



Published in final edited form as:

Nat Chem Biol. 2008 December ; 4(12): 728–737. doi:10.1038/nchembio.127.

Malleable machines take shape in eukaryotic transcriptional regulation

Monika Fuxreiter¹, Peter Tompa¹, István Simon¹, Vladimir N Uversky², Jeffrey C Hansen³, and Francisco J Asturias⁴

¹Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Karolina ut 29, H-1113, H-1518 Budapest, Hungary.

²Institute for Intrinsically Disordered Protein Research, Department of Biochemistry and Molecular Biology, Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana 46202-2111, USA.

³Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523-1870, USA.

⁴Department of Cell Biology, The Scripps Research Institute, La Jolla, California 92037, USA.

Abstract

Transcriptional control requires the spatially and temporally coordinated action of many macromolecular complexes. Chromosomal proteins, transcription factors, co-activators and components of the general transcription machinery, including RNA polymerases, often use structurally or stoichiometrically ill-defined regions for interactions that convey regulatory information in processes ranging from chromatin remodeling to mRNA processing. Determining the functional significance of intrinsically disordered protein regions and developing conceptual models of their action will help to illuminate their key role in transcription regulation. Complexes comprising disordered regions often display short recognition elements embedded in flexible and sequentially variable environments that can lead to structural and functional malleability. This provides versatility to recognize multiple targets having different structures, facilitate conformational rearrangements and physically communicate with many partners in response to environmental changes. All these features expand the capacities of ordered complexes and give rise to efficient regulatory mechanisms.

At the critical intersection between signaling pathways and the process of gene expression, transcription regulation controls and influences many cellular and physiological processes, from cell differentiation and development to metabolic responses to environmental stimuli. Transcription regulation depends on communication and interaction between many large multiprotein complexes, which results in transmission of regulatory information to the RNA polymerases that carry out the synthesis of mRNA from a chromosomal DNA template. In eukaryotic organisms, diverse arrays of proteins are involved in every aspect of transcription regulation, from chromatin remodeling to mRNA processing. These include (i) the histones, chromatin and macromolecular complexes that modify chromatin structure and provide access to the DNA, (ii) transcription factors that bind to upstream regulatory sequences, (iii) co-activators that communicate signals to the core transcription machinery and (iv) general

transcription factors that facilitate formation of the pre-initiation complex (Box 1 and Fig. 1).

Early observations about two decades ago indicated that various components of the transcription machinery, such as the transactivator domains (TADs) of the GCN4 and GAL4 transcription factors and the C-terminal domain (CTD) of RNA polymerase II (RNAP II), cannot be characterized by a well-defined three-dimensional structure, and that their functions may even be independent of their actual sequences¹. These domains were envisaged to act as charged ‘blobs’ or ‘noodles’ that would allow recognition of interacting partners of variable architectures and form nontraditional assemblages. Although this scenario has been refined subsequently², it has been recognized only recently that structurally ill-defined proteins and protein segments, termed intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs), are abundant in eukaryotic proteomes. For example, >50% of proteins are predicted to contain at least one IDR that is 30 amino acids in length³. The fraction of proteins with long IDRs increases with the complexity of the organism, which suggests an evolutionary benefit related to protein disorder.

In the past decade, rapidly accumulating experimental evidence has pointed to the frequent occurrence of disordered or unstructured regions in proteins involved in every aspect of transcription, thus providing flexible pieces of the transcriptional puzzle⁴. This review discusses recent advances in the IDP field related to components of the transcription machinery and demonstrates how structural properties and recognition mechanisms of IDPs are ideally suited to implement complex communication pathways among multiprotein regulatory assemblies. A wide range of examples that highlight how IDRs can uniquely enable a combination of specific yet versatile molecular interactions are presented.

IDPs function as a conformational ensemble

A wealth of experimental and theoretical data demonstrate that many proteins or protein segments do not adopt a unique equilibrium structure under native conditions but exist in a rapidly fluctuating ensemble of conformations⁵ that may persist even in the bound form⁶. These results conflict with the generally accepted idea that a well-defined structure is a prerequisite for protein function and have led to a reassessment of the structure-function paradigm⁷. Currently, there are over 500 IDP examples assembled in the DisProt database⁸ whose disordered state is experimentally supported by biophysical data. The relevance of structural disorder *in vivo* has been corroborated experimentally—for example, by in-cell NMR measurements, which indicate the persistence of a disordered state in crowding conditions^{9,10}.

IDPs comprise a variety of broad structural categories¹¹. These range from completely unstructured proteins (“random coils”) that resemble the denatured states of globular proteins (for example, the p160 steroid receptor coactivator ACTR; ref. 12) to partially structured (“pre-molten globules”) or more compact ensembles (“molten globules”)¹³ that may have some secondary structure (for example, p27^{Kip1} (ref. 14) or MeCP2 (ref. 15)). Disorder can also be present in locally disordered N- or C-terminal tails or internal linkers¹¹.

The disordered state of IDPs is intimately linked to their unusual amino acid composition—they are enriched in polar and charged residues (lysine, arginine, glutamate, glutamine and serine) and depleted in hydrophobic residues (tryptophan, phenylalanine, tyrosine, leucine, valine, isoleucine and methionine) that form the hydrophobic core in a conventional folded protein¹⁶. IDPs are often associated with low-complexity regions and include sequence

repeats¹⁷. IDRs often have distinctive, specific amino acid compositions for given functions¹⁸.

The conformational malleability of IDRs allows a diversity of binding modes to other proteins, RNA, DNA or small ligands, resulting in combinatorial molecular recognition and functional versatility¹⁹. IDR-mediated interactions include simple association, chemical modifications, interference with enzymatic activities, structural rearrangements, assistance of folding or even transitions to amyloid forms²⁰. IDPs can be classified into seven general functional categories in terms of their mechanism of action²¹. Within these categories, function stems directly from disorder in only one (entropic chain; for example, nucleoporin FG repeat²²), whereas function is realized via binding to partner molecules in the other six (display sites²³, scavengers²¹, effectors¹⁴, assemblers²⁴, chaperones²⁵ and prions²⁶). Structural disorder is prevalent in proteins involved in signal transduction and regulation of gene expression²⁷, which likely reflects the significance of IDRs in mediating interactions within and between complex molecular assemblies.

Box 1 Basic elements of the eukaryotic transcription machinery

Transcription of protein-encoding genes is one of the first steps in deciphering the genetic material. Fundamentally, it is the synthesis of mRNA from a DNA template, carried out in eukaryotes by the enzyme RNA polymerase II. Transcription is a hierarchical process involving many different macromolecular assemblies regulated by numerous protein-protein and protein-DNA interactions (Fig. 1).

Chromatin and chromatin modifying enzymes function to regulate accessibility of the DNA at the global (that is, multiple-kilobase-pair) level (Fig. 1a). The subunit of chromatin is the nucleosome, which is a complex of 146-base-pair chromosomal DNA with an octamer of core histones. Arrays of nucleosomes spaced at roughly 200-base-pair intervals make up chromatin fibers, which are structurally dynamic and can condense locally and globally into chromosomal domains. There are two major classes of chromatin modifying enzymes: those that add or remove specific post-translational modifications (for example, the Gcn5 acetyltransferase) and those that use the energy of ATP hydrolysis to alter the structure of nucleosomes and specific chromatin regions (for example, SWI/SNF). Chromatin modifying enzymes are large multi component assemblies that have elongated, flexible shapes (Fig. 1a). Collectively, chromatin and chromatin modifying enzymes operate at the epigenetic level and coordinate global accessibility of promoter DNA.

Once the chromatin environment becomes accessible, regulation of the promoter involves the action of regulatory transcription factors, co-activators/co-repressors and the basal transcription machinery. Upstream DNA sequences are targeted by specific DNA binding proteins, called transcription factors (Fig. 1b), which are responsible for controlling (activation/inhibition) specific gene expression. Transcription factors have a modular architecture with a DNA binding domain and a transactivator domain (TAD). TADs communicate with other regulatory transcriptional proteins and have an important role in orchestrating the transcriptional assemblies. Co-activators provide a link between chromatin, transcription factors and the basal transcription machinery (Fig. 1c). Co-activators generally have multiple transcription factor binding sites and are thus able to process multiple transcriptional regulatory inputs. The co-activator p300 is known to interact with over 50 proteins and has histone acetyltransferase activity. Co-activators can also adopt a variety of structural forms needed for different stages of transcription. For example, the Mediator co-activator complex undergoes a large conformational transition from a closed form to an open form to be able to accommodate RNAP II. The actual recruitment of RNAP II to promoter DNA is accomplished by the general transcription

factors (GTFs): TFIIA, TFIIB, TFIID, TFIIE, TFIIIF and TFIIH (Fig. 1d). The TATA DNA element at the promoter is recognized by the TATA box binding protein (TBP), which significantly bends DNA. This TBP-DNA association provides a platform for the assembly of associated factors. The assembly of RNAP II, GTFs and co-activators is called the pre-initiation complex (PIC). The C-terminal domain (CTD) of RNAP II is heavily phosphorylated during transcription (for example, by TFIIH or by the kinase subunit of the Mediator), and its phosphorylation stage has a critical role in pre-mRNA processing and termination of transcription.

Molecular recognition by IDPs

Binding of IDPs in general is characterized by high specificity for given (even multiple) partners as compared to nonspecific targets, and high k_{on} and k_{off} rates that enable rapid association with the partner without an excessive binding strength. The latter results from the large entropic penalty accompanying binding, due to the partial or full folding of the IDR (ref. 28). Currently, key molecular aspects of IDP recognition are captured by four related interaction models, as follows.

Induced folding of IDPs can be facilitated by stabilizing transient structural elements (residual structures), toward which the conformational equilibrium is shifted upon binding²⁹. Experimental data in more than a dozen cases (for example, p27^{Kip1} (ref. 14) and p53 (ref. 30)) support a model in which such transiently populated preformed elements (i) of IDPs are critical for recognition. A similar principle underlies the concept of molecular recognition features (ii)³¹, where protein segments biased for a given conformation are distinguished in recognition and undergo an induced folding process. Even in the absence of transient secondary structures, the accessibility of different segments of the IDP chain can vary along its sequence³². Transient exposure of specific sites can facilitate recognition by serving as primary contact sites (iii) for a partner. As IDPs are predominantly constituted by polar and charged residues, hydrophobic amino acids can be used for this purpose. Short linear motifs (iv) distinguished in specific protein-protein interactions³³ were found to be preferentially embedded in disordered environments³⁴. Linear motifs are frequently associated with signaling processes, as favored target sequences of SH3 domains, 14-3-3 proteins and calmodulin, as well as sites for post-translational modification, such as phosphorylation³⁵, acetylation or ubiquitination³⁶, are found on these motifs.

All four concepts emphasize the segmental nature of IDPs, which has three direct consequences for recognition. First, it increases the association rates by anchoring only a small, exposed binding region, with further contacts facilitated by flexible linkers via a “fly-casting mechanism”³⁷. Second, as recognition is primarily determined by short segments, part or even the majority of the protein may remain disordered in the bound form, resulting in structurally ambiguous, ‘fuzzy’ complexes⁶. Third, recognition by short motifs can be achieved by a very few specificity-determining residues, which leaves the rest of the motif, and the entire flanking region, rather free to change in sequence³⁴.

These features of IDPs and IDRs allow them to perform unorthodox functions that require extreme adaptability (Fig. 2). For example, they can accommodate more than one partner (promiscuity) (Fig. 2a), like the C-terminal region of p53, which adopts different conformations upon interacting with four different partners (S100 β , sirtuin, CBP and cyclin A; Fig. 2b)³⁸. When bound in different conformations, IDPs can serve more than one (sometimes opposing) function (moonlighting)³⁹. IDPs can also be involved in multiple activities even at the same time (one-to-many signaling), as exemplified by the high mobility group (HMGA) proteins that serve as hubs of nuclear function and sensors of a wide variety of signaling pathways⁴⁰. Action of IDPs can also be fine-tuned by gradual

modifications (ultra-sensitivity), for example, phosphorylation at multiple sites that individually are suboptimal for the partner⁴¹.

IDPs within the transcriptional machinery

Bioinformatics analysis predicts that most proteins involved in transcription regulation contain at least one long (>30 amino acids) intrinsically disordered segment (with many predicted to be far more disordered), and biophysical evidence to substantiate a number of specific instances is available. A summary of well-characterized examples of IDPs involved in transcriptional regulation is presented in Table 1.

In general, IDRs and IDPs can contribute to the functioning of the transcription machinery (Fig. 1) in three ways. First, IDRs can serve as linkers connecting structured globular domains and thus extend modularity of function. Fluctuation of IDRs between widely differing conformational states imparts structural variability on the connected recognition elements or folded segments, thereby enabling recognition of partners with variable structure, such as recognition of promoters with variable organization by transcription factors⁴². Second, molecular associations involving IDRs result in pliable complexes that can undergo conformational changes in response to regulatory signals. For example, the Mediator co-activator complex transforms from a closed to an open form to accommodate RNAP II (ref. 43). Third, IDRs can also facilitate assembly or disassembly of large complexes coupled to functional transitions—for example, interaction of the RNAP II CTD with different partners during a transcription cycle⁴⁴. In the following sections, we will describe how these principles are exploited in the action of four major components of the transcription machinery (Box 1): proteins in chromatin remodeling, transcription factors, co-activators and the basal machinery (Fig. 1).

IDPs and chromatin structure

Genomic DNA in eukaryotes is complexed with core and linker histones to form nucleosomes and chromatin fibers⁴⁵, with most genes also assembled into higher order chromatin structures (Fig. 1a). The higher order condensation and conformational dynamics of chromatin fibers are mediated by the disordered core histone N-terminal ‘tails’. These IDRs are not visible in the crystal structure of the nucleosome⁴⁶ and have also been shown to be disordered in solution⁴⁷. In addition to chromatin condensation, a primary role of the disordered N-terminal tails is to specifically interact with nonhistone proteins and multiprotein complexes involved in gene activation or repression⁴⁸. For example, the N-terminal domain (NTD) of the core histone H4 functions as an assembly platform for different chromatin remodeling complexes, such as ISWI (ref. 49) and NURF (ref. 50). The H4 NTD also helps target ISWI to specific loci⁵¹. The many interactions involving the N-terminal tails are likely of transient nature and require rapid association and dissociation with their partners. Although both core and linker histone tails are predominantly (~40%) composed of positively charged residues, their functions cannot be recapitulated by simply neutralizing DNA charges at high ionic strength⁴⁵. These observations might imply that local organization of IDRs can be a functional determinant for histone proteins that can be perturbed by post-translational modifications (for example, methylation and acetylation)⁵². These covalent modifications have been proposed to function as a “histone code” that influences and regulates specific macromolecular interactions and cell functions^{48,53}. Although in most cases the structural consequences of these covalent modifications have not been elucidated, they might alter secondary structure preferences of IDRs, which could abolish nucleosome-nucleosome interactions or modulate interactions with chromatin remodeling complexes⁴⁹ and other chromatin proteins.

The intrinsically disordered C-terminal tail of linker histones undergoes a disorder-to-helix transition upon binding to DNA (ref. 54) and contributes to stabilization of condensed chromatin structures. The linker histone H1 °CTD also binds to and activates the DFF40 (also known as CAD) apoptotic nuclease. Interestingly, different CTD regions are used for chromatin condensation and nuclease binding. These H1 °CTD regions have distinct, highly conserved amino acid compositions¹⁸. Accordingly, scrambling the sequence of these segments does not alter CTD-dependent stabilization of higher order chromatin structures (X. Lu and J.C. Hansen, Colorado State University, personal communication). Although possible changes in the interaction and activation mechanism have not been elucidated, the IDR had to be sufficiently long (at least 47 residues). These results emphasize that there is no specific sequence requirement for the IDRs—only some spatial and electrostatic conditions have to be fulfilled.

IDPs also seem to have important roles in the function of ATP-dependent chromatin remodeling complexes, such as ISWI, CHRAC, NURF, RSC and SWI/SNF, which regulate accessibility of the genomic DNA (Fig. 1a). All these complexes facilitate sliding and transfer of histone octamers along the DNA and expose targeted DNA segments to nucleases and other probes. Gel mobility analysis suggests that most of these remodeling complexes exhibit significant deviations from globularity⁴⁹·⁵⁵–⁵⁸, as they migrate at higher molecular weight than that expected from the mass of the complex. IDRs in chromatin remodeling complexes likely mediate low-affinity interactions with DNA and establish variable contact patterns required for sliding. Thus disordered regions enhance mobility along DNA (ref. 55), and their removal significantly impairs the sliding process⁵⁰·⁵⁵·⁵⁶. Structural characterization by electron microscopy of the yeast SWI/SNF remodeler revealed a structure composed of eight subunits assembled into a modular and highly irregular structure⁵⁷ (Fig. 1). Gel mobility analysis and disorder predictions also suggest that the Snf5 and Swi3 subunits of the SWI/SNF remodeling complex are rich in IDRs. Removal of the N-terminal domain of Swi3 (predicted to be poorly ordered) results in the remainder of the protein migrating at its actual molecular weight (C. Peterson, University of Massachusetts Medical School, personal communication). Swi3 serves as an assembly scaffold and is involved in histone binding using IDRs to interact with multiple partners. Snf5 has a role in recruitment of SWI/SNF to specific genomic regions—another process with coordinated changes in macromolecular interactions. We hypothesize that IDPs allow malleability in the structure of chromatin remodeling and modifying complexes, which facilitates their interaction with the equally structurally malleable chromatin fiber.

Modification of chromatin structure may also be facilitated by increased DNA distortion induced by “architectural transcription factors”⁴⁰. For example, the disordered high-mobility group (HMG) proteins involved in chromatin structure regulation contain multiple copies of short motifs (called AT hooks) that tend to bind to the minor groove and significantly bend the DNA (ref. 59). The AT hooks are connected by flexible linkers that result in diverse modular binding patterns that neutralize charges and induce specific conformational changes of DNA. Besides DNA, HMG proteins can also interact with nucleosome particles and a large number of transcription factors, inducing formation of assemblies (enhanceosomes) at promoter or enhancer regions of inducible genes⁴⁰. These interactions also facilitate contacts among the regulatory proteins and cross-associations with DNA that may result in complex regulatory pathways. As observed for the histone tails, HMG function can be regulated by modulating binding affinity to both DNA and nucleosomes via post-translational modifications such as phosphorylation and methylation, which are facilitated by structural disorder³⁵.

A range of recent observations underscores that a variety of largely disordered proteins are involved in regulation of higher level chromatin structural organization (Table 1), the multifunctionality of which is enabled by structural disorder.

Transcription factors: disordered transactivator domains

Transcription factors serve as sensors of specific DNA sequences that code for activation or repression of transcription. This is facilitated by a modular architecture in which DNA binding domains (DBDs) contact enhancer or suppressor sequences and transactivator domains (TADs) interact with other proteins that influence recruitment and assembly of the transcription machinery (Fig. 1b). Peculiar structural properties of TADs and their resistance to crystallization attempts have long been noted. Bioinformatics analysis reveals that ~90% of transcription factors contain extended IDRs⁶⁰, out of which a few specific examples are listed in Table 1. Perhaps as a result of disobeying strict geometrical complementarity, TADs might also exhibit weak sequence specificity. The classical example is GCN4, in which the acidic TAD can be replaced with short random acidic segments without significant loss of activity⁶¹. It is not known, however, whether these scrambled sequences act by the same mechanism and whether their interactions with other co-activators remain unperturbed. For example, the disordered C-terminal region of the yeast activator Gal4 has dual functions: it elicits gene activation and is also capable of binding to Gal80. Mutations of cysteine or proline residues abolish interactions with Gal80 without affecting transcriptional activity, which suggests that different functions have different sequence and structural requirements². It has also been shown that short hydrophobic motifs that might provide an underlying “organization” appear to be critical for TAD activity⁶²⁻⁶³. For example, the repeat units of the transactivator domain of the oncogenic EFP proteins were found to function even when randomized, reversed or interchanged, but the presence of a given arrangement of tyrosine residues⁶⁴ was found to be an essential requirement. These hydrophobic residues may serve as primary contact sites, for example, as seen in the folding-coupled binding of the KID domain of cAMP response element-binding (CREB) protein to the KIX domain of CREB-binding protein (CBP) induced by phosphorylation of KID (ref. 65). Hydrophobic patches can also act like molecular recognition features that trigger the conversion from the initial, low-affinity complex to the high-affinity complex⁶⁶. Preformed elements of TADs can also facilitate binding their interactions, as in the case of the α -helical element in binding of p53 to Mdm2 (ref. 30).

Post-translational modification sites are frequently located in disordered regions, perhaps enabling them to function as molecular switches within TADs. A prime example is ubiquitination, with its far-reaching functional consequences. Ubiquitination of a TAD not only induces degradation of the transcription factor but also signals for activation (a process called “licensing”)²³. Thus, disorder of TADs may be involved in a very specific regulatory feature in which ubiquitination induces activation through destruction of the targeted protein.

In contrast with TADs, DBDs are mostly structured. However, certain transcription factors, such as GCN4 or GAGA, undergo an induced folding process upon binding to DNA, with recognition facilitated by preformed secondary structure elements⁶⁷. Such disorder-to-order transitions have been shown to enhance interaction specificity⁶⁸.

Co-activators: linking chromatin structure to the core machinery

Highly complex, modular proteins and multiprotein complexes are responsible for communicating signals from enhancer- and repressor-bound factors to the core transcription machinery (Fig. 1c). In eukaryotic cells, co-activators are involved in multiple functions, ranging from modifying chromatin structure to interacting with a variety of regulatory

proteins, general transcription factors and RNAP II. IDPs appear to perform many critical functions that enable co-activators to transduce regulatory information.

The co-activators CBP/p300 affect chromatin structure through their intrinsic histone acetyltransferase activity, and can also serve as a scaffold for the assembly of much of the transcription machinery²⁴. Approximately half of its 2,442 residues are found in disordered regions, including the NCBR domain and linkers between six folded domains¹¹. The six globular domains serve as templates for induced folding of many disordered transcription factors—for example, TAZ1 for HIF1- α (ref. 69) and CITED2 (ref. 70), Bromo domain for p53 CTD (ref. 71), TAZ2 for E1A (ref. 11) and KIX for the pKID domain of CREB (ref. 72) (Fig. 3). The disordered NCBR domain undergoes a synergistic folding upon binding to the ACTR domain of the p160 co-activator that is facilitated by the presence of transient helical structures in NCBR (ref. 12). These interactions contribute to the recruitment of RNAP II and the basal transcription machinery. In addition to enabling conformational flexibility, disordered linker regions harbor various linear motifs that serve as potential binding sites for regulatory proteins (for example, KHKXLXXLL for nuclear receptors⁷³) or that serve as post-translational modification sites (for example, SUMOylation sites⁷⁴). IDRs have a critical role in determining the properties of CBP and p300, which have been described as a “molecular interpreter” of different combinations of transcriptional signals owing to their highly adaptable architectures.

Coordination between chromatin remodeling and the next step in the gene expression process, assembly of the pre-initiation complex (PIC), might occur through the interplay between CBP/p300 and Mediator⁷⁵, a multisubunit assembly comprising about 25 different proteins that plays an essential role in regulation from yeast to humans⁷⁶. The two co-activators act synergistically, but p300 competes with the general transcription factor TFIID for binding to Mediator at the promoter. Autoacetylation of p300 in a disordered loop region provides a catalytic switch⁷⁷ by initiating a conformational change that results in dissociation of p300 from the promoter. This process allows the association of TFIID with Mediator, triggering assembly of the PIC. Biochemical⁷⁸ and structural⁷⁹ analyses have revealed that Mediator has a modular architecture, with three structural domains that in broad terms appear to have specific functional roles. Significant conformational changes are prompted by interaction of the human Mediator complex with activators and the RNAP II CTD (refs. 80·81). In yeast, interaction of Mediator with RNAP II to form a regulation-responsive holoenzyme requires a substantial structural rearrangement of Mediator that exposes a cryptic RNAP II binding site⁴³.

It has been suggested that the rearrangement of Mediator required for holoenzyme formation might constitute a general and essential aspect of the regulatory mechanism⁸², and may also be required in other organisms^{81·83}. It seems likely that the complex conformational behavior exhibited by the Mediator complex might be facilitated by disordered regions acting as “molecular hinges” in response to various regulatory signals. Several Mediator subunits were predicted to contain a significant level of disorder, and the identified IDRs were found to be conserved, which supports this molecular hinge hypothesis⁸⁴. Although structural data of the whole Mediator is limited to low-resolution reconstructed electron microscopy images, the presence of disordered regions was experimentally corroborated in the case of the Med8-Med18-Med20 submodule, which contains multiple binding sites for the TATA box binding protein (TBP) complex⁸⁵. In the crystal structure, only a short α -helical region of Med8 has been observed⁸⁶, whereas the linker between the C- and N-terminal regions of Med8 exhibits enhanced sensitivity to proteolytic digestion in the free protein. Thus it appears that the short structured segment of Med8 serves as a molecular recognition feature to promote recognition of a larger disordered region. The activities of

different Mediator subunits are also substantially influenced by post-translational modifications⁸⁷ facilitated by intrinsic disorder.

Core machinery: hinges and wobbling parts

In eukaryotes, the series of molecular events that leads to gene expression culminates in initiation of transcription by RNA polymerase and the general transcription factors (GTFs) (Fig. 1d). RNA polymerases are for the most part well-structured enzymes, although flexibility of specific domains is of functional significance. In the eukaryotic RNAP II, the CTD of the largest subunit is highly disordered and contains 25–52 tandem repeats of the sequence YSPTSPS (ref. 88). The CTD serves as a scaffold for the highly orchestrated assembly of a range of complexes involved in the initiation, elongation and termination of transcription, thereby linking these steps directly to mRNA maturation⁸⁹. Owing to its malleability, the CTD is capable of interacting with multiple partners and adopting different conformations upon binding⁹⁰. Its function is tightly regulated by phosphorylation, and the concerted action of a variety of CTD kinases and phosphatases signals its transitions from states competent for PIC assembly to initiation, elongation and then termination of transcription⁸⁹. Phosphorylation of the CTD results in significant changes in its structure and charge pattern, but CTD activity depends only weakly on its actual amino acid sequence.

Disordered regions also influence the functions of the GTFs in the assembly of the pre-initiation complex and in interactions with promoters and co-activators. In TFIIB (Fig. 2), a globular CTD that contacts the TBP at the TATA box is connected to the N-terminal RNAP II–interacting region by a linker region that is disordered in solution. Part of this linker folds into a “B finger” upon interacting with RNAP II, reaching into the active site of the enzyme and playing a crucial role in determining the transcription start site⁹¹. The conformational malleability of the B finger segment is likely required to facilitate entrance of the domain into the active site cleft of polymerase. The rest of the linker remains disordered even in the presence of RNAP II and might allow for different modes of interaction between the C-terminal portion of TFIIB and polymerase⁹². Cryo-electron microscopy analysis of a yeast RNAP II–TFIIF complex has revealed that TFIIF exhibits scattered density around the RNAP II active site cleft, with a considerable fraction of the factor appearing disordered⁹³. This suggests a structural organization of TFIIF that, in agreement with predictions based on sequence analysis, consists of globular domains connected by disordered segments. The central region of Tfg1, the largest yeast TFIIF subunit, is highly charged and hypersensitive to proteolysis, which also suggests a poorly ordered structure⁹⁴. Interestingly, this part of TFIIF is expected to be largely exposed in the RNAP II–TFIIF complex, and TFIIF has been shown to play a critical role in promoting association of Mediator with polymerase and the PIC (ref. 85). Although structural characterization of the basal transcription machinery is still at a very early stage, the information available so far suggests that flexibility and disorder are prevalent and likely play a crucial role in enabling RNA polymerase to interact with promoters that display a wide range of sequence and structural variability. It is tempting to suggest that solutions that have evolved to address promoter heterogeneity are likely to also be involved in regulation.

Concluding remarks

The early idea that transcription regulation requires malleable architectures rather than rigid assemblages is bolstered by a growing body of experimental evidence. Assembly of the transcription machinery involves a complex set of highly specific, transient interactions with multiple partners that often have variable architecture. In principle, the controversial requirements for specificity and structural adaptability in these interactions can only be

satisfied if short recognition sites are embedded in a malleable environment capable of accommodating different structures. IDRs that function through short recognition elements or act as linkers connecting globular domains fit naturally in this context. As shown for co-activators, IDR linkers can provide enormous structural flexibility and modularity to the complex. As IDRs can interact with many partners simultaneously, they can help in the assembly and disassembly of large complexes in response to the needs of the cell (for example, Mediator). Within this framework, a network of highly specific yet transient contacts between disordered regions enables fast information flow and rapid responses to diverse regulatory signals by enabling large-scale conformational rearrangements of the transcription machinery. All these events take place against a background of chromatin structural transitions that are mediated by the intrinsically disordered histone tail domains and regulated by structurally complex chromatin modifying enzymes. At least one likely reason why there is so much intrinsic disorder in transcriptional proteins is because of the structural and functional flexibility needed to precisely coordinate so many transient, high-specificity interactions.

In regards to possible evolutionary consequences, this model can reconcile the opposing demands of conservation of regulatory mechanisms and an ease of adaptation to environmental changes. Large disordered regions that host short functional sites are more tolerant to mutations than globular proteins, thereby exhibiting ‘robustness’ in both structural and functional senses^{18,61,64}. Indeed, the rapid evolution of IDPs has been demonstrated in a number of cases^{17,95}. On the other hand, short motifs embedded in disordered regions can provide specific interaction sites with other proteins or serve as post-translational modification sites, yet they are easily turned on and off by evolution³³, which lends the capacity to rapidly create or abolish specific functionality. In summary, the benefits of the structural flexibility, adaptability and evolvability provided by protein disorder are apparent at all levels of gene transcription regulation.

Future implications

In this review we demonstrated that IDPs and IDRs are involved in all stages of the transcription process and appear to be critically important in regulatory functions. We are just at the beginning, however, of understanding the molecular mechanisms that underlie the many functions of IDPs, and their contribution to transcription regulation. Our limitations are two-fold. High-resolution structural imaging of large, flexible and transient assemblies is a major bottleneck in understanding the transcriptional machinery. In addition, the paucity of structural data on IDPs and IDRs *in vivo* curbs extrapolations from *in vitro* observations to conditions within the cell. Whereas these limitations restrain our understanding of the role of disorder in transcription regulation, the major benefit of the recognition of the importance of disorder is that it suggests new approaches toward transcription regulation that include the detailed structural and functional description of the IDRs.

Currently there are two major areas where characterization of the structural ensemble of IDRs and IDPs could significantly improve our understanding of transcription. One is the mechanisms of huge assemblages, such as chromatin remodeling machines or co-activator complexes (for example, Mediator). Anomalous gel electrophoretic mobility or cryo-electron microscopy could be used to detect the presence of disordered regions that can trigger conformational changes. Biochemical experiments, such as limited proteolysis at very low protease concentrations, can identify transiently exposed segments of the IDRs. These regions usually serve as interaction sites that contribute to self-organization of the complexes or mediate interactions with other partners. *In vivo* or *in vitro* functional assays can validate the functional roles of these regions. Bioinformatics methods are extremely useful to complement experimental data to pinpoint IDRs and locate possible functional

sites (for example, for post-translational modifications) and initiate targeted studies. Further development of these tools may allow estimation of the degree of folding that is induced upon binding to a partner or a possible architecture of a large assembly.

The other area is the mechanisms of nonconventional interactions, such as those that exhibit weak dependence on the actual sequence of one of the participating macromolecules or regions. A pertinent example of such a binding event is that of histone tails to DNA or regulatory proteins. Analyzing changes in disorder upon sequence modifications by NMR or CD (and also by bioinformatics methods) may help to identify those properties that are required for the interactions. Furthermore, these methods can be useful to understand the effect of epigenetic factors, such as post-translational modifications of histone tails. Recognition of disorder (and low-complexity regions) may also provide possible connections between elements of the transcription machineries of different organisms, which could in turn help to illuminate their function.

Although much remains to be learned about the function and behavior of IDPs in transcription regulation, recognition of their prevalence and importance in controlling gene expression constitutes a new perspective that opens up new research avenues in the study of transcription regulation.

Acknowledgments

M.F. is supported by MRTN-CT-2005-019566 of the European FP6 and by the Bolyai János fellowship; P.T. is supported by Országos Tudományos Kutatási Alapprogramok NK 71582, ETT 245/2006; and I.S. is supported by Országos Tudományos Kutatási Alapprogramok K72569. V.N.U. is supported by the grants R01 LM007688-01A1 (to A.K. Dunker and V.N.U.) and GM071714-01A2 (to A.K. Dunker and V.N.U.) from the US National Institutes of Health. We gratefully acknowledge the support of the Indiana University, Purdue University at Indianapolis Signature Centers Initiative. J.C.H. is supported by US National Institutes of Health grants GM45916 and GM66834. F.J.A. is supported by funds from the US National Institute of General Medical Sciences and the Leukemia and Lymphoma Society of America. The authors are grateful to C. Oldfield for his contribution to Figure 2.

References

1. Sigler PB. Transcriptional activation. Acid blobs and negative noodles. *Nature* 1988;333:210–212. [PubMed: 3367995]
2. Ansari AZ, Reece RJ, Ptashne M. A transcriptional activating region with two contrasting modes of protein interaction. *Proc. Natl. Acad. Sci. USA* 1998;95:13543–13548. [PubMed: 9811836]
3. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. *Genome Inform. Ser. Workshop Genome Inform* 2000;11:161–171.
4. Asturias FJ. Another piece in the transcription initiation puzzle. *Nat. Struct. Mol. Biol* 2004;11:1031–1033. [PubMed: 15523475]
5. Romero P, et al. Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput* 1998;3:437–448. [PubMed: 9697202]
6. Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci* 2008;33:2–8. [PubMed: 18054235]
7. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol* 1999;293:321–331. [PubMed: 10550212]
8. Sickmeier M, et al. DisProt: the database of disordered proteins. *Nucleic Acids Res* 2007;35:D786–D793. [PubMed: 17145717]
9. Flaugh SL, Lumb KJ. Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27(Kip1). *Biomacromolecules* 2001;2:538–540. [PubMed: 11749217]
10. McNulty BC, Young GB, Pielak GJ. Macromolecular crowding in the *Escherichia coli* periplasm maintains alpha-synuclein disorder. *J. Mol. Biol* 2006;355:893–897. [PubMed: 16343531]

11. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol* 2005;6:197–208. [PubMed: 15738986]
12. Demarest SJ, et al. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 2002;415:549–553. [PubMed: 11823864]
13. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 2002;11:739–756. [PubMed: 11910019]
14. Lacy ER, et al. p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat. Struct. Mol. Biol* 2004;11:358–364. [PubMed: 15024385]
15. Adams VH, McBryant SJ, Wade PA, Woodcock CL, Hansen JC. Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J. Biol. Chem* 2007;282:15057–15064. [PubMed: 17371874]
16. Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 2000;41:415–427. [PubMed: 11025552]
17. Tompa P. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays* 2003;25:847–855. [PubMed: 12938174]
18. Hansen JC, Lu X, Ross ED, Woody RW. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem* 2006;281:1853–1856. [PubMed: 16301309]
19. Dunker AK, et al. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac. Symp. Biocomput* 1998;3:473–484. [PubMed: 9697205]
20. Tompa P. Intrinsically unstructured proteins. *Trends Biochem. Sci* 2002;27:527–533. [PubMed: 12368089]
21. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* 2005;579:3346–3354. [PubMed: 15943980]
22. Denning DP, Patel SS, Uversky V, Fink AL, Rexach M. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl. Acad. Sci. USA* 2003;100:2450–2455. [PubMed: 12604785]
23. Salghetti SE, Caudy AA, Chenoweth JG, Tansey WP. Regulation of transcriptional activation domain function by ubiquitin. *Science* 2001;293:1651–1653. [PubMed: 11463878]
24. Goodman RH, Smolik S. CBP/p300 in cell growth, transformation, and development. *Genes Dev* 2000;14:1553–1577. [PubMed: 10887150]
25. Tompa P, Csermely P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J* 2004;18:1169–1175. [PubMed: 15284216]
26. Pierce MM, Baxa U, Steven AC, Bax A, Wickner RB. Is the prion domain of soluble Ure2p unstructured? *Biochemistry* 2005;44:321–328. [PubMed: 15628874]
27. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit* 2005;18:343–384. [PubMed: 16094605]
28. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol* 2002;12:54–60. [PubMed: 11839490]
29. Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol* 2004;338:1015–1026. [PubMed: 15111064]
30. Lee H, et al. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J. Biol. Chem* 2000;275:29426–29432. [PubMed: 10884388]
31. Oldfield CJ, et al. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 2005;44:12454–12470. [PubMed: 16156658]
32. Csizmok V, et al. Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. *Biochemistry* 2005;44:3955–3964. [PubMed: 15751971]
33. Neduva V, Russell RB. Linear motifs: evolutionary interaction switches. *FEBS Lett* 2005;579:3342–3345. [PubMed: 15943979]
34. Fuxreiter M, Tompa P, Simon I. Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* 2007;23:950–956. [PubMed: 17387114]

35. Iakoucheva LM, et al. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 2004;32:1037–1049. [PubMed: 14960716]
36. Cox CJ, et al. The regions of securin and cyclin B proteins recognized by the ubiquitination machinery are natively unfolded. *FEBS Lett* 2002;527:303–308. [PubMed: 12220679]
37. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. USA* 2000;97:8868–8873. [PubMed: 10908673]
38. Oldfield CJ, et al. Flexible nets: disorder and induced fit in the associations of p53 and 14–3-3 with their partners. *BMC Genomics* 2008;9 suppl. 1:S1.
39. Tompa P, Szasz C, Buday L. Structural disorder throws new light on moonlighting. *Trends Biochem. Sci* 2005;30:484–489. [PubMed: 16054818]
40. Reeves R, Beckerbauer L. HMGI/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim. Biophys. Acta* 2001;1519:13–29. [PubMed: 11406267]
41. Nash P, et al. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* 2001;414:514–521. [PubMed: 11734846]
42. Tjian R, Maniatis T. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 1994;77:5–8. [PubMed: 8156597]
43. Davis JA, Takagi Y, Kornberg RD, Asturias FA. Structure of the yeast RNA polymerase II holoenzyme: mediator conformation and polymerase interaction. *Mol. Cell* 2002;10:409–415. [PubMed: 12191485]
44. Kostek SA, et al. Molecular architecture and conformational flexibility of human RNA polymerase II. *Structure* 2006;14:1691–1700. [PubMed: 17098194]
45. Hansen JC. Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms and functions. *Annu. Rev. Biophys. Biomol. Struct* 2002;31:361–392. [PubMed: 11988475]
46. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 1997;389:251–260. [PubMed: 9305837]
47. Hansen JC, Tse C, Wolffe AP. Structure and function of the core histone N-termini: more than meets the eye. *Biochemistry* 1998;37:17637–17641. [PubMed: 9922128]
48. Ruthenburg AJ, Li H, Patel DJ, Allis CD. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol* 2007;8:983–994. [PubMed: 18037899]
49. Clapier CR, Langst G, Corona DF, Becker PB, Nightingale KP. Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI. *Mol. Cell. Biol* 2001;21:875–883. [PubMed: 11154274]
50. Xiao H, et al. Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* 2001;8:531–543. [PubMed: 11583616]
51. Fazio TG, Gelbart ME, Tsukiyama T. Two distinct mechanisms of chromatin interaction by the Isw2 chromatin remodeling complex in vivo. *Mol. Cell. Biol* 2005;25:9165–9174. [PubMed: 16227570]
52. Ferreira H, Flaus A, Owen-Hughes T. Histone modifications influence the action of Snf2 family remodelling enzymes by different mechanisms. *J. Mol. Biol* 2007;374:563–579. [PubMed: 17949749]
53. Jenuwein T, Allis CD. Translating the histone code. *Science* 2001;293:1074–1080. [PubMed: 11498575]
54. Baneres JL, Martin A, Parello J. The N tails of histones H3 and H4 adopt a highly structured conformation in the nucleosome. *J. Mol. Biol* 1997;273:503–508. [PubMed: 9356240]
55. Hartlepp KF, et al. The histone fold subunits of *Drosophila* CHRAC facilitate nucleosome sliding through dynamic DNA interactions. *Mol. Cell. Biol* 2005;25:9886–9896. [PubMed: 16260604]
56. Hamiche A, Kang JG, Dennis C, Xiao H, Wu C. Histone tails modulate nucleosome mobility and regulate ATP-dependent nucleosome sliding by NURF. *Proc. Natl. Acad. Sci. USA* 2001;98:14316–14321. [PubMed: 11724935]

57. Smith CL, Horowitz-Scherer R, Flanagan JF, Woodcock CL, Peterson CL. Structural analysis of the yeast SWI/SNF chromatin remodeling complex. *Nat. Struct. Biol* 2003;10:141–145. [PubMed: 12524530]
58. Wu PY, Ruhlmann C, Winston F, Schultz P. Molecular architecture of the *S. cerevisiae* SAGA complex. *Mol. Cell* 2004;15:199–208. [PubMed: 15260971]
59. Huth JR, et al. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol* 1997;4:657–665. [PubMed: 9253416]
60. Liu J, et al. Intrinsic disorder in transcription factors. *Biochemistry* 2006;45:6873–6888. [PubMed: 16734424]
61. Hope IA, Mahadevan S, Struhl K. Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. *Nature* 1988;333:635–640. [PubMed: 3287180]
62. Cress WD, Triezenberg SJ. Critical structural elements of the VP16 transcriptional activation domain. *Science* 1991;251:87–90. [PubMed: 1846049]
63. Drysdale CM, et al. The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. *Mol. Cell. Biol* 1995;15:1220–1233. [PubMed: 7862116]
64. Ng KP, et al. Multiple aromatic side chains within a disordered structure are critical for transcription and transforming activity of EWS family oncoproteins. *Proc. Natl. Acad. Sci. USA* 2007;104:479–484. [PubMed: 17202261]
65. Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 2007;447:1021–1025. [PubMed: 17522630]
66. Ferreira ME, et al. Mechanism of transcription factor recruitment by acidic activators. *J. Biol. Chem* 2005;280:21779–21784. [PubMed: 15826952]
67. Zitzewitz JA, Ibarra-Molero B, Fishel DR, Terry KL, Matthews CR. Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J. Mol. Biol* 2000;296:1105–1116. [PubMed: 10686107]
68. Spolar RS, Record MT Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science* 1994;263:777–784. comment 263, 769-770 (1994). [PubMed: 8303294]
69. Dames SA, Martinez-Yamout M, De Guzman RN, Dyson HJ, Wright PE. Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response. *Proc. Natl. Acad. Sci. USA* 2002;99:5271–5276. [PubMed: 11959977]
70. De Guzman RN, Martinez-Yamout MA, Dyson HJ, Wright PE. Interaction of the TAZ1 domain of the CREB-binding protein with the activation domain of CITED2: regulation by competition between intrinsically unstructured ligands for non-identical binding sites. *J. Biol. Chem* 2004;279:3042–3049. [PubMed: 14594809]
71. Mujtaba S, et al. Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol. Cell* 2004;13:251–263. [PubMed: 14759370]
72. Radhakrishnan I, et al. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* 1997;91:741–752. [PubMed: 9413984]
73. Heery DM, Kalkhoven E, Hoare S, Parker MG. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 1997;387:733–736. [PubMed: 9192902]
74. Girdwood D, et al. P300 transcriptional repression is mediated by SUMO modification. *Mol. Cell* 2003;11:1043–1054. [PubMed: 12718889]
75. Black JC, Choi JE, Lombardo SR, Carey M. A mechanism for coordinating chromatin modification and preinitiation complex assembly. *Mol. Cell* 2006;23:809–818. [PubMed: 16973433]
76. Kornberg RD. Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci* 2005;30:235–239. [PubMed: 15896740]
77. Thompson PR, et al. Regulation of the p300 HAT domain via a novel activation loop. *Nat. Struct. Mol. Biol* 2004;11:308–315. [PubMed: 15004546]

78. Myers LC, Gustafsson CM, Hayashibara KC, Brown PO, Kornberg RD. Mediator protein mutations that selectively abolish activated transcription. *Proc. Natl. Acad. Sci. USA* 1999;96:67–72. [PubMed: 9874773]
79. Asturias FJ, Jiang YW, Myers LC, Gustafsson CM, Kornberg RD. Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* 1999;283:985–987. [PubMed: 9974391]
80. Taatjes DJ, Naar AM, Andel F III, Nogales E, Tjian R. Structure, function, and activator-induced conformations of the CRSP coactivator. *Science* 2002;295:1058–1062. [PubMed: 11834832]
81. Taatjes DJ, Schneider-Poetsch T, Tjian R. Distinct conformational states of nuclear receptor-bound CRSP-Med complexes. *Nat. Struct. Mol. Biol* 2004;11:664–671. [PubMed: 15195149]
82. Chadick JZ, Asturias FJ. Structure of eukaryotic Mediator complexes. *Trends Biochem. Sci* 2005;30:264–271. [PubMed: 15896745]
83. Elmlund H, et al. The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. *Proc. Natl. Acad. Sci. USA* 2006;103:15788–15793. [PubMed: 17043218]
84. Tóth-Petróczy A, et al. Malleable machines in transcription regulation: the Mediator complex. *PLoS Comput. Biol.* (in the press).
85. Takagi Y, et al. Head module control of mediator interactions. *Mol. Cell* 2006;23:355–364. [PubMed: 16885025]
86. Lariviere L, et al. Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nat. Struct. Mol. Biol* 2006;13:895–901. [PubMed: 16964259]
87. van de Peppel J, et al. Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol. Cell* 2005;19:511–522. [PubMed: 16109375]
88. Corden JL. Tails of RNA polymerase II. *Trends Biochem. Sci* 1990;15:383–387. [PubMed: 2251729]
89. Proudfoot NJ, Furger A, Dye MJ. Integrating mRNA processing with transcription. *Cell* 2002;108:501–512. [PubMed: 11909521]
90. Fabrega C, Shen V, Shuman S, Lima CD. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol. Cell* 2003;11:1549–1561. [PubMed: 12820968]
91. Bushnell DA, Westover KD, Davis RE, Kornberg RD. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Å. *Science* 2004;303:983–988. [PubMed: 14963322]
92. Chen HT, Hahn S. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* 2004;119:169–180. [PubMed: 15479635]
93. Chung WH, et al. RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. *Mol. Cell* 2003;12:1003–1013. [PubMed: 14580350]
94. Yong C, et al. Structure of the human transcription factor TFIIF revealed by limited proteolysis with trypsin. *FEBS Lett* 1998;435:191–194. [PubMed: 9762906]
95. Brown CJ, et al. Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol* 2002;55:104–110. [PubMed: 12165847]
96. Zor T, De Guzman RN, Dyson HJ, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J. Mol. Biol* 2004;337:521–534. [PubMed: 15019774]
97. Lu X, Hansen JC. Identification of specific functional subdomains within the linker histone H10 C-terminal domain. *J. Biol. Chem* 2004;279:8701–8707. [PubMed: 14668337]
98. McBryant SJ, Krause C, Hansen JC. Domain organization and quaternary structure of the *Saccharomyces cerevisiae* silent information regulator 3 protein, Sir3p. *Biochemistry* 2006;45:15941–15948. [PubMed: 17176117]
99. Szollosi E, et al. Intrinsic structural disorder of DF31, a *Drosophila* protein of chromatin decondensation and remodeling activities. *J. Proteome Res* 2008;7:2291–2299. [PubMed: 18484763]

100. Radhakrishnan I, Perez-Alvarado GC, Dyson HJ, Wright PE. Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Lett* 1998;430:317–322. [PubMed: 9688563]
101. Hon WC, et al. Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL. *Nature* 2002;417:975–978. [PubMed: 12050673]
102. Cramer P, et al. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 2000;288:640–649. [PubMed: 10784442]

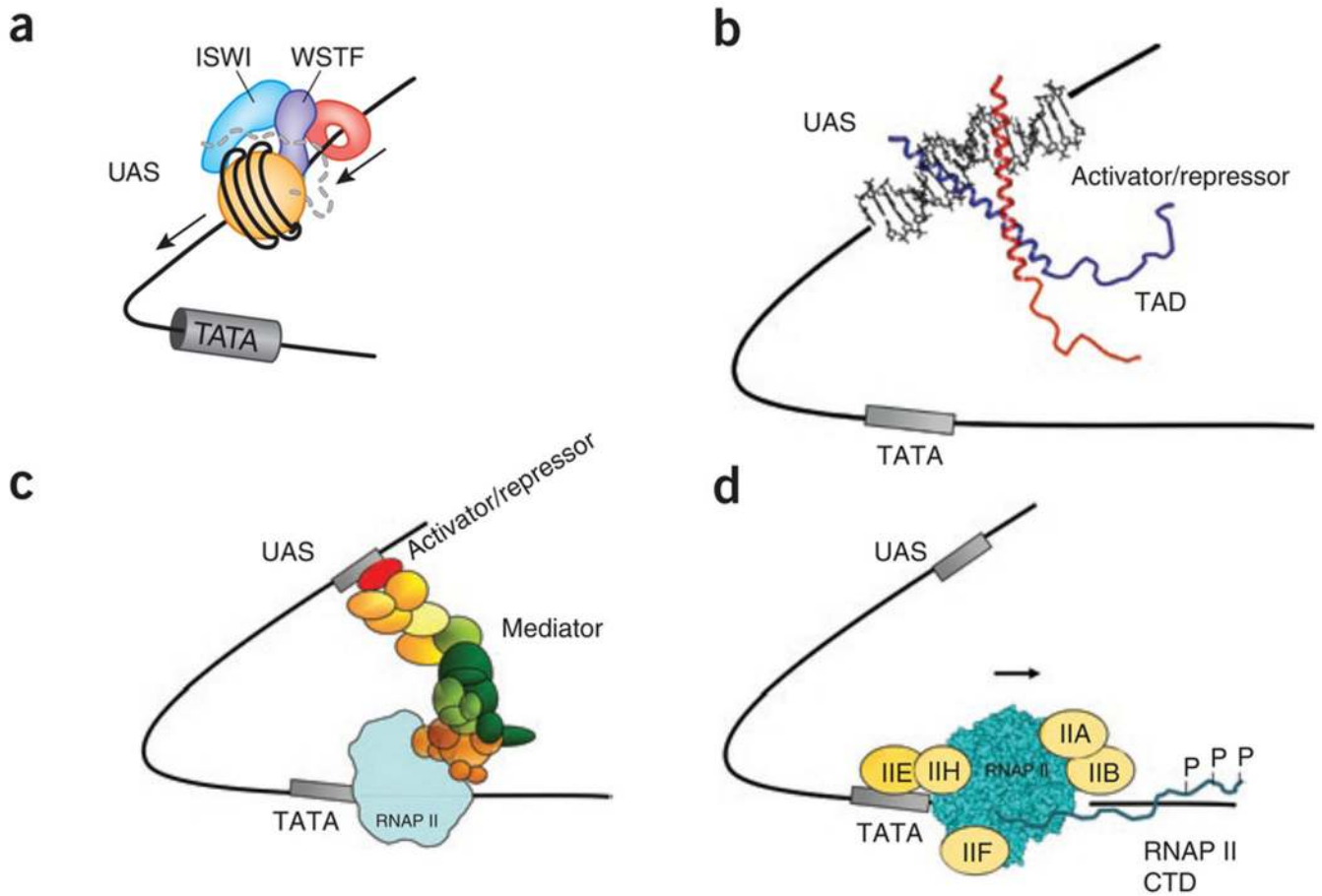


Figure 1. Schematic representation of the four main (highly interdependent) components of the eukaryotic transcription machinery. **(a–d)** Chromatin remodeling **(a)**, transcription factors **(b)**, co-activators **(c)** and basal machinery **(d)**. Disordered histone (yellow) tails (dashed line) assist the assembly of nucleosomes and provide platforms for chromatin remodeling complexes. Chromatin remodeling complexes such as SWI/SNF regulate the accessibility of the DNA, and their mobility is enhanced by disordered regions. Transcription factors decipher regulatory information encoded in enhancer regions (UAS, upstream activating sequence), and they interact with other proteins via disordered TADs. Large co-activator complexes such as the Mediator (head, middle and tail modules are shown in orange, green and yellow, respectively) transmit signals from enhancer- and repressor-bound factors to the core machinery. Transcription initiation is achieved by RNAP II assisted by five general transcription factors (TFIIA, TFIIB, TFIIE, TFIIIF and TFIIF). The disordered CTD of RNAP II serves as a scaffold for a range of complexes involved in different stages of transcription, and it functions in a phosphorylation-dependent manner.

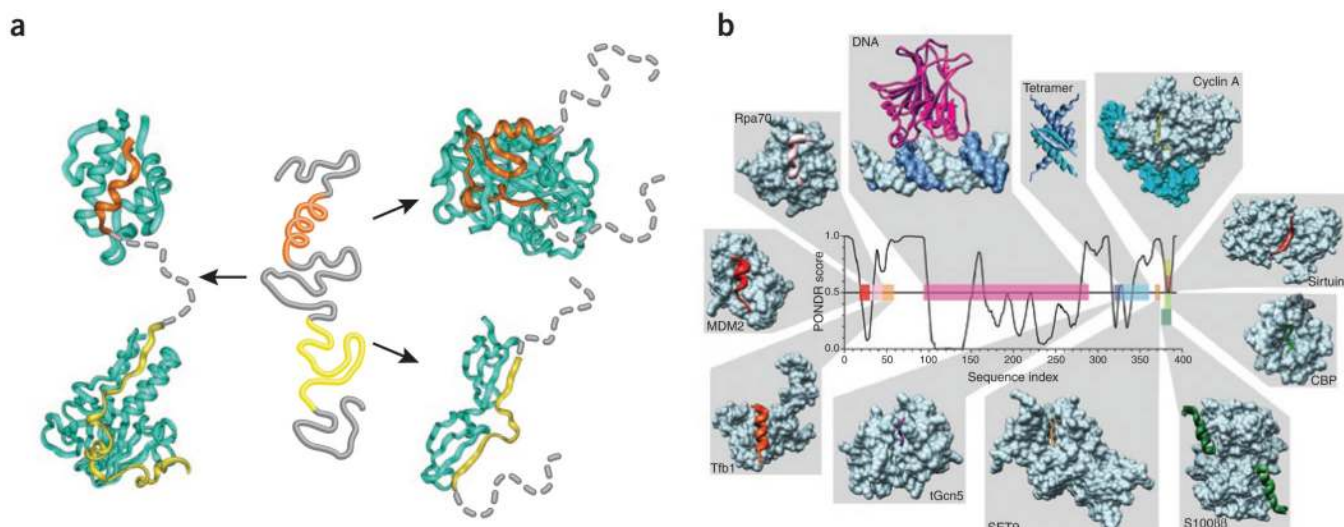


Figure 2.

Molecular recognition by intrinsically disordered proteins. **(a)** Binding of IDPs to their partners is often coupled to disorder-to-order transition. IDP segments might have transient secondary structures that are preserved in the bound state (preformed element, orange), whereas other segments may exhibit different conformations (yellow) with different partners (promiscuity). Owing to this segmental mode of binding, some parts of IDPs may remain disordered (fuzzy, dashed lines) in the bound state, which contributes to the multifunctionality (for example, moonlighting) of these proteins. **(b)** Disorder profile of the p53 transcription factor. p53 is at the center of a large signaling network regulating expression of genes involved in a variety of cellular processes, such as cell cycle progression, apoptosis induction, DNA repair and response to cellular stress. p53 interacts with a large number of other proteins, and the interaction sites are signaled by downward spikes in the plot of disorder. Disorder predictions were performed by PONDNR VL-XT (<http://www.pondr.com/>); segments with scores above 0.5 correspond to disordered regions, while those below 0.5 correspond to ordered regions or binding sites. The structures of the complexes representing illustrative examples of various molecular recognition features (MoRFs) are displayed around the predicted disorder pattern.

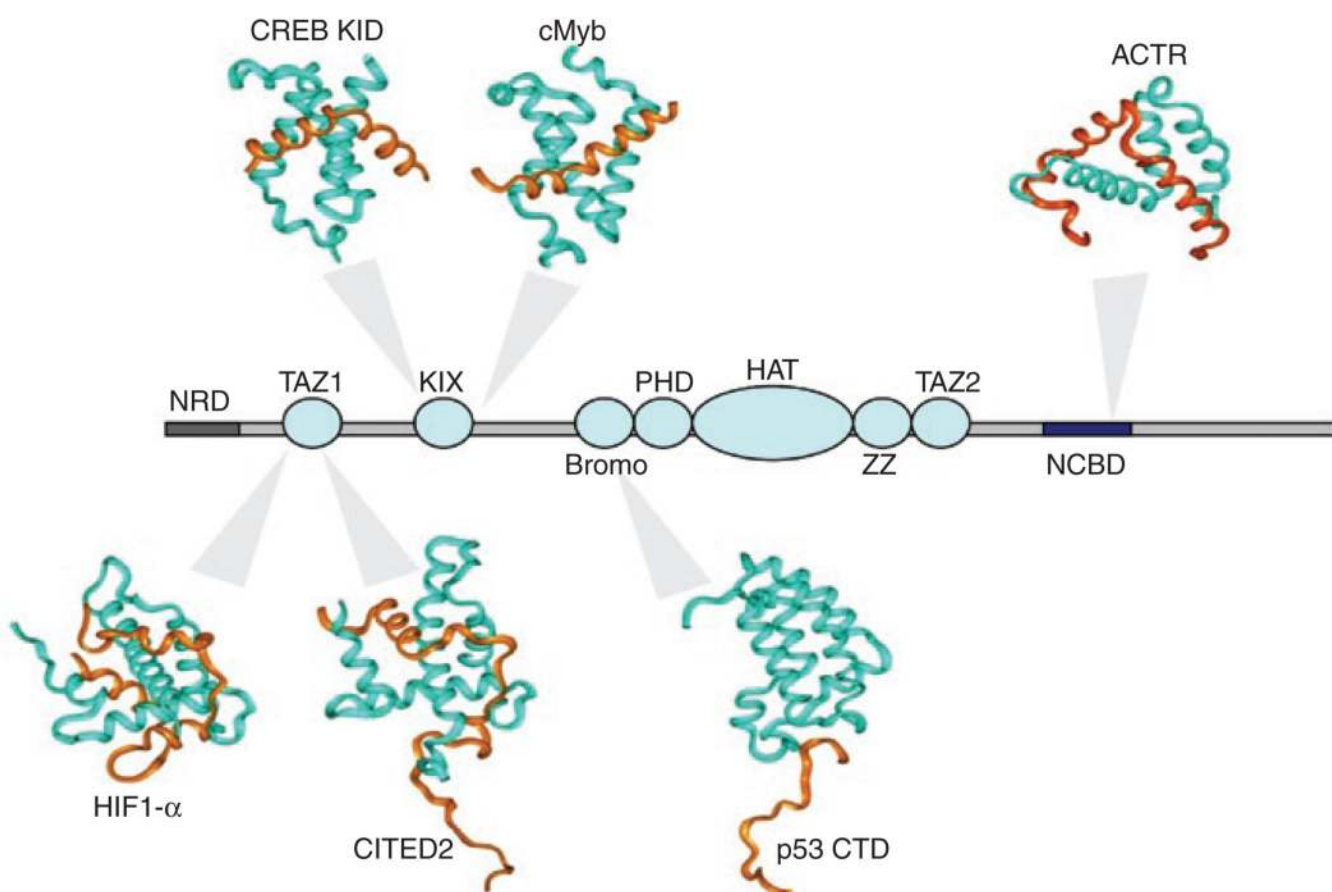


Figure 3. Domain structure of CBP/p300 transcriptional co-activator and three-dimensional structures of complexes of globular CBP domains with disordered transcription factors bound. TAZ1 domain with HIF1- α (Protein Data Bank (PDB) code 1L8C)69 and CITED2 (PDB code 1R8U)70; KIX domain with KID of CREB (PDB code 1KDX)72 and cMyb (PDB code 1SB0)96; Bromo domain with p53 CTD (PDB code 1JSP)71; disordered NCBR domain with ACTR (PDB code 1KBH)12.

Table 1

Representative IDPs within the transcriptional machinery

Protein/complex	Function	Disordered region	Evidence	DisProt	Reference
Chromatin remodeling					
Linker histone (H1)	Chromatin condensation, DFF40/CAD activation	CTD	Missing from crystal structure; sequence-independent binding	DP00097, DP00136	97
Core histone (H3, H4)	Internucleosomal interactions, chromatin fiber condensation	NTD	Missing from crystal structure		18,47
Snf5 and Swi3 of Swi/Snf	Scaffold in histone binding, interactions with specific genomic regions	NTD	Irregular EM structure, anomalous migration on SDS gel		57
NurF301 of NURF	Nucleosome binding and sliding, NURF assembly, stimulation of ATPase activity	NTD, Q-rich domain	Anomalous migration on SDS gel, sequence analysis		50
CHRAC14-CHRAC16	CHRAC assembly, DNA binding, nucleosome sliding	NTD, CTD	Missing density in X-ray structure		55
Spt7, Spt8 and Ada3 of SAGA	Interactions with histone tails, histone acetylation		Missing density from EM structure, anomalous migration on SDS gel		58
Sir3p	Transcriptionally "silent" chromatin binding, platform for protein-protein interactions, oligomerization	Linker (amino acids 215–550) connecting CTD and NTD	CD spectroscopy, proteolytic sensitivity		98
MeCP2	Maintenance of condensed chromosomal superstructures, recognize methylated DNA	Linkers connecting folded domains (amino acids 8–68, 75–120, 160–200, 247–272, 327–377)	CD spectroscopy, proteolytic sensitivity	DP00539	15
Df31	Chromatin stabilization, histone chaperone	Full length	CD and NMR spectroscopy, gel-filtration, limited proteolysis		99
HMG proteins	Nucleosome binding and sliding, enhanceosome formation, "nuclear hub"	Full length	CD and NMR spectroscopy	DP00038, DP00039, DP00040	40
Transcription factors					
p53	Tumor suppressor, interactions with a large number of transcriptional regulators	NTD (amino acids 1–73) and linker to CTD (amino acids 291–312)	NMR spectroscopy, bioinformatics	DP00086	30,38
GCN4	Eukaryotic transcriptional activator, creates basic leucine zipper on DNA	CTD (amino acids 221–281)	CD spectroscopy	DP00083	67
EFP	Transcriptional activator, interactions with proteins in mRNA biogenesis	Full length	Sequence-independent binding, bioinformatics		64
CREB	Transcriptional response to cAMP,	KID (amino acids 101–160)	NMR spectroscopy	DP00080	100

Protein/complex	Function	Disordered region	Evidence	DisProt	Reference
Chromatin remodeling					
HIF1- α	binding to CBP Hydroxic response, interaction with VHL tumor suppressor	CTD (amino acids 403–698, 776–826)	CD and NMR spectroscopy, analytical ultracentrifugation, fluorescent spectroscopy	DP00262	101
Co-activators					
CBP/p300	HAT activity, assembly of the transcription machinery	NCBR domain, linkers between folded domains (amino acids 1–60, 76–347, 439–587, 674–1080, 1752–1762, 1855–2057, 2117–2442)	NMR spectroscopy	DP00348	11
Mediator	Transmitting regulatory signals to the core machinery	Med8 linker between NTD and CTD (amino acids 129–210)	Proteolytic sensitivity, missing density in X-ray structure		86
Core machinery					
RNAP	Assembly of complexes at different stages of transcription, connecting transcription to mRNA processing	CTD	NMR spectroscopy, missing density in X-ray structure	DP00504	88,102
TFIIB	Interactions with TBP and the active site of RNAP	Linker connecting NTD and CTD (amino acids 55–126)	Missing density in X-ray structure, cross-linking and hydrolysis studies		91,92
TFIIF	Promoter recognition, communication between RNAP and the Mediator head	Central part (amino acids 280–630 in yeast)	Missing density in EM structure, proteolytic sensitivity		93,94