

Minireview

Regulating highly dynamic unstructured proteins and their coding mRNAs

Buyong Ma* and Ruth Nussinov*[†]

Addresses: *Basic Research Program, SAIC-Frederick, Inc, Center for Cancer Research Nanobiology Program, NCI-Frederick, Frederick, MD 21702, USA. [†]Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel.

Correspondence: Ruth Nussinov. Email: ruthnu@helix.nih.gov

Published: 28 January 2009

Genome Biology 2009, **10**:204 (doi:10.1186/gb-2009-10-1-204)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2009/10/1/204>

© 2009 BioMed Central Ltd

Abstract

The lifetimes and conformations of intrinsically unstructured proteins (IUPs) and their mRNAs are orchestrated to ensure precision, speed and flexibility in biological control.

The complexity of a protein sequence - that is, its information content - is related to structure and function [1,2]. As far as we know, sequences of proteins with defined structures tend to have higher sequence complexity, whereas sequences of intrinsically unstructured proteins (IUPs) are of lower complexity. A significant part of an IUP is devoid of a stable three-dimensional structure when free (unbound) in solution. Unstructured or disordered proteins are known to have numerous vital functions [2], and simple sequences apparently evolve more rapidly than those of highly structured proteins [3].

Living systems have either adapted to IUPs very early in evolution or have evolved complex mechanisms to take advantage of their properties at a later stage. A recent report in *Science* by Gsponer *et al.* [4] indicates that in yeast, regardless of evolutionary time scale, the regulation of the production, maintenance and function of unstructured proteins can occur at multiple levels: during mRNA transcription and degradation, during protein translation and degradation, and by controlling the fidelity of transcription and translation. Such regulation of IUPs at nearly every stage of transcription and translation may be warranted to ensure precision, speed and flexibility in biological control [5]. An intriguing question is how the cell coordinates the DNA → RNA → protein sequence → structure → function paradigm to orchestrate IUP lifetimes. While specific mechanisms and pathways may vary for different IUPs, analysis of the *Saccharomyces cerevisiae* proteome illustrates the

range of molecular strategies that control the availability of such proteins within the cell.

Both mRNA and protein sequence can affect mRNA stability and translation rates

The mRNA nucleotide sequence provides the codons specifying the amino acid sequence of the encoded protein; thus, the two sequences are not independent of each other. So, even though the degeneracy of the genetic code prevents a one-to-one sequence relationship, it is expected that simple low-complexity protein sequences would enforce some constraints on the encoding mRNA sequences, although it is still unclear to what extent. Such relationships have been observed; for example, GC-rich genomic regions encode some simple protein repeats [3]. DNA sequence analysis also shows that dinucleotide occurrences are remarkably non-random, thus biasing codon frequencies [6]. Codon usage also reflects a correlation with GC content, a correlation probably resulting from constraints on the primary genetic structure [7]. More directly relevant to disordered protein sequences is the possibility that α -helices and β -strands could be preferentially 'coded' by stems in mRNA secondary structure, and coils by mRNA loops [8]. Statistical analysis of retroviral mRNA supports a relationship between mRNA secondary structure and the proteins they encode [9]. However, a comprehensive analysis of the sequences of IUP mRNAs and their potential secondary structures is needed.

Less structured mRNAs are intrinsically less stable and more easily degradable. Jeff Ross has argued that it would make little sense to synthesize very stable proteins from unstable mRNAs, and that it makes more sense to have unstable mRNAs encode unstable proteins [10]. mRNAs that encode proteins produced only in short bursts in response to internal or external stimuli have short half-lives [10]. Nevertheless, for short-lived IUPs, the degradation of mRNA due to less structure may not be as important as the transcript degradation signal encoded by poly(A) tail length. Indeed, Gsponer *et al.* [4] found that 60% of the IUPs in the U group (highly unstructured proteins with 30-100% of the sequence unstructured) have a short poly(A) tail compared with only 20% in the S group (highly structured with less than 10% of the sequence unstructured). This large difference strongly suggests that the length of poly(A) tail is a signal for mRNA degradation in IUP-coding mRNAs. The minimum length of a poly(A) tail is around 22-33 adenines to allow its efficient interactions with the 5' cap sequence, with other proteins to protect against 5' and 3' degradation, and to form a stable translation complex [11].

Less structured mRNAs are *a priori* expected to have faster translation rates as they do not incur the energy penalty of having to open up RNA secondary structure. Such high translation rates may not always be desirable. In principle, disordered regions with low sequence complexity can be coded to decrease translation efficiency. Even without a protein-mRNA correlation, the sequence of the coding regions can affect mRNA secondary structure [12] and thus help control protein synthesis. However, secondary structure can have different effects: in the hepatitis C virus, the stable RNA structure may prevent translation mediated by the internal ribosome entry site [13]; on the other hand, a purine-overloaded virus-encoded mRNA lacking secondary structure also had low efficiency of translation, preventing protein synthesis and thus endogenous antigen presentation [14]. Remarkably, reducing the purine bias through constructs that expressed codon-modified sequences while maintaining the encoded protein sequence increased the amount of stem-loop structure in the corresponding mRNA and dramatically enhanced synthesis of the viral protein [14].

Therefore, to ensure slow synthesis of IUPs and thus avoid protein aggregation (to which IUPs are prone), there could be a mechanism for overwriting possible interference from mRNA secondary structure; this might comprise a dual poly(A) tail function to regulate both mRNA degradation and translation, with a shorter poly(A) tail being less efficient at ribosome binding [15]. Thus, with short poly(A) tails, the mRNAs of IUPs could ensure low ribosomal density and slower translation rates. Although this possibility was not explicitly discussed by Gsponer *et al.*, it could also underlie the lower ribosomal density shown in one of their schematic figures.

Protein population shift and conformational selection due to post-translational modification

Molecular disorder has been viewed as local or global instability. Yet, even when proteins appear disordered, there are preferred conformational states, with higher population times [16]. Thus, IUP conformations that potentially bind to a variety of binding partners can be hidden in the illusion of seeming disorder. As they are unstable, they might not be observed by experiment.

The definition of an 'unstructured' or 'disordered' protein is based on current experimental timescales for protein structure characterization. IUPs are highly dynamic, however, and advances in analytical techniques have revealed previously unobserved details of the ensemble of structures they adopt. For example, upon binding to the KIX domain of the CREB-binding protein, the folding and binding of the intrinsically unstructured phosphorylated kinase-inducible activation domain (pKID) of the transcription factor CREB results in an ensemble of transient encounter complexes [17]. This ensemble is at least partially produced by selection among pre-existing pKID conformations. In another example, a structural ensemble of ubiquitin with solution dynamics up to microseconds has been revealed to cover the complete structural heterogeneity observed in 46 ubiquitin crystal structures, validating a molecular recognition mechanism of conformational selection [18] rather than induced-fit for ubiquitin [19]. The heterodimeric FACT (facilitates chromatin transcription) protein is predicted to have large IUP regions in each subunit. Successive high-speed atomic force microscopy (AFM) images of FACT on a mica surface clearly reveal two distinct tail-like IUP regions that protrude from the main body of FACT and fluctuate in position [20].

IUPs are on average twice as likely [4] as other proteins to be substrates of kinases, highlighting the importance of post-translational modification in fine-tuning IUP function. Post-translational modifications of IUPs serve as important modulators of the conformational energy landscape, which in turn regulates IUP binding. An example illustrating the importance of post-translational modifications in IUPs is the p53 protein, which has more than a dozen phosphorylation and acetylation sites, conferring different biological signals [21]. As illustrated in Figure 1, ensembles may have clusters of geometrically similar conformational substates separated by low energy barriers. A post-translational modification can bias this distribution, increasing the population time of a cluster that preferentially binds a specific partner. Post-translational modification is an allosteric switch, which can turn on or off an IUP's binding potential (Figure 1), with a consequent binding and population shift.

Post-translational modifications of IUPs similarly serve as on/off signals for their own degradation. In the case of p53, phosphorylation at Ser20 turns off binding to the protein MDM2, with a consequent increase in p53 concentration,

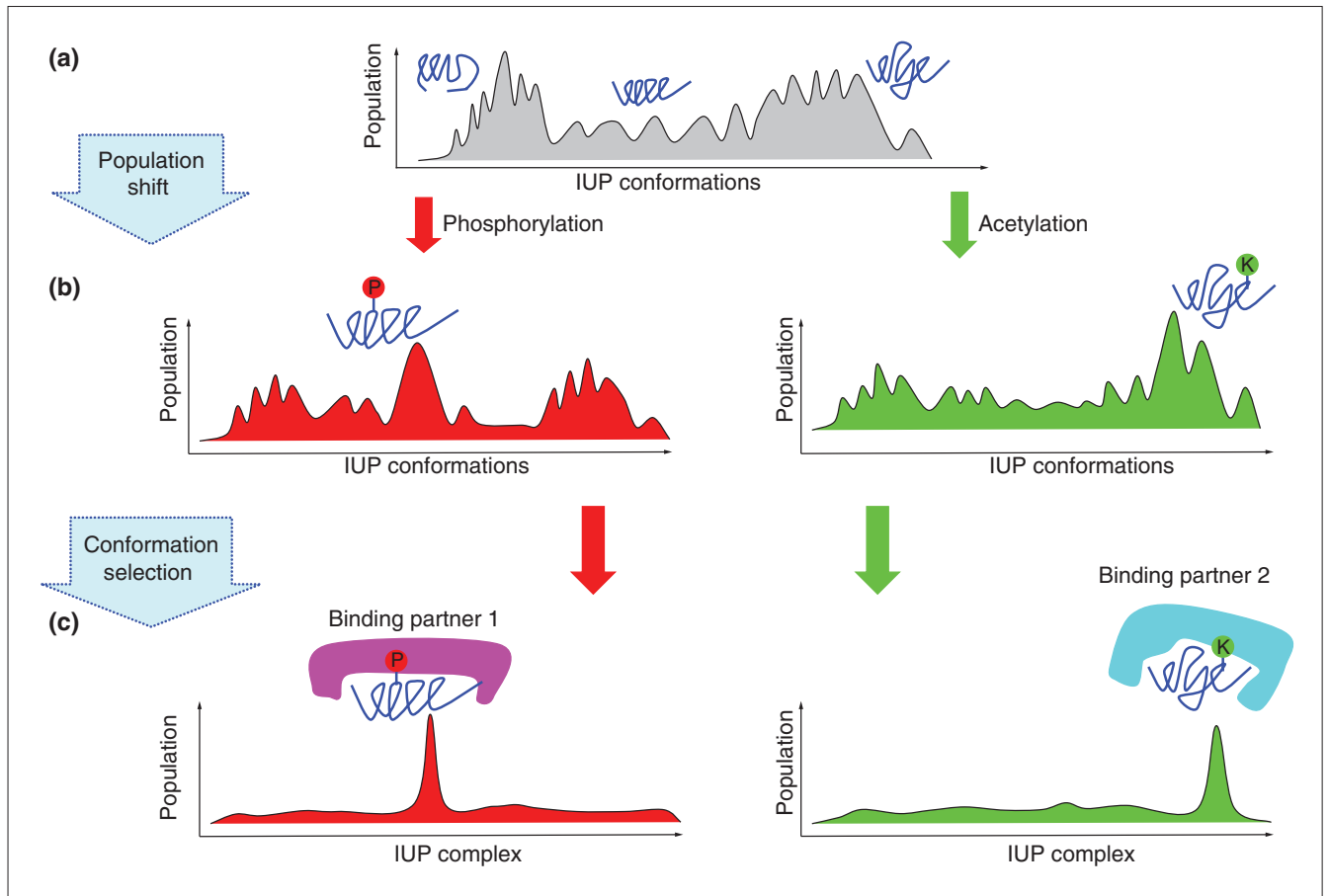


Figure 1
 The energy landscape of IUP conformations, the effects of post-translational modifications and their relationship to function. **(a)** The x-axis depicts the conformational ensemble. Conformations that are geometrically similar lie close to each other. The y-axis depicts the population size. **(b)** The dynamic conformational selection of IUPs through post-translational modifications and molecular interactions. Here two post-translational modifications are shown: phosphorylation (P) and acetylation (K). Both result in conformational selection and population shift in the ensemble of structures. Many structural clusters coexist for a seemingly unstructured protein. Post-translational modifications create allosteric perturbation sites, propagating through the structures like waves. The observable outcome is a shift in the distribution of the population, biasing the ensemble towards conformers whose structures are favored to bind specific partners. **(c)** A specific conformation is selected by a binding partner with best complementarity to the IUP binding site.

whereas phosphorylation at Thr155 targets p53 to degradation via the ubiquitin system (reviewed in [21]). Hence, selective post-translational modification modulates the ensemble distribution via a dynamic conformational selection mechanism [18,22], tuning it to functional need.

Precision control of the abundance and dynamics of IUPs by protein-mRNA interactions

Transcription factors are enriched in IUPs, and many IUPs are hubs in the cellular gene interaction network. This network can be disrupted by changes in the abundance of IUPs or by mutations introduced during transcription or translation. For p53, whose concentration has to be low in normal cells, the majority of cancer-related mutations occur in the folded core domain that is responsible for DNA

recognition; the disordered amino and carboxyl termini have considerably fewer cancer-related mutations. This could be explained by these regions being less critical for function, but it also reflects the fact that they are disordered regions that already have broadly distributed conformational ensembles and are thus less prone to disturbance.

Achieving a pre-existing steady-state production of a protein is a prerequisite for an optimal dynamic response to a cellular signal. Even though a rate of expression (transcription and translation) can relate to fluctuation in protein production, Raser and O'Shea concluded that stochasticity in protein production is intrinsic to promoter-specific gene expression and does not depend on the rate of expression [23]. Gsponer *et al.* [4] have followed the Raser and O'Shea argument: they investigated whether IUPs have lower

transcriptional stochasticity than other proteins because of a lower percentage of TATA box sequence in their promoters, and observed this to be the case. In addition, the authors also observed a lower stochasticity in the translation of IUPs. If degenerate codon usage is similar for the same amino acids, one might expect that the low complexity of IUP protein sequences could lead to a more uniform translation rate. However, the lower translational stochasticity found by Gsponer *et al.* could also reflect additional regulation mechanisms involving protein-mRNA interaction [24,25], which could be optimized to maintain either constant or oscillating protein levels.

Recent studies of the p53 system provide an insight into the protein-mRNA regulation problem. The interaction of p53 and MDM2 is a typical feedback system. p53 transactivates MDM2, and binding of MDM2 in turn leads to p53 degradation (which can be turned off by p53 phosphorylation at Ser20). However, post-translational modifications and an on/off degradation switch are insufficient to guarantee an efficient response by p53 to cell stress. For additional translational control, p53 binds specifically to the 5' untranslated region of its own mRNA, thus preventing p53 mRNA translation. As a result, the higher the p53 concentration, the lower the p53 mRNA translation [24]. Also, MDM2 interacts with p53 mRNA; the RING domain of MDM2 binds to a stem-loop structure in p53 mRNA at the Leu22 codon, thus impairing p53-MDM2 binding, which mediates p53 degradation [25].

The broad picture emerging from the accumulating data on the sequence and structure of IUPs and their regulation by protein-mRNA interactions vividly illustrates the molecular strategies that nature has designed to efficiently control the life of IUPs and the life of the cell. As a typical IUP that regulates hundreds of genes, the p53 protein and its mRNA serve as a paradigm of these sequence-structure-function and cross-regulation relationships. Nature has optimized IUPs to perform complex cellular functions, enforcing low sequence complexity with consequent highly dynamic protein conformation. As Gsponer *et al.* [4] show, IUPs have evolved to be under tight regulation to minimize their own half-lives and those of their mRNAs. Yet, since the sequences of mRNAs and the protein sequences they encode are not independent of each other, the lower sequence complexity of IUPs may already imply lower structural stability and thus shorter mRNA half-life. However, even if the lower stability, in terms of the lower secondary structure content of the mRNA, indeed derives from the lower complexity of the IUP sequences, the stronger poly(A) length is an independent degradation signal ensuring short mRNA lifetime. Post-translational modifications can also serve as degradation signals for IUPs by allosterically shifting the population to states that bind proteins targeted for degradation. IUPs also contain degradation-sensitive unstable hydrophobic-poor PEST regions (enriched in Pro, Glu, Ser and Thr). Precision

control of transcription can be achieved by the TATA box length and mRNA translational cross-regulation can be attained by interaction with the encoded protein.

Acknowledgements

This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

- Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK: **Sequence complexity of disordered protein.** *Proteins* 2001, **42**:38-48.
- Dyson HJ, Wright PE: **Intrinsically unstructured proteins and their functions.** *Nat Rev Mol Cell Biol* 2005, **6**:197-208.
- Alba MM, Tompa P, Veitia RA: **Amino acid repeats and the structure and evolution of proteins.** *Genome Dyn* 2007, **3**:119-130.
- Gsponer J, Futschik ME, Teichmann SA, Babu MM: **Tight regulation of unstructured proteins: from transcript synthesis to protein degradation.** *Science* 2008, **322**:1365-1368.
- Shu Y, Lin H: **Transcription, translation, degradation, and circadian clock.** *Biochem Biophys Res Commun* 2004, **321**:1-6.
- Nussinov R: **Eukaryotic dinucleotide preference rules and their implications for degenerate codon usage.** *J Mol Biol* 1981, **149**:125-131.
- Antezana MA, Jordan IK: **Highly conserved regimes of neighbor-base-dependent mutation generated the background primary-structural heterogeneities along vertebrate chromosomes.** *PLoS ONE* 2008, **3**:e2145.
- Jia M, Luo L: **The relation between mRNA folding and protein structure.** *Biochem Biophys Res Commun* 2006, **343**:177-182.
- Konecny J, Schoniger M, Hofacker I, Weitz MD, Hofacker GL: **Concurrent neutral evolution of mRNA secondary structures and encoded proteins.** *J Mol Evol* 2000, **50**:238-242.
- Ross J: **mRNA stability in mammalian cells.** *Microbiol Rev* 1995, **59**:423-450.
- Amrani N, Ghosh S, Mangus DA, Jacobson A: **Translation factors promote the formation of two states of the closed-loop mRNP.** *Nature* 2008, **453**:1276-1280.
- Katz L, Burge CB: **Widespread selection for local RNA secondary structure in coding regions of bacterial genes.** *Genome Res* 2003, **13**:2042-2051.
- Rijnbrand R, Bredenbeek PJ, Haasnoot PC, Kieft JS, Spaan WJ, Lemon SM: **The influence of downstream protein-coding sequence on internal ribosome entry on hepatitis C virus and other flavivirus RNAs.** *RNA* 2001, **7**:585-597.
- Tellam J, Smith C, Rist M, Webb N, Cooper L, Vuocolo T, Connolly G, Tschärke DC, Devoy MP, Khanna R: **Regulation of protein translation through mRNA structure influences MHC class I loading and T cell recognition.** *Proc Natl Acad Sci USA* 2008, **105**:9319-9324.
- Preiss T, Hentze MW: **Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast.** *Nature* 1998, **392**:516-520.
- Tsai CJ, Ma B, Sham YY, Kumar S, Nussinov R: **Structured disorder and conformational selection.** *Proteins* 2001, **44**:418-427.
- Sugase K, Dyson HJ, Wright PE: **Mechanism of coupled folding and binding of an intrinsically disordered protein.** *Nature* 2007, **447**:1021-1025.
- Ma B, Shatsky M, Wolfson HJ, Nussinov R: **Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations.** *Protein Sci* 2002, **11**:184-197.
- Boehr DD, Wright PE: **Biochemistry. How do proteins interact?** *Science* 2008, **320**:1429-1430.
- Miyagi A, Tsunaka Y, Uchihashi T, Mayanagi K, Hirose S, Morikawa K, Ando T: **Visualization of intrinsically disordered regions of proteins by high-speed atomic force microscopy.** *Chemphyschem* 2008, **9**:1859-1866.
- Bode AM, Dong Z: **Post-translational modification of p53 in tumorigenesis.** *Nat Rev Cancer* 2004, **4**:793-805.

22. Latzer J, Shen T, Wolynes PG: **Conformational switching upon phosphorylation: a predictive framework based on energy landscape principles.** *Biochemistry* 2008, **47**:2110-2122.
23. Raser JM, O'Shea EK: **Control of stochasticity in eukaryotic gene expression.** *Science* 2004, **304**:1811-1814.
24. Halaby MJ, Yang DQ: **p53 translational control: a new facet of p53 regulation and its implication for tumorigenesis and cancer therapeutics.** *Gene* 2007, **395**:1-7.
25. Candeias MM, Malbert-Colas L, Powell DJ, Daskalogianni C, Maslon MM, Naski N, Bourougaa K, Calvo F, Fähræus R: **p53 mRNA controls p53 activity by managing Mdm2 functions.** *Nat Cell Biol* 2008, **10**:1098-1105.