Review

Cell PRESS

Dynamic protein–DNA recognition: beyond what can be seen

Monika Fuxreiter[1], Istvan Simon[2] and Sarah Bondos[3]

[1] Department of Biological Chemistry, Weizmann Institute of Science, 7600 Rehovot, Israel
[2] Institute of Enzymology, BRC, H-1113 Budapest, Hungary
[3] Department of Molecular and Cellular Medicine, Texas A&M Health Science Center, College Station, Texas 77845, USA

Traditionally, specific DNA recognition is thought to rely on static contacts with the bases or phosphates. Recent results, however, indicate that residues far outside the binding context can crucially influence selectivity or binding affinity via transient, dynamic interactions with the DNA binding interface. These regions usually do not adopt a well-defined structure, even when bound to DNA, and thus form a fuzzy complex. Here, we propose the existence of a dynamic DNA readout mechanism, wherein distant segments modulate conformational preferences, flexibility or spacing of the DNA binding motifs or serve as competitive partners. Despite their low sequence similarity, these intrinsically disordered regions are often conserved at the structural level, and exploited for regulation of the transcription machinery via protein–protein interactions, post-translational modifications or alternative splicing.

DNA readout mechanisms

Specific DNA recognition by proteins is fundamental to many regulatory processes that control the flow of genetic information. Traditionally, DNA readout has been thought to be achieved by direct and indirect mechanisms [1,2]. The characteristic pattern of hydrogen bonding interactions with the bases provides a direct readout, which can serve as a fingerprint of the cognate DNA sequence. Electrostatic interactions with the phosphates are referred to as an indirect readout because they mostly depend on sequence-specific local variations in the DNA geometry [3]. Proteins themselves can also facilitate deviations from the ideal B-form conformation [4]. As part of the indirect readout, the sequence-specific water structure around the DNA is utilized to establish water-mediated contacts with the bases or phosphates [5]. Although it is clear that no simple code exists [6], recent results now indicate that the underlying principles are even more complicated than initially anticipated.

Conversion from a nonspecific to a specific complex requires the adaptability of the interface [7] and is usually accompanied by significant changes in flexibility. In the 'DNA shape' readout this process is modulated by the groove-width-dependent electrostatic potential; ultimately, the DNA geometry [8]. The majority of DNA-binding proteins are equipped with intrinsically disordered (ID) segments that contribute to DNA recognition at various levels (Box 1). ID regions can facilitate diffusion along DNA [9,10] or play a role in the transition from a nonspecific to a specific complex [11]. ID segments can also modulate selectivity by forming specific interactions with DNA [12–14]. Most ID regions are thought to fold upon contact with DNA, and function similarly to globular segments.

ID regions, however, can also remain invisible in the structure of the protein–DNA complex, yet contribute to selectivity or binding affinity. For protein–protein interactions, the importance of structurally ambiguous regions has been recognized and fuzzy complexes have been defined [15]. ID segments in complexes can form a set of alternative structures or a dynamic ensemble of conformations that enables the formation of alternative or dynamically varying interactions [16,17]. In general, the lack of folding reduces the loss of conformational entropy upon binding [18]. The presence of ID segments thus increases mobility and improves binding affinity.

The influence of bound ID regions, which retain their flexibility in complex form, on protein–DNA interactions has not been characterized in detail. Nearly all binding studies of eukaryotic transcription factors utilize only the DNA binding domain, rather than the full-length protein. This is caused, in part, by the increased susceptibility of ID regions to proteolysis and aggregation [19,20]. Furthermore, highly flexible/ID regions far from the binding interface are not considered to be crucial for the formation of the final complex. Accumulating experimental evidence, however, points to the importance of dynamic factors in determining DNA binding specificity/affinity.

Through several examples (Table 1), we demonstrate that ID segments outside the binding context can influence DNA recognition. Transient contacts with the DNA binding site can modulate the ionization status or conformation of residues that are available to interact with DNA, thereby influencing the pathway and/or thermodynamics of the binding process [21,22]. Pliable segments could even act nonspecifically as a charged cloud simply fill the space between the protein and DNA, thus reducing the probability of forming productive, specific interactions [11,23]. These highly flexible/ID regions can also affect the spacing between structured binding motifs or increase the lifetime of the encounter complex [24]. To provide a framework for these observations, we propose the existence of a dynamic DNA readout, in which dynamic interactions established by highly flexible/ID regions are integral parts of the DNA recognition process. Although these contacts are not visible

Corresponding author: Fuxreiter, M. (monika@enzim.hu).

0968-0004/$ – see front matter doi:10.1016/j.tibs.2011.04.006 Trends in Biochemical Sciences xx (2011) 1–9

Box 1. ID regions in DNA binding

The role of ID segments in both nonspecific and specific DNA binding has long been recognized [18]. Approximately 70% of DNA-binding proteins harbor ID tails [87]. These charged tails facilitate target search along the DNA through nonspecific electrostatic interactions, thus reducing the concentration of the free protein and thereby lowering the dimensions of the diffusion along the DNA. ID tails also allow simultaneous contacts with two DNA strands, referred as a 'monkey bar' mechanism, [87]. The efficiency of this mechanism depends on the optimal charge distribution of the tail and not necessarily on its actual sequence [68]. ID linkers of multidomain proteins increase the rate of intersegmental transfer by serving as a bridge between two DNA fragments [10,46]. ID linkers also promote proper orientation of the DNA reading heads by protein–protein interactions once the cognate sequence is encountered. In Cys_2–His_2 zinc finger proteins, for example, the ID linkers lock the DNA binding motifs and thereby contribute to the conversion from the nonspecific to the specific complex, via a 'snap-lock' mechanism [11,32].

Flexible loops or ID regions can also directly contribute to the specificity of protein–DNA interactions by forming direct contacts with the DNA [48]. DNA-binding proteins with an immunoglobulin-like fold insert sequentially diverse loops into the major groove, such as for p53-family transcription factors [88] or Rel homology domains [89]. The variability of the loop enables diverse orientations of the β-sandwich domains. The majority of disordered tails or arms prefer to interact with the minor groove [3]. The Arg-rich motifs establish direct hydrogen bonds with the bases and neutralize phosphate charges. This results in bending of the DNA, as in the case of the HMG boxes or AT hooks. ID tails not only accelerate the formation of specific contacts, but they are also favorable for binding affinity [13,83].

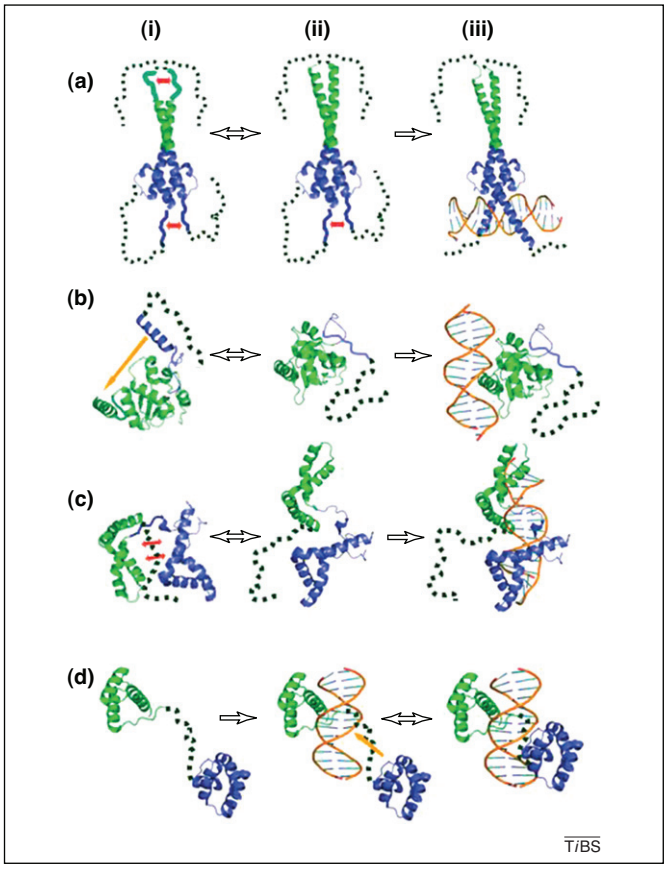

Figure 1. Four schemes of dynamic DNA recognition. Transient, dynamic interactions by flexible/ID regions outside the interface modulate the structural transitions that are required for optimal DNA contacts. In the conformational selection **(a)**, flexibility modulation **(b)** and competitive binding **(c)** mechanisms, the dynamic conformational equilibrium of the DNA binding region is influenced by the ID region. In the tethering mechanism **(d)**, the equilibrium of the DNA-bound form is affected. **(a)** Conformational selection. The Max transcription factor (PDB code: 1NKP) binds DNA as a dimer. The disordered N-terminal region (upper dotted line) reduces the electrostatic repulsion (red arrows) between the two monomers, and increases the population of the folded state at the flanking leucine zipper (green) (I→II). This also stabilizes the bHLH region (blue) and thus improves binding affinity for DNA (II→III). **(b)** Flexibility modulation. Binding of the Ets-1 transcription factor to DNA (PDB code: 1MDM) is coupled to the unfolding of the HI-1 autoinhibitory helix (blue), which interacts (yellow arrow) with the recognition helices (green). This process is modulated by the disordered SRR region (dotted) (I→II). The SRR region enhances the dynamics of the hydrophobic network, which includes the recognition helices. Increased mobility of the interface residues (green) favors DNA binding (II→III). **(c)** Competitive binding. The HMG boxes (blue and green) of HMGB1 (PDB code 2GZK) are in dynamic equilibrium between an open and closed form (I→II). Repulsion between the two boxes (red arrows) is masked by the disordered C-terminal tail (I). Opening the structure exposes crucial residues for DNA binding (II→III). **(d)** Tethering. The Oct-1 transcription factor (PDB code: 1HK0) interacts with a bipartite DNA sequence. The first binding event occurs via one of the globular POU domains (I→II). The local concentration of the second POU domain near the DNA is increased by the disordered linker (II→III), which also regulates the separation of the two structured domains.

in the structure of the final complex, they are required to shape the interface and enable/disable particular side chains for DNA interactions. Furthermore, these malleable segments can respond to protein–protein interactions [25–28], post-translational modifications [29–31], or splicing [32–34] to affect DNA binding in response to cellular cues.

Dynamic mechanisms of DNA recognition

We define four schemes how distant ID segments affect DNA binding (Figure 1). Via transient interactions, ID regions can directly modulate the conformational preferences (i) or flexibility (ii) of the binding interface. Rapidly fluctuating ID segments can also influence DNA contacts via screening/competitive binding (iii) or tether structured binding domains (iv). These are not distinct categories; multiple mechanisms might cooperate in some cases, even within the same protein [22,35]. In this way, different ID regions can have different effects on DNA binding and fine-tune affinity. In each category, a few examples are described in detail; all other instances are listed in Table 1.

Conformational selection

ID regions can shift conformational equilibrium of the DNA-contacting region within the same protein and facilitate formation of the structures required for binding. The Max transcription factor interacts with E-box (CACGTG) DNA sequences via a basic helix–loop–helix (bHLH) and a leucine zipper (LZ) segment to repress transcription of Myc target genes. In its free form, Max dominantly exists as a homodimer, with the LZ domain adopting a helical structure (Figure 1a). The LZ region, however, is unstable, and undergoes folding–unfolding transitions. This equilibrium is shifted towards the folded state by the N- and C-terminal segments, which remain as a random coil even in the dimer [21]. The acidic N terminus masks the destabilizing electrostatic potential present at the LZ region. Owing to a strong cooperativity between the bHLH and LZ domains, the N- and C-terminal regions also facilitate the formation of the recognition helices in the bHLH region, thereby decreasing the dissociation constant with the E-box by 10 to 100-fold.

A connection between ID-induced secondary structure and improved binding affinity has also been observed for methyl CpG binding protein 2 (MeCP2). MeCP2 reads

Table 1. Examples of *fuzzy* protein–DNA complexes

Model	IDP	ID region	Conservation	Post-translational modification	Experimental evidence	Ref.
Conformational selection	Max	N-tail (1–18), C-tail (93–132)	*[a]	Increased phosphorylation	Crystallography, CD[b], NMR	21, 29
	MeCP2	NTD (1–75), ID[c] (164–210), CTD (330–486)		–	CD, fluorescence anisotropy	22, 95
	TDG[d]	CTD (340–410)		Decreased acetylation	NMR	60
	Neurogenin	Basic motif (90–104)		–	NMR, CD, fluorescence anisotropy	96
	ApLLP[e]	N-tail (1–11), C-tail (99–120)	*	–	CD	97
Flexibility modulation	Ets-1	SRR[f] (244–300)		Decreased phosphorylation	NMR	39, 40
	SSB[g]	C-tail (113–177)	*	–	Crystallography, fluorescence anisotropy	28, 98
Competitive binding	PC4	NTD (1–60)		Decreased phosphorylation; increased acetylation	NMR	27, 58
	FACT	NTD (1–186), CTD (625–723)	*	Decreased phosphorylation	High-speed AFM[h], CD	30
	HMGB1	C-tail (186–215)	*	Decreased phosphorylation	NMR, SAXS[i], PRE[j]	43
	Ubx	I1 (235–286), I2 (174–216), R[k] (1–174)	*	–	CD, proteolysis	35
	DSS1[l]	1–70		–	Crystallography	44
	NKX3.1[m]	AD[n] (85–96), SI[o] (99–105)	*	Increased phosphorylation	NMR	31
	PPAR-γ[p]	NTD (9–110)		–	Crystallography, H/D exchange	45
	UvrD[q]	C-tail (645–720)	*	–	Crystallography	99
	β-telomere	CTD		Phosphorylation	Crystallography	62
Tethering	OCT1	76–101	*	–	Crystallography, NMR	51
	RPA	IULD (106–180)	*	Phosphorylation	NMR	26, 73
	KorB[r]	NTD (1–64), linker (253–293)		–	CD,AU, SAXS	63

[a]Regions with conserved ID character.
[b]Circular dichroism.
[c]Intervening domain.
[d]Thymine DNA glycosylase.
[e]*Aplysia* LAPS18-like protein.
[f]Ser-rich region.
[g]ssDNA binding protein.
[h]Atomic force microscopy.
[i]Small angle X-ray scattering.
[j]Paramagnetic resonance enhancement.
[k]Regulatory domain.
[l]Brh2 interacting protein.
[m]NK 2 homeodomain.
[n]Acidic domain.
[o]SRF interacting motif.
[p]Peroxisome proliferator-activated receptor gamma.
[q]Helicase of the 1A superfamily.
[r]Plasmid partition protein KorB.

epigenetic information encoded in DNA methylation patterns [36]. MeCP2 has several autonomous DNA binding domains intercalated between ID regions, and mutations within these regions are associated with Rett syndrome. The propensity of regular secondary structures increases by 7% upon binding to DNA, but MeCP2 still remains substantially disordered in the complex [22]. Fusing the structured methylated DNA binding domain (MBD) with the disordered N-terminal domain (NTD) decreases the K_D for DNA from 8.5 nM to 0.8 nM. The NTD does not harbor a DNA binding site; instead, through interdomain interactions, it populates those MBD conformations that favor DNA interactions [22]. The NTD destabilizes the overall MBD structure, but the stability of the MBD–DNA and MBD–transcriptional regulatory domain (TRD)–DNA complexes are comparable. Hence, the ID region frustrates the native fold of the structured MBD domain, which is relieved by DNA interactions and improves binding affinity. Although its mechanism is far from being understood, fusion of the disordered MeCP2 C-terminal domain (CTD) to the TRD also increases binding affinity by ~30-fold, even though the CTD does not adopt a well-defined structure in the complex.

Box 2. Dynamic allostery in structured protein–DNA complexes

Protein–DNA complexes without any apparent degree of disorder can be modulated by altering flexibility along the protein chain upon DNA binding. Interactions between cAMP and catabolic activator protein (CAP) induce a conformational transition in CAP by 60° rotation of the F helices, which are inserted into the major groove of the DNA. The S62F CAP mutant, however, remains in a binding-incompatible conformation even in the presence of cAMP, even though it exhibits a similar affinity for DNA. The binding of the S62F CAP mutant to DNA binding is entropically driven ($T\Delta S$=13.2 kcal/mol), whereas for the wild-type protein, DNA binding is enthalpically favored ($T\Delta S$= –14.3 kcal/mol, ΔH= –23.2 kcal/mol) as measured by high-sensitivity isothermal calorimetry [38]. NMR chemical shift differences indicate a population shift in S62F CAP on the μs/ms timescale upon interacting with DNA. Based on the bond order parameters, a rearrangement around the cAMP binding pocket takes place in S62F CAP upon contact with DNA that influences local packing density and increases mobility. By contrast, the wild-type CAP becomes more rigid upon interaction with DNA. Thereby, the conformational entropy increases in S62F CAP, but decreases in wild-type CAP. This finding illustrates how redistribution of states that affect flexibility/mobility can influence the thermodynamics of DNA binding.

Altering flexibility can also modulate the specificity of DNA binding. The linker region in cytidine repressor adopts various conformations upon binding to different operator sequences, which are associated with different degrees of flexibility [90]. In the case of the lactose repressor, a dramatic difference between the heat capacity difference towards specific and nonspecific operators (–1100 cal/mol/K versus –200 cal/mol/K) has been observed, despite the comparable buried surface area [7]. Flexibility of the specific and noncognate complexes, however, deviates considerably and is responsible for the different thermodynamics. Increased mobility in the nonspecific complex facilitates probing different sets of interactions to localize the correct operator site. DNA binding narrows the native state conformational ensemble and significantly reduces flexibility of the protein [7]. This mechanism is probably common to many ID proteins that fold upon interaction with DNA [91].

Flexibility modulation

Long-range changes in structure and dynamics can also regulate DNA binding (Box 2). Remote regions can interfere with motions on the μs/ms timescale and modulate the flexibility of the residues at the protein–DNA interface [37], thus altering the conformational entropy of binding [38]. Most examples discussed in Box 2 utilize ID segments, which fold upon interacting with DNA.

Remote ID regions that preserve their conformational heterogeneity in complex with DNA can also modulate flexibility. The Ets-1 transcription factor regulates various genes, and is involved in stem cell development and tumorigenesis. Ets-1–DNA binding is regulated by an autoinhibitory region and requires the HI-1 helix to unfold [39] (Figure 1b). Interactions with DNA are further attenuated by a Ser-rich region (SRR), which is disordered in both the free and complex form. Phosphorylation of five sites within the SRR region gradually reduces binding affinity up to ~1000-fold. Although phosphorylation itself has minimal impact on the secondary structure of SRRs [40], it interferes with the formation of transient intraprotein contacts. The most pronounced differences are in the dynamic properties of the HI-1 autoinhibitory helix and recognition helices H1 and H3. These units form a hydrophobic network, whose motions are dampened by SRR phosphorylation. Truncation of the SRR region gradually increases the mobility of these residues and facilitates HI-1 unfolding; a process that is required for DNA binding [39]. Thus, the distant ID region in Ets-1 perturbs the dynamics of the protein–DNA interface and modulates the conformational transition that leads to the tight, specific complex.

Competitive binding

Rapidly fluctuating chains of ID regions in a protein–DNA complex can screen electrostatic attraction between the protein and DNA or make specificity-determining residues inaccessible to DNA. The human positive cofactor 4 (PC4) recruits general transcription factors and stimulates RNA polymerase II (RNAP II). The activity of PC4 is regulated by a disordered NTD, which consists of a Ser- and acidic-rich region and a Lys-rich region. The NTD alone lacks considerable affinity for either ssDNA or dsDNA. Instead, this ID region establishes transient, dynamic interactions with the structured CTD and these compete with the ssDNA binding sites. Consequently, the ssDNA binding affinity and the DNA unwinding activity of PC4 are both reduced [27].

A similar competitive mechanism is observed for the high mobility group (HMG) protein B1. HMGB1 participates in various nuclear processes by contacting its target DNA via two HMG boxes; a process that is negatively regulated by the disordered C-terminal tail [41]. The tail stabilizes HMGB1, but does not perturb its secondary or tertiary structure [42]. In the absence of DNA, the two HMG boxes assemble on the acidic C-terminal tail and the binding surfaces are only transiently exposed [43]. The tail-bound collapsed form is in dynamic equilibrium with an extended DNA binding competent form, in which the tail remains disordered (Figure 1c). The tail screens interactions between the two HMG boxes, therefore, it affects DNA recognition in a length-dependent manner [42].

In additional examples, intramolecular protein–protein interactions involving an ID region compete for DNA interactions (Table 1) [31,44,45]. For instance, the ultrabithorax (Ubx) homeotic transcription factor contains several ID regions to fine-tune DNA binding [35] (Box 3).

Tethering

The separate DNA-binding motifs in multidomain proteins are often connected by highly flexible/ID linkers, which facilitate the target search along the DNA [10,46] (Box 1). The linkers are usually absent from the crystal structures of complexes [47,48], and thus preserve their conformational heterogeneity when bound to DNA.

Pit-1, Oct-1, Unc-86 (POU) domain transcription factors recognize different bipartite DNA motifs, depending on the length and sequence of the connecting region. The linker of the octamer binding factor 1 (Oct-1) connects two helix–turn–helix motifs (Figure 1d). It cannot be seen in the complex [49] and is also sensitive to proteolysis [50]. Shortening this segment or modifying its amino acid composition cause dramatic decreases in binding to either sites [51]. Charge distribution of the linker, including an essential negatively charged Glu, is crucial for DNA recognition.

The human replication protein A (RPA) participates in both nucleotide excision repair and combinatorial repair using multiple ssDNA binding domains. The 70-kDa sub-

Box 3. Ubx: multiple ID regions to fine-tune DNA binding

The structured homeodomain (HD) of the Hox transcription factor Ubx binds its optimal DNA sequence with very high affinity (K_d = 63±24 pM), yet with little ability to distinguish target DNA sequences *in vivo* (~3-fold variation in affinity). By contrast, the full-length Ubx has lower affinity, yet higher selectivity for the optimal DNA sequence (K_d = 160±33 pM, 12-fold affinity variation), which indicates that regions outside the HD are important for DNA affinity and sequence selectivity [34,35] (Figure 2a). Most of the non-HD regulatory regions have ID character. The I1 region of Ubx reduces affinity twofold, and the I2 region reduces affinity a further 40-fold (Figure 2) [35]; both via the competitive binding mechanism. The R region restores much of this loss in affinity.

In contrast to the full-length Ubx, Ubx HD–DNA complex formation is strongly pH dependent [92]. I1, but not I2, reduces the pH-dependence of DNA binding, which indicates that I1 must directly interact with the HD to shift the pK_as of residues that are crucial for DNA interactions. By contrast, I2 does not establish direct contact with the HD; instead, it rapidly fluctuates to hinder DNA access. The R region either directly interacts with I1 and I2 to perturb their conformational equilibrium, or nonspecifically blocks their access to DNA.

Protein–protein interactions or alternative splicing enable 'context specific gene regulation' of Ubx; that is, cognate DNA sequences depend on the cellular context. The I1 region contains a YPWM motif, which mediates communication with the Hox cofactor Extradenticle (Exd) [25], which relieves repression of DNA binding by I1 [34] (Figure 2a).

Different *Ubx* splicing isoforms are produced in a stage- and tissue-specific manner by alternative splicing of three microexons that are located in the region that links the YPWM motif to the HD (Figure 2b). Removal of the microexons (alone or in combination) affects Ubx DNA affinity and selectivity [34] both *in vitro* and *in vivo* [93,94].

unit of human RPA (hRPA70) contains weak and high-affinity DNA binding domains (DBDs), connected by a 78-amino-acid intrinsically unstructured linker domain (IULD) [52]. RPA makes its first contacts with ssDNA using the high-affinity DBDs. The highly flexible IULD anchors the weak affinity DBD to DNA, thereby increasing its local concentration near the substrate. Conversely, the loss of the transient IULD–weak affinity DBD interaction results in a threefold reduction in binding [26].

Regulation of DNA binding via ID regions

ID regions are also often the sites of regulatory actions of signaling pathways [53]. Post-translational modifications, protein–protein interactions or alternative splicing [54–57] offer opportunities for the cell to exploit distant ID domains to regulate DNA binding by a structured domain in response to single or multiple cellular cues, by enhancing or interfering with any of the four mechanisms described above.

Post-translational modifications

Phosphorylation usually perturbs the positively charged DNA binding interface via electrostatic interactions. FACT (facilitates chromatin transcription) displaces histone H2A–H2B dimers from nucleosomes and facilitates RNAP II-mediated transcription. To interact with DNA, FACT utilizes an HMG domain flanked by a basic and an acidic ID region. The acidic ID region forms intramolecular interactions with both the HMG domain and the basic ID segment, which compete for DNA contacts [30]. These interactions, which mask nucleotide-binding elements, are strengthened by phosphorylation of the acidic ID region. Consequently, phosphorylation blocks DNA binding, without inducing folding of the ID regions.

Similar to FACT, phosphorylation of PC4 gradually decreases its binding affinity for dsDNA [58]. Progressive phosphorylation of eight Ser residues in the disordered PC4 NTD shields the neighboring Lys-rich region, which mediates communication between the disordered N-terminal and the ordered C-terminal part of PC4. Hence, DNA binding of the structured region is affected indirectly via an ID segment and not directly via phosphoserine contacts.

In both examples, phosphorylation conforms to the 'incremental rheostat' type regulatory mechanism, which was originally proposed for Ets-1 [39]. In this model, gradual changes in DNA binding affinity take place via multiple phosphorylation events, in contrast to a phosphorylation-dependent on/off switch mechanism. The incremental rheostat regulation provides a sensitive mechanism by which to control transcription in response to different environmental signals (e.g. Ca^{2+} signaling).

Phosphorylation can also act via perturbing structural elements that are required for DNA binding. For example, phosphorylation of the Max transcription factor modulates the structure/disorder balance of the basic region of the DBD, thereby influencing its affinity for DNA [29]. The behavior of ID regions can also be regulated by other post-translational modifications (Table 1), such as acetylation [59] and sumoylation [60]; the structural details, however, are not as well characterized.

Protein–protein interactions

Protein–protein interactions often utilize ID regions by targeting short, low-complexity motifs [57]. Intramolecular contacts of ID segments with DNA recognition elements frequently compete with these intermolecular interactions for the same binding site. Thus, the affinity for DNA is regulated via a competitive binding mechanism; often referred to as 'cooperativity' [61].

Transcriptional activation of the tumor suppressor p53 depends on its interactions with hRPA70. Binding sites for ssDNA and p53 overlap on the weak affinity DBD of hRPA70 [26]. The affinity of hRPA70 for ssDNA is higher than that for p53, and is primarily controlled by the disordered IULD. The ID region increases the local concentration of the weak-affinity DBD near its cognate site, which results in the occlusion of p53 from the basic cleft of hRPA70. Thus, depending on its orientation, flexibility and length, the hRPA70 IULD regulates the balance between binding p53 and damaged DNA.

Activities of additional protein–DNA complexes can be modulated by protein–protein interactions via regions that retain their ID character in the bound state [25,27,28,62,63] (Table 1). Although no direct structural evidence is available, disordered transactivation domains of transcription factors [64] probably function via this competitive binding mechanism to activate or repress transcription via protein–protein interactions.

Alternative splicing

Alternative splicing provides a third opportunity for the cell to modulate DNA binding by altering ID regulatory

regions. Alternative splicing affects the degree of disorder of the ID segment and can insert/delete post-translational modification sites or motifs for protein–protein interactions. Notably, alternative translation start sites in murine MeCP2 increase the length and flexibility of the disordered NTD. Given the role of the NTD in modulating the recognition of methylated CpG islands by transient interactions with the MBD, the different isoforms result in different gene expression patterns in mouse brain [33].

Alternative splicing of *ETS1* removes the entire disordered SRR region; phosphorylation of this region reduces affinity for DNA via modulating the flexibility of the interface. Hence, the activity of human Ets-1 is differentially regulated by two distinct mechanisms: phosphorylation and alternative splicing [65]. By contrast, these two regulatory mechanisms are synergistic in the Max transcription factor. The p22 Max isoform contains a nine-residue long insertion in the disordered N-terminal region, as compared to the shorter p21 isoform. The longer NTD is more effective in stabilizing the DBD and results in higher DNA binding affinity [29]. Phosphorylation enhances this process, but it is only effective in the long p22 isoform.

Alternative splicing also affects ID regions that fold upon DNA binding. For example, different isoforms of the Wilms tumor suppressor protein have distinct linkers that connect the zinc fingers of the protein, and which display variability in flexibility and length [32]. These differences provide a 10–20-fold variation in DNA binding of the different gene products.

Conservation of ID character of regulatory regions

What evolutionary pressures are placed on ID regions that differentially regulate DNA binding affinity? The lack of structural constraints enables ID proteins/regions to undergo large sequence drifts [66], thus making sequence comparisons of evolutionarily related proteins difficult. The ID character of these regions, however, can be conserved despite the fact that the amino acid sequence varies. For example, the N- and C-terminal tails are disordered in many DNA-binding proteins that lack apparent sequence similarity [67,68]. In these cases, the amino acid composition is conserved, which results in a similar structural character, that is, a certain degree of flexibility/disorder [51,69]. The core histone H4 N-terminal tail [70] and the linker histone H1 C-terminal tail, for example [71], remain active even upon scrambling their sequences, while maintaining their amino acid composition. Sequence conservation in ID regions usually appears at the level of short, low-complexity motifs, which are responsible for similar functions [55,72].

The disordered linker in RPA is crucial for ssDNA binding [26]; nevertheless, it lacks a detectable level of sequence similarity in different organisms. Notably, these highly divergent sequences fail to fold into a well-defined structure. In five organisms from three kingdoms of life, the backbone flexibilities of RPA IULDs reveal similar dynamics [73]. Thus sequence divergence is neutral for linker function, given that the dynamic behavior of RPA has been evolutionarily preserved.

Functionally relevant conserved ID character has also been found in Ubx. Sequence alignment of Ubx proteins from six organisms that represent 540 million years of evolution indicates that the ID character, but not the sequence, of regions that regulate DNA binding in *Drosophila melanogaster* Ubx is conserved [35] (Figure 2).

Concluding remarks and future perspectives

The role of flexible or dynamic protein segments in DNA binding has long been recognized [74] (Box 1); the factors underlying specificity or affinity, however, are considered to be primarily static. The available examples (Table 1) argue that dynamic, transient contacts can be utilized to fine-tune static interactions between protein and DNA. The dynamic DNA readout integrates all traditional, static specificity-determining factors, but underscores that global dynamic factors also contribute to the recognition process. Flexible regions distant from the protein–DNA interface can be targeted by diverse signaling pathways to regulate DNA interactions. Hence, specificity/affinity of DNA binding cannot be assigned solely to a few residues, but instead the impact of the whole protein must be considered.

The dynamic regulatory regions discussed in this review lie outside the binding context, so this appears to be a

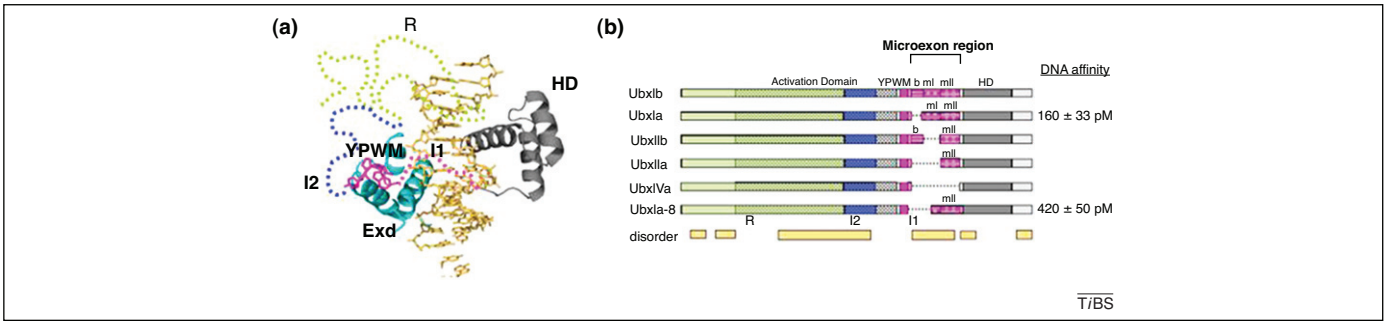

Figure 2. DNA-binding of Ubx is regulated by multiple ID regions. **(a)** Ubx contacts DNA via a structured HD (residues 295–354, dark grey). DNA-binding affinity is influenced by various ID regions (dotted lines) outside the HD: I2 inhibitory region (residues 175–216, light blue), I1 inhibitory region (residues 235–286, magenta), and R regulatory region (residues 1–174, light green). R and I2 are parts of the activation domain. The I1 region contains a YPWM motif (residues 240–243, shown by sticks), which mediates interactions with the HD of the Exd transcription factor (cyan) (PDB code: 1B8I). These contacts relieve the repression of transcription caused by the YPWM motif. **(b)** Ubx activity is regulated by alternative splicing in a tissue-specific manner. The invisible linker that connects the YPWM motif (magenta) with the HD (grey) contains three microexons, b, ml, and mll; the removal of these microexons varies the YPWM–HD distance from 10 to 53 amino acids. Isoforms that contain the nine-residue 'b element' are a minor component in all tissues and are less efficient in Exd binding. Inclusion of the two 17-amino-acid microexons, ml and mll, is determined by tissue identity, germ layer, and developmental stage. Based on these complex expression patterns, Ubx DNA binding affinity varies over a wide range in a context-dependent manner. The R region is shown by light green, the I2 region by blue, and conserved ID regions are displayed in yellow.

special case of allostery. Allosteric regulation of transcription factors via modulation of DBD conformation by the substrate DNA was first advocated more than a decade ago [75]. ID regions have also been proposed to amplify signals and mediate allosteric responses [76,77].

In the dynamic readout model, the protein–DNA complex is viewed as a conformational ensemble, even in active forms. In contrast to the Monod–Wyman–Changeux [78] or Koshland–Nemethy–Filmer [79] models for protein allostery, no unique conformation, which could be selected from a pre-existing equilibrium or formed via induced fit, is required for optimal activity. The entire protein–DNA complex exists as a multiplicity of states, and modulating its conformational equilibrium provides a means to regulate DNA binding activity. All the examples discussed herein (Table 1) are fuzzy, where the ID segments preserve heterogeneity even in their bound form, and the dynamic nature of the ID segments is crucial for influencing DNA binding. The dynamic readout model, however, is also applicable for complexes with local or limited disorder (Box 2).

There are two direct implications of the proposed model. First, it illustrates that specificity/affinity determinants cannot be deduced from the static structure of the final complex. Weak, transient interactions formed with the DNA binding region must also be considered. They can be investigated using rapidly developing experimental techniques that are able to handle low-population states and transient contacts [80–82] (Table 1). Thermodynamic data, in combination with targeted mutagenesis and/or deletion constructs, can also be informative regarding the contribution of dynamic factors [83]. Even in the absence of actual structural disorder, large changes in flexibility or conformational entropy can be crucial for specific DNA recognition [84]. Second, the model suggests new approaches to modulate DNA binding. Owing to the lack of structural constraints, manipulating distant ID regions might be easier than modifying residues that participate in the tight DNA binding interface. Furthermore, ID regions provide several ways in which to alter DNA binding in a cell context-specific manner. Changes in the length or amino acid composition of the ID segments directly affect the degree of flexibility/disorder required for a given function. Insertion/deletion of short functional motifs in ID regions can alter interactions with other protein partners, and thereby modulate DNA binding. Post-translational modifications of these ID segments can respond to extracellular signals and are often exploited in natural regulatory pathways. All these ID regions and their functional sites can be easily predicted from the primary sequence without any additional structural information [85,86] and subsequently subjected to experimental studies.

Although the dynamic DNA readout model provides a more realistic view of specific DNA binding, the ultimate question is whether we will ever be able to understand how given sequences are recognized, especially in the context of the living cell. Recent experimental results described in this review urge us to move from a deterministic to a more stochastic description of protein functionality. Indeed, we have offered a framework to explain multifunctionality of the same protein and the versatility of responses to environmental signals. In the long term, targeting these mechanisms might also provide more efficient means to intervene with gene expression.

Acknowledgments

This was supported by Human-MB08-A 80125 of the Marie Curie Actions of the European FP7 (M.F.). Fruitful comments on the manuscript by Peter Tompa, Jeff Hansen, Gary Daughdrill and Agnes Toth-Petroczy are gratefully acknowledged. M.F. thanks L. Máthé for inspiration and stimulating discussions.

References

1 Luscombe, N.M. *et al.* (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* 29, 2860–2874
2 Gromiha, M. *et al.* (2004) Intermolecular and intramolecular readout mechanisms in protein–DNA recognition. *J. Mol. Biol.* 337, 285–294
3 Rohs, R. *et al.* (2010) Origins of specificity in protein–DNA recognition. *Annu. Rev. Biochem.* 79, 233–269
4 Kim, Y. *et al.* (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365, 512–520
5 Shakked, Z. *et al.* (1994) Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature* 368, 469–473
6 Matthews, B.W. (1988) No code for recognition. *Nature* 335, 294–295
7 Kalodimos, C.G. *et al.* (2004) Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes. *Science* 305, 386–389
8 Rohs, R. *et al.* (2009) The role of DNA shape in protein–DNA recognition. *Nature* 461, 1248–1253
9 Shoemaker, B.A. *et al.* (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8868–8873
10 Doucleff, M. and Clore, G.M. (2008) Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13871–13876
11 Laity, J.H. *et al.* (2000) DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers. *J. Mol. Biol.* 295, 719–727
12 Billeter, M. *et al.* (1993) Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. *J. Mol. Biol.* 234, 1084–1093
13 Toth-Petroczy, A. *et al.* (2009) Disordered tails of homeodomains facilitate DNA recognition by providing a trade-off between folding and specific binding. *J. Am. Chem. Soc.* 131, 15084–15085
14 Boch, J. *et al.* (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512
15 Tompa, P. and Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* 33, 2–8
16 Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208
17 Mittag, T. *et al.* (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17772–17777
18 Spolar, R.S. and Record, M.T., Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263, 777–784
19 Tsvetkov, P. *et al.* (2008) Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome. *Proteins* 70, 1357–1366
20 Fiumara, F. *et al.* (2010) Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* 143, 1121–1135
21 Naud, J.F. *et al.* (2005) Structural and thermodynamical characterization of the complete p21 gene product of Max. *Biochemistry* 44, 12746–12758
22 Ghosh, R.P. *et al.* (2010) Unique physical properties and interactions of the domains of methylated DNA binding protein 2. *Biochemistry* 49, 4395–4410
23 Mittag, T. *et al.* (2010) Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* 23, 105–116
24 Pontius, B.W. (1993) Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem. Sci.* 18, 181–186

25 Passner, J.M. *et al.* (1999) Structure of a DNA-bound Ultrabithorax–Extradenticle homeodomain complex. *Nature* 397, 714–719
26 Vise, P.D. *et al.* (2005) NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. *Nucleic Acids Res.* 33, 2061–2077
27 Jonker, H.R. *et al.* (2006) The intrinsically unstructured domain of PC4 modulates the activity of the structured core through inter- and intramolecular interactions. *Biochemistry* 45, 5067–5081
28 Kozlov, A.G. *et al.* (2010) Regulation of single-stranded DNA binding by the C termini of *Escherichia coli* single-stranded DNA-binding (SSB) protein. *J. Biol. Chem.* 285, 17246–17252
29 Pursglove, S.E. *et al.* (2004) Biophysical properties of regions flanking the bHLH-Zip motif in the p22 Max protein. *Biochem. Biophys. Res. Commun.* 323, 750–759
30 Tsunaka, Y. *et al.* (2009) Phosphorylated intrinsically disordered region of FACT masks its nucleosomal DNA binding elements. *J. Biol. Chem.* 284, 24610–24621
31 Ju, J.H. *et al.* (2009) Interactions of the acidic domain and SRF interacting motifs with the NKX3.1 homeodomain. *Biochemistry* 48, 10601–10607
32 Laity, J.H. *et al.* (2000) Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11932–11935
33 Dragich, J.M. *et al.* (2007) Differential distribution of the MeCP2 splice variants in the postnatal mouse brain. *J. Comp. Neurol.* 501, 526–542
34 Liu, Y. *et al.* (2009) Internal regulatory interactions determine DNA binding specificity by a Hox transcription factor. *J. Mol. Biol.* 390, 760–774
35 Liu, Y. *et al.* (2008) Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the *Drosophila* hox protein ultrabithorax. *J. Biol. Chem.* 283, 20874–20887
36 Hansen, J.C. *et al.* (2010) Binding of the Rett syndrome protein, MeCP2, to methylated and unmethylated DNA and chromatin. *IUBMB Life* 62, 732–738
37 Gunther, S. *et al.* (2006) Molecular flexibility in protein–DNA interactions. *Biosystems* 85, 126–136
38 Tzeng, S.R. and Kalodimos, C.G. (2009) Dynamic activation of an allosteric regulatory protein. *Nature* 462, 368–372
39 Pufall, M.A. *et al.* (2005) Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* 309, 142–145
40 Lee, G.M. *et al.* (2008) The affinity of Ets-1 for DNA is modulated by phosphorylation through transient interactions of an unstructured region. *J. Mol. Biol.* 382, 1014–1030
41 Stros, M. *et al.* (1994) DNA looping by the HMG-box domains of HMG1 and modulation of DNA binding by the acidic C-terminal domain. *Nucleic Acids Res.* 22, 1044–1051
42 Watson, M. *et al.* (2007) Mapping intramolecular interactions between domains in HMGB1 using a tail-truncation approach. *J. Mol. Biol.* 374, 1286–1297
43 Stott, K. *et al.* (2010) Tail-mediated collapse of HMGB1 is dynamic and occurs via differential binding of the acidic tail to the A and B domains. *J. Mol. Biol.* 403, 706–722
44 Zhou, Q. *et al.* (2009) Dss1 regulates interaction of Brh2 with DNA. *Biochemistry* 48, 11929–11938
45 Chandra, V. *et al.* (2008) Structure of the intact PPAR-gamma-RXR-alpha nuclear receptor complex on DNA. *Nature* 456, 350–356
46 Vuzman, D. *et al.* (2010) Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. *Biophys. J.* 99, 1202–1211
47 Hassler, M. and Richmond, T.J. (2001) The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex. *EMBO J.* 20, 3018–3028
48 Garvie, C.W. and Wolberger, C. (2001) Recognition of specific DNA sequences. *Mol. Cell* 8, 937–946
49 Klemm, J.D. *et al.* (1994) Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* 77, 21–32
50 Botfield, M.C. *et al.* (1992) Biochemical characterization of the Oct-2 POU domain with implications for bipartite DNA recognition. *Biochemistry* 31, 5841–5848
51 van Leeuwen, H.C. *et al.* (1997) Linker length and composition influence the flexibility of Oct-1 DNA binding. *EMBO J.* 16, 2043–2053
52 Olson, K.E. *et al.* (2005) Secondary structure and dynamics of an intrinsically unstructured linker domain. *J. Biomol. Struct. Dyn.* 23, 113–124
53 Dunker, A.K. *et al.* (2008) Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* 18, 756–764
54 Iakoucheva, L.M. *et al.* (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* 32, 1037–1049
55 Fuxreiter, M. *et al.* (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* 23, 950–956
56 Romero, P.R. *et al.* (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8390–8395
57 Diella, F. *et al.* (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.* 13, 6580–6603
58 Jonker, H.R. *et al.* (2006) Gradual phosphorylation regulates PC4 coactivator function. *FEBS J.* 273, 1430–1444
59 Kumar, B.R. *et al.* (2001) p300-mediated acetylation of human transcriptional coactivator PC4 is inhibited by phosphorylation. *J. Biol. Chem.* 276, 16804–16809
60 Smet-Nocca, C. *et al.* (2008) The thymine–DNA glycosylase regulatory domain: residual structure and DNA binding. *Biochemistry* 47, 6519–6530
61 Pan, Y. *et al.* (2009) How do transcription factors select specific binding sites in the genome? *Nat. Struct. Mol. Biol.* 16, 1118–1120
62 Buczek, P. and Horvath, M.P. (2006) Structural reorganization and the cooperative binding of single-stranded telomere DNA in *Sterkiella nova*. *J. Biol. Chem.* 281, 40124–40134
63 Rajasekar, K. *et al.* (2010) Order and disorder in the domain organization of the plasmid partition protein KorB. *J. Biol. Chem.* 285, 15440–15449
64 Minezaki, Y. *et al.* (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* 359, 1137–1149
65 Koizumi, S. *et al.* (1990) Isoforms of the human ets-1 protein: generation by alternative splicing and differential phosphorylation. *Oncogene* 5, 675–681
66 Brown, C.J. *et al.* (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol.* 55, 104–110
67 Kelman, Z. *et al.* (1998) Devoted to the lagging strand-the subunit of DNA polymerase III holoenzyme contacts SSB to promote processive elongation and sliding clamp assembly. *EMBO J.* 17, 2436–2449
68 Vuzman, D. and Levy, Y. (2010) DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21004–21009
69 Hansen, J.C. *et al.* (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem.* 281, 1853–1856
70 McBryant, S.J. *et al.* (2009) Determinants of histone H4 N-terminal domain function during nucleosomal array oligomerization: roles of amino acid sequence, domain length, and charge density. *J. Biol. Chem.* 284, 16716–16722
71 Lu, X. *et al.* (2009) Chromatin condensing functions of the linker histone C-terminal domain are mediated by specific amino acid composition and intrinsic protein disorder. *Biochemistry* 48, 164–172
72 Neduva, V. and Russell, R.B. (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett.* 579, 3342–3345
73 Daughdrill, G.W. *et al.* (2007) Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *J. Mol. Evol.* 65, 277–288
74 Pabo, C.O. and Sauer, R.T. (1984) Protein–DNA recognition. *Annu. Rev. Biochem.* 53, 293–321
75 Lefstin, J.A. and Yamamoto, K.R. (1998) Allosteric effects of DNA on transcriptional regulators. *Nature* 392, 885–888
76 Ma, B. and Nussinov, R. (2009) Amplification of signaling via cellular allosteric relay and protein disorder. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6887–6888
77 Galea, C.A. *et al.* (2008) Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1. *J. Mol. Biol.* 376, 827–838
78 Monod, J. *et al.* (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12, 88–118

79 Koshland, D.E., Jr *et al.* (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5, 365–385
80 Marsh, J.A. and Forman-Kay, J.D. (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J. Mol. Biol.* 391, 359–374
81 Jensen, M.R. *et al.* (2010) Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J. Am. Chem. Soc.* 132, 1270–1272
82 Bertini, I. *et al.* (2010) Conformational space of flexible biological macromolecules from average data. *J. Am. Chem. Soc.* 132, 13553–13558
83 Crane-Robinson, C. *et al.* (2006) The extended arms of DNA-binding domains: a tale of tails. *Trends Biochem. Sci.* 31, 547–552
84 Popovych, N. *et al.* (2006) Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* 13, 831–838
85 He, B. *et al.* (2009) Predicting intrinsic disorder in proteins: an overview. *Cell Res.* 19, 929–949
86 Meszaros, B. *et al.* (2009) Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.* 5, e1000376
87 Vuzman, D. *et al.* (2010) Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J. Mol. Biol.* 396, 674–684
88 Cho, Y. *et al.* (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* 265, 346–355
89 Ghosh, G. *et al.* (1995) Structure of NF-kappa B p50 homodimer bound to a kappa B site. *Nature* 373, 303–310
90 Tretyachenko-Ladokhina, V. *et al.* (2006) Flexibility and adaptability in binding of *E. coli* cytidine repressor to different operators suggests a role in differential gene regulation. *J. Mol. Biol.* 362, 271–286
91 Levy, Y. *et al.* (2007) Fly-casting in protein–DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* 129, 738–739
92 Li, L. *et al.* (1996) pH-dependent enhancement of DNA binding by the ultrabithorax homeodomain. *Biochemistry* 35, 9832–9839
93 Reed, H.C. *et al.* (2010) Alternative splicing modulates Ubx protein function in Drosophila melanogaster. *Genetics* 184, 745–758
94 Tour, E. *et al.* (2005) Evolutionarily conserved domains required for activation and repression functions of the *Drosophila* Hox protein Ultrabithorax. *Development* 132, 5271–5281
95 Adams, V.H. *et al.* (2007) Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J. Biol. Chem.* 282, 15057–15064
96 Aguado-Llera, D. *et al.* (2010) The basic helix–loop–helix region of human neurogenin 1 is a monomeric natively unfolded protein which forms a "fuzzy" complex upon DNA binding. *Biochemistry* 49, 1577–1589
97 Liu, J. and Song, J. (2008) A novel nucleolar transcriptional activator ApLLP for long-term memory formation is intrinsically unstructured but functionally active. *Biochem. Biophys. Res. Commun.* 366, 585–591
98 Savvides, S.N. *et al.* (2004) The C-terminal domain of full-length *E. coli* SSB is disordered even when bound to DNA. *Protein Sci.* 13, 1942–1947
99 Manelyte, L. *et al.* (2009) The unstructured C-terminal extension of UvrD interacts with UvrB, but is dispensable for nucleotide excision repair. *DNA Repair (Amst.)* 8, 1300–1310