doi:10.1016/j.jmb.2011.07.024 J. Mol. Biol. (2011) 412, 319–324

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

COMMUNICATION

Intrinsic Disorder in Ubiquitination Substrates

Tzachi Hagai, Ariel Azia, Ágnes Tóth-Petróczy and Yaakov Levy*

Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Received 23 April 2011;
received in revised form
30 June 2011;
accepted 14 July 2011
Available online
23 July 2011

Edited by P. Wright

Keywords:
intrinsically disordered;
proteasomal degradation;
ubiquitination;
SUMOylation;
acetylation

The ubiquitin–proteasome system is responsible for the degradation of numerous proteins in eukaryotes. Degradation is an essential process in many cellular pathways and involves the proteasome degrading a wide variety of unrelated substrates while retaining specificity in terms of its targets for destruction and avoiding unneeded proteolysis. How the proteasome achieves this task is the subject of intensive research. Many proteins are targeted for degradation by being covalently attached to a poly-ubiquitin chain. Several studies have indicated the importance of a disordered region for efficient degradation. Here, we analyze a data set of 482 *in vivo* ubiquitinated substrates and a subset in which ubiquitination is known to mediate degradation. We show that, in contrast to phosphorylation sites and other regulatory regions, ubiquitination sites do not tend to be located in disordered regions and that a large number of substrates are modified at structured regions. In degradation-mediated ubiquitination, there is a significant bias of ubiquitination sites to be in disordered regions; however, a significant number is still found in ordered regions. Moreover, in many cases, disordered regions are absent from ubiquitinated substrates or are located far away from the modified region. These surprising findings raise the question of how these proteins are successfully unfolded and ultimately degraded by the proteasome. They indicate that the folded domain must be perturbed by some additional factor, such as the p97 complex, or that ubiquitination may induce unfolding.

Protein degradation plays a central role in regulating various cellular functions. Proteins are often targeted for proteasomal degradation by tagging them with a poly-ubiquitin chain. However, this attachment may not be sufficient, and it was shown that an additional component, namely, an unstructured region, is essential for efficient degradation.[1] The disordered region's functionality was studied and was shown to be important in the initial stages of the binding of the substrate to the proteasome and its commitment to degradation.[2] Recently, the characteristics of this unstructured region have been investigated in detail. In several natural[3] and designed[4] proteasome substrates, a disordered region at either the N- or the C-terminus was found to promote degradation; however, an internal disordered region was also reported to assist the proteasome machinery.[3,5,6] It was observed that the unstructured tail must have a minimal length of 20–30 amino acids in order to support efficient degradation of poly-ubiquitinated substrates, while internal disordered regions must be much longer.[7] Additionally, the space separation between the ubiquitin moiety and the unstructured region imposes restrictions on degradation efficiency.

In order to substantiate the reported results and explore how these characteristics are realized in various proteins, we investigated the structural properties of natural ubiquitination substrates. We composed a data set of 465 mammalian and 233 yeast proteins whose *in vivo* ubiquitination and SUMOylation sites have been determined experimentally

*Corresponding author. E-mail address:
koby.levy@weizmann.ac.il.

(Tables S1 and S2). Since ubiquitination mediates various pathways in addition to proteasomal degradation,[8] we composed a subset of 42 proteins that are known experimentally to undergo ubiquitination, which mediates their proteasomal degradation.

We analyzed their sequence properties using the disorder predictors IUPred[9] and PONDR-FIT[10] as well as secondary structure assignment tools. The structural properties of the modification sites were compared to the corresponding properties of the unmodified residues in the same set. In addition, we analyzed other posttranslational modifications, namely, acetylation and phosphorylation, that occur in the same data set.

As observed previously,[11] phosphorylation sites tend to be located predominantly in disordered regions (68.4%, 60.1% and 26.9% for Ser, Thr and Tyr phosphosites, respectively), with only a minority of them found in structured domains. This tendency is particularly strong for serine and threonine residues and less marked for tyrosine (Fig. 1a). Unlike phosphorylation sites and other regulatory regions,[12–16] lysine residues that serve as modification sites are often located in structured regions (Fig. 1b and Figs. S1 and S2). In agreement with previous results,[17] lysine residues that serve as acetylation sites tend to be located in structured regions (67.7% of acetylated and 75.0% of unmodified lysine residues are predicted to be in structured regions). Similarly to acetylation and in contrast to SUMOylation (for which a larger fraction of sites are located in disordered regions), ubiquitination sites are mostly located in structured regions (73.0% of ubiquitinated lysine residues in the proteins studied are in structured regions) and do not show a bias in comparison to unmodified residues ($P=0.1$, Fisher exact test).

When the data set was separated into its yeast and mammalian components, we observe that, in the yeast set, Lys residues that serve as ubiquitination sites show a small but significant bias toward disordered regions in comparison to unmodified Lys residues ($P=7\times10^{-5}$, Fisher exact test), in agreement with previous suggestions;[18] however, the large majority of yeast ubiquitination sites are still found in structured regions (75.4% for ubiquitination sites *versus* 83.7% for unmodified residues).

Analyzing the secondary structure regions in which these sites are found yields results consistent with those shown in Fig. 1. Ser and Thr phosphorylation sites occur more often in coils than in other secondary structures both in terms of absolute percentages and in comparison to their unmodified counterparts, while Tyr phosphorylation sites are more evenly distributed across secondary structural

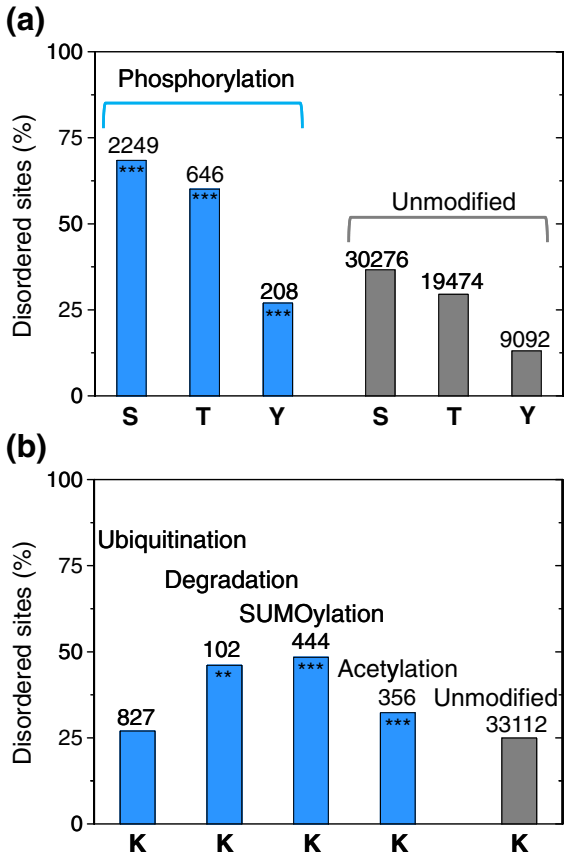

Fig. 1. The tendency of various modification sites to be disordered. Percentages of residues predicted by the IUPred algorithm[9] to be located in disordered regions in a set of 698 mammalian and yeast proteins. (a) Serine, threonine and tyrosine residues that are known to function as phosphorylation sites are significantly more likely ($P<10^{-6}$, Fisher exact test) to be located in disordered regions (68.4%, 60.1% and 26.9%, which correspond to 1539, 388 and 56 sites, respectively) than their counterparts that are not known to serve this function (unmodified residues: 36.7%, 29.5% and 13.1%, respectively, which correspond to 11,097, 5748 and 1189 residues, respectively). (b) The percentages of lysine residues predicted to be located in disordered regions: ubiquitination sites (27.0%, 223 sites), degradation-related ubiquitination sites (46.1%, 47 sites), SUMOylation sites (48.4%, 215 sites) and acetylation sites (32.3%, 115 sites) and lysine residues that are not known to be modified (25.0%, 8262 residues). SUMOylation, acetylation and degradation-related ubiquitination sites are significantly biased toward disordered regions compared to unmodified lysine residues ($P=1\times10^{-10}$, 1×10^{-3} and 2.9×10^{-6}, respectively; Fisher exact test), but the majority of sites are still in structured regions. Ubiquitination sites in the entire set are not significantly biased ($P=0.1$, Fisher exact test). Similar results are obtained using the PONDR-FIT algorithm10 (Fig. S2), although the numbers of residues predicted to be located in disordered regions are higher by up to 10%. The statistical significance of the disordered tendency in modified compared to the unmodified sites is marked with asterisks (single asterisk shows a significance of $P<0.05$, double asterisks show a significance of $P<0.01$ and triple asterisks show a significance of $P<0.001$).

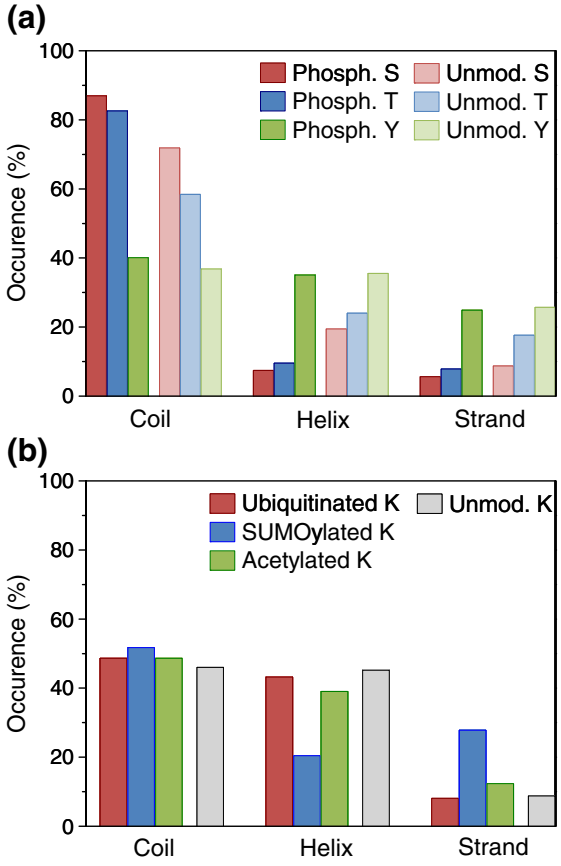

Fig. 2. Secondary structure preferences of various modification sites. (a) The secondary structure preferences of phosphorylated serine and threonine are biased toward coil regions in comparison to all residues ($P<10^{-15}$, Fisher exact test). Phosphorylated tyrosine residues are not significantly biased toward residing in coils ($P=0.17$). (b) Secondary structure preferences of ubiquitinated, SUMOylated, acetylated and all lysine residues. Lysine residues that are involved in modifications are distributed similarly to those that are not ($P=0.174$ for acetylation sites residues and $P=0.174$ for ubiquitination sites; Fisher exact test), with the exception of SUMOylation sites, which tend to occur more in coils ($P=0.009$, Fisher exact test).

elements (Fig. 2a). With respect to Lys modification, ubiquitination and acetylation sites clearly favor location in a coil or a helix over a strand, and in this, their distribution parallels that of their unmodified counterparts (Fig. 2b). The amino acid composition surrounding ubiquitination sites that reside on coils is similar to disordered regions, while ubiquitination sites on helices or strands tend to be surrounded by ordered residues (Table S4).

Our finding that a large fraction of lysine modification sites are located in ordered regions is surprising in light of previous suggestions concerning the structural preferences of posttranslational sites and considering the discussion above regarding the importance of disorder in the vicinity of ubiquitination sites for efficient degradation.

Since some of the ubiquitination sites in our protein data set are related to pathways other than proteasomal degradation or have uncharacterized functions, we formulated a smaller subset composed of mammalian proteins whose ubiquitination sites are linked to their degradation (42 proteins, 109 sites). In this subset of degradation-related ubiquitination proteins, a smaller percentage of the modification sites tend to be in structured domains (Fig. 1b) (and a correspondingly larger percentage, 55.1%, of the sites are on coils); however, although this bias is significant ($P=2\times10^{-4}$, Fisher exact test), only 46.1% of the sites are predicted to be in disordered regions. Thus, ubiquitination sites, even those that are linked to degradation, are often located in structured regions.

We next examined the existence of disordered regions, that is, long stretches of at least 20 consecutive residues that are predicted to be disordered, either at the protein termini (tails) or in internal regions. These disordered regions were defined as continuous stretches of residues that are predicted to be disordered and can be separated by less than three consecutive structured residues. The presence of these regions was shown to be imperative for degradation in several studies,[3,4] and the absence of them was suggested to enable the degradation resistance of Rad23.[7] In the overall data set, we found that ~70% of the ubiquitination substrates in yeast (Fig. S3) and in mammalian substrates (Fig. 3a) lack unstructured tails of that length. In the smaller set of ubiquitination sites that are specifically linked to degradation, a similar fraction of proteins have no such tail (Fig. 3a, inset). Since SUMOylation is a ubiquitin-like modification that is not as strongly related to proteasomal degradation, we examined the frequency of disordered regions in SUMOylation substrates and observed comparable numbers (Fig. 3a, inset). Several studies have suggested that the proteasome may degrade or process proteins from internal regions,[3,5,6] and thus, we repeated this analysis including both disordered tails and internal regions. In all cases, we observe a larger number of substrates having disordered regions (70–80%), but still a significant portion is predicted to lack a sufficiently long region. Moreover, we note that internal disordered regions may need to be much longer than 20–30 residues to assist degradation,[7] and thus, these numbers may overestimate the fraction of substrates that have inherently disordered regions that can be efficiently utilized by the proteasome. These results are similar to the observed percentages of proteins classified as "intrinsically disordered" in entire eukaryote proteomes.[19–21]

We analyzed the sequence separation between the ubiquitination sites and the disordered tails and

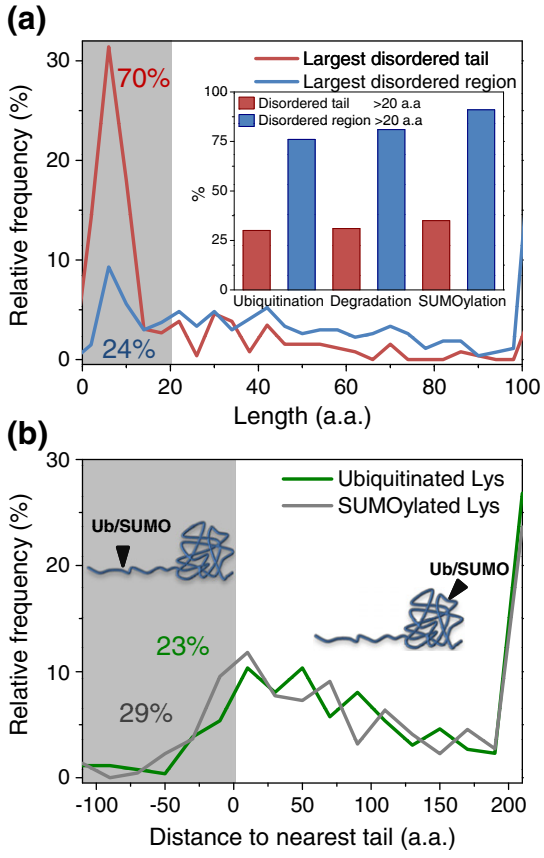

Fig. 3. Presence of disordered regions in substrates for ubiquitination and SUMOylation. (a) Distribution of ubiquitination substrates based on the longest disordered tail (red) and disordered region (which include both tails and internal regions) (blue) in mammalian proteins that undergo ubiquitination. Seventy percent of the proteins (184 proteins) do not have a disordered tail, and 24% (65 proteins) do not have any disordered region (internally or at a terminal) that is at least 20 amino acids long (gray band). Inset: percentages of proteins having a disordered tail (red) and disordered region (blue) in ubiquitination substrates, degradation-l$2inked ubiquitination substrates and SUMOylation substrates. In yeast proteins, similar results are observed (Fig. S3). (b) Sequence separation between the modification site (a lysine residue) and the nearest disordered tail in ubiquitinated (green) and SUMOylated (gray) mammalian proteins. A negative number (gray band) indicates that the site is localized at the tail, and its value corresponds to the distance between the modification site and the end of the tail. Twenty-three percent of ubiquitination sites and 29% of SUMOylation sites are within disordered tails. Similar results are seen in yeast proteins.

found that 23% of the sites in mammalian proteins are located within the tail, whereas the remainder may be located much further along the sequence (Fig. 3b). However, sequence separation does not imply proximity in space, and for efficient degradation, the disordered region and the ubiquitination site need to be spatially close.[4] To examine the spatial separation between the unstructured region and the ubiquitination site, we collected all substrates whose solved three-dimensional structures include the degradation-linked ubiquitination site and for which the structure includes at least part of the unstructured region to allow a plausible distance calculation. Currently, there are only a few substrates that fulfill all these requirements (three proteins, seven sites; see Table S5). When the spatial separation between the unstructured region and the ubiquitination site in this set was calculated, we observe distances of 12–28 Å, in agreement with a recent study conducted on an artificial substrate.[4] From this analysis, we can infer the required distance between the two regions that enables successful proteasomal degradation. We then conducted the same analysis on a larger set of ubiquitinated proteins for which not all sites are known to be targeting substrates to degradation. In this set (12 proteins, 20 sites; Table S5), we observe a large variation in the distance between the disordered region and the ubiquitinated lysine (10–73 Å). In some of the proteins, we observe a distance that is far greater than the estimated distance to coordinate proteasomal engagement. These results indicate that some of these regions may not be implicated in proteasomal degradation or that additional conformational changes or unfolding events are needed within the targeted substrate.

The presence of a disordered region in the vicinity of the ubiquitination site was shown to be important for efficient degradation of poly-ubiquitinated substrates in the 26S proteasome.[1,2] However, an interesting picture emerges from comprehensively analyzing the structural attributes of a large set of natural ubiquitination substrates: ubiquitination sites that are involved in proteasomal degradation are relatively enriched in disordered regions; however, ~50% of these sites are located in structured domains on the surface of the modified substrate. Moreover, disordered regions, which are needed for efficient degradation, are sometimes too short, completely absent or located far away from the ubiquitin tag.

How can proteins that do not possess inherently unstructured regions in proximity to the ubiquitin tag be efficiently degraded by the proteasome? Tightly folded domains can be difficult to unfold and degrade slowly.[22] It is possible that local unfolding events that expose a temporarily disordered region can be utilized by the proteasome. Alternatively, external factors may promote at least partial unfolding of the substrate prior to its engagement by the proteasome. Recently, it has been suggested that the p97 complex assists in unfolding substrates that lack disordered regions.[23] Another possible means to disrupt the folded structure and induce substrate unfolding prior to its binding to the proteasome is the use of

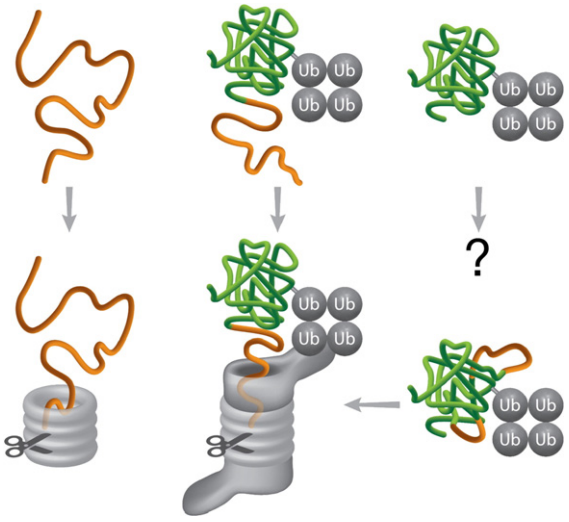

Fig. 4. Structural characteristics and proteasomal degradation pathways. (a) Proteins containing a high fraction of intrinsically disordered regions can be degraded successfully by the 20S catalytic particle (shown as a cylinder) of the 26S proteasome without prior unfolding in a ubiquitin-independent process. (b) Most proteins are thought to be tagged for degradation by ubiquitin attachment and, subsequently, are unfolded and degraded in the 26S proteasome. In many cases, these proteins contain disordered regions in the vicinity of the ubiquitin tagging that enable successful engagement by the proteasome. (c) Proteins lacking an inherently disordered region long enough for proteasomal engagement may use additional mechanisms to induce partial unfolding prior to degradation. These mechanisms may be external factors, such as the p97 complex, or posttranslational modifications that can perturb the folded structure such that a sufficiently long disordered region is exposed.

posttranslational modifications that may alter the biophysical properties of many substrates. Phosphorylation is known to be capable of inducing order-to-disorder transitions,[24] which may expose unstructured regions that would suffice for efficient degradation. Since many substrates are phosphorylated prior to their degradation,[25] it is possible that, in some of these proteins, the phosphorylation may also induce unfolding. For example, phosphorylation of the cyclin-dependent kinase inhibitor $p19^{4inkd}$ has been suggested to destabilize its structure and to assist in its degradation.[26] Another modification that may induce unfolding is the ubiquitin moiety itself, which is commonly attached to many substrates. Since the covalent conjugation of one protein to another was shown to significantly alter the physical properties of the conjugate proteins, in both natural[27] and artificial systems,[28] it is possible that the attachment of ubiquitin chains may assist local unfolding that would introduce the required disordered region. This concept was suggested many years ago;[29] however, due to technical difficulties, it was never tested experimentally. A recent computational study[30] showed that the yeast Ubc7p, a natural ubiquitination substrate that is successfully degraded despite lacking any disordered region,[31] undergoes a significant thermal destabilization and perturbation of the structured domain upon ubiquitination with a Lys48-linked poly-ubiquitin chain. The potentially active role of the ubiquitin conjugates in assisting degradation by inducing structural destabilization should be examined for other systems and clearly awaits experimental verification.

Overall, we can classify three different types of proteasomal engagement based on the protein's structural properties (Fig. 4). Some substrates, such as α-synuclein and p21, are inherently disordered and can sometimes be degraded in a ubiquitin-independent manner by the 20S proteasome.[32] These substrates may even need protection to avoid rapid degradation, as proposed in the "nanny model."[33]

Other proteins that contain a greater content of structured regions often require ubiquitination and unfolding before degradation in the 26S proteasome. These substrates may already contain the inherent disordered regions needed for efficient degradation (Fig. 4b), as in p53 and β-catenin. Other substrates, such as p19 and Ubc7p, lack these unstructured regions, and their structural dynamics can be modified by either posttranslational modifications or external factors (Fig. 4c).

Acknowledgements

We would like to thank Vladimir Uversky for supplying disorder predictions using the PONDR-FIT algorithm and Monika Fuxreiter for helpful discussions. This work was supported by the Kimmelman Center for Macromolecular Assemblies and the Israel Science Foundation.

Supplementary Data

Supplementary data to this article can be found online at doi:10.1016/j.jmb.2011.07.024

References

1. Schrader, E. K., Harstad, K. G. & Matouschek, A. (2009). Targeting proteins for degradation. *Nat. Chem. Biol.* **5**, 815–822.
2. Peth, A., Uchiki, T. & Goldberg, A. L. (2010). ATP-dependent steps in the binding of ubiquitin conjugates to the 26S proteasome that commit to degradation. *Mol. Cell*, **40**, 671–681.
3. Zhao, M., Zhang, N. Y., Zurawel, A., Hansen, K. C. & Liu, C. W. (2010). Degradation of some polyubiquitinated

proteins requires an intrinsic proteasomal binding element in the substrates. *J. Biol. Chem.* **285**, 4771–4780.
4. Inobe, T., Fishbain, S., Prakash, S. & Matouschek, A. (2011). Defining the geometry of the two-component proteasome degron. *Nat. Chem. Biol.* **7**, 161–167.
5. Liu, C. W., Corboy, M. J., DeMartino, G. N. & Thomas, P. J. (2003). Endoproteolytic activity of the proteasome. *Science*, **299**, 408–411.
6. Piwko, W. & Jentsch, S. (2006). Proteasome-mediated protein processing by bidirectional degradation initiated from an internal site. *Nat. Struct. Mol. Biol.* **13**, 691–697.
7. Fishbain, S., Prakash, S., Herrig, A., Elsasser, S. & Matouschek, A. (2011). Rad23 escapes degradation because it lacks a proteasome initiation region. *Nat. Commun.* **2**, 192.
8. Grabbe, C., Husnjak, K. & Dikic, I. (2011). The spatial and temporal organization of ubiquitin networks. *Nat. Rev., Mol. Cell Biol.* **12**, 295–307.
9. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.
10. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. (2010). PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta*, **1804**, 996–1010.
11. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–1049.
12. Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev., Mol. Cell Biol.* **6**, 197–208.
13. Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527–533.
14. Tompa, P. (Ed.). (2009). *Structure and Function of Intrinsically Disordered Proteins*, 1st edit., Taylor & Francis Group, Boca Raton, FL.
15. Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N. & Dunker, A. K. (2007). Intrinsic disorder and functional proteomics. *Biophys. J.* **92**, 1439–1456.
16. Babu, M. M., van der Lee, R., de Groot, N. S. & Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* **21**, 432–440.
17. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C. *et al.* (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*, **325**, 834–840.
18. Radivojac, P., Vacic, V., Haynes, C., Cocklin, R. R., Mohan, A., Heyen, J. W. *et al.* (2009). Identification, analysis, and prediction of protein ubiquitination sites. *Proteins*, **78**, 365–380.
19. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635–645.
20. Vucetic, S., Brown, C. J., Dunker, A. K. & Obradovic, Z. (2003). Flavors of protein disorder. *Proteins*, **52**, 573–584.
21. Gsponer, J., Futschik, M. E., Teichmann, S. A. & Babu, M. M. (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science*, **322**, 1365–1368.
22. Henderson, A., Erales, J., Hoyt, M. A. & Coffino, P. (2011). Dependence of proteasome processing rate on substrate unfolding. *J. Biol. Chem.* **286**, 17495–17502.
23. Beskow, A., Grimberg, K. B., Bott, L. C., Salomons, F. A., Dantuma, N. P. & Young, P. (2009). A conserved unfoldase activity for the p97 AAA-ATPase in proteasomal degradation. *J. Mol. Biol.* **394**, 732–746.
24. Johnson, L. N. & Lewis, R. J. (2001). Structural basis for control by phosphorylation. *Chem. Rev.* **101**, 2209–2242.
25. Hunter, T. (2007). The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol. Cell*, **28**, 730–738.
26. Barrick, D. (2009). Biological regulation via ankyrin repeat folding. *ACS Chem. Biol.* **4**, 19–22.
27. Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A. & Clarke, J. (2007). The folding and evolution of multidomain proteins. *Nat. Rev., Mol. Cell Biol.* **8**, 319–330.
28. Kim, Y. H. & Stites, W. E. (2008). Effects of excluded volume upon protein stability in covalently cross-linked proteins with variable linker lengths. *Biochemistry*, **47**, 8804–8814.
29. Varshavsky, A. (1992). The N-end rule. *Cell*, **69**, 725–735.
30. Hagai, T. & Levy, Y. (2010). Ubiquitin not only serves as a tag but also assists degradation by inducing protein unfolding. *Proc. Natl Acad. Sci. USA*, **107**, 2001–2006.
31. Ravid, T. & Hochstrasser, M. (2007). Autoregulation of an E2 enzyme by ubiquitin-chain assembly on its catalytic residue. *Nat. Cell Biol.* **9**, 422–427.
32. Baugh, J. M., Viktorova, E. G. & Pilipenko, E. V. (2009). Proteasomes can degrade a significant proportion of cellular proteins independent of ubiquitination. *J. Mol. Biol.* **386**, 814–827.
33. Tsvetkov, P., Reuven, N. & Shaul, Y. (2009). The nanny model for IDPs. *Nat. Chem. Biol.* **5**, 778–781.