</sup> pET28-human NL3-cyt (encoding amino-acids 731–848), pRSET-human S-SCAM (encoding amino-acids 302–510), pET28-human ER (encoding amino-acids 296–555), pET28-HNF-4 $\alpha$  (encoding amino-acids 141–464), pET28-human acid- $\beta$ -glucosidase (encoding amino-acids 20–516), pET28-GOT (full-length), pRSET-PSD95 (encoding amino-acids 65–393), pET28-VirE2,<sup>31</sup> pCDNA3-HDMX, and pET21-p21.

### Proteinase K digestion

Proteinase K (Sigma) digestion of [<sup>35</sup>S]-methionine-labeled proteins was carried out in 6  $\mu$ L digestion buffer (150 mM NaCl/100 mM Tris HCl, pH 7.5), at a proteinase K concentration of 100 ng/mL, at 37°C, essentially as described by Fontana *et al.*<sup>38</sup> Proteinase K digestion of recombinant and purified proteins was done at an *E*:*S* ratio of 1:500 (w:w), at 37°C, except for *TcAChE*, for which digestion was at 24°C, at an *E*:*S* ratio of 1:50, due to its low thermal stability.<sup>54</sup> Samples were then mixed with Laemmli sample buffer containing 5 mM PMSE, heated at 95°C for 5 min, subjected to SDS-PAGE, and detected either by Gelcode staining or by autoradiography.

### *In vitro* 20S proteasomal degradation assay

Purified 20S proteasomes were generated as described previously.<sup>50</sup> Degradation of [<sup>35</sup>S]-methionine-labeled proteins translated *in vitro* was carried out with 1  $\mu$ g of the purified 20S proteasomes in 30  $\mu$ L of the same degradation buffer as used for proteinase K digestion, at 37°C for 1 h. Sample preparation and SDS-PAGE were as described earlier, and detection was by autoradiography.

### Generation of the *TcAChE* MG

A MG species of *TcAChE* was generated by treatment with NEM, essentially as described.<sup>44</sup> Briefly, 200  $\mu$ g of purified *TcAChE* in 100  $\mu$ L of 5 mM NEM/100 mM NaCl/50 mM sodium phosphate, pH 7.0, were incubated for 20 h at 4°C. Excess NEM was separated from the modified protein on a 5-mL desalting column (GE Healthcare).

### ACKNOWLEDGMENTS

We thank Drs. Shira Albeck, Yoav Peleg, and Tamar Unger, of the Israel Structural Proteomics Center (ISPC), for generously making available a series of plasmids, Tzviya Zeev-Ben-Mordehai for valuable discussions, Michal Cohen and Lilly Toker for the purified *TcAChE*, Dr. Orit Leitner for the anti-*TcAChE* antibodies, Prof. Moshe Oren (Department of Molecular Cell Biology, WIS) for the pCDNA3-HDMX plasmid and Dr. Michele Pagano (Department of Pathology, NYU Medical Center, New York) for the pET21-p21 plasmid.

### REFERENCES

1. Kuwajima K. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* 1989;6:87–103.
2. Fink AL. Chaperone-mediated protein folding. *Physiol Rev* 1999;79:425–449.
3. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. *Biochemistry* 2002;41:6573–6582.
4. Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 2000;41:415–427.
5. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 2005;6:197–208.
6. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* 2005;579:3346–3354.
7. Sussman JL, Harel M, Frolow F, Oefner C, Goldman A, Toker L, Silman I. Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science* 1991;253:872–879.
8. 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.
9. Friedler A, Veprintsev DB, Freund SM, von Glos KI, Fersht AR. Modulation of binding of DNA to the C-terminal domain of p53 by acetylation. *Structure* 2005;13:629–636.
10. Smet C, Leroy A, Sillen A, Wieruszkeski JM, Landrieu I, Lippens G. Accepting its random coil nature allows a partial NMR assignment of the neuronal Tau protein. *Chembiochem* 2004;5:1639–1646.
11. Bruschweiler R, Liao X, Wright PE. Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* 1995;268:886–889.
12. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 2005;21:3435–3438.
13. Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guillot S, Dunker AK. Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput* 1998:437–448.

14. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinform Comput Biol* 2005;3:35–60.
15. Ferron F, Longhi S, Canard B, Karlin D. A practical overview of protein disorder prediction methods. *Proteins* 2006;65:1–14.
16. Bracken C, Iakoucheva LM, Romero PR, Dunker AK. Combining prediction, computation and experiment for the characterization of protein disorder. *Curr Opin Struct Biol* 2004;14:570–576.
17. Bochicchio B, Tamburro AM. Polyproline II structure in proteins: identification by chiroptical spectroscopies, stability, and functions. *Chirality* 2002;14:782–792.
18. Barron LD. Structure and behaviour of biomolecules from Raman optical activity. *Curr Opin Struct Biol* 2006;16:638–643.
19. 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.
20. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK. DisProt: the database of disordered proteins. *Nucleic Acids Res* 2007;35(Database issue):D786–D793.
21. Hershko A, Ciechanover A. The ubiquitin system. *Annu Rev Biochem* 1998;67:425–479.
22. Goldberg AL. Protein degradation and protection against misfolded or damaged proteins. *Nature* 2003;426:895–899.
23. Orlowski M, Wilk S. Ubiquitin-independent proteolytic functions of the proteasome. *Arch Biochem Biophys* 2003;415:1–5.
24. Voges D, Zwickl P, Baumeister W. The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu Rev Biochem* 1999;68:1015–1068.
25. Navon A, Goldberg AL. Proteins are unfolded on the surface of the ATPase ring before transport into the proteasome. *Mol Cell* 2001;8:1339–1349.
26. Liu CW, Corboy MJ, DeMartino GN, Thomas PJ. Endoproteolytic activity of the proteasome. *Science* 2003;299:408–411.
27. Coux O, Tanaka K, Goldberg AL. Structure and functions of the 20S and 26S proteasomes. *Annu Rev Biochem* 1996;65:801–847.
28. Gilbert MM, Auld VJ. Evolution of clams (cholinesterase-like adhesion molecules): structure and function during development. *Front Biosci* 2005;10:2177–2192.
29. Zeev-Ben-Mordehai T, Rydberg EH, Solomon A, Toker L, Auld VJ, Silman I, Botti S, Sussman JL. The intracellular domain of the *Drosophila* cholinesterase-like neural adhesion protein, gliotactin, is natively unfolded. *Proteins* 2003;53:758–767.
30. Bloom J, Pagano M. To be or not to be ubiquitinated? *Cell Cycle* 2004;3:138–140.
31. Frenkiel-Krispin D, Wolf SG, Albeck S, Unger T, Peleg Y, Jacobovitch J, Michael Y, Daube S, Sharon M, Robinson CV, Svergun DI, Fass D, Tzfira T, Elbaum M. Plant Transformation by *Agrobacterium tumefaciens*: modulation of single-stranded DNA-VirE2 complex assembly by VirE1. *J Biol Chem* 2007;282:3458–3464.
32. Hirao K, Hata Y, Ide N, Takeuchi M, Irie M, Yao I, Deguchi M, Toyoda A, Sudhof TC, Takai Y. A novel multiple PDZ domain-containing molecule interacting with *N*-methyl-D-aspartate receptors and neuronal cell adhesion proteins. *J Biol Chem* 1998;273:21105–21110.
33. Katzenellenbogen BS, Katzenellenbogen JA. Estrogen receptor transcription and transactivation: Estrogen receptor alpha and estrogen receptor beta: regulation by selective estrogen receptor modulators and importance in breast cancer. *Breast Cancer Res* 2000;2:335–344.
34. Parviz F, Matullo C, Garrison WD, Savatski L, Adamson JW, Ning G, Kaestner KH, Rossi JM, Zaret KS, Duncan SA. Hepatocyte nuclear factor 4 $\alpha$  controls the development of a hepatic epithelium and liver morphogenesis. *Nat Genet* 2003;34:292–296.
35. Futerman AH, Sussman JL, Horowitz M, Silman I, Zimran A. New directions in the treatment of Gaucher disease. *Trends Pharmacol Sci* 2004;25:147–151.
36. Zlotnik A, Gurevich B, Tkachov S, Maoz I, Shapira Y, Teichberg VI. Brain neuroprotection by scavenging blood glutamate. *Exp Neurol* 2007;203:213–220.
37. Tsai CJ, Polverino de Lauro P, Fontana A, Nussinov R. Comparison of protein fragments identified by limited proteolysis and by computational cutting of proteins. *Protein Sci* 2002;11:1753–1770.
38. Fontana A, de Lauro PP, Spolaore B, Frare E, Picotti P, Zambonin M. Probing protein structure by limited proteolysis. *Acta Biochim Pol* 2004;51:299–321.
39. Ptitsyn OB. Molten globule and protein folding. *Adv Protein Chem* 1995;47:83–229.
40. Eichler J, Kreimer DI, Varon L, Silman I, Weiner L. A “molten globule” of *Torpedo* acetylcholinesterase undergoes thiol-disulfide exchange. *J Biol Chem* 1994;269:30093–30096.
41. Ren J, Stuart DI, Acharya KR. Alpha-lactalbumin possesses a distinct zinc binding site. *J Biol Chem* 1993;268:19292–19298.
42. Kuwajima K. The molten globule state of alpha-lactalbumin. *FASEB J* 1996;10:102–109.
43. Wenzel T, Baumeister W. Conformational constraints in protein degradation by the 20S proteasome. *Nat Struct Biol* 1995;2:199–204.
44. Dolginova EA, Roth E, Silman I, Weiner LM. Chemical modification of *Torpedo* acetylcholinesterase by disulfides: appearance of a “molten globule” state. *Biochemistry* 1992;31:12248–12254.
45. Irie M, Hata Y, Takeuchi M, Ichtchenko K, Toyoda A, Hirao K, Takai Y, Rosahl TW, Sudhof TC. Binding of neuroligins to PSD-95. *Science* 1997;277:1511–1515.
46. Wade M, Wong ET, Tang M, Stommel JM, Wahl GM. Hdmx modulates the outcome of p53 activation in human tumor cells. *J Biol Chem* 2006;281:33036–33044.
47. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* 2000;11:161–171.
48. Sheng M, Hoogenraad CC. The postsynaptic architecture of excitatory synapses: a more quantitative view. *Annu Rev Biochem* 2007;76:823–847.
49. Schlessinger A, Liu J, Rost B. Natively unstructured loops differ from other loops. *PLoS Comput Biol* 2007;3:e140.
50. Asher G, Tsvetkov P, Kahana C, Shaul Y. A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. *Genes Dev* 2005;19:316–321.
51. Venkatraman P, Wetzel R, Tanaka M, Nukina N, Goldberg AL. Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Mol Cell* 2004;14:95–104.
52. Asher G, Reuven N, Shaul Y. 20S proteasomes and protein degradation “by default”. *Bioessays* 2006;28:844–849.
53. Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* 2006;5:2985–2995.
54. Kreimer DI, Shnyrov VL, Villar E, Silman I, Weiner L. Irreversible thermal denaturation of *Torpedo californica* acetylcholinesterase. *Protein Sci* 1995;4:2349–2357.