Alternative splicing of intrinsically disordered regions and rewiring of protein interactions
Marija Buljan1, Guilhem Chalancon1, A Keith Dunker2, Alex Bateman3, S Balaji1, Monika Fuxreiter4 and M Madan Babu1

Introduction
One of the major challenges in biology is to understand how novel phenotypic traits emerge and to describe their underlying molecular mechanisms. Numerous studies have highlighted the importance of gene duplication and recombination on one side [1,2] and alterations in gene expression regulation on the other side [3] as fundamental for the evolution of new phenotypic traits. Recently, alternative splicing (AS) has been shown to contribute significantly to changes in proteome composition between organisms and to the increase in cellular complexity and diversity of phenotypes among eukaryotes [4,5,6,7].

While AS plays a key role in the expansion of proteomic and regulatory complexity, the molecular and biochemical functions of differentially spliced exons remain largely uncharacterized [8]. AS tends to avoid protein domains but more often affect intrinsically disordered (ID) regions [9–12]. However, the general principles behind how these different isoforms contribute to the functional versatility and how they mediate the emergence of novel phenotypes is just beginning to be grasped [5,6,13,14]. ID regions are polypeptide segments that do not acquire a defined tertiary structure autonomously but adopt diverse interconverting conformational states [15–18]. Frequently, disordered regions fold completely upon binding [19] while in some cases such regions remain disordered even in the bound state [20,21]. Here, we synthesize emerging concepts to shed light on how AS of disordered regions might rewire molecular interactions, and how this may facilitate the emergence of novel phenotypes. In particular, we discuss how such events could be crucial for defining tissue-specific (TS) and organism-specific traits. While AS of structured domains may lead to similar effects, these are not discussed here.

How do disordered regions increase the functional versatility of proteins?
ID segments increase the functional versatility of proteins by: (a) providing conformational heterogeneity between structured domains, thus increasing the number of conformational states (Figure 1a); (b) presenting linear peptide motifs [22] that bind to other proteins, nucleic acids and small molecule ligands, thereby enabling larger protein interaction landscapes (Figure 1b) and (c) presenting amino acids that become post-translationally modified (PTM), thereby altering the chemical nature of the polypeptide and further increasing its diversity of interactions (Figure 1c). In addition, the overall net charge of the disordered segment, the modification state, and the amino acid composition can all influence the conformational ensemble and the chemical states of the entire protein, thereby altering the functions of the full-length proteins [18,23–32].

Splicing of disordered segments rewrites interaction networks
Spliced ID regions often harbor functionally important sites
Earlier computational studies of alternative protein isoforms from the UniProt database (http://www.uniprot.org/) as well as structures of different isoforms in the PDB...
Disordered segments increase the functional versatility of proteins. Disordered segments are represented in white, and structured domains as gray and blue blobs. A disordered segment can influence the functional versatility of a protein by increasing the conformational heterogeneity between domains and the number of conformational states achieved by the protein (left panel). Linear peptide motifs (green rectangles) embedded in disordered segments can tune the specificity, selectivity, kinetics, affinity and avidity of interactions with distinct partners (blue blobs) (middle panel). Post-translational modification sites embedded in disordered segments can modify the function by altering the segment’s chemical composition and/or by serving as regulators of diverse interaction properties (right panel).

(www.rcsb.org) suggest that AS exons in mRNA often code for ID regions [9,12]. Calculations on a much larger set of 15,678 proteins with 36,320 alternatively spliced regions confirms these observations and suggest that a higher portion of alternatively spliced mRNA codes for ID regions as compared to that which codes for structured regions (Dunker et al., unpublished results). AS protein segments in general [33**] and tissue-specific AS segments in particular [13**,14**] frequently contain disordered regions with embedded linear interaction motifs. This suggests that the alternative inclusion of such functional elements can trigger changes in protein function by altering its stability, subcellular location or by influencing its molecular interactions [13**,14**]. For example, the alternative inclusion of peptide KEN-motifs affect protein half-life by targeting specific isoforms for degradation, the alternative inclusion of NLS signal peptides affect the subcellular location of proteins, or the alternative inclusion of motifs that bind PDZ, PTB, SH2, SH3 or WW domains modulate interactions with these signaling domains [13**,33**,34,35].

AS disordered regions can also display the following characteristics: the segment (a) lacks linear interaction motifs or PTM sites [12,13**], thus inclusion of such segments may increase the conformational entropy or may alter the binding affinity of the full-length protein due to multiple weak transient interactions [20,21]; (b) contains multiple occurrences of the same motif [33**], thus altering the avidity of an interaction [36] or increasing the local concentration of an interacting protein within specific regions in a cell [37]; (c) contains distinct, possibly overlapping, motifs or Molecular Recognition Features (MoRFs) [38*], thus conferring the property of an interaction hub, and thereby contributing to multiple functional outcomes, including moonlighting [39]; and (d) contains multiple occurrences of PTM sites [13**], thus changing the overall charge distribution and influencing the conformational states and fine-tuning the interactions with other molecules [25,27] including the alteration of partner selection preferences [38*].

Tissue-specific rewiring of interactions through spliced ID segments
Buljan et al. [13**] recently analyzed published transcriptome data from several different human tissues and cancer cell lines [40] and found that tissue-specific AS protein segments frequently contain ID regions that embed functional elements such as evolutionarily conserved interaction motifs and PTM sites. Furthermore, genes with TS exons tend to occupy central positions in interaction networks and tend to display distinct interaction partners in the respective tissues. Many of these genes have roles in signaling and development and are associated with cancer and other human diseases. In an independent study, Ellis et al. [14**] reported similar findings, and, using an automated protein–protein interaction assay, they demonstrated a direct connection between the AS disordered segments and protein binding properties. Specifically, approximately one-third of the
analyzed neural-regulated exons were observed to affect protein–protein interactions. These two studies establish the concept that AS of ID segments with embedded functional elements can rewire protein interactions, regulatory networks and signaling pathways in a tissue-specific manner, thus conferring functional versatility to proteins and increasing interaction network diversity across tissues [13**] (Figures 2 and 3).

Not only disordered regions but also structured domains can undergo AS. AS affecting structured domains have been observed to rewire biomolecular interactions [41**,42**,43,44]. For instance, an evolutionarily conserved embryonic stem cell specific AS of mutually exclusive exons changes the DNA-binding preference of the forkhead transcription factor FOXP1. The altered residues in FOXP1’s DNA-binding domain affect specificity of DNA binding and hence influence pluripotency and differentiation of Embryonic Stem cells [42**].

**Molecular mechanism of interaction rewiring via splicing of disordered segments**

Splicing of disordered segments can affect a number of properties of protein interactions. Below, we discuss how such AS events influence intra-molecular and inter-molecular interactions [13**,45]. Note that through similar

---

**Figure 2**

Molecular mechanisms by which alternatively spliced ID segments influence protein interactions. Alternative splicing of disordered segments (ASDS) can regulate protein interactions via distinct mechanisms. (a) ASDS can influence the interaction of full-length proteins with other biomolecules by masking or competing for the interaction interface via auto-inhibition. For example, a segment with the Acid Box (AB; red oval) in Fibroblast Growth Factor (FGF) receptor prevents the protein from interacting with its ligand FGF1 and is subjected to alternative splicing. In the long isoform, this segment connects the immunoglobulin domains D1 and D2 (cyan, surface and cartoon representation; D1:PDB:2CKN and D2:PDB:1RY7). (b) ASDS can directly affect binary interactions, as shown for the binding of the human peroxin PEX3 (blue, surface representation) with the PEX19 binding motif (red helix within a disordered segment) (PDB:3MK4). (c) ASDS can influence dynamics and kinetics of protein complexes. The Ca2+/calmodulin-dependent protein kinase II (CAMKII) forms a dodecameric complex. Each holoenzyme subunit (in cyan) comprises of an N-terminal kinase domain and a C-terminal hub domain. These two domains are connected by a disordered linker (red), the length of which varies among splice isoforms. Variation in the linker length directly impacts the sensitivity of CAMKII for calcium gradients (PDB:3BHH).

Auto-inhibitory interactions

Auto-inhibition modules (IMs) influence interaction of full-length proteins with other biomolecules by masking or competing for the interaction interface (Figure 2a). Trudeau et al. [46**] analyzed auto-inhibited proteins and demonstrated that such IMs are enriched in intrinsic disorder. By comparing the properties of proteins with disordered IMs to those with structured IMs, they show that ID IMs permit fine-tuning of the equilibrium between the active and inactive states. The disordered IMs were observed to be more highly phosphorylated, more frequently AS, and to contain evolutionarily conserved linear peptide motifs that often changed their secondary structure during activation [46**]. Such IMs may not only be regulated by splicing but may also emerge independently in organisms during the course of evolution. For instance, an interaction between Homeobox HoxA11 and Forkhead box 01A (Foxo1a) in mammals has evolved only after a previously present, but masked, binding site in HoxA11 became exposed through mutations in a disordered region that mediated self-interaction [47]. Thus, inclusion or skipping of such segments may contribute to selectivity and altered kinetics in partner recognition. Such properties may be further modulated through TS splicing and hence contribute to fine-tuning of molecular interactions in a spatially and temporally regulated manner [13**].

Binary interactions

Linear motifs, MoRFs and PTMs within ID regions enable molecular interactions. Thus the AS of ID segments containing these elements can form new molecular interactions and thereby lead to the recruitment of the same biochemical or molecular functions (often mediated by structured domains in the same polypeptide) into a different context (Figure 2b) [13**]. For instance, differential inclusion of disordered segments with interaction motifs in a kinase protein could cause the resulting kinase to recruit completely different protein substrates. Similarly, the differential inclusion of disordered PTM sites (e.g. phosphorylation sites) could alter whether a given protein becomes the substrate for a
modifying enzyme (e.g. kinase). Thus AS of such disordered regions has the potential to rewire molecular interaction networks and alter the specificity, selectivity, affinity and kinetics of such molecular interactions [13**]. Indeed, Merkin et al. [6**] recently investigated AS in equivalent tissues across diverse organisms and found that AS often alters whether a protein can be phosphorylated, thereby delimiting the scope of kinase signaling. Similarly, changing the length of a disordered region can interfere allosterically with interface formation elsewhere due to increased conformational entropy [48,49]. For example, alternative isoforms of the homomeric transcription factor Ultrabithorax (UBX) gene, which are expressed in a developmental and in a TS manner, differ only in the length of the disordered segment adjacent to the DNA-binding homeodomain [48,50–52]. This AS region in UBX is not directly involved in binding, but rather modulates the protein’s dynamics and determines its DNA binding affinity. Thus AS of the ID region enables UBX isoforms to recognize different DNA sequences in distinct contexts and to bind to the same DNA sequence differently. Hence, alternative inclusion of ID regions can have a significant influence on the affinity and kinetics of protein interactions [13**].

Protein complexes
Many proteins form stable homo-oligomeric or hetero-oligomeric protein complexes in the cell [53,54]. ID regions in the interaction interface and their PTM often influence complex formation [55,56]. Expression of the subunit isoforms with differentially included ID segments can influence complex formation and cause tunable protein properties [57] (Figure 2c). Moreover, the simultaneous expression of multiple isoforms that form homo-complexes or hetero-complexes may lead to the sequestration of functional isoforms into ‘heterogeneous’ oligomeric complexes, thereby leading to transient loss or gain of function [58]. Depending on the expression level of the isoform that encodes TS segments, the equilibrium between the different oligomeric ‘protein-complex states’ will be influenced, thereby leading to competition for a common interaction partner [59] or altered kinetics in response to a signal. This may result in dominant negative effects or ultra-sensitivity due to molecular titration effects [60,61]. For example, the increased diversity of cell signaling options caused by receptor dimerization is further enhanced by the generation of splice variants in the case of GPCRs, one of the largest families of membrane signaling proteins [62].

Implications for evolution, development and disease
Evolution and development
During evolution, ID regions contribute more extensively than structured regions to the rewiring of protein interaction networks [63–65] by means of short insertions/deletions or point mutations that modulate interactions (Figure 3). Fine-tuning of gained or rewired interactions can also result from mutations of flanking residues around an interaction motif [66,67]. Furthermore, species-specific evolution of AS [68] of such altered disordered regions may further expand the functional repertoire of orthologous proteins by rewiring protein interaction networks in space (e.g. across tissues) or time (e.g. during development or evolution) [68]. Indeed, the work of Merkin et al. [6**] and Barbosa-Morais et al. [5**] revealed the following trends: (a) AS is well conserved only for a subset of exons and is frequently lineage-specific; (b) TS spliced segments are enriched in PTM sites and disordered regions; (c) tissue-specific AS often alters protein phosphorylation sites, delimiting the scope of kinase signaling; (d) certain splicing events likely remodel protein interactions involving orthologous genes in equivalent tissues across different organisms; and (e) segments spliced in such species-specific manner are enriched in ID regions and frequently occur in regulatory proteins. Taken together, these studies suggest that diversification of splicing and sequence mutations in ID regions during the course of evolution may underlie the emergence of novel phenotypic traits. These could include both tissue differentiation and differences among the extant organisms [5**,6**,7,13**] (Figure 3). We speculate that similar principles might underlie the growing evidence for population-specific [69–71] or sex-specific traits within a species [72].

Disease
While molecular interactions can evolve through a small number of mutations within ID segments due to the emergence or loss of binding motifs, this property also raises the risk of rewiring interactions that lead to cancer or genetic diseases [73,74] (Figure 3). Indeed, TS exons are significantly enriched in genes associated with cancer and embryonic lethality [13**], thus under-scoring the importance of TS exons in disease development and the associated regulatory pathways. Since TS isoforms likely have distinct molecular interactions, this may explain how the same mutation may lead to different phenotypes across tissue or organ types even when gene expression levels across the tissues are similar. Such tissue-specific AS-dependent gain or loss of specific interactions could lead to selective manifestation of disease in particular organs or tissue types [74–76]. The flood of mutation data from the cancer genome sequencing projects can be better interpreted by bearing in mind, not only which genes are expressed but also which isoforms are expressed in the tissue of interest. Exciting insights are already being provided into how mutations that affect splicing events or cause altered expression of certain isoforms — thus leading to mis-regulation of protein function — can cause diseases, such as cancer, in certain tissue types [77,78].
Conclusion and future directions

Rapidly advancing technology makes it possible to obtain sequences and expression patterns of different isoforms in diverse tissue types, disease states and different developmental stages. Thus, some of the major challenges are to understand how sequence variation and differential inclusion of exons leads to the rewiring of molecular interactions at the genomic and molecular level, and then to predict when these alterations affect phenotype or cause disease. While interpreting the impact of differential inclusion of structured domains is more feasible, understanding the impact of addition or removal of ID segments remains a challenge. Hence, an important direction for future research is to develop a comprehensive characterization and categorization of PTM and molecular interaction sites in ID regions and how the former alters the latter. These improvements will involve development and expansion of databases such as ELM [79], PhosphoELM [80], iELM [81], and improved predictions of sites for PTMs and molecular interactions. This knowledge will provide the molecular basis for interpreting how and when the differential inclusion of disordered segments can contribute to the generation of novel phenotypic traits in molecular terms. Such improved understanding could lead to therapeutic strategies tailored to altering the functions of specific exons, for example by developing small molecules that target AS disordered segments (e.g. via antisense oligonucleotides or targeted delivery of small RNA or small molecules in specific tissues) while avoiding constitutive exons present in all isoforms [58,82–86]. Finally, as sequencing individual human genomes at affordable cost is becoming a reality, a better understanding of how natural variation influences disordered regions, AS patterns and the rewiring of molecular interaction networks will have a significant impact for personalized medicine and for bettering human health.

Acknowledgements

This work was supported by the Medical Research Council (U105185859; M.B., G.C., M.F., S.B., and M.M.B.), the Wellcome Trust (A.B.), HFSP (RGY0073/2010; M.B. and M.M.B.), the EMBO Young Investigator Program (M.M.B.), ERASvsBio+ (GRAPPLE; M.M.B.), the Gates Cambridge Scholarship and the Knox Trinity Studentship (G.C.) and the Momentum program (LP2012-41) of the Hungarian Academy of Sciences (M.F.).

Note added in proof

The reader is referred to two relevant articles that were published after our review was accepted for publication: Van Roey et al., Science Signaling 2013, http://dx.doi.org/10.1126/scisignal.2003345 and Talavera et al., Nature Biotechnology 2013, http://dx.doi.org/10.1038/nbt.2540.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

Rewiring protein interactions by splicing ID regions Buljan et al. 449


34. The authors perform a comprehensive study of the influence of AS and alternative promoter usage on inclusion or skipping of well-annotated short linear-motif interaction modules (SLiMs). They show that different isoforms of the same gene frequently differ in whether or not they contain a particular linear motif. Based on the SLiM’s role in the cell, they describe the effect of differential inclusion of these elements on the protein’s function for a number of cases.


The authors of this study focus on Molecular Recognition Features (MoRFs), which are able to bind different interaction partners. MoRFs are intrinsically disordered regions that undergo disorder-to-order transition after binding. They show that some MoRFs are able to attain different folds upon binding different partners and that AS and PTMs within MoRFs are able to contribute to distinct partner binding.


The authors of this study analyze the human transcriptome data and predict the likely influence of AS on protein’s connectivity. They focus on the AS effects on protein domains and show that in nearly 20% of genes, AS of well-established interacting domains could modify protein’s interaction potential.


In this elegant study, the authors use a large-scale approach for identifying ESC-specific splice events. The authors focus on an ESC-specific exon in the FOXP1 gene. The FOXP1 protein isoform, which is expressed in ESCs differs from the canonical isoform in crucial residues in the DNA binding domain. The authors further demonstrate how this changes the protein’s binding affinity for DNA and results in the distinct pattern of gene regulation, which plays a major role in maintaining stem cell pluripotency and differentiation.


In this work, the authors analyze inhibition modules in auto-inhibited proteins and illustrate how the observed enrichment of these regions in intrinsically disorder confers several advantages. Such regions are more often phosphorylated, alternatively spliced and undergo structure changes. They suggest that these properties allow fine-tuning the inhibitory properties of the segment.


