

Flexibility and Disorder in Gene Regulation: LacI/GalR and Hox Proteins*

Published, JBC Papers in Press, September 4, 2015, DOI 10.1074/jbc.R115.685032

Sarah E. Bondos^{†,1}, Liskin Swint-Kruse^{§,1},
and Kathleen S. Matthews^{¶,2}

From the [†]Department of Molecular and Cellular Medicine, Texas A&M Health Science Center, College Station, Texas 77843, the [§]Department of Biochemistry and Molecular Biology, the University of Kansas Medical Center, Kansas City, Kansas 66160, and the [¶]Department of Biosciences, Rice University, Houston, Texas 77251

To modulate transcription, a variety of input signals must be sensed by genetic regulatory proteins. In these proteins, flexibility and disorder are emerging as common themes. Prokaryotic regulators generally have short, flexible segments, whereas eukaryotic regulators have extended regions that lack predicted secondary structure (intrinsic disorder). Two examples illustrate the impact of flexibility and disorder on gene regulation: the prokaryotic LacI/GalR family, with detailed information from studies on LacI, and the eukaryotic family of Hox proteins, with specific insights from investigations of Ultrabithorax (Ubx). The widespread importance of structural disorder in gene regulatory proteins may derive from the need for flexibility in signal response and, particularly in eukaryotes, in protein partner selection.

Over the past two decades, molecular flexibility has emerged as critical to protein function. Although not readily apparent in crystal structures, a variety of computer simulations and solution experiments, including NMR, fluorescence, and small angle x-ray scattering, demonstrate widespread flexibility in protein molecules (1). This plasticity ranges across side chain fluctuations, domain motions, and folding transitions to the extreme pliability of intrinsically disordered regions (2).

Although universal, flexibility and disorder are present to different extents in prokaryotic and eukaryotic organisms (3–5). These differences are well illustrated by genetic regulatory proteins; in prokaryotic regulators, flexibility primarily occurs in short regions around specific functional sites. Examples that have been studied in some detail include the biotin

repressor BirA (6), lambda repressor (7), tetracycline repressor family (8), MerR family (9), and the LacI/GalR³ family (10). In contrast, extended regions of disorder are found in genetic regulatory proteins ranging from yeast to humans (e.g. GCN4, p53, BRCA1, and Hox proteins) and are especially evident in the activation domains of eukaryotic transcription regulators (5, 11). As paradigms for these two categories—localized flexibility and intrinsic disorder—we review these roles in two families of transcription regulators that have been extensively studied: LacI/GalR (prokaryotic) and Hox (eukaryotic).

Overview of Structures and Functions

LacI/GalR Proteins

The LacI/GalR family of transcription regulators comprises >4000 homologs; all members of this family are found exclusively in bacteria (10, 12, 13). The common structure of this family is a homodimer that contains one DNA-binding site and two binding sites for small-molecule, allosteric ligands (10). Some members form tetramers by a variety of mechanisms, whereas other homologs bind heteroproteins as part of the regulatory cycle (10). Fig. 1 (A–C) shows the tetrameric structure for the paradigmatic lactose repressor protein (LacI), which we use here to provide an overview of the flexible regions required for transcription regulation by LacI/GalR homologs.

First, a flexible linker connects the DNA- and ligand-binding domains (Fig. 1, A–C) (14, 15). In ~60% of LacI/GalR homologs (13), this linker includes a conserved motif that forms a “hinge helix” in known structures. The side chains of the hinge helices interact with the minor groove at the center of the two DNA half-sites, bending the operator by ~45° (Fig. 1B) (14–17). In this complex, various linker side chains form specific, hydrophobic interactions with operator DNA; thus, the linker-DNA interactions appear to be critical for recognizing specific LacI/GalR operator sequences (14–17). For LacI, the hinge helices remain compact (and presumably folded) even when the LacI-operator complex is bound to its allosteric ligand, inducer IPTG (18). However, when bound to nonspecific DNA, the NMR structure of LacI DNA-binding domains/linkers shows that the hinge helix is unfolded (16). In the absence of any DNA, both NMR and small angle x-ray scattering of full-length LacI show high mobility for the N-terminal DNA-binding domain that accompanies unfolding of the hinge helix (18, 19).

The second flexible region in the LacI/GalR proteins is a three-stranded “pivot” between the N- and C-subdomains of the regulatory domains (20, 21). Changes at this pivot occur when small, allosteric ligands bind the regulatory domain. Binding therefore alters the juxtaposition of the N-subdomains, which “pulls” the hinge helices and provides a key mechanism for altering their orientation and contacts to DNA (14, 21–23).

The third flexible region is unique to *Escherichia coli* LacI. This protein has an additional C-terminal sequence that com-

* This work was supported by grants from the National Science Foundation (NSF) and the Ted Nash Long Life Foundation (to S. E. B.); a Lied Basic Science Grant (to L. S. K.) from the University of Kansas Medical Research Institute Health Clinical and Translational Science Award (National Institutes of Health Grant UL1TR000001, formerly UL1RR033179); and grants from National Institutes of Health (GM22441) and The Robert A. Welch Foundation (C-0576) (to K. S. M.). This minireview is derived in part from the William C. Rose Award Lecture at the American Society for Biochemistry and Molecular Biology 2015 Annual Meeting at Boston, MA, March 30, 2015. The authors declare that they have no conflicts of interest with the contents of this article.

[†] Both authors contributed equally to this work.

² To whom correspondence should be addressed: Dept. of Biosciences, Rice University, MS-140, Houston, Texas 77251. Tel.: 713-348-4871; Fax: 713-522-6782; E-mail: ksm@rice.edu.

³ The abbreviations used are: LacI, lactose repressor protein; GalR, galactose repressor protein; HD, homeodomain; IPTG, isopropyl-β,D-thiogalactoside; Ubx, Ultrabithorax; aa, amino acids.

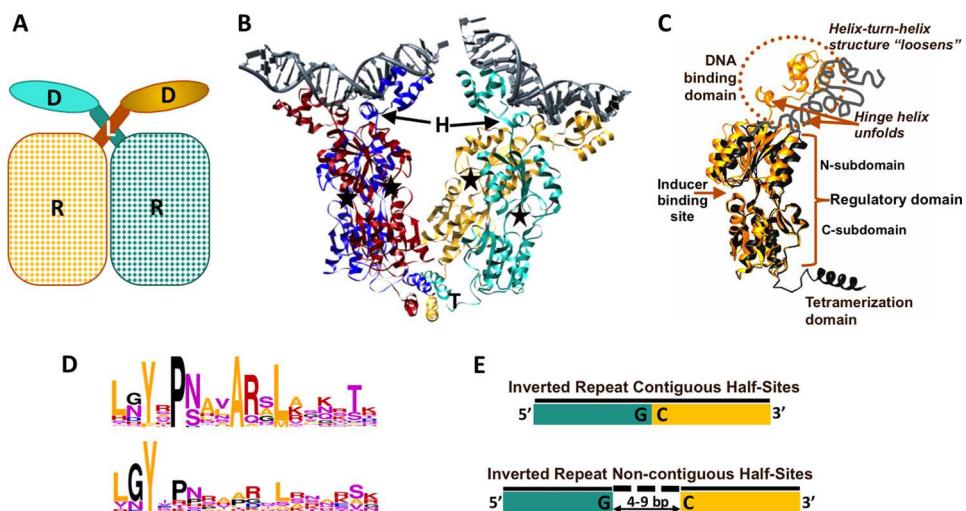


FIGURE 1. LacI/GalR protein structure and DNA targets. *A*, graphic of a LacI/GalR homodimer. *D* indicates DNA-binding domains (gold/cyan ovals); *L* indicates linkers that contain the hinge helix (dark gold/dark cyan bars); and *R* indicates regulatory domains (large, stippled shapes). *B*, structure of LacI tetramer. Each monomer is a different color. The arrows labeled *H* point to the linker hinge helices. Inducer-binding sites are labeled with black stars; DNA is shown at the top of the figure as a twisted ladder. The tetramerization domain (*T*) is a four-helix bundle at the bottom of the figure; note the flexible linkers connecting this domain to each regulatory domain. *C*, overlay of monomers from DNA-bound (Protein Data Bank (PDB) 1LGB, gold) and inducer (IPTG)-bound (PDB 1LBH, black) LacI structures (14). Note that the hinge helix and N-terminal DNA-binding domain are not resolved in the presence of inducer, presumably due to flexibility that arises from hinge helix unfolding; the DNA-binding domain may also become less structured. *D* and *E*, linker sequence variation and DNA sites for subsets of LacI/GalR homologs that contain (top) and lack (bottom) the YPAL motif (13). In YPAL homologs, structures show that amino acids in positions P + 1 through L + 2 fold into an α helix (14, 15, 17). YPAL homologs recognize operators with contiguous DNA half-sites (panel *E*, top). Homologs that lack the YPAL motif (e.g. *E. coli* CytR) bind DNA with half-sites that are more widely spaced (panel *E*, bottom) (13, 27). When examined individually, the sequences of non-YPAL linkers resemble those of intrinsically disordered proteins (13). Logos were created with WebLogo (95).

prises the highly stable tetramerization domain (Fig. 1, *B* and *C*). Flexible linkers join the tetramerization domain to the regulatory domain, allowing the angle between the two dimers to vary (18, 24, 25). For this region, freedom of motion is essential for DNA looping and is discussed further below.

The sequences and roles of these flexible regions vary significantly among LacI/GalR homologs to generate functional diversity (reviewed in Ref. 10). For example, differences in the pivot and N-subdomain interface can lead to alternative regulatory outcomes. LacI is inducible—the consequence of binding its natural allosteric effector is to reduce DNA affinity and hence relieve repression of downstream genes (Fig. 1C). In contrast, PurR is repressible—the consequence of binding its allosteric ligand is to enhance DNA binding and repression (15). In addition, for \sim 40% of homologs, the \sim 18-amino acid linker that connects the core domain to the DNA-binding domain appears to be completely disordered, lacking a hinge helix (13). Similar to eukaryotic intrinsically disordered proteins (26), the linker sequence in these proteins has a high density of charge and/or prolines, although the specific positions vary (Fig. 1D). In these homologs, disorder in the linker appears to have arisen to facilitate binding DNA operators with varied spacing between half-sites (Fig. 1E) (13, 27).

Of the homologs with disordered linkers, *E. coli* CytR is the best studied. For high affinity DNA binding, CytR requires cooperative binding of flanking catabolite repressor proteins (CRPs) (10, 28). The unfolded linkers in CytR allow its two N-terminal DNA-binding domains to bind operators with varied half-site spacing (Fig. 1E) (28). Notably, the disordered linkers do not propagate allosteric information to the DNA-binding domains as found for LacI. Instead, the conformational

change precludes simultaneous binding to catabolite repressor protein and target DNA (29).

The range of functional differences among LacI/GalR family members illustrates how sequence changes in flexible protein regions can introduce functional variation without affecting the overall fold.

Hox Proteins

Within multi-cellular organisms, the family of Hox transcription regulators specifies the identities of many tissues (30, 31). Each Hox homolog regulates a different set of target genes during development to specify cellular position within the organism (e.g. various head or cardiac substructures) and to determine cellular function (30). All Hox proteins contain (i) a conserved DNA-binding domain (“homeodomain”) (32–35) and (ii) a hexapeptide motif that mediates interactions with the Exd/Pbx class of Hox co-factors (Fig. 2A) (31). Hox proteins also contain transcription activation and repression domains that influence functional specificity (e.g. Ref. 36). Large regions of the Hox proteins are intrinsically disordered, as reflected by sequence analyses, striking protease sensitivity, and challenges in protein purification (32) (Fig. 2B). Unlike the LacI/GalR homologs, both domain organization and the locations of regulatory sites (e.g. phosphorylation and splicing sites) vary considerably among Hox family proteins (Fig. 2A) (33, 37–40).

In all Hox proteins, the 60-amino acid DNA-binding homeodomain accounts for only a small fraction of the total sequence (Fig. 2A). Homeodomains contain three helices, the third of which binds the DNA major groove and is stabilized by the other two helices (34). At its N terminus, the homeodomain contains a dynamic, disordered “N-terminal arm” of 9 amino

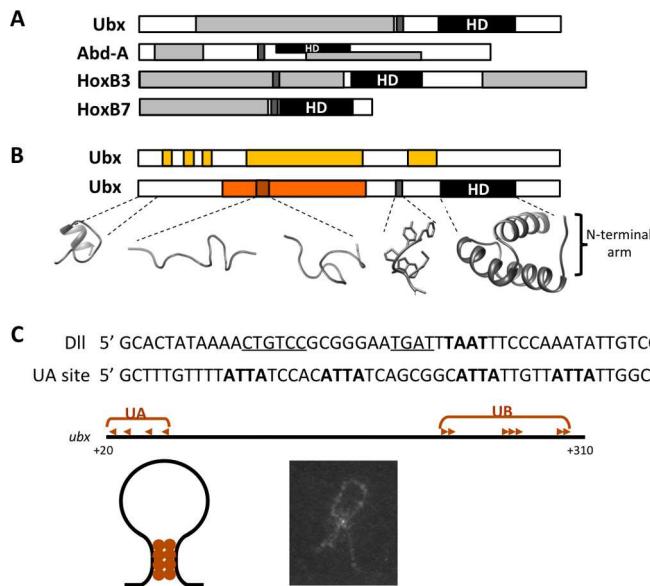


FIGURE 2. Hox protein structure and DNA targets. *A*, bar schematics depicting the functional domains of representative Hox proteins; note that domain organization differs for each protein. *Black regions* represent the homeodomain (HD); *light gray regions* indicate the activation domain(s); *dark gray regions* show the conserved hexapeptide motif. Ubx (389 aa) and Abd-A (330 aa) are derived from *D. melanogaster*; HoxB7 (431 aa), and HoxB3 (217 aa) are from *Homo sapiens* (96) (adapted from Ref. 37). *B*, bar schematics depicting structural and functional domains in Ubx. *Yellow bars* (top) indicate disordered regions of Ubx (32). *Orange bars* (bottom) represent a region of Ubx that is 35% glycine, including 13 consecutive glycines (dark orange) (32). The HD is shown as a *black bar*, and the hexapeptide motif is shown as a *dark gray bar*. Various domains are depicted as structures below. From *left* to *right*: (i) model of N terminus (91); (ii) molecular models, based on molecular dynamics simulations of 13 glycines in a free peptide, for the conserved Ubx polyglycine sequence that demonstrates the range of possible conformations (Justin Drake and B. Montgomery Pettitt, UTMB-Galveston, personal communication); (iii) conserved FYPWMA hexapeptide motif (from PDB file 1B8I) (34); and (iv) Ubx homeodomain (from PDB file 1B8I) (34). *C*, examples of Ubx DNA-binding sites. DII is a composite site bound by a single Ubx protein with partner proteins Homothorax (Hth) and Exd to regulate the *distalless* gene (44). Sequences recognized by Ubx are in **bold**, and sequences recognized by Hth (left *underline*) and Exd (right *underline*) are indicated. The UA-binding region contains four Ubx-binding sites (**bold**) and is part of the *ubx* gene; a linear schematic of a portion of the *ubx* gene is shown beneath, showing a second cluster of binding sites designated as UB (66). When multiple Ubx proteins bind to UA and UB sequences in the *ubx* gene, they interact to create a DNA loop. A loop schematic is shown on the *lower left*, and an electron micrograph of a loop is shown at the *lower right* (reprinted with permission from Ref. 66).

acids. In DNA-bound homeodomains, the N-terminal arm interacts with both bases and backbone phosphates in the DNA minor groove (34, 35). Although the N-terminal arm never adopts a regular secondary structure in this complex, DNA interactions restrict its motion (35). The disordered N-terminal arm facilitates DNA sequence recognition by detecting small, sequence-specific variations in the phosphate positions (35, 41). Finally, the N-terminal arm can also influence contacts between Helix 3 and the major groove (42). Both theoretical and experimental results reveal that binding affinity is highly influenced by the disordered N-terminal arm (e.g. Ref. 43).

One of the best-studied Hox proteins is Ultrabithorax (Ubx) from *Drosophila melanogaster* (Fig. 2B). The Ubx transcription activation domain is glycine-rich (33% *versus* 7% natural abundance generally in proteins), including 13 glycine residues in a row; not surprisingly, this region is extremely disordered (32, 37). Genetic studies have identified numerous DNA sequences that are bound by Ubx *in vivo*. Biochemical studies of Ubx, one

of the few full-length Hox proteins that have been purified, have provided a structure of its DNA-bound homeodomain and identified regions of Ubx that regulate DNA binding (30, 32, 34, 44).

Most Hox proteins, including Ubx, have DNA target sequences that contain a 5'-TAAT-3' sequence (5'-ATTA-3' on the complementary strand) (Fig. 2C) (46, 47). Despite the short length of this sequence, Ubx binds specific sites with high affinity (32, 47). Disordered regions outside the homeodomain can profoundly impact DNA binding and sequence selection, providing an effective mechanism to diversify binding (32, 44). As a consequence, full-length Ubx *in vivo* binds alternative DNA sequences with a much wider array of affinities than does the isolated Ubx homeodomain (44).

Flexibility Enables the Search for DNA-binding Sites

All transcription regulators must recognize their specific cognate DNA sequence among myriad nonspecific sites (48). The strategies used for this process are similar for prokaryotes and eukaryotes, although the latter environment is further complicated by the presence and packing of nucleosomes (49). Nevertheless, all regulatory proteins carry out this task more rapidly than predicted for diffusional search (50). For both prokaryotes and eukaryotes, combinations of sliding, hopping, intersegment transfer (brachiation), and looping yield the most efficient search process (51, 52). As discussed further below, protein flexibility is key to several of these processes. Discerning the modes of transfer can be complex, giving rise to divergent views on search mechanisms (e.g. Ref. 52).

Sliding

Once a protein associates with nonspecific DNA, sliding reduces the dimensionality of the search and thereby enhances association rates for specific sites (Fig. 3A) (48, 50, 53). As a specific example from prokaryotes, *in vivo* experiments with LacI indicate that (i) sliding distances are ~45 bp before dissociation from DNA, consistent with theoretical analysis (52), and (ii) obstruction by other DNA-bound proteins occurs (54). The flexibility of the LacI hinge helices appears to be critical to the sliding process because these domains are unfolded when complexed with nonspecific DNA but folded in the operator-bound form in NMR studies (16).

Despite the presence of chromatin structure, sliding is also effective in eukaryotes. For Hox homeodomains, the disordered N-terminal arms play key roles in sliding, with the length and charge of this region driving sliding dynamics (55). Electrostatic interactions dominate binding in the nonspecific complex (51, 53), although the orientation and mode of homeodomain-nonspecific DNA interaction are otherwise similar to the specific complex (unlike other transcription factor families; e.g. Ref. 56).

Hopping

In this mode of transfer, proteins bind to DNA, dissociate, and then rebind DNA at another site (Fig. 3B) (53). The length of the “hop” may be quite short or can cover long distances (49). For both prokaryotic and eukaryotic transcription factors, hopping appears to increase the speed of the search process (51, 53). In addition, hopping provides a mechanism for some eukary-

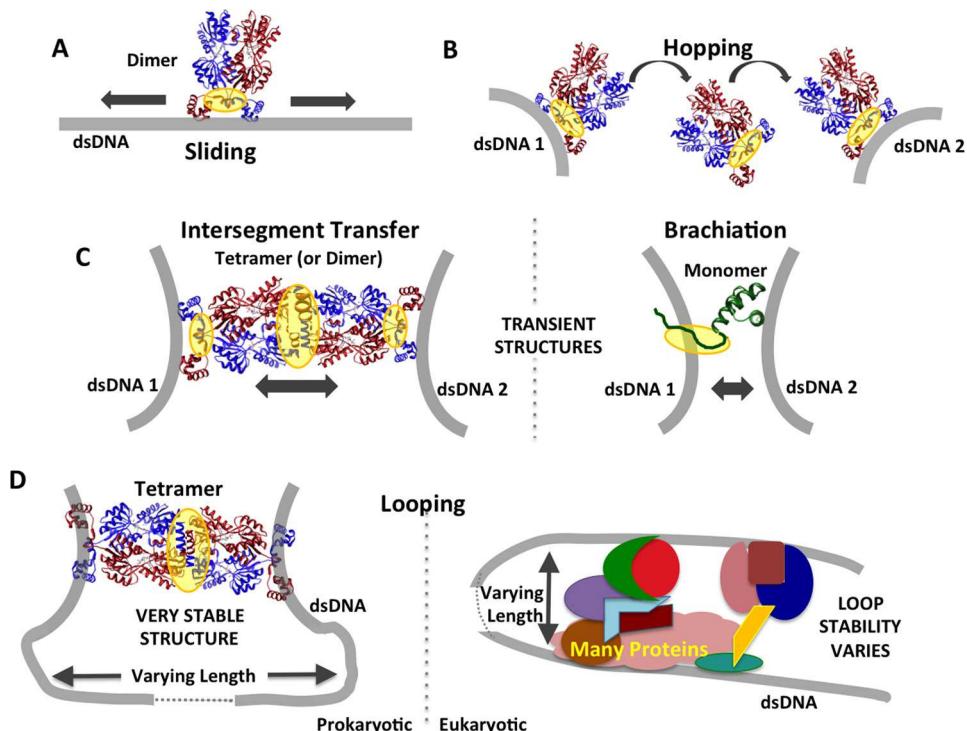


FIGURE 3. Schematics for different search modes of DNA binding. The LacI structure (dimer or tetramer) is used for most of these examples, with requisite flexible regions highlighted with gold ovals. *A*, proteins can slide along the DNA backbone in search of specific binding sites. *B*, proteins can dissociate from DNA and reassociate with the same or a different DNA in a “hopping” process. *C*, intersegment transfer and brachiation allow proteins with multiple binding sites to move from one DNA segment to another via transient contacts to both DNA strands. This type of movement can be accomplished by an oligomeric protein with two DNA-binding sites (e.g. LacI tetramer, *left*) or by a monomer with two separate DNA-binding regions within a single domain (e.g. Ubx homeodomain, *right*, green with flexible N-terminal arm in gold oval). *D*, stable loops can be formed when two DNA-binding domains simultaneously form specific complexes at DNA target sites. The two sites can be separated by stretches of DNA that vary widely in length. Prokaryotic looping (*left*) generally involves single proteins (e.g. LacI tetramer) or a protein assisted by a nearby DNA bending protein (e.g. two GalR dimers and the bending protein HU) (60, 65). In contrast, multi-protein complexes at eukaryotic promoters (*right*) can be highly complex and comprise multiple loops of varying stability that can encompass up to $\sim 10^6$ bp (as in for example Ref. 97).

otic transcription factors to bypass nucleosomes when sliding along DNA (49).

Intersegment Transfer/Brachiation

Intersegment transfer, also called “brachiation” (using appendages to swing from object to object), allows movement from one DNA segment to the next (Fig. 3C) (55). This mechanism is distinct from hopping and is more prominent at high concentrations of DNA, as found *in vivo* (53). Intersegment transfer facilitates searches over long stretches of DNA because regions that are far in sequence space can be close in cellular space (as occurs via extensive packing in many eukaryotic systems) (57). This mechanism requires that two segments of DNA be simultaneously bound by protein. Hence, at least two DNA-binding interfaces are needed on the protein, and sufficient protein flexibility is required (58). The two interfaces can be provided by multimeric assembly, by multiple DNA-binding domains within a monomer, or by monomers with a single, bipartite DNA-binding domain.

For tetrameric LacI, the two dimers provide the requisite two binding interfaces, and flexibility in the segments that link the regulatory domains to the C-terminal tetramerization domain allows variation in dimer orientation (18, 24, 25). For the homeodomain, the intrinsically disordered N-terminal arm, which binds the minor groove, and the third helix, which binds the major groove, provide the two protein-DNA interfaces (55).

This type of interaction accelerated the rate of target recognition by the HoxD9 homeodomain by more than 3 orders of magnitude (59). Thus, the flexibility of the N-terminal arm plus the flexible “joint” between the N-terminal arm and the helical portion of the homeodomain play a critical role in enhancing the rate of searching.

Looping

DNA looping occurs when regulatory proteins or their complexes simultaneously bind two DNA sites (Fig. 3D). Transient looping may occur during brachiation/intersegment transfer, but stable loops persist and impact transcription (60, 61). For example, LacI looped complexes are significantly more stable than LacI bound at a single site (62). In eukaryotes, looping can place enhancers and promoters in direct physical contact (61). Loop formation is influenced by DNA sequence and/or the presence of ancillary proteins (61, 63).

For *E. coli* tetrameric LacI, distances between target operator-binding sites can vary from hundreds to more than a thousand base pairs (62, 64). The natural *lac* operon has a spacing of ~ 400 bp between operators O1 and O2 and ~ 100 bp between O1 and O3 (64). The distances between binding sites, as well as their relative rotation around the DNA helix, can greatly alter transcription (64). In addition, protein flexibility is critical to forming looped structures (24). For tetrameric LacI binding to two operators, the two dimers adopt an “open” conformation

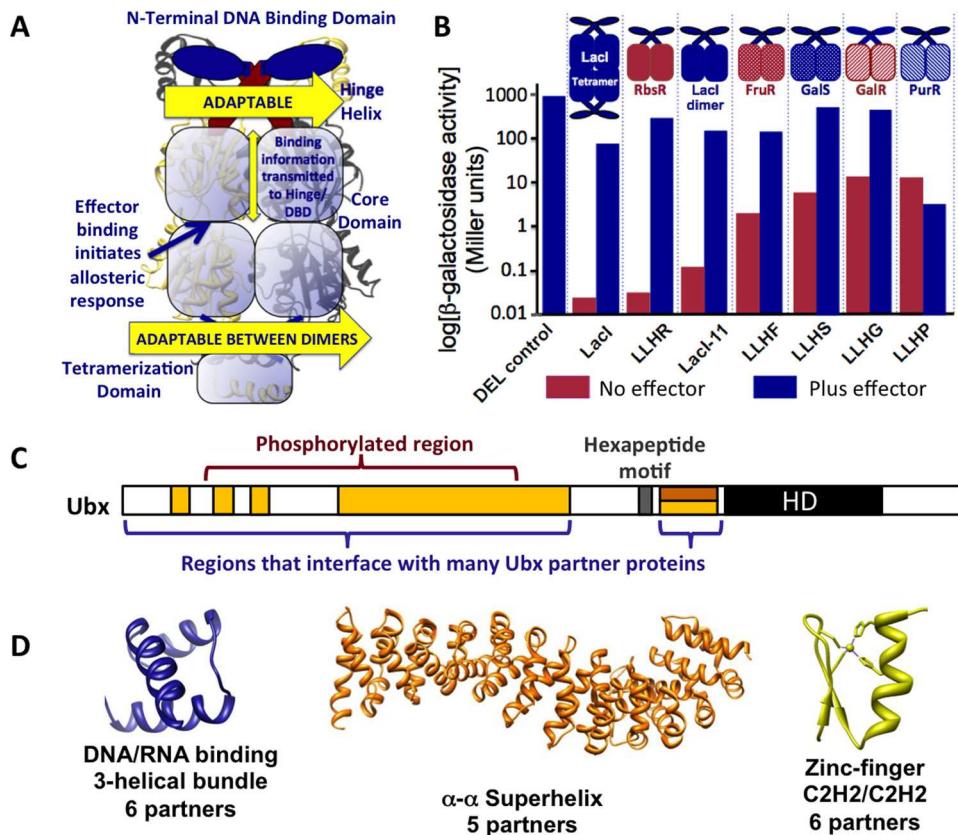


FIGURE 4. Functional regions in LacI/GalR and Ubx. *A*, adaptable regions of the LacI/GalR proteins are shown on a graphic of the homodimer superimposed on a crystal structure (PDB file 1LBG) (14). The flexibility of these regions is critical to strong transcription repression and allosteric regulation. *B*, the flexible interface between the LacI/GalR linkers and regulatory domains can facilitate allosteric response to multiple ligands (79). The LacI DNA-binding domain/linker can be fused to the regulatory domains of other homologs (shown as schematic dimers) to create functional chimeric repressors (red bars in graph) with intact allosteric response to small effector ligands (blue bars). Note that the "LLHP" chimera has the opposite allosteric response of LacI (79). *DEL control* indicates the activity of reporter enzyme in the absence of repressor. *C*, regions within Ubx important for modulating its DNA binding. The upper red bracket indicates the Ubx region that contains phosphorylation sites (38). Blue brackets below indicate sequences that interact with Ubx partner proteins (69). Yellow regions are intrinsically disordered (32). The gold and brown striped region is both spliced and disordered (32, 38). The conserved hexapeptide motif important for Exd interaction is dark gray, and the HD is black. *D*, example structures for three families of proteins that interact with Ubx. For the six partners of the DNA/RNA-binding three-helical bundle family, the engrailed homeodomain is shown (PDB file 1ENH) (98); for the five partners of the α-α superhelix family, β-catenin is shown (PDB file 1QZ7) (99); and for the six partners of the zinc finger C2H2/C2H2 family, Zif268 is shown (PDB file 1A11) (100). Intrinsic disorder of Ubx regions may be key for recognizing this wide range of partners (69).

(i.e. the angle between the two dimers is increased relative to the crystal structure (18)). Chemically cross-linking LacI N termini across two dimers limits dimer-dimer mobility and precludes looping (24). An alternative approach to effect looping is utilized by the homolog GalR, which forms highly stable loops with the assistance of protein HU to facilitate DNA bending (65).

The substantial intrinsic disorder found in eukaryotic regulatory proteins greatly facilitates loop formation. Many eukaryotic transcription regulators, including the Hox proteins, bind to clusters of DNA sites (66). Both side-to-side cooperative Hox binding to Hox-site clusters and back-to-back Hox-Hox interactions between two clusters can enable looped structures (Fig. 2C) (66). Hox proteins can either form loops themselves or recruit large protein complexes, such as the polycomb group proteins, the cohesion complex, and the condensing complex, to bridge distant DNA sequences (67) (Fig. 3D). In addition, Ubx binds other transcription factors that have their own DNA-binding sites near those of Ubx target DNA sequences (45, 68). This arrangement provides opportunities for creating combinatorial loops that are sensitive to cellular conditions and

allow response to cell-signaling stimuli. Importantly, the intrinsically disordered regions of Ubx are required for these heterologous protein interactions (69).

Regulatory Mechanisms Exploit Protein Flexibility/Disorder

Transcription regulation often requires that regulatory proteins alter their DNA binding in response to external signals. Both the LacI/GalR and the Hox proteins utilize flexibility and disorder to transmit this incoming information to the DNA-binding domain.

Allosteric Communication in the LacI/GalR Proteins

Effector binding to LacI/GalR proteins impacts several flexible regions. For *E. coli* LacI, structures of free, DNA-bound, and IPTG-bound protein (14, 70), along with molecular dynamics simulations (23, 71, 72), have been used to study these adaptable regions. The largest changes are found in the linker region and in the N-subdomain interface of LacI (14, 23, 70). Inducer binding alters the juxtaposition of the LacI N-subdomains to bring them into closer contact (Fig. 4A) (73). The

required flexibility in this region has been explored via mutagenesis (74). A key residue is Lys-84, which is buried within the otherwise apolar interface between the N-subdomains and changes positions in the bound and unbound structures (14). When Lys-84 was substituted with Leu or Ala, the allosteric response was diminished to ≤ 10 -fold (as compared with $> 10^4$ -fold for wild type), the kinetics of inducer binding were greatly slowed, and protein stability was significantly enhanced (74, 75).

The motion at the N-subdomain also alters the linker/hinge helix of LacI. The point of closest approach between the two linkers of a dimer is the side chain of Val-52. When this residue was mutated to cysteine, a disulfide bond could be formed that blocked allosteric response to inducer binding (76). Other substitutions at position 52 showed that extrinsic interactions, such as interactions with operator DNA, had more influence on LacI function and allosteric response than did the intrinsic propensity of amino acids for folding the hinge helix (77). The length of the linker region is also important. When 1–3 Glu residues were inserted after the hinge helix, LacI showed progressive decreases in DNA binding affinity and allosteric response (78). Thus, this flexible linker region must be precisely positioned (i) to allow communication between the DNA-binding and regulatory domains and (ii) to align the DNA-binding domains within each dimer.

Nevertheless, linker flexibility facilitates tolerance to significant sequence diversity (Fig. 4B). In fact, fully functional hybrids were created by fusing the LacI DNA-binding domain/linker to regulatory domains from other homologs. Each chimera has the DNA binding specificity of LacI, ligand binding of the parent regulatory domain, and allosteric response defined by the regulatory domain (79). Thus, the interface between the linker and regulatory domains is highly adaptable.

Hox Regulatory Mechanisms

Prokaryotic repressors are generally designed to respond to a limited number of signals, often only one. In contrast, eukaryotic Hox proteins integrate multiple input signals to generate highly specific outcomes unique to the tissue and organism (80). Further, these proteins must differentiate a plethora of DNA sites with both cellular and tissue specificity (30). To that end, many Hox proteins have several splice isoforms (*e.g.* Refs. 44 and 81), a variety of modification sites (*e.g.* phosphorylation) (38, 82, 83), and a number of protein partners (Fig. 4C) (31, 45, 68). These regulatory mechanisms are frequently used to diversify the functions of transcription factors (84). Although these processes typically occur within intrinsically disordered regions, their locations vary among Hox proteins.

In Ubx, all of the regulatory processes are associated with intrinsically disordered regions that also regulate DNA binding (32, 38, 44). To provide an example in each category: (i) when Ubx interacts with partner protein DIP1 via these disordered regions, Ubx transcription activation is precluded *in vivo* (68); (ii) the conserved hexapeptide, which alters DNA binding specificity, and the homeodomain are connected by a disordered linker that varies from 7 to 50 amino acids in length in alternatively spliced isoforms, with the result that Ubx splicing isoforms regulate different genes and construct different tissues *in*

vivo (39, 40, 85); and (iii) Ubx is phosphorylated within the disordered region of the transcription activation domain in a tissue-specific manner, suggesting a regulatory function (37, 38).

The disordered regions also mediate Ubx binding to a variety of heteroprotein partners, a critical element in Hox protein function (45, 68, 69, 86). Hox proteins bind to components of the transcription machinery (45, 87), as well as to other specific transcription factors, to facilitate Hox regulation of the correct subset of genes in different tissues (Fig. 4D) (45, 88, 89). For Ubx partners identified by yeast two-hybrid methods, two key elements have emerged: (i) binding to many of these partners requires the disordered regions within Ubx and (ii) partners can be classified into specific “folds” (69). Indeed, of the selected topologies, three folds include at least five Ubx partners, jointly representing more than half of known Ubx partner proteins (Fig. 4D). Different structural families preferentially bind different disordered segments and splice isoforms of Ubx (69).

These regulatory mechanisms can influence one another (80). For example, alternative splicing impacts Hox binding to other proteins (39, 69). Likewise, phosphorylation of Hox proteins can impact protein interactions and cooperative DNA binding (90). Thus, regions that exhibit intrinsic disorder have the potential to integrate multiple sources of information to regulate and coordinate Hox functions.

Interestingly, the various disordered regions of Ubx can be distorted to allow formation of biomaterials (91). Deleting the disordered regions precludes self-assembly (92, 93). Two consequences of intrinsic disorder have the potential to make these materials commercially useful: (i) Ubx fibers are remarkably strong and extensible (91, 92) and (ii) the disordered regions allow fiber formation to accommodate a wide range of other proteins fused to the Ubx sequence (94).

Conclusion

Although prokaryotic and eukaryotic proteins exhibit many unique features, flexibility has emerged as key to transcription regulation in both kingdoms. This feature of proteins permits regulatory proteins to adapt to varied spacing among DNA-binding sites and to engage multiple mechanisms of searching for and binding to DNA target sites. Flexibility allows the variety of protein interactions required to construct complex DNA structures such as loops, either by direct binding or through interactions with other proteins. Finally, flexibility, and indeed in some cases, extensive disorder are required for regulation of transcription factor function through allosteric ligand binding, protein sequence alterations (splicing and/or posttranslational modifications), and/or protein–protein interactions. The multiple modes by which flexibility enables transcription regulation generate both diverse and highly effective mechanisms for an organism to respond to a varied local cellular environment as well as features essential for the development and function of multicellular organisms.

Acknowledgments—We express our appreciation for molecular models derived from simulations provided by Justin Drake and B. Montgomery Pettitt, University of Texas Medical Branch-Galveston.

References

- DuBay, K. H., Bowman, G. R., and Geissler, P. L. (2015) Fluctuations within folded proteins: implications for thermodynamic and allosteric regulation. *Acc. Chem. Res.* **48**, 1098–1105
- Tompa, P., and Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* **33**, 2–8
- Pancsa, R., and Tompa, P. (2012) Structural disorder in eukaryotes. *PLoS ONE* **7**, e34687
- Xue, B., Dunker, A. K., and Uversky, V. N. (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* **30**, 137–149
- Guo, X., Bulyk, M. L., and Hartemink, A. J. (2012) Intrinsic disorder within and flanking the DNA-binding domains of human transcription factors. *Pac. Symp. Biocomput.* 104–115, 10.1142/9789814366496_0011
- Eginton, C., Cressman, W. J., Bachas, S., Wade, H., and Beckett, D. (2015) Allosteric coupling via distant disorder-to-order transitions. *J. Mol. Biol.* **427**, 1695–1704
- Batabyal, S., Choudhury, S., Sao, D., Mondol, T., and Kumar Pal, S. (2014) Dynamical perspective of protein-DNA interaction. *Biomol. Concepts* **5**, 21–43
- Dover, L. G., Corsino, P. E., Daniels, I. R., Cocklin, S. L., Tatituri, V., Besra, G. S., and Fütterer, K. (2004) Crystal structure of the TetR/CamR family repressor *Mycobacterium tuberculosis* EthR implicated in ethionamide resistance. *J. Mol. Biol.* **340**, 1095–1105
- Chang, C.-C., Lin, L.-Y., Zou, X.-W., Huang, C.-C., and Chan, N.-L. (2015) Structural basis of the mercury(II)-mediated conformational switching of the dual-function transcriptional regulator MerR. *Nucleic Acids Res.* **43**, 7612–7623
- Swint-Kruse, L., and Matthews, K. S. (2009) Allostery in the LacI/GalR family: variations on a theme. *Curr. Opin. Microbiol.* **12**, 129–137
- Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873–6888
- Weickert, M. J., and Adhya, S. (1992) A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.* **267**, 15869–15874
- Tungtr, S., Parente, D. J., and Swint-Kruse, L. (2011) Functionally important positions can comprise the majority of a protein's architecture. *Proteins* **79**, 1589–1608
- Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247–1254
- Schumacher, M. A., Choi, K. Y., Lu, F., Zalkin, H., and Brennan, R. G. (1995) Mechanism of corepressor-mediated specific DNA binding by the purine repressor. *Cell* **83**, 147–155
- Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386–389
- Schumacher, M. A., Allen, G. S., Diel, M., Seidel, G., Hillen, W., and Brennan, R. G. (2004) Structural basis for allosteric control of the transcription regulator CcpA by the phosphoprotein HPr-Ser46-P. *Cell* **118**, 731–741
- Taraban, M., Zhan, H., Whitten, A. E., Langley, D. B., Matthews, K. S., Swint-Kruse, L., and Trehewella, J. (2008) Ligand-induced conformational changes and conformational dynamics in the solution structure of the lactose repressor protein. *J. Mol. Biol.* **376**, 466–481
- Wade-Jardetzky, N., Bray, R. P., Conover, W. W., Jardetzky, O., Geisler, N., and Weber, K. (1979) Differential mobility of the N-terminal headpiece in the Lac-repressor protein. *J. Mol. Biol.* **128**, 259–264
- Swint-Kruse, L., Zhan, H., Fairbanks, B. M., Maheshwari, A., and Matthews, K. S. (2003) Perturbation from a distance: mutations that alter LacI function through long-range effects. *Biochemistry* **42**, 14004–14016
- Mowbray, S. L., and Björkman, A. J. (1999) Conformational changes of ribose-binding protein and two related repressors are tailored to fit the functional need. *J. Mol. Biol.* **294**, 487–499
- Xu, J., and Matthews, K. S. (2009) Flexibility in the inducer binding region is crucial for allostery in the *Escherichia coli* lactose repressor. *Biochemistry* **48**, 4988–4998
- Flynn, T. C., Swint-Kruse, L., Kong, Y., Booth, C., Matthews, K. S., and Ma, J. (2003) Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. *Protein Sci.* **12**, 2523–2541
- Rutkuskas, D., Zhan, H., Matthews, K. S., Pavone, F. S., and Vanzi, F. (2009) Tetramer opening in LacI-mediated DNA looping. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16627–16632
- Vanzi, F., Broggio, C., Sacconi, L., and Pavone, F. S. (2006) Lac repressor hinge flexibility and DNA looping: single molecule kinetics by tethered particle motion. *Nucleic Acids Res.* **34**, 3409–3420
- Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427
- Francke, C., Kerkhoven, R., Wels, M., and Siezen, R. J. (2008) A generic approach to identify transcription factor-specific operator motifs; inferences for LacI-family mediated regulation in *Lactobacillus plantarum* WCF51. *BMC Genomics* **9**, 145
- Tretyachenko-Ladokhina, V., Cocco, M. J., and Senevir, D. F. (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
- Kallipolitis, B. H., and Valentin-Hansen, P. (2004) A role for the interdomain linker region of the *Escherichia coli* CytR regulator in repression complex formation. *J. Mol. Biol.* **342**, 1–7
- Pearson, J. C., Lemons, D., and McGinnis, W. (2005) Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* **6**, 893–904
- Moens, C. B., and Selleri, L. (2006) Hox cofactors in vertebrate development. *Dev. Biol.* **291**, 193–206
- Liu, Y., Matthews, K. S., and Bondos, S. E. (2008) Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the *Drosophila* Hox protein Ultrabithorax. *J. Biol. Chem.* **283**, 20874–20887
- McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., and Gehring, W. J. (1984) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* **37**, 403–408
- Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999) Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature* **397**, 714–719
- Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A., Jacob, V., Aggarwal, A. K., Honig, B., and Mann, R. S. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* **131**, 530–543
- Galant, R., and Carroll, S. B. (2002) Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* **415**, 910–913
- Tan, X. X., Bondos, S., Li, L., and Matthews, K. S. (2002) Transcription activation by Ultrabithorax Ib protein requires a predicted α -helical region. *Biochemistry* **41**, 2774–2785
- Gavis, E. R., and Hogness, D. S. (1991) Phosphorylation, expression and function of the Ultrabithorax protein family in *Drosophila melanogaster*. *Development* **112**, 1077–1093
- Johnson, F. B., Parker, E., and Krasnow, M. A. (1995) Extradenticle protein is a selective cofactor for the *Drosophila* homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 739–743
- Mann, R. S., and Hogness, D. S. (1990) Functional dissection of Ultrabithorax proteins in *D. melanogaster*. *Cell* **60**, 597–610
- Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. (2009) The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248–1253
- Garvie, C. W., and Wolberger, C. (2001) Recognition of specific DNA sequences. *Mol. Cell* **8**, 937–946
- Tóth-Petróczy, A., Simon, I., Fuxreiter, M., and Levy, Y. (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

44. Liu, Y., Matthews, K. S., and Bondos, S. E. (2009) Internal regulatory interactions determine DNA binding specificity by a Hox transcription factor. *J. Mol. Biol.* **390**, 760–774

45. Bondos, S. E., Tan, X. X., and Matthews, K. S. (2006) Physical and genetic interactions link Hox function with diverse transcription factors and cell signaling proteins. *Mol. Cell. Proteomics* **5**, 824–834

46. Berger, M. F., Badis, G., Gehrke, A. R., Talukder, S., Philippakis, A. A., Peña-Castillo, L., Alleyne, T. M., Mnaimneh, S., Botvinnik, O. B., Chan, E. T., Khalid, F., Zhang, W., Newburger, D., Jaeger, S. A., Morris, Q. D., Bulyk, M. L., and Hughes, T. R. (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276

47. Ekker, S. C., Young, K. E., von Kessler, D. P., and Beachy, P. A. (1991) Optimal DNA sequence recognition by the Ultrabithorax homeodomain of *Drosophila*. *EMBO J.* **10**, 1179–1186

48. Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052

49. Gorman, J., Plys, A. J., Visnapuu, M. L., Alani, E., and Greene, E. C. (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.* **17**, 932–938

50. Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948

51. Givaty, O., and Levy, Y. (2009) Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.* **385**, 1087–1097

52. Mahmutovic, A., Berg, O. G., and Elf, J. (2015) What matters for Lac repressor search *in vivo*—sliding, hopping, intersegment transfer, crowding on DNA or recognition? *Nucleic Acids Res.* **43**, 3454–3464

53. Iwahara, J., Zweckstetter, M., and Clore, G. M. (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15062–15067

54. Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E. G., Berg, O. G., and Elf, J. (2012) The Lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598

55. Vuzman, D., Azia, A., and Levy, Y. (2010) Searching DNA via a “Monkey Bar” mechanism: the significance of disordered tails. *J. Mol. Biol.* **396**, 674–684

56. Cuculis, L., Abil, Z., Zhao, H., and Schroeder, C. M. (2015) Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat. Commun.* **6**, 7277

57. Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., and Aiden, E. L. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680

58. Doucette, 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

59. Iwahara, J., and Clore, G. M. (2006) Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *J. Am. Chem. Soc.* **128**, 404–405

60. Becker, N. A., Greiner, A. M., Peters, J. P., and Maher, L. J., 3rd. (2014) Bacterial promoter repression by DNA looping without protein–protein binding competition. *Nucleic Acids Res.* **42**, 5495–5504

61. Erokhin, M., Vassetzky, Y., Georgiev, P., and Chetverina, D. (2015) Eukaryotic enhancers: common features, regulation, and participation in diseases. *Cell. Mol. Life Sci.* **72**, 2361–2375

62. Whitson, P. A., Hsieh, W. T., Wells, R. D., and Matthews, K. S. (1987) Influence of supercoiling and sequence context on operator DNA binding with Lac repressor. *J. Biol. Chem.* **262**, 14592–14599

63. Becker, N. A., and Maher, L. J., 3rd. (2015) High-resolution mapping of architectural DNA binding protein facilitation of a DNA repression loop in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7177–7182

64. Müller, J., Oehler, S., and Müller-Hill, B. (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J. Mol. Biol.* **257**, 21–29

65. Semsey, S., Virnik, K., and Adhya, S. (2006) Three-stage regulation of the amphibolic gal operon: from repressosome to GalR-free DNA. *J. Mol. Biol.* **358**, 355–363

66. Beachy, P. A., Varkey, J., Young, K. E., von Kessler, D. P., Sun, B. I., and Ekker, S. C. (1993) Cooperative binding of an Ultrabithorax homeodomain protein to nearby and distant DNA sites. *Mol. Cell. Biol.* **13**, 6941–6956

67. Luo, L., Yang, X., Takihara, Y., Knoetgen, H., and Kessel, M. (2004) The cell-cycle regulator Geminin inhibits Hox function through direct and polycomb-mediated interactions. *Nature* **427**, 749–753

68. Bondos, S. E., Catanese, D. J., Jr., Tan, X. X., Bicknell, A., Li, L., and Matthews, K. S. (2004) Hox transcription factor Ultrabithorax Ib physically and genetically interacts with Disconnected Interacting Protein 1, a double-stranded RNA-binding protein. *J. Biol. Chem.* **279**, 26433–26444

69. Hsiao, H. C., Gonzalez, K. L., Catanese, D. J., Jr., Jordy, K. E., Matthews, K. S., and Bondos, S. E. (2014) The intrinsically disordered regions of the *Drosophila melanogaster* Hox protein Ultrabithorax select interacting proteins based on partner topology. *PLoS ONE* **9**, e108217

70. Daber, R., Stayrook, S., Rosenberg, A., and Lewis, M. (2007) Structural analysis of Lac repressor bound to allosteric effectors. *J. Mol. Biol.* **370**, 609–619

71. Xu, J., Liu, S., Chen, M., Ma, J., and Matthews, K. S. (2011) Altering residues N125 and D149 impacts sugar effector binding and allosteric parameters in *Escherichia coli* lactose repressor. *Biochemistry* **50**, 9002–9013

72. Swint-Kruse, L., Larson, C., Pettitt, B. M., and Matthews, K. S. (2002) Fine-tuning function: correlation of hinge domain interactions with functional distinctions between LacI and PurR. *Protein Sci.* **11**, 778–794

73. Swint-Kruse, L. (2004) Using networks to identify fine structural differences between functionally distinct protein states. *Biochemistry* **43**, 10886–10895

74. Swint-Kruse, L., Zhan, H., and Matthews, K. S. (2005) Integrated insights from simulation, experiment, and mutational analysis yield new details of LacI function. *Biochemistry* **44**, 11201–11213

75. Bell, C. E., Barry, J., Matthews, K. S., and Lewis, M. (2001) Structure of a variant of Lac repressor with increased thermostability and decreased affinity for operator. *J. Mol. Biol.* **313**, 99–109

76. Falcon, C. M., and Matthews, K. S. (2001) Engineered disulfide linking the hinge regions within lactose repressor dimer increases operator affinity, decreases sequence selectivity, and alters allostery. *Biochemistry* **40**, 15650–15659

77. Zhan, H., Swint-Kruse, L., and Matthews, K. S. (2006) Extrinsic interactions dominate helical propensity in coupled binding and folding of the lactose repressor protein hinge helix. *Biochemistry* **45**, 5896–5906

78. Falcon, C. M., and Matthews, K. S. (2000) Operator DNA sequence variation enhances high affinity binding by hinge helix mutants of lactose repressor protein. *Biochemistry* **39**, 11074–11083

79. Meinhardt, S., Manley, M. W., Jr., Becker, N. A., Hessman, J. A., Maher, L. J., 3rd, and Swint-Kruse, L. (2012) Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. *Nucleic Acids Res.* **40**, 11139–11154

80. Niklas, K. J., Bondos, S. E., Dunker, A. K., and Newman, S. A. (2015) Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. *Front. Cell Dev. Biol.* **3**, 8

81. Stadler, C. R., Vegi, N., Mulaw, M. A., Edmaier, K. E., Rawat, V. P., Dolnik, A., Bullinger, L., Heilmeier, B., Quintanilla-Fend, L., Spiekermann, K., Hiddemann, W., Döhner, K., Döhner, H., Feuring-Buske, M., and Buske, C. (2014) The leukemogenicity of Hoxa9 depends on alternative splicing. *Leukemia* **28**, 1838–1843

82. Wang, H., Bei, L., Shah, C. A., Hu, L., and Eklund, E. A. (2015) HoxA10 terminates emergency granulopoiesis by increasing expression of Triad1. *J. Immunol.* **194**, 5375–5387

83. Taghli-Lamallem, O., Hsia, C., Ronshaugen, M., and McGinnis, W. (2008) Context-dependent regulation of Hox protein functions by CK2 phosphorylation sites. *Dev. Genes Evol.* **218**, 321–332

84. Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M., and Iakoucheva, L. M. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes.

PLoS Comput. Biol. **2**, e100

85. Reed, H. C., Hoare, T., Thomsen, S., Weaver, T. A., White, R. A., Akam, M., and Alonso, C. R. (2010) Alternative splicing modulates Ubx protein function in *Drosophila melanogaster*. *Genetics* **184**, 745–758

86. Baéza, M., Viala, S., Heim, M., Dard, A., Hudry, B., Duffraisse, M., Rogulja-Ortmann, A., Brun, C., and Merabet, S. (2015) Inhibitory activities of short linear motifs underlie Hox interactome specificity *in vivo*. *eLife* **4**, e06034, 10.7554/eLife.06034

87. Boube, M., Hudry, B., Immarigeon, C., Carrier, Y., Bernat-Fabre, S., Merabet, S., Graba, Y., Bourbon, H. M., and Cribbs, D. L. (2014) *Drosophila melanogaster* Hox transcription factors access the RNA polymerase II machinery through direct homeodomain binding to a conserved motif of mediator subunit Med19. *PLoS Genet.* **10**, e1004303

88. Slattery, M., Ma, L., Négre, N., White, K. P., and Mann, R. S. (2011) Genome-wide tissue-specific occupancy of the Hox protein Ultrabithorax and Hox cofactor Homothorax in *Drosophila*. *PLoS ONE* **6**, e14686

89. Merabet, S., and Dard, A. (2014) Tracking context-specific transcription factors regulating Hox activity. *Dev. Dyn.* **243**, 16–23

90. Berry, M., and Gehring, W. (2000) Phosphorylation status of the SCR homeodomain determines its functional activity: essential role for protein phosphatase 2A,B'. *EMBO J.* **19**, 2946–2957

91. Howell, D., Tsai, S., Churion, K., Patterson, J., Abbey, C., Atkinson, J., Porterpan, D., You, Y., Meissner, K., Bayless, K., and Bondos, S. (2015) Identification of multiple dityrosine bonds in materials composed of the *Drosophila* protein Ultrabithorax. *Adv. Funct. Mater.* 10.1002/adfm.201502852

92. Huang, Z., Lu, Y., Majithia, R., Shah, J., Meissner, K., Matthews, K. S., Bondos, S. E., and Lou, J. (2010) Size dictates mechanical properties for protein fibers self-assembled by the *Drosophila* Hox transcription factor Ultrabithorax. *Biomacromolecules* **11**, 3644–3651

93. Greer, A. M., Huang, Z., Oriakhi, A., Lu, Y., Lou, J., Matthews, K. S., and Bondos, S. E. (2009) The *Drosophila* transcription factor Ultrabithorax self-assembles into protein-based biomaterials with multiple morphologies. *Biomacromolecules* **10**, 829–837

94. Tsai, S.-P., Howell, D. W., Huang, Z., Hsiao, H.-C., Lu, Y., Matthews, K. S., Lou, J., and Bondos, S. E. (2015) The effect of protein fusions on the production and mechanical properties of protein-based materials. *Adv. Funct. Mater.* **25**, 1442–1450, 10.1002/adfm.201402997

95. Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190

96. Viganò, M. A., Di Rocco, G., Zappavigna, V., and Mavilio, F. (1998) Definition of the transcriptional activation domains of three human HOX proteins depends on the DNA-binding context. *Mol. Cell. Biol.* **18**, 6201–6212

97. Atchison, M. L. (2014) Function of YY1 in long-distance DNA interactions. *Front. Immunol.* **5**, 45

98. Clarke, N. D., Kissinger, C. R., Desjarlais, J., Gilliland, G. L., and Pabo, C. O. (1994) Structural studies of the engrailed homeodomain. *Protein Sci.* **3**, 1779–1787

99. Xing, Y., Clements, W. K., Kimelman, D., and Xu, W. (2003) Crystal structure of a β -catenin/axin complex suggests a mechanism for the β -catenin destruction complex. *Genes Dev.* **17**, 2753–2764

100. Elrod-Erickson, M., Benson, T. E., and Pabo, C. O. (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6**, 451–464