



www.sciencemag.org/cgi/content/full/322/5906/1365/DC1

Supporting Online Material for

Tight Regulation of Unstructured Proteins: from Transcript Synthesis to Protein Degradation

Jörg Gsponer,* Matthias E. Futschik, Sarah A. Teichmann, M. Madan Babu*

*To whom correspondence should be addressed. E-mail: madanm@mrc-lmb.cam.ac.uk (M.M.B.); jgsponer@mrc-lmb.cam.ac.uk (J.G.)

Published 28 November 2008, *Science* **322**, 1365 (2008)

DOI: 10.1126/science.1163581

This PDF file includes:

Materials and Methods
SOM Text S1 to S5
Figs. S1 to S6
Tables S1 to S12
References

Supplementary Information

for

Tight regulation of unstructured proteins: from transcript
synthesis to protein degradation

Jörg Gsponer^{1,*}, Matthias E. Futschik^{1,2,3}, Sarah A. Teichmann¹ and M. Madan Babu^{1,*}

¹MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, United Kingdom

²Institute for Theoretical Biology, Charité, Humboldt-University, Berlin, Germany

³Centre of Molecular and Structural Biomedicine, University of Algarve, Faro, Portugal

*To whom correspondence should be addressed.

Email: MMB (madanm@mrc-lmb.cam.ac.uk) or JG (jgsponer@mrc-lmb.cam.ac.uk)

Supporting Online Material

Materials and Methods

SOM Text S1 to S5

Figures S1 to S6

Tables S1 to S12

References S1 to S104

Table of contents

Materials and Methods	3
Supplementary Online Material Text	4
Text S1: Extended introduction.....	4
Text S2: The differences in protein abundance are not affected by the tagging procedure.....	5
Text S3: Analysis of the N-end rule pathway.....	5
Text S4: Kinase-substrate phosphorylation network analysis	6
Text S5: Extended discussion and conclusion.....	6
Supplementary Figures S1 to S6	10
Figure S1: Framework to investigate the regulation of unstructured proteins	10
Figure S2: Protein abundance of S, M and U using different tagging procedures.....	10
Figure S3: Distribution of N-terminal residues	11
Figure S4: Enrichment of N-terminal residues in the S, M and U group of proteins.....	11
Figure S5: Phosphorylation of S, M and U proteins.....	12
Figure S6: Mechanisms to ensure fidelity in signalling, regulation and recognition	12
Supplementary Tables S1 to S12.....	13
Table S1: Compendium of datasets used in our study	13
Table S2: List of kinases with over 50% of the targets that are IUPs.....	14
Table S3: p-values for the various cellular parameters in <i>S. cerevisiae</i>	15
Table S4: p-values for the various cellular parameters in <i>S. pombe</i> and Humans.....	16
Table S5: p-values using Kolmogorov-Smirnov test.....	17
Table S6: p-values using FoldIndex for predicting IUPs	18
Table S7: p-values using IUPred for predicting IUPs	18
Table S8: Division of the proteome into two groups (S and U) does not affect the results.....	19
Table S9: The results are independent of the number of interaction partners of the IUP	19
Table S10: The results are independent of the major sub-cellular locations	20
Table S11: The reported observations are independent of protein length	20
Table S12: GO functional analysis of the S and U group of proteins	21
Supplementary References S1 to S104.....	21

Materials and Methods

Materials

The complete proteome sequences of *S. cerevisiae*, *S. pombe*, and *H. sapiens* were obtained from the NCBI and ENSEMBL websites. Information on transcriptional complexity, mRNA abundance, transcriptional rate, transcript half-life, transcripts bound to the Puf RNA binding proteins and transcript poly-A length for *S. cerevisiae* was obtained from Balaji *et al.* (S1), Holstege *et al.* (S2), Wang *et al.* (S3), Gerber *et al.* (S4) and Beilharz and Preiss (S5). Data on protein abundance, protein half-life and translational rate for *S. cerevisiae* was obtained from Ghaemmaghami *et al.* (S6), Belle *et al.* (S7) and Arava *et al.* (S8). Information on translational noise for *S. cerevisiae* was obtained from Newman *et al.* (S9). Data on the presence or absence of TATA box, which was used as a substitute to infer transcriptional noise, for *S. cerevisiae* and *H. sapiens* were obtained from Tirosch *et al.* (S10) and Basehoar *et al.* (S11). Data on mRNA half-life, weighted poly-A tail status and protein abundance for *S. pombe* was obtained from Lackner *et al.* (S12) and Matsuyama *et al.* (S13). Information on the decay rate for human transcripts was obtained from Yang *et al.* (S14). The targets of *S. cerevisiae* kinases were extracted from the phosphorylation network determined by Ptacek *et al.* (S15) (see **Table S1**).

Methods

The prediction of intrinsic disorder was carried out using Disopred2 (S16), which is a support vector machine based prediction method. For every protein, we filtered out coiled-coil regions and transmembrane segments using the program *pfilt* (<http://bioinf.cs.ucl.ac.uk/downloads/pfilt>). We then calculated the fraction of the sequence that was predicted to be unstructured. Depending on this fraction, we classified each protein of the *S. cerevisiae* proteome into one of the three classes: Highly structured, S (1971 sequences; 0 – 10% of the sequence is unstructured), moderately unstructured, M (2711 sequences; 10% – 30%) and highly unstructured, U (2020 sequences; 30% - 100%). Using the same criteria, we obtained three groups for *S. pombe* (S=1576 sequences, M=2005 sequences, and U=1409 sequences). We also obtained comparable sized groups in human: Highly structured, (7466 sequences; 0 – 14% is unstructured), moderately unstructured, (8371 sequences; 14% – 35%) and highly unstructured, (7145 sequences; 35% - 100%). Though using only two structural classes may be more straight-forward than three groups as presented in the main text, dividing the proteome into three groups and investigating the differences between the U and the S group will ensure that we are unambiguously dealing with highly unstructured and highly structured proteins. To ensure that the results we obtained are independent of the prediction method used, we also determined the degree of intrinsic disorder of *S. cerevisiae* proteins with FoldIndex (S17) and IUPRED (S18). PEST motif prediction was carried out using the *epestfind* algorithm of the EMBOSS package. A protein was defined as Q/N-rich if it contained at least 20 glutamines and/or asparagines in any continuous segment of 80 residues (S19).

All statistical analyses to estimate significance (Wilcoxon rank-sum, Kolmogorov-Smirnov and Fisher's exact test) were carried out using the R statistical analysis package. Wilcoxon rank-sum and Kolmogorov-Smirnov tests are non-parametric tests that assesses whether two samples of observations come from the same distribution or not. Importantly, both do not require assumptions about the form of the distribution of the measurements. As most datasets used in this analysis are not normally distributed, we determined the significance of the difference between the S and U groups by using Wilcoxon's test. Consistently, we find similar results by using the Kolmogorov-Smirnov test (**Table S5**). In order to evaluate the significance of the enrichment for certain properties which don't have distributions but only have percentages (TATA box, short poly(A)-tail, PEST sequence, etc) in the S and U groups, we used Fisher's exact test.

To determine the percentages of S, M and U group proteins among the substrates of all kinases in the phosphorylation network, the fraction of S, M, and U, targets of each kinase was initially calculated and then the averages taken over all the kinases. To test whether the 227 transcripts that Puf5p binds show any enrichment to encode for IUPs, the expected percentage of unstructured proteins among 227 transcripts was determined by calculating the average number of unstructured proteins in 1000 sets of 227 proteins each, which were randomly selected among all *S. cerevisiae* proteins. A z-score was then determined as the ratio of the difference between the observed and expected average values from 1000 random trials to the standard deviation. This z-score was then transformed into a *p*-value using *pnorm* in R.

Supplementary Online Material Text

Text S1: Extended introduction

Traditionally, the understanding of molecular mechanisms in biological processes has involved the determination of three-dimensional structure of the proteins involved. Structure determination methods such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have proven to be very successful in the investigation of the structure-function relationship of globular proteins and protein complexes (S20). Recent studies, enabled by advances in biochemical methodologies, indicate that many proteins in eukaryotic cells are intrinsically unstructured (will be hereafter referred to as IUPs), but are still functional (S21-23). Such studies, aided by computational analyses, have shown that over half of the eukaryotic proteins contain intrinsically unstructured regions in over 30% of their residues (S16, S24).

Properties, advantages and functions of unstructured proteins

IUPs lack a unique 3D structure, either entirely or in parts, when alone in solution (S21, S23). This lack of structure in a protein is believed to provide several advantages. For instance (i) IUPs provide a larger interaction surface area than globular proteins of a similar size, (ii) the conformational flexibility allows IUPs to recognize and interact with several different targets and (iii) regulation is facilitated via diverse post-translational modifications such as phosphorylation and several short linear peptide motifs, commonly referred to as short linear interaction motifs or eukaryotic linear motifs (S25-37).

Moreover, as IUPs frequently fold upon binding, they can interact with their targets with high specificity and relatively low affinity, which allows them to specifically associate to initiate a process and at the same time dissociate easily when the task is accomplished (S21, S26, S38-40). Though the mechanism of coupled-folding upon binding has now been known for a while (S21), the detailed molecular mechanism and the thermodynamic principles of this process is only recently beginning to be understood (S41). Interaction-prone segments of disordered proteins that undergo disorder-to-order transitions upon specific binding have been referred to as Molecular Recognition Features (MoRFs). The MoRFs from the Protein Data Bank, which were indicated to be intrinsically disordered in the absence of their binding partners by several criteria, have been analyzed and classified into three subtypes according to their structures in the bound state: alpha-MoRFs that form alpha-helices, beta-MoRFs that form beta-strands, and iota-MoRFs that form structures without a regular pattern of backbone hydrogen bonds (S38, S42, S43).

The properties listed above are ideal for proteins which mediate specific recognition of interaction partners and precise co-ordination of regulatory events both in space and time. Accordingly it has been reported that proteins, which participate in binding, regulatory and signalling functions such as transcription factors and signalling proteins, are enriched for intrinsically unstructured regions (S16, S33, S44-49)(Table S12).

The need for regulating unstructured proteins

Due to their unusual structural features and important functional properties, the presence of IUPs in a cell may need to be carefully monitored. In fact, it has been demonstrated that mutations or altering the expression levels of IUPs are associated with several disease conditions such as cancer and neuro-degeneration. For instance, over-expression of several IUPs such as TC-1 (S50), Stathmin (S51) or under-expression of Arf (S52) and p27 (S53) has been linked with various types of cancer (S49). Similarly, over-expression and mutations in α -synuclein, ataxin-1, tau and huntingtin are known to increase the risk of misfolding and fibrillar aggregate formation and have been linked to serious pathological conditions in humans such as Parkinson's disease, Spinocerebellar ataxia type 1, Alzheimer's disease and Huntington's disease (S54-60). Importantly, it has been recently shown that the loss of miRNA based regulation of levels of Atrophin, an IUP, leads to neuro-degeneration in *Drosophila melanogaster* (S61), clearly demonstrating that maintaining the right endogenous levels of IUPs and the integrity of the associated regulatory mechanisms is crucial for the survival of an organism.

Given their critical role for the functioning of eukaryotic cells, it becomes important to understand how cells have evolved to precisely regulate the availability of IUPs. In this study, we ask whether there is evidence for specific regulatory mechanisms that affect the availability of IUPs within a cell, *i.e.*, their abundance and residence time (amount of time spent in the cell). Such regulatory mechanisms would ensure that IUPs are not present longer than required, thereby minimizing the possibility of undesirable ectopic interactions between components of different pathways and the possibility of aggregation. The abundance and the time spent by IUPs in the cell, both of which depend on the rate of production and degradation, can be regulated at several stages during transcription and translation (*see Fig S1*). This framework, coupled with the recent availability of genome-scale data from several eukaryotes on each of these stages (*see Table S1*), motivated us to carry out a systematic investigation to find out which of the regulatory steps, or a subset of them, contribute significantly to control the availability of IUPs.

Text S2: The reported differences in protein abundance are not affected by the tagging procedure used to measure protein levels

The protein abundance values were obtained from published data described by the Weismann laboratory. The two datasets include protein abundance measurement by Tap-tagging and GFP-tagging of genes. As seen from **Fig S2**, the tagging procedure does not affect measurement of the relative abundance levels and still provides consistent results (Abundance measurement by TAP-tagging method: average value for structured group μ_S : 100% (4282 a.u.), for the unstructured group μ_U : 64% (2741 a.u.), $p=10^{-16}$; Fisher's test; Abundance measurement by GFP-tagging method: μ_S : 100% (6723 a.u.), μ_U : 60% (4073 a.u.), $p<2\times 10^{-16}$; Fisher's exact test). The differences between the abundances of the structured and unstructured proteins are of equal magnitude in both datasets. Moreover, the tagging procedure did not influence the ability to measure levels of proteins from the different structural classes. This is supported by the fact that we had comparable number of data points for all the three structural classes (S, M, and U; **Fig S2**) in both datasets. Therefore, the findings are unlikely to be biased through the effect of tagging.

Text S3: Analysis of the N-end rule pathway

Several studies have revealed molecular determinants within proteins that target them for degradation. One such signal, which has been shown to be linked to the *in vivo* protein stability, is the N-degradon, which relates the identity of the N-terminal residue of any protein to its *in vivo* half-life (S62, S63). According to this N-end rule, which is a ubiquitin dependent pathway, the protein is stable if the exposed N-terminal residue is a small amino-acid and unstable if it is large and bulky. Though the N-terminal amino acid in most cellular systems is primarily a methionine residue (a stabilizing residue according to the N-end rule), the action of the methionine amino-peptidase via the N-terminal methionine amino-peptidase pathway or the activity of proteases such as signal peptidases and caspases, removes the first methionine of most proteins or cleaves an internal recognition site and exposes the amino-acid next to the methionine or any internal residue next to a cleavage site, respectively (S63, S64). Therefore one could adopt a simplistic view in assuming that the exposed N-terminal residue is largely the amino-acid next to the first methionine or any other residue next to a protease cleavage site.

Analysis of the distribution of the N-terminal residues (**Fig S3, S4**) next to the initiator methionine of all the proteins in the group of structured and unstructured proteins showed that Glu, a residue which is associated with short protein half-life in yeast, is enriched in the U group and Cys, a residue linked to long protein half-life, is enriched in the S group (**Fig S3, S4**). All the other residues occurred in proportions whose differences were not statistically significant enough to make a meaningful interpretation. One reason why we may not see a strong signal for the N-end rule pathway may be because the exposed residue can undergo further modifications such as N-terminal acetylation, myristoylation, deamidation of Gln and Asn to Glu and Asp, ligation with an Arg by the Arginine-tRNA transferase, cleavage by proteases and so on, thereby making it relatively difficult to predict the stability of the protein purely based on the predicted proteome sequence (S63, S64).

Text S4: Kinase-substrate phosphorylation network analysis

Computational analysis using phosphorylation site prediction methods have suggested that intrinsically unstructured regions are enriched for sites that can be post-translationally modified (S34). To test whether IUPs generally tend to be phosphorylated more often than structured proteins, we analyzed the experimentally determined yeast kinase-phosphorylation network of Ptacek *et al* (S15). In this experiment, the targets of 87 of the 122 *S. cerevisiae* kinases were identified by incubating two yeast proteome microarrays in the presence of labelled ATP with and without each kinase individually. This allowed identification of auto-phosphorylation events and the targets of each of the 87 kinases. This information was further processed by the authors to identify a set of high-confidence targets of each kinase by integrating their primary data with sub-cellular location data. This dataset consists of over 4000 phosphorylation events involving 87 kinases and 1325 protein targets.

We investigated the average number of kinases that target the highly structured group of proteins, the moderately unstructured group and the highly unstructured group of proteins (**Fig S5**). Our analysis shows that on average the unstructured group of proteins are targeted by twice as many kinases as the structured group of proteins (μ_U :2.2 kinases; μ_U :4.0; expected average is 3 kinases per substrate) and that the distributions of the number of kinases targetting structured and unstructured proteins are significantly different ($p=1\times 10^{-12}$; Wilcoxon test). This suggests that unstructured proteins are more often phosphorylated by several different kinases than structured proteins, which may potentially result in fine-tuning their function and availability. An analysis of the targets of each of the kinases revealed another important trend: for several kinases in the dataset, their targets were highly enriched in intrinsically unstructured proteins. To investigate this further, we calculated the average fraction of highly structured, moderately unstructured and highly unstructured proteins among the targets of each kinase (**Fig S5b**). Our analysis (see **Materials and Methods**) reveals that the targets of several kinases show a significant enrichment in highly unstructured proteins (μ_S :19% \pm 13% (SD) of all targets; μ_U :51 % \pm 19% (SD); $p < 10^{-16}$ comparing μ_U and μ_S ; please note that the remaining 30% \pm 14% (SD) are moderately unstructured proteins). The expected value from the genome-wide distribution is \sim 30% for each group (**Fig S5b**, dashed line). It is thus clear that unstructured proteins are enriched to be substrates for the kinases. A careful investigation of the kinases that show a significant enrichment for their targets to be unstructured revealed that nearly all these kinases are regulated in a cell-cycle dependent manner (*e.g.*, Cdc28; **Table S2**), or activated upon exposure to particular stimuli (*e.g.*, Fus3; **Table S2**) or stress (*e.g.*, Atg1; **Table S2**). An analysis of the set of IUPs that are highly targeted by several different kinases revealed that there was no single trend for abundance or half-life, which may be because phosphorylation can either stabilize or destabilize a protein. An analysis of the set of the set of IUPs targeted by these kinases revealed the absence of a trend for abundance or half-life. This maybe because phosphorylation can either stabilize or de-stabilize a protein (see below).

These results along with our observations that most IUPs tend to be tightly regulated during transcription and translation suggests that apart from basal transcriptional and translational regulation, post-translational modification of IUPs via phosphorylation and subsequent modulation of the regulatory steps could be an important mechanism in fine-tuning their availability under different conditions.

Text S5: Extended discussion and conclusion

In order to decipher the complexity of cellular systems, it is essential to elucidate the interplay between the different cellular processes in an integrative manner. Many previous studies have successfully addressed isolated aspects of cellular regulation at a genome-wide level (S2-4, S6, S7, S9, S11, S65, S66). In this study, we take a systems approach and integrate structural data with information on different levels of cellular regulation to understand how the availability of IUPs is orchestrated in a living cell. Though we initially set out to identify the key regulatory steps that control availability of IUPs, we surprisingly found that living systems do not use a single, or a subset of regulatory mechanisms. Instead, the availability of a major fraction of IUPs is tightly regulated at nearly every stage during transcription and translation. Such a combination of different regulatory mechanisms provides precision, speed, and flexibility in biological control beyond what is possible with any single subset of these regulatory mechanisms (S67).

At the level of messenger RNA, we observe that most transcripts which code for highly unstructured proteins are degraded rapidly in comparison to transcripts that encode highly structured proteins. Our observation that transcripts encoding IUPs tend to have a short poly-A tail and that RNA binding proteins, such as Puf5p, show an enrichment to bind to these mRNA that encode IUPs suggests that regulation of poly-A tail length and targeting by specific RNA binding proteins could be key mechanisms to regulate the availability of several IUPs in a cell. Indeed, the importance of regulation of mRNA stability via RNA binding proteins is becoming increasingly recognized in a wide range of organisms and cellular conditions (S68-73).

Although the availability (in terms of higher abundance and longer half-life) of IUPs may increase due to stabilization or protection from the degradation machinery upon post-translational modification or interaction with other stabilizing proteins via short linear motifs (S23, S25, S53, S74-78), our analysis reveals that most IUPs are on average short-lived and less abundant than structured proteins. The availability of a significant fraction of the IUPs in the cell appears to be tightly regulated and this is partially due to accelerated proteolytic degradation (S79) (possibly via PEST sequences), and a reduced translational rate of their transcripts.

It should be noted that Tompa *et al* (S79) arrive at a conclusion that intrinsic disorder serves as a weak signal for protein degradation. The crucial difference between our study and that of Tompa *et al* is that they test for a linear relationship between the number of residues that are unstructured, or the number of segments which are unstructured, and protein half-life. We do not test for such a relationship but simply investigate whether the group of highly unstructured proteins has a shorter half-life when compared to the group of highly structured proteins. Tompa *et al* use the slope of the linear fit, and the r^2 values as a measure of relationship and statistical association, whereas we calculate differences between the average values in the different structural groups and measure p -values to assess the statistical significance of the difference. Grouping of data points into structural classes and assessing the difference between the distributions is a powerful way to identify the existence of a relationship without assuming any underlying model of correlation. Even if one assumes a linear relationship, it should be noted that calculating the best linear fit for a large number of experimentally determined individual data points is unlikely to yield high r^2 values. Therefore, the p -value and the sign of the r value are more informative. Indeed, upon performing such a calculation using IUPred or Disopred as predictors, we find very high p -values and negative r values suggesting an inverse relationship between protein disorder and half-life ($p=2 \times 10^{-10}$ and $r = -0.11$ for IUPred and $p=2 \times 10^{-11}$ and $r = -0.12$ for Disopred), which is consistent with our observations.

Our observation that IUPs tend to display less stochasticity both at the level of gene transcription and protein translation, suggests that the presence of most IUPs in a cell is finely tuned. This suggestion is supported by our finding that IUPs are highly targeted by several kinases whose levels are themselves modulated in response to specific stimuli, stress or cell-cycle (SOM text S4 and Table S2).

Tight control of the availability of intrinsically unstructured proteins increases fidelity in signalling, regulation and recognition

Our finding that most IUPs have short protein half-life and are present in low abundance within the cell could simply be explained by the possibility that proteins which contain unstructured regions may be more prone to degradation by the proteasome and proteases (S80, S81). However, our observation that transcripts that encode unstructured proteins are likely to be more tightly controlled and that this phenomenon is evolutionarily conserved suggests other possibilities. It could well mean that in addition to the biophysical property of unstructured regions that make them susceptible for degradation (S80, S81), there are selective forces that operate on such proteins which ensure that there is a tight basal level control at almost every level of transcription and translation to precisely orchestrate their availability within a cell. In the following section, we provide an explanation for this tight control and discuss it in detail (Fig S6).

Proteins with unstructured regions predominantly have signaling and regulatory roles (S16, S26, S33, S44-49). The presence of unstructured regions has several advantages which make them appropriate for these functions. These include: (i) coupled

binding and folding upon interaction with the right partner thereby providing specificity (S26, S41), (ii) presence of several linear motifs that facilitates readout by specific domains that recognize particular post-translational modifications on the motif (S82-84) and (iii) multi-site protein modification on the same protein which permits integration of signalling pathways with distinct outcomes (S84, S85). Thus, altering the levels or mis-regulation of such unstructured proteins could result in unwanted ectopic interactions thereby disturbing the fine balance of the signalling networks (S84). This problem becomes even more important given that several IUPs, which are signalling proteins, are significantly re-used in multiple pathways that result in several different outcomes *e.g.*, proteins involved in epigenetic regulation and cellular differentiation such as Oct4, Nanog, Sox2 (S86), components of the MAP kinase pathway (S87), and the NFkB/P53/CBP-P300 pathway (S88).

Two major solutions to how cells can reuse the same components in multiple pathways and yet avoid erroneous cross-talk or undesirable outcomes have been proposed in the literature (**Fig S6**). These are (i) spatial and temporal segregation of signalling proteins within a cell and (ii) increased signalling complexity, both of which could provide fidelity in regulation (see legend of **Fig S6** for details). The results from our analysis suggest a third major solution to this problem (**Fig S6**): The tight regulation of signalling and regulatory IUPs minimizes the potentially harmful effects of ectopic interactions and provides robustness to the signalling process by ensuring that such proteins are present in appropriate amounts and for the relevant amount of time. It should be stressed that these three major strategies to prevent erroneous cross-talk are not mutually exclusive and are likely to be employed in combination to achieve specificity. Thus, regulating availability will directly affect the spatio-temporal sequestration and the connectedness and complexity of different signalling pathways.

Our finding that most IUPs, which are enriched for signalling and regulatory proteins, are short-lived and less prone to noise, suggests that propagation of noise in the cell could be reduced through this evolutionary design. Such low noise ensures that the proteins are present in precise amounts in a population of cells and this would have a direct consequence on the kinetics of the reaction involving activation, inhibition or recognition by a specific read-out protein. In this context, it is interesting to note the results of three independent studies which demonstrate the importance of kinetics of post-translational modification and availability of signalling proteins on distinct phenotypic outcomes. In one study, Kumar *et. al.* (S89) performed a systematic set of perturbation experiments on the B-cell antigen-receptor dependent intracellular signalling network, which involves activation and inhibition of shared components in multiple pathways. Their study demonstrates that the kinetics of post-translational modification on specific sites on a protein results in distinct outcomes such as apoptosis and proliferation. The other study by O'Dea *et. al.* (S90) investigated how degradation rates of I κ B (which is an IUP and physically interacts and inhibits the NF- κ B transcription factor) can affect signalling. The authors found that an intrinsically high degradation rate of free I κ B was critical to maintain homeostasis between the different signalling states and make the pathway sensitive for a signalling input (S90). In yet another study, the authors investigated what determines a mammalian cell to selectively adopt an apoptotic or an adaptive response upon inflicting mild ER stress. They discovered that survival is favoured as a direct consequence of the intrinsic instability of mRNAs and proteins that promote apoptosis (*e.g.*, Chop, an IUP) compared to those that facilitate adaptation (*e.g.*, BiP, a structured protein) (S91). These studies provide evidence for the proposed mechanism that the tight control of the availability IUPs at mRNA and protein levels can allow cells to achieve specificity, sensitivity and fidelity in signalling.

Though such a tight control results in low basal availability of a large fraction of IUPs, certain IUPs need to be present at high levels or for long periods of time in response to specific stimuli, stress or cell-cycle. In fact, in some cases (*e.g.*, titin), high levels may be required throughout the life-time of a cell. Fine-tuning of IUP availability can be achieved by post-translational modifications and interactions with other factors (S23, S25, S53, S74). Both mechanisms, either alone or in combination, can promote stabilization (*i.e.*, increased abundance and longer half-life) via change in cellular localization or by protection from the degradation machinery (*e.g.*, certain phosphorylated forms of p27^{kip1} and ataxin-1) (S75, S76). Although association with other proteins may increase their stability, free IUPs are likely to be rapidly degraded by the 20S proteasome via degradation by default (S80) as has been shown *in vivo* and *in vitro* for the unbound forms of p21^{cip1} (S92), p27^{kip1} (S93), α -synuclein (S94) and tau (S95). In addition, certain post-translational modifications may promote regulated degradation (*e.g.*, p27^{kip1}) (S53, S74). In this context, our finding that a large fraction of IUPs tend to be phosphorylated by

several kinases and display low noise in transcription and translation suggests that their levels and the time spent are finely tuned in a cell.

Interestingly, we find that the difference in levels of control is even more pronounced between structured proteins and the sub-group of IUPs that contain poly-Q and poly-N stretches. A tighter control of this sub-group of IUPs might prevent protein aggregation by ensuring that they are not present in higher amounts or longer than required in the cell. However, further detailed investigation is needed to test the generality of this hypothesis.

Extended conclusion

In conclusion, our investigation reveals that a tight control of synthesis and clearance of a significant fraction of intrinsically unstructured proteins and their transcripts results in low noise and precise availability in cellular systems. Our results, which demonstrate the presence of such a broad level of control, are important not only for understanding how the residence time of IUPs is regulated but should also serve as a guide to identify additional molecular strategies that nature has designed to precisely control availability of proteins within a cell. Therefore, we believe that the findings and the framework presented here raise several new and important questions, particularly, how the availability of specific IUPs is controlled in different cell-types. The impact of additional control mechanisms such as miRNA based regulation, targeting elements with specific secondary structures in transcripts and nucleosome remodelling on the availability of IUPs should be fruitful lines of investigation. Elucidating the details of these mechanisms on specific proteins will be essential to gain a deeper understanding of the pathological conditions that arise when such regulatory mechanisms become defective. A recent example of this is the loss of a miRNA based regulation of Atrophin in a *Drosophila* neuro-degeneration model. Taken together, our results can be directly translated into possible intervention strategies to manipulate the cellular proteostasis network, which is the set of co-ordinated regulatory mechanisms that maintain protein homeostasis, *i.e.*, the appropriate concentration, localization and residence time of all proteins within a cell (S96, S97). We suggest that manipulating protein concentrations by affecting the tight regulation of key IUPs can be exploited to induce or accentuate proteostatic deficiencies that contribute to human diseases (S96). Most importantly, we anticipate that by manipulating this tight control (expression levels of uncharacterized IUPs or by modulating the regulatory steps which affect the availability of IUPs), one could discover novel pathological conditions that have never been described before.

The identification of molecular strategies that regulate availability of transcripts or proteins, such as those presented here, could have direct implications in synthetic biology experiments. These experiments aim to rationally rewire signal transduction pathways to dictate specific cellular outcomes and engineer autonomous gene regulatory circuits for logical control of gene expression (S98). Finally, the fact that the distributed regulation of IUPs occurs in several divergent organisms implies that it is an important general mechanism. We suggest that this mechanism could provide robustness to cellular systems by minimizing the potentially harmful effects of IUPs and at the same time permits their vital contribution to the functioning of a cell.

Supplementary Figures S1 to S6

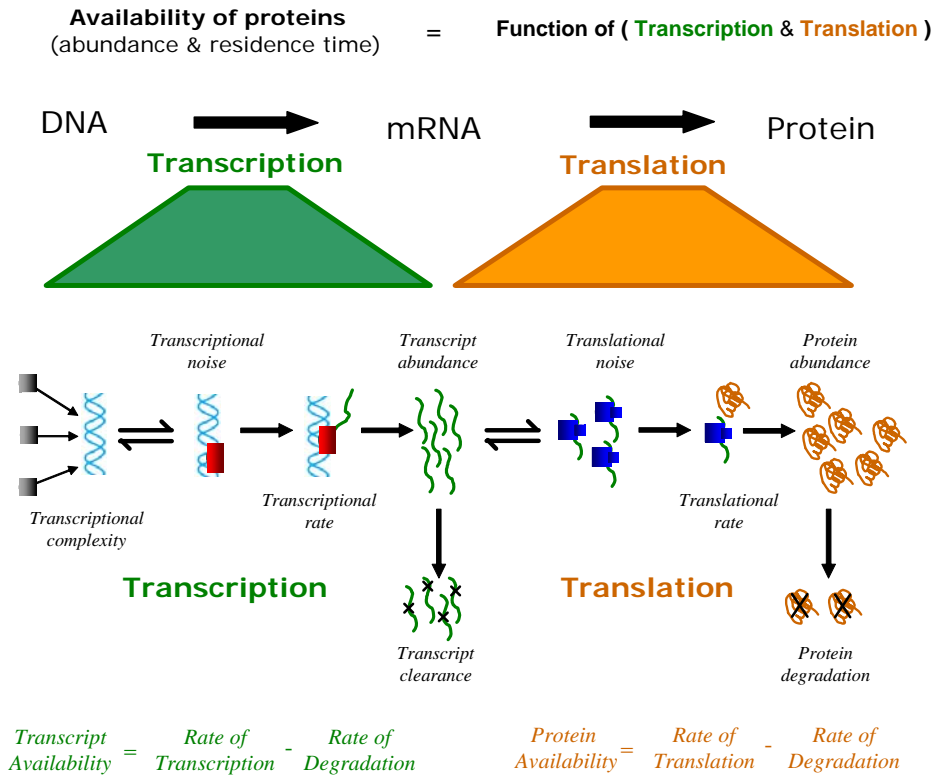


Figure S1: Framework used in this study to investigate the regulation of unstructured proteins.

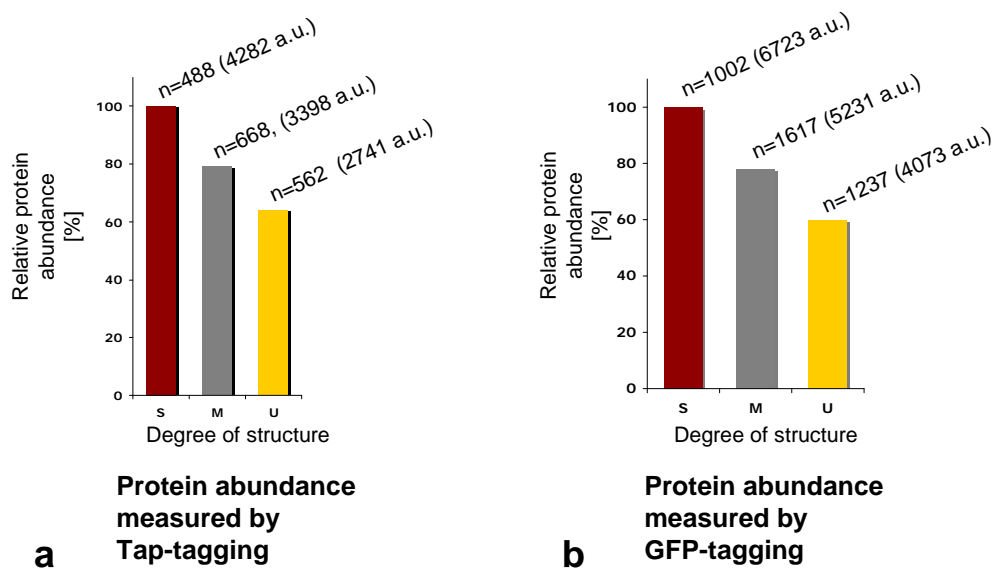


Figure S2: Relative protein abundance of the three groups of proteins normalized according to the values of the structured group for (a) TAP-tagged abundance values and (b) GFP-tagged abundance values. ‘n’ is the number of proteins in each structural group and the values in parantheses represent the average abundance (a.u. - intensity values measured in arbitrary units).

Half-life	Amino acid	All residues	Structured	Unstructured	Moderately structured	Half-life
l	A	532	126	183	223	>20 hour
l	C	60	32	12	16	>20 hour
s	D	319	80	125	114	3 min
s	E	353	66	169	118	30 min
s	F	240	85	52	103	3 min
l	G	308	79	115	114	>20 hour
s	H	95	33	28	34	10 min
s	I	243	95	53	95	30 min
s	K	394	127	84	183	3 min
s	L	436	161	90	185	3 min
l	M	121	20	35	66	>20 hour
s	N	339	88	113	138	3 min
l	P	291	91	85	115	>20 hour
s	Q	163	35	53	75	10 min
s	R	211	72	52	87	2 min
l	S	1505	392	503	610	>20 hour
l	T	519	178	138	203	>20 hour
l	V	415	154	89	172	>20 hour
s	W	45	18	9	18	3 min
s	Y	113	39	32	42	10 min
		6702	1971	2020	2711	

Figure S3: Distribution of residues in the N-terminal regions of the three groups of proteins and their corresponding theoretical half-life values as obtained from Brachmair *et al.*, Science 1987 (S62).

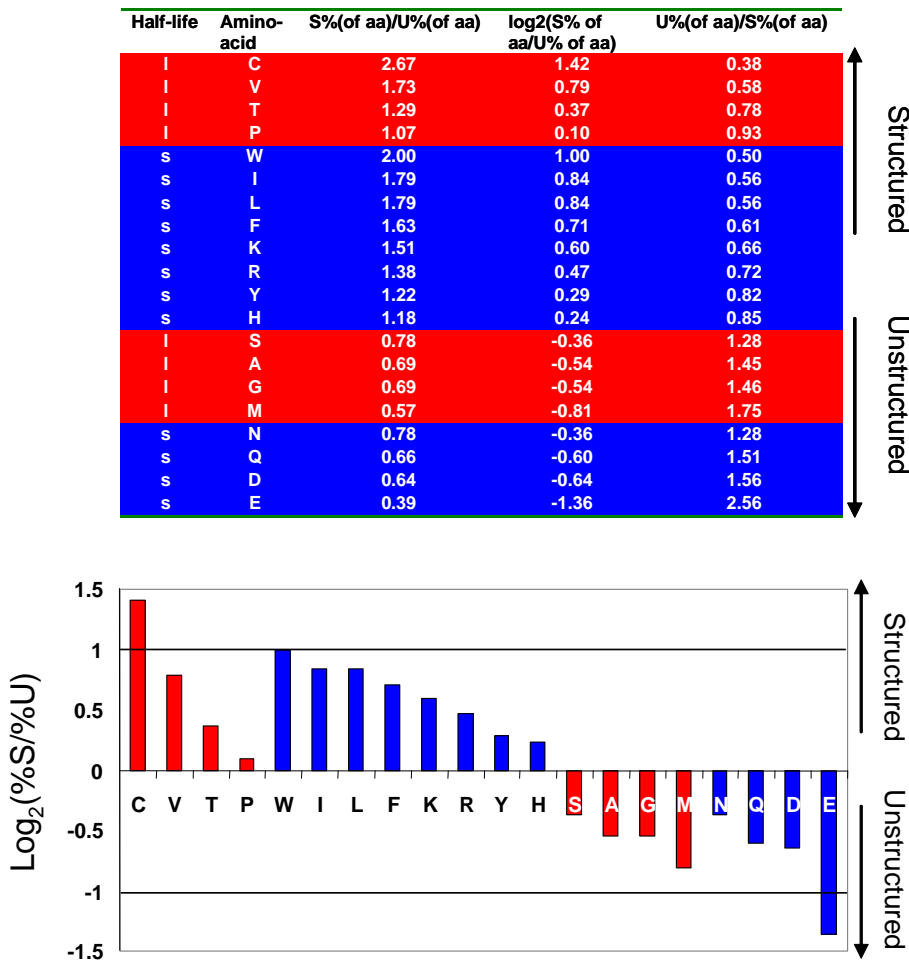


Figure S4: Enrichment of different amino-acids in the two groups of proteins to occur in the N-terminus. L (red) or S (blue) in the first column represents long and short *in vivo* protein half-life. Glu, a residue associated with short protein half-life in yeast is significantly enriched in the U group and Cys, linked to long protein half-life, is significantly enriched in the S group. The residues whose propensities are between 1 and -1 mean that their occurrence is not significantly different between the two groups. One reason for the absence of a strong signal for the N-end rule may be because the exposed residue can undergo further modifications such as acetylation, myristoylation, deamidation of Gln and Asn to Glu and Asp, ligation with an Arg by the Arginine-tRNA transferase, cleavage by proteases, etc thereby making it difficult to predict the stability of the protein based on sequence (S63, S64).

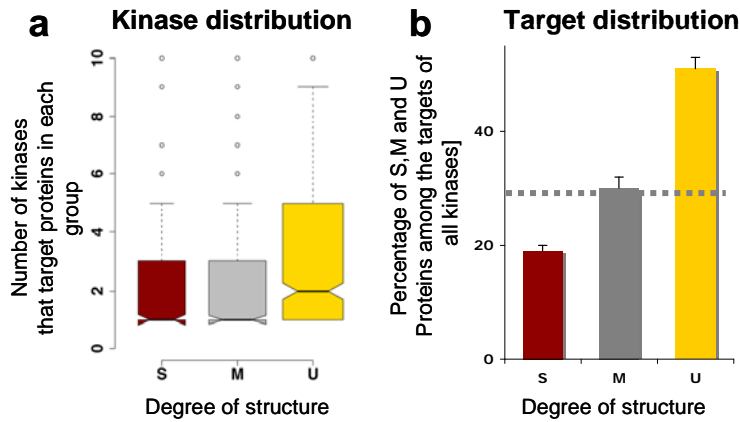


Figure S5: Comparison of (a) box-plot of the distribution of the number of kinases that target proteins in each group and (b) the percentage of S, M and U proteins among the substrates of all kinases. The expected average is shown as a broken line.

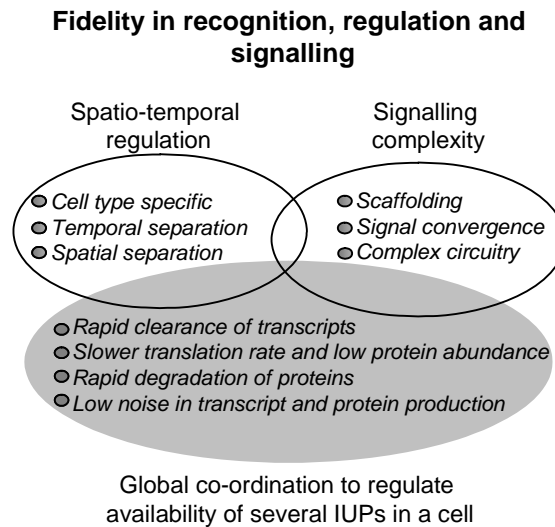


Figure S6: General mechanisms to prevent erroneous cross-talk in signalling and regulation are spatio-temporal segregation and increased signalling complexity. Specific mechanisms of spatio-temporal segregation include (i) cell-type specific expression of the relevant components, (ii) temporal expression of specific components, (iii) sub-cellular compartmentalization and (iv) spatial gradients of signalling activity (S99, S100). Increased signalling complexity can be achieved by (i) scaffolding proteins that tether different signalling proteins and thereby isolate one pathway from another (S100) (ii) convergence on a relatively small number of interaction mechanisms to provide context-dependent signalling (S101) (iii) an intricate connectedness of a large number of interacting constituents that involves complex circuitry such as feed-back and bi-stable switches (S102, S103) and (iv) multi-site protein modification that dictate specific outcomes (S82, S84, S85). In addition to spatio-temporal segregation and increased signalling complexity, our results indicate that global co-ordination of various regulatory steps ensures that IUPs are available for the right amount of time and therefore minimizes erroneous cross-talk or ectopic interactions.

Supplementary Tables S1 to S12

Table S1: Compendium of datasets used in our study

Type of information [source]	Description of the method used to obtain the data
Transcriptional noise (S11) (S10)	The presence or absence of TATA box was used as a proxy for transcriptional noise. The authors identified the genes in yeast and humans that have a TATA box by scanning promoter regions using a position weight matrix.
Transcriptional complexity (S1)	The transcriptional regulatory network for yeast was constructed by compiling high-confidence protein-DNA interactions from several published ChIP-chip experiments that involved 156 transcription factors and over 5000 genes across several conditions.
Transcriptional rate (S2)	Transcriptional rates for yeast grown in YPD were calculated by the authors based on the transcript abundances and mRNA half-lives. These were in turn determined by obtaining and comparing transcript levels of the wild-type and the temperature sensitive RNA polymerase <i>rpb1-1</i> mutant strains using an Affymetrix microarray.
Transcript abundance (S2) (S12)	Transcript abundances for yeast grown in YPD (<i>S. cerevisiae</i>) and Edinburgh minimal medium (<i>S. pombe</i>) were determined by using an Affymetrix high density oligonucleotide array.
Transcript half-life (S3) (S12) (S14)	Transcript half-lives were determined by obtaining transcript levels over several minutes after inhibiting transcription. This was done using the temperature sensitive RNA polymerase <i>rpb1-1</i> mutant <i>S. cerevisiae</i> strain, by adding 1,10-phenanthroline to <i>S. pombe</i> and by adding Actinomycin D to human HepG2 cells.
Transcript degradation (S5) (S12) (S4)	The authors used binding to a poly-U sepharose column at low temperature followed by stepwise thermal elution to fractionate cellular mRNA by poly-A tail length. mRNA that gets eluted at low temperatures have short poly-A tail, whereas those that elute at high temperatures have a long poly-A tail. Microarray hybridization of the mRNA in each fraction revealed the poly-A status (long or short) of the transcripts. Gerber <i>et al</i> used formaldehyde assisted cross-linking, followed by microarray hybridization, of the five Puf proteins in yeast to identify the transcripts that were bound to these proteins.
Translational noise (S9)	Measurements of noise in protein levels in a population of cells were obtained in a single-cell proteomics study that coupled high-throughput flow cytometry and a library of GFP tagged yeast strains. Noise was calculated as the ratio of the standard deviation to its mean abundance. For every protein, DM represents the difference between the noise value of that protein and the median over all proteins. Lower values represent less stochasticity.
Translational rate (S8)	To obtain a profile of ribosome association for the yeast transcriptome, which is an indicator for translational rate, the authors fractionated polysomes using velocity sedimentation. Following this, a quantitative microarray analysis of several fractions across the gradient was used to estimate the translational status of each mRNA.
Protein abundance (S6) (S9) (S13)	Estimates of the endogenous protein expression levels during log-phase were obtained by tagging every yeast protein with TAP-tag or GFP and measuring the intensity for <i>S. cerevisiae</i> or with a His-tag followed by immunoblotting using anti-His antibody for <i>S. pombe</i> .
Protein half-life (S7)	Protein half-lives were determined by first inhibiting protein synthesis via the addition of cyclohexamide and by monitoring the abundance of each TAP-tagged protein in the yeast genome as a function of time.
Protein degradation (S104) (S62)	The data on N-end rule pathway was obtained from published literature and the occurrence of PEST motifs in the yeast proteome was determined by using the <i>pepfind</i> algorithm of the EMBOSS package.
Kinase-target network (S15)	The targets of 87 kinases, representing most of the 122 <i>S. cerevisiae</i> kinases, were identified by incubating each kinase individually with two yeast proteome microarrays in the presence of labelled ATP.

Table S2: List of kinases with significant fraction of the targets enriched for IUPs

Kinase	Function	N_T	N_U	P-val
60-100% of the targets are IUPs, i.e., from the U group				
Rck2	Protein kinase involved in the response to oxidative and osmotic stress	25	20	3x10 ⁻⁶
Rim11	Protein kinase required for signal transduction during entry into meiosis	17	13	6x10 ⁻⁴
Atg1	Regulator of autophagy genes upon starvation for nutrients such as carbon; nitrogen; sulfur; and various amino acids or upon endoplasmic reticulum stress	187	138	<10 ⁻¹⁶
Snf1	Important for phosphorylating genes involved in the regulation of transcription; translation and general stress response	80	58	1x10 ⁻¹⁵
Mek1	Meiotic kinase	31	22	1x10 ⁻⁵
Cdc28C1b5	Cell cycle dependent kinase	43	30	3x10 ⁻⁷
Rad53	DNA damage response kinase	33	23	1x10 ⁻⁵
Ire1	ER stress kinase. Inactive Ire1 is bound by Kar2p and released upon ER stress	78	52	4x10 ⁻¹¹
Kin2	Kinase involved in exocytosis	28	18	7x10 ⁻⁴
Swel	Cell cycle dependent kinase (G2/M transition)	119	75	4x10 ⁻¹⁴
Sky1	SR protein kinase (SRPK) involved in regulating proteins involved in mRNA metabolism and cation homeostasis	40	25	7x10 ⁻⁵
Yck2	Kinase involved in response to glucose stimulus	221	138	<10 ⁻¹⁶
50-60% of the targets are IUPs, i.e., from the U group				
Tos3	Kinase involved in response to nutrient stress	81	48	6x10 ⁻⁸
Ime2	Kinase involved in activation of Meiosis. It associates with Ime1p and mediates its stability	61	36	5x10 ⁻⁶
Tpk1	Kinase regulating transcription & growth upon response to nutrients and stress	256	150	<10 ⁻¹⁶
Fus3	Map Kinase pathway - involved in regulating cell proliferation upon peptide pheromone stimulation	76	44	7x10 ⁻⁷
Fun31	Regulator of protein synthesis upon nutrient stimuli or stress	19	11	0.03
Akl1	Endocytosis and cytoskeletal organization	42	24	6x10 ⁻⁵
Hal5	Response to metal ion stress and salt	27	15	0.01
YGL059W	Mitochondrial kinase involved in regulating pyruvate dehydrogenase	92	51	4x10 ⁻⁷
Tpk2	Kinase regulating transcription & growth upon response to nutrients and stress	29	16	0.01
Ste20	Map Kinase pathway - involved in regulating cell proliferation upon peptide pheromone stimulation	70	38	4x10 ⁻⁵
Bck1	Kinase essential for growth and the integrity of proliferating cells. Deletion renders loss of osmotic stability	71	37	1x10 ⁻⁴
Dbf2	Cell cycle dependent kinase important for exit from mitosis and stress response	62	32	1x10 ⁻⁵
Prk1	Regulation of actin cytoskeleton	41	21	8x10 ⁻³
YPL141C	Uncharacterized kinase	63	32	8x10 ⁻⁴
IPL1	Kinase important for mitotic activity	34	17	0.03
Kin82	Uncharacterized kinase	28	14	0.05

N_T: Total number of targets phosphorylated by a specific kinase.

N_U: Number of highly unstructured proteins (U group) phosphorylated by a specific kinase

P-val: Statistical significance of the enrichment for IUPs among the targets of a kinase. The significance of the enrichment of IUPs among the targets of each kinase was calculated using the Fisher's exact t-test and taking the distribution of S, M and U in the entire proteome as reference.

Table S3: p-values for the various cellular parameters in *S. cerevisiae*

Cellular quantity / Regulatory mechanism	x_S		x_M		x_U		x_{UQN}	
	\bar{x} or \tilde{x}	n	\bar{x} or \tilde{x}	n	\bar{x} or \tilde{x}	n	\bar{x} or \tilde{x}	n
Transcriptional noise [% with TATA box] [*]	26	1517	18 (2×10^{-7})	2308	16 (8×10^{-7})	1803	11 (1×10^{-4})	74
Transcriptional complexity (TFs/gene) [†]	2.0±0.1	1236	2.0±0.1 (0.5)	1758	2.0±0.1 (0.55)	1262	2.0±0.1 (0.11)	52
Transcriptional rate [mRNAs/hr] [†]	2.4±0.1	1399	2.0±0.1 (0.04)	2037	1.7±0.1 (6×10^{-8})	1543	1.9±0.3 (0.13)	66
Transcript abundance [copies/cell] [†]	1.0±0.09	1500	0.8±0.05 (0.01)	2189	0.7±0.04 (1×10^{-6})	1237	0.9±0.1 (0.4)	74
Transcript half-life [min] [†]	23±0.8	1367	21±0.6 (1×10^{-7})	1931	19±0.6 ($<10^{-16}$)	1361	18±2 (7×10^{-4})	56
Transcript degradation [% with short poly(A) tail] [*]	24	357	37 (6×10^{-5})	508	66 ($<10^{-16}$)	430	90 (1×10^{-8})	17
Translational noise [DM] [†]	0.43±0.22	602	0.05±0.15 (1×10^{-5})	859	-0.28±0.14 (3×10^{-11})	683	-0.52±0.2 (1×10^{-3})	44
Translational rate [ribosomes/ORF] [†]	0.52±0.01	1635	0.43±0.01 (1×10^{-13})	2314	0.35±0.01 ($<10^{-16}$)	1728	0.26±0.06 (1×10^{-9})	70
Protein abundance [proteins/cell] [†]	3220±378	1002	2440±209 (1×10^{-9})	1617	1590±135 ($<10^{-16}$)	1237	1370 (2×10^{-6})	66
Protein half-life [min] [†]	49±3.1	793	40±2.0 (3×10^{-8})	1320	35±2.1 (1×10^{-15})	1086	25±6.1 (1×10^{-7})	58
Protein degradation [% with PEST] [*]	9	1971	31 ($<10^{-16}$)	2711	49 ($<10^{-16}$)	2020	50 ($<10^{-16}$)	74

x_S : mean (\bar{x}) or median (\tilde{x}) of the group of highly structured proteins, x_M : mean or median of the group of moderately unstructured proteins, x_U : mean or median of the group of highly unstructured proteins and x_{UQN} : mean or median of the group of highly unstructured proteins containing poly (Q/N) segments. ‘n’ is the number of datapoints. * Mean values are reported and the statistical significance was calculated with Fisher’s exact test. † Median values and their confidence intervals ($=1.58(IQR)/\sqrt{n}$, where IQR is the interquartile range and n the group sample size) are reported. Statistical significance was calculated with the Wilcoxon test.

Table S4: p-values for the various cellular parameters in *S. pombe* and Humans

Cellular quantity	x_S		x_U		x_{UQN}	
	\bar{x} or \tilde{x}	n	\bar{x} or \tilde{x}	n	\bar{x} or \tilde{x}	n
<i>S. pombe</i>						
Transcript abundance [signal intensity] [†]	2119±150	1504	1455±73 ($<10^{-16}$)	1388	1496 (2×10^{-4})	95
Transcript half-life [% long] [*]	66	565	39 ($<10^{-16}$)	541	23 (5×10^{-8})	43
Transcript degradation [A.U.] [‡]	0.032±0.001	887	0.035±0.002 ($<10^{-16}$)	747	0.037±0.002 (9×10^{-6})	42
Protein abundance [signal intensity] [†]	0.3±0.04	1362	0.1±0.01 ($<10^{-16}$)	1192	0.08±0.03 (5×10^{-11})	88
Protein degradation [% with PEST] [*]	8	1578	48 ($<10^{-16}$)	1410	58 ($<10^{-16}$)	97
<i>H. sapiens</i>						
Transcriptional noise [% TATA box] [*]	15	1168	10 (2×10^{-6})	1546	Insuf. data	
Transcript decay rate [h^{-1}] [†]	0.08	426	0.19 ($<10^{-16}$)	609	0.26 ($<10^{-16}$)	92
Protein degradation [% with PEST] [*]	10	5402	50 ($<10^{-16}$)	8713	67 ($<10^{-16}$)	435

x_S : mean (\bar{x}) or median (\tilde{x}) of the group of highly structured proteins, x_U : mean or median of the group of highly unstructured proteins and x_{UQN} : mean or median of the group of highly unstructured proteins containing poly (Q/N) segments. ‘n’ is the number of datapoints. ^{*} Mean values are reported and the statistical significance was calculated with Fisher’s exact test. [†] Median values and their confidence intervals ($=1.58(IQR/\sqrt{n})$, where IQR is the interquartile range and n the group sample size) are reported. Statistical significance was calculated with the Wilcoxon test. [‡] Mean values are reported and the statistical significance was calculated with *t*-test.

Table S5: p-values for the various cellular parameters in *S. cerevisiae* using the Kolmogorov-Smirnov test

Cellular quantity	<i>p</i> -value
Transcriptional complexity (TFs/gene)	0.1
Transcriptional rate [mRNAs/hr]	1×10^{-9}
Transcript abundance [copies/cell]	6×10^{-11}
Transcript half-life [min]	4×10^{-13}
Translational noise [DM]	2×10^{-10}
Translational rate [ribosomes/ORF]	$< 10^{-16}$
Protein abundance [proteins/cell]	$< 10^{-16}$
Protein half-life [min]	4×10^{-13}

Table S6: The results are independent of the IUP prediction method used**p-values using FoldIndex for predicting IUPs**

Cellular quantity	x_S	x_U	p -value
Transcriptional noise [% with TATA box] *	28	15	$<10^{-16}$
Transcriptional complexity (TFs/gene) †	2.0	2.0	0.06
Transcriptional rate [mRNAs/hr] †	2.3	1.8	7×10^{-3}
Transcript abundance [copies/cell] †	0.9	0.8	5×10^{-7}
Transcript half-life [min] †	23.0	19.0	$<10^{-16}$
Transcript degradation [% with short poly(A) tail] *	23	63	$<10^{-16}$
Translational noise [DM] †	0.35	-0.25	$<10^{-16}$
Translational rate [ribosomes/ORF] †	0.52	0.39	$<10^{-16}$
Protein abundance [proteins/cell] †	2990	1780	$<10^{-16}$
Protein half-life [min] †	45	36	1×10^{-8}
Protein degradation [% with PEST] *	13	42	$<10^{-16}$

Mean or median and p -values of different cellular quantities for proteins that are structured (S) and highly unstructured (U) in *S. cerevisiae*. The degree of structure was determined using the program FoldIndex. To get equally sized bins, S was defined as proteins having less than 20% of their residues disordered according to FoldIndex and U as proteins with more than 45% of the residues disordered. * Mean values are reported and the statistical significance was calculated with Fisher's exact test. † Median values are reported and the statistical significance was calculated with the Wilcoxon test.

Table S7: p-values using IUPred for predicting IUPs

Cellular quantity	x_S	x_U	p -value
Transcriptional noise [% with TATA box] *	23	15	1×10^{-9}
Transcriptional complexity (TFs/gene) †	2.0	2.0	0.1
Transcriptional rate [mRNAs/hr] †	1.9	1.8	0.6
Transcript abundance [copies/cell] †	0.8	0.8	0.7
Transcript half-life [min] †	24.0	19.0	$<10^{-16}$
Transcript degradation [% with short poly(A) tail] *	24	67	$<10^{-16}$
Translational noise [DM] †	0.37	-0.24	1×10^{-7}
Translational rate [ribosomes/ORF] †	0.57	0.34	$<10^{-16}$
Protein abundance [proteins/cell] †	2430	1780	4×10^{-9}
Protein half-life [min] †	47	35	2×10^{-11}
Protein degradation [% with PEST] *	20	54	$<10^{-16}$

Mean or median and p -values of different cellular quantities for proteins that are structured (S) and highly unstructured (U) in *S. cerevisiae*. The degree of structure was determined using the program IUPRED. To get equally sized bins, S was defined as proteins having less than 2% of their residues disordered according to IUPRED and U as proteins with more than 18% of the residues disordered. * Mean values are reported and the statistical significance was calculated with Fisher's exact test. † Median values are reported and the statistical significance was calculated with the Wilcoxon test.

Table S8: Division of the proteome into two groups (S and U) does not affect the results

p-values after grouping proteins into two classes only

Cellular quantity	x_S	x_U	p -value
Transcriptional noise [% with TATA box] *	23	16	3×10^{-12}
Transcriptional complexity (TFs/gene) †	2.0	2.0	0.6
Transcriptional rate [mRNAs/hr] †	2.2	1.8	0.003
Transcript abundance [copies/cell] †	0.9	0.8	0.001
Transcript half-life [min] †	23.0	19.0	$< 10^{-16}$
Transcript degradation [% with short poly(A) tail] *	28	56	$< 10^{-16}$
Translational noise [DM] †	0.22	-0.15	1×10^{-5}
Translational rate [ribosomes/ORF] †	0.50	0.37	$< 10^{-16}$
Protein abundance [proteins/cell] †	2900	1860	$< 10^{-16}$
Protein half-life [min] †	45	37	1×10^{-12}
Protein degradation [% with PEST] *	19	42	$< 10^{-16}$

Mean or median and p -values of different cellular parameters for proteins that are structured (S) and highly unstructured (U) in *S. cerevisiae*. The proteins were divided into two groups only: (S = 0-20% unstructured; U =20-100% unstructured). * Mean values are reported and the statistical significance was calculated with Fisher's exact test. † Median values are reported and the statistical significance was calculated with the Wilcoxon test.

Table S9: The results are independent of the number of interaction partners of the IUP

p-values of non-hub proteins only

Cellular quantity	x_S	x_U	p -value
Transcriptional noise [% with TATA box] *	26	10	4×10^{-4}
Transcriptional complexity (TFs/gene) †	2.0	2.0	0.2
Transcriptional rate [mRNAs/hr] †	2.1	1.5	2×10^{-6}
Transcript abundance [copies/cell] †	0.8	0.7	4×10^{-4}
Transcript half-life [min] †	23.0	20.0	1×10^{-7}
Transcript degradation [% with short poly(A) tail] *	16	68	$< 10^{-16}$
Translational noise [DM] †	0.27	0.02	1×10^{-3}
Translational rate [ribosomes/ORF] †	0.45	0.32	2×10^{-10}
Protein abundance [proteins/cell] †	2560	1330	$< 10^{-16}$
Protein half-life [min] †	45	31	1×10^{-8}
Protein degradation [% with PEST] *	10	49	$< 10^{-16}$

Mean or median and p -values of different cellular quantities for proteins that are structured (S) and highly unstructured (U) in *S. cerevisiae*. Only non-hub proteins - proteins that have less than 10 interaction partners in the Yeast BioGRID protein-protein interaction dataset (version 31) – were selected. * Mean values are reported and the statistical significance was calculated with Fisher's exact test. † Median values are reported and the statistical significance was calculated with the Wilcoxon test.

Table S10: The reported results are independent of the major sub-cellular locations

p-values of S and U proteins in the major sub-cellular locations

Cellular quantity	cytoplasm		nucleus	
	x_S	x_U	x_S	x_U
Transcriptional noise [% with TATA box] *	27	14 (2×10^{-8})	21	15 (5×10^{-3})
Transcriptional complexity (TFs/gene) †	2.0	2.0 (0.1)	2.0	2.0 (0.3)
Transcriptional rate [mRNAs/hr] †	4.1	1.6 ($<10^{-16}$)	3.6	1.7 (4×10^{-16})
Transcript abundance [copies/cell] †	1.6	0.7 ($<10^{-16}$)	1.3	0.7 ($<10^{-16}$)
Transcript half-life [min] †	22.0	19.0 (3×10^{-8})	19	17 (4×10^{-5})
Transcript degradation [% short poly(A)] *	19	65 (2×10^{-10})	19	74 ($<10^{-16}$)
Translational noise [DM] †	0.30	-0.52 (1×10^{-10})	0.27	-0.29 (9×10^{-5})
Translational rate [ribosomes/ORF] †	0.56	0.32 ($<10^{-16}$)	0.54	0.32 ($<10^{-16}$)
Protein abundance [proteins/cell] †	4800	1540 ($<10^{-16}$)	3970	1545 ($<10^{-16}$)
Protein half-life [min] †	54	35 (3×10^{-15})	51	31 (2×10^{-12})
Protein degradation [% with PEST] *	10	48 ($<10^{-16}$)	16	47 ($<10^{-16}$)

Mean or median and p -values of different cellular quantities for cytoplasmic and nuclear proteins that are structured (S) and highly unstructured (U) in *S. cerevisiae*. To get equally sized bins, S was defined as proteins having less than 15% of their residues disordered and U as proteins with more than 35% of the residues disordered. * Mean values are reported and the statistical significance was calculated with Fisher's exact test. † Median values are reported and the statistical significance was calculated with the Wilcoxon test.

Table S11: The reported observations are independent of protein length

p-values of S and U proteins after grouping according to protein length

Cellular quantity / Protein Length	150-350 aa		350-600 aa		600+ aa	
	x_S	x_U	x_S	x_U	x_S	x_U
Transcriptional noise [% with TATA box] *	27	17 (4×10^{-5})	28	14 (3×10^{-8})	19	12 (4×10^{-8})
Transcriptional complexity (TFs/gene) †	2.0	2.0 (0.04)	2.0	2.0 (0.2)	2.0	2.0 (0.2)
Transcriptional rate [mRNAs/hr] †	2.8	2.3 (0.1)	2.4	1.7 (6×10^{-5})	2.0	1.3 (1×10^{-7})
Transcript abundance [copies/cell] †	1.1	1.0 (0.1)	0.9	0.7 (4×10^{-6})	0.9	0.6 (2×10^{-5})
Transcript half-life [min] †	21.0	18.0 (2×10^{-9})	23.0	17.0 (2×10^{-10})	23.0	19.0 (9×10^{-4})
Transcript degradation [% short poly(A)] *	10	30 (4×10^{-4})	25	73 (1×10^{-14})	31	95 (1×10^{-8})
Translational noise [DM] †	0.31	-0.20 (0.01)	0.97	-0.16 (1×10^{-7})	0.37	-0.61 (4×10^{-7})
Translational rate [ribosomes/ORF] †	0.58	0.55 (0.07)	0.43	0.30 (1×10^{-8})	0.24	0.14 (3×10^{-8})
Protein abundance [proteins/cell] †	3060	2120 (2×10^{-5})	3270	1830 (2×10^{-7})	3975	861 ($<10^{-16}$)
Protein half-life [min] †	55	52 (0.02)	53	34 (3×10^{-10})	38	27 (3×10^{-7})
Protein degradation [% with PEST] *	7	29 ($<10^{-16}$)	10	57 ($<10^{-16}$)	20	72 ($<10^{-16}$)

Mean or median and p -values of different cellular quantities for proteins that are structured (S) and highly unstructured (U) in *S. cerevisiae*. Proteins were divided according to their length in three equally sized groups: (i) 150-350 amino acids, (ii) 350-600 amino acids, and (iii) more than 600 amino acids. The mean length of structured and unstructured proteins is not significantly different ($p > 0.1$) within each group of proteins. * Mean values are reported and the statistical significance was calculated with Fisher's exact test. † Median values are reported and the statistical significance was calculated with the Wilcoxon test.

Table S12: GO functional analysis of the S and U group of proteins

GO functional categories enriched in the structured and unstructured proteins

GO biological process over-represented among structured proteins (S)	p-value	GO biological process over-represented among unstructured proteins (U)	p-value
organic acid metabolism	1.9E-42	cell organization and biogenesis	4.7E-24
carboxylic acid metabolism	1.9E-42	regulation of biological process	1.2E-20
Biosynthesis	2.8E-41	regulation of physiological process	4.9E-20
cellular biosynthesis	9.4E-40	organelle organization and biogenesis	2.1E-19
Metabolism	3.5E-37	transcription	2.4E-18
cellular metabolism	5.3E-37	regulation of cellular physiological process	1.0E-17
amino acid and derivative metabolism	9.5E-32	regulation of cellular process	1.2E-17
amine metabolism	4.5E-30	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism	6.2E-16
nitrogen compound metabolism	6.3E-29	regulation of metabolism	1.0E-15
amino acid metabolism	3.2E-27	regulation of transcription	1.5E-15

Supplementary References S1 to S104

- S1. Balaji S, Babu MM, Iyer LM, Luscombe NM, & Aravind L (2006) Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J Mol Biol* 360(1):213-227.
- S2. Holstege FC, *et al.* (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95(5):717-728.
- S3. Wang Y, *et al.* (2002) Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A* 99(9):5860-5865.
- S4. Gerber AP, Herschlag D, & Brown PO (2004) Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol* 2(3):E79.
- S5. Beilharz TH & Preiss T (2007) Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *Rna* 13(7):982-997.
- S6. Ghaemmaghani S, *et al.* (2003) Global analysis of protein expression in yeast. *Nature* 425(6959):737-741.
- S7. Belle A, Tanay A, Bitincka L, Shamir R, & O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A* 103(35):13004-13009.
- S8. Arava Y, *et al.* (2003) Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 100(7):3889-3894.
- S9. Newman JR, *et al.* (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441(7095):840-846.
- S10. Tirosh I, Weinberger A, Carmi M, & Barkai N (2006) A genetic signature of interspecies variations in gene expression. *Nat Genet* 38(7):830-834.
- S11. Basehoar AD, Zanton SJ, & Pugh BF (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* 116(5):699-709.
- S12. Lackner DH, *et al.* (2007) A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol Cell* 26(1):145-155.
- S13. Matsuyama A, *et al.* (2006) ORFeome cloning and global analysis of protein localization in the fission yeast *Schizosaccharomyces pombe*. *Nat Biotechnol* 24(7):841-847.
- S14. Yang E, *et al.* (2003) Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Res* 13(8):1863-1872.
- S15. Ptacek J, *et al.* (2005) Global analysis of protein phosphorylation in yeast. *Nature* 438(7068):679-684.
- S16. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, & Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337(3):635-645.
- S17. Prilusky J, *et al.* (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21(16):3435-3438.
- S18. 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(16):3433-3434.
- S19. Michelitsch MD & Weissman JS (2000) A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci U S A* 97(22):11910-11915.
- S20. Laskowski RA & Thornton JM (2008) Understanding the molecular machinery of genetics through 3D structures. *Nat Rev Genet* 9(2):141-151.
- S21. Wright PE & Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293(2):321-331.
- S22. Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27(10):527-533.
- S23. Dunker AK, *et al.* (2001) Intrinsically disordered protein. *J Mol Graph Model* 19(1):26-59.
- S24. Dunker AK, Obradovic Z, Romero P, Garner EC, & Brown CJ (2000) Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* 11:161-171.
- S25. Kriwacki RW, Hengst L, Tennant L, Reed SI, & Wright PE (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A* 93(21):11504-11509.
- S26. Dyson HJ & Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6(3):197-208.
- S27. Tompa P (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* 579(15):3346-3354.
- S28. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, & Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* 272(20):5129-5148.
- S29. Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, & Nussinov R (2003) Extended disordered proteins: targeting function with less scaffold. *Trends Biochem Sci* 28(2):81-85.
- S30. Daughdrill GW, Chadsey MS, Karlinsey JE, Hughes KT, & Dahlquist FW (1997) The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28. *Nat Struct Biol* 4(4):285-291.
- S31. Lacy ER, *et al.* (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat Struct Mol Biol* 11(4):358-364.
- S32. Fuxreiter M, Tompa P, & Simon I (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* 23(8):950-956.
- S33. Xie H, *et al.* (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res* 6(5):1917-1932.
- S34. Iakoucheva LM, *et al.* (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 32(3):1037-1049.
- S35. Meszaros B, Tompa P, Simon I, & Dosztanyi Z (2007) Molecular principles of the interactions of disordered proteins. *J Mol Biol* 372(2):549-561.
- S36. Puntervoll P, *et al.* (2003) ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res* 31(13):3625-3630.
- S37. Davey NE, Shields DC, & Edwards RJ (2006) SLiMDisc: short, linear motif discovery, correcting for common evolutionary descent. *Nucleic Acids Res* 34(12):3546-3554.
- S38. Oldfield CJ, *et al.* (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 44(37):12454-12470.
- S39. Dyson HJ & Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 12(1):54-60.
- S40. Fuxreiter M, Simon I, Friedrich P, & Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol* 338(5):1015-1026.
- S41. Sugase K, Dyson HJ, & Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447(7147):1021-1025.
- S42. Mohan A, *et al.* (2006) Analysis of molecular recognition features (MoRFs). *J Mol Biol* 362(5):1043-1059.
- S43. Vacic V, *et al.* (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* 6(6):2351-2366.
- S44. Lobley A, Swindells MB, Orengo CA, & Jones DT (2007) Inferring function using patterns of native disorder in proteins. *PLoS Comput Biol* 3(8):e162.
- S45. Chen JW, Romero P, Uversky VN, & Dunker AK (2006) Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. *J Proteome Res* 5(4):888-898.
- S46. Uversky VN, Oldfield CJ, & Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* 18(5):343-384.
- S47. Xie H, *et al.* (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res* 6(5):1882-1898.
- S48. Liu J, *et al.* (2006) Intrinsic disorder in transcription factors. *Biochemistry* 45(22):6873-6888.
- S49. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, & Dunker AK (2002) Intrinsic disorder in cell-signaling and

- cancer-associated proteins. *J Mol Biol* 323(3):573-584.
- S50. Sunde M, *et al.* (2004) TC-1 is a novel tumorigenic and natively disordered protein associated with thyroid cancer. *Cancer Res* 64(8):2766-2773.
- S51. Baldassarre G, *et al.* (2005) p27(Kip1)-stathmin interaction influences sarcoma cell migration and invasion. *Cancer Cell* 7(1):51-63.
- S52. Sherr CJ (2006) Divorcing ARF and p53: an unsettled case. *Nat Rev Cancer* 6(9):663-673.
- S53. Grimm M, *et al.* (2007) Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases. *Cell* 128(2):269-280.
- S54. Stefani M & Dobson CM (2003) Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* 81(11):678-699.
- S55. Uversky VN & Fink AL (2004) Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim Biophys Acta* 1698(2):131-153.
- S56. Chiti F & Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333-366.
- S57. Eisenberg D, *et al.* (2006) The structural biology of protein aggregation diseases: Fundamental questions and some answers. *Acc Chem Res* 39(9):568-575.
- S58. Goedert M (2001) Alpha-synuclein and neurodegenerative diseases. *Nat Rev Neurosci* 2(7):492-501.
- S59. Fernandez-Funez P, *et al.* (2000) Identification of genes that modify ataxin-1-induced neurodegeneration. *Nature* 408(6808):101-106.
- S60. Hardy J & Gwinn-Hardy K (1998) Genetic classification of primary neurodegenerative disease. *Science* 282(5391):1075-1079.
- S61. Karres JS, Hilgers V, Carrera I, Treisman J, & Cohen SM (2007) The conserved microRNA miR-8 tunes atrophin levels to prevent neurodegeneration in Drosophila. *Cell* 131(1):136-145.
- S62. Bachmair A, Finley D, & Varshavsky A (1986) In vivo half-life of a protein is a function of its amino-terminal residue. *Science* 234(4773):179-186.
- S63. Meinnel T, Serero A, & Giglione C (2006) Impact of the N-terminal amino acid on targeted protein degradation. *Biol Chem* 387(7):839-851.
- S64. Dice JF (1987) Molecular determinants of protein half-lives in eukaryotic cells. *Faseb J* 1(5):349-357.
- S65. Tartaglia GG, Pechmann S, Dobson CM, & Vendruscolo M (2007) Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem Sci* 32(5):204-206.
- S66. Singh GP, Ganapathi M, Sandhu KS, & Dash D (2006) Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins* 62(2):309-315.
- S67. Hargrove JL & Schmidt FH (1989) The role of mRNA and protein stability in gene expression. *Faseb J* 3(12):2360-2370.
- S68. Duttagupta R, *et al.* (2005) Global analysis of Pub1p targets reveals a coordinate control of gene expression through modulation of binding and stability. *Mol Cell Biol* 25(13):5499-5513.
- S69. Rendl LM, Bieman MA, & Smibert CA (2008) *S. cerevisiae* Vts1p induces deadenylation-dependent transcript degradation and interacts with the Ccr4p-Pop2p-Not deadenylase complex. *Rna* 14(7):1328-1336.
- S70. Ulbricht RJ & Olivas WM (2008) Puf1p acts in combination with other yeast Puf proteins to control mRNA stability. *Rna* 14(2):246-262.
- S71. Vasudevan S & Peltz SW (2001) Regulated ARE-mediated mRNA decay in *Saccharomyces cerevisiae*. *Mol Cell* 7(6):1191-1200.
- S72. Zaidi SH, Denman R, & Malter JS (1994) Multiple proteins interact at a unique cis-element in the 3'-untranslated region of amyloid precursor protein mRNA. *J Biol Chem* 269(39):24000-24006.
- S73. Zalfa F, *et al.* (2007) A new function for the fragile X mental retardation protein in regulation of PSD-95 mRNA stability. *Nat Neurosci* 10(5):578-587.
- S74. Chu I, *et al.* (2007) p27 phosphorylation by Src regulates inhibition of cyclin E-Cdk2. *Cell* 128(2):281-294.
- S75. Chen HK, *et al.* (2003) Interaction of Akt-phosphorylated ataxin-1 with 14-3-3 mediates neurodegeneration in spinocerebellar ataxia type 1. *Cell* 113(4):457-468.
- S76. Chu IM, Hengst L, & Slingerland JM (2008) The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat Rev Cancer* 8(4):253-267.
- S77. Miller S, Lesk AM, Janin J, & Chothia C (1987) The accessible surface area and stability of oligomeric proteins. *Nature* 328(6133):834-836.
- S78. Goldberg AL (2003) Protein degradation and protection against misfolded or damaged proteins. *Nature* 426(6968):895-899.
- S79. Tompa P, Prilusky J, Silman I, & Sussman JL (2007) Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins*.
- S80. Asher G, Reuven N, & Shaul Y (2006) 20S proteasomes and protein degradation "by default". *Bioessays* 28(8):844-849.

- S81. Tsvetkov P, *et al.* (2007) Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome. *Proteins*.
- S82. Seet BT, Dikic I, Zhou MM, & Pawson T (2006) Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* 7(7):473-483.
- S83. Bhattacharyya RP, Remenyi A, Yeh BJ, & Lim WA (2006) Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu Rev Biochem* 75:655-680.
- S84. Pawson T & Warner N (2007) Oncogenic re-wiring of cellular signaling pathways. *Oncogene* 26(9):1268-1275.
- S85. Yang XJ (2005) Multisite protein modification and intramolecular signaling. *Oncogene* 24(10):1653-1662.
- S86. Boiani M & Scholer HR (2005) Regulatory networks in embryo-derived pluripotent stem cells. *Nat Rev Mol Cell Biol* 6(11):872-884.
- S87. Chang L & Karin M (2001) Mammalian MAP kinase signalling cascades. *Nature* 410(6824):37-40.
- S88. Goodman RH & Smolik S (2000) CBP/p300 in cell growth, transformation, and development. *Genes Dev* 14(13):1553-1577.
- S89. Kumar D, Srikanth R, Ahlfors H, Lahesmaa R, & Rao KV (2007) Capturing cell-fate decisions from the molecular signatures of a receptor-dependent signaling response. *Mol Syst Biol* 3:150.
- S90. O'Dea EL, *et al.* (2007) A homeostatic model of IkappaB metabolism to control constitutive NF-kappaB activity. *Mol Syst Biol* 3:111.
- S91. Rutkowski DT, *et al.* (2006) Adaptation to ER stress is mediated by differential stabilities of pro-survival and pro-apoptotic mRNAs and proteins. *PLoS Biol* 4(11):e374.
- S92. Li X, *et al.* (2007) Ubiquitin- and ATP-independent proteolytic turnover of p21 by the REGgamma-proteasome pathway. *Mol Cell* 26(6):831-842.
- S93. Tambyrajah WS, Bowler LD, Medina-Palazon C, & Sinclair AJ (2007) Cell cycle-dependent caspase-like activity that cleaves p27(KIP1) is the beta(1) subunit of the 20S proteasome. *Arch Biochem Biophys* 466(2):186-193.
- S94. Tofaris GK, Layfield R, & Spillantini MG (2001) alpha-synuclein metabolism and aggregation is linked to ubiquitin-independent degradation by the proteasome. *FEBS Lett* 509(1):22-26.
- S95. David DC, *et al.* (2002) Proteasomal degradation of tau protein. *J Neurochem* 83(1):176-185.
- S96. Balch WE, Morimoto RI, Dillin A, & Kelly JW (2008) Adapting Proteostasis for Disease Intervention. *Science* 319(5865):916-919.
- S97. Wiseman RL, Powers ET, Buxbaum JN, Kelly JW, & Balch WE (2007) An adaptable standard for protein export from the endoplasmic reticulum. *Cell* 131(4):809-821.
- S98. Chin JW (2006) Modular approaches to expanding the functions of living matter. *Nat Chem Biol* 2(6):304-311.
- S99. Kholodenko BN (2006) Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol* 7(3):165-176.
- S100. Schwartz MA & Madhani HD (2004) Principles of MAP kinase signaling specificity in *Saccharomyces cerevisiae*. *Annu Rev Genet* 38:725-748.
- S101. Natarajan M, Lin KM, Hsueh RC, Sternweis PC, & Ranganathan R (2006) A global analysis of cross-talk in a mammalian cellular signalling network. *Nat Cell Biol* 8(6):571-580.
- S102. Oda K & Kitano H (2006) A comprehensive map of the toll-like receptor signaling network. *Mol Syst Biol* 2:20060015.
- S103. Xia Y, *et al.* (2004) Analyzing cellular biochemistry in terms of molecular networks. *Annu Rev Biochem* 73:1051-1087.
- S104. Rogers S, Wells R, & Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 234(4774):364-368.