Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Order through disorder: The role of intrinsically disordered regions in transcription factor binding specificity

Sagie Brodsky[a], Tamar Jana[a] and Naama Barkai

Abstract

Transcription factors (TFs) must bind at specific genomic locations to accurately regulate gene expression. The ability of TFs to recognize specific DNA sequence motifs arises from the inherent preferences of their globular DNA-binding domains (DBDs). Yet, these preferences are insufficient to explain the *in vivo* TF binding site selection. TFs are enriched with intrinsically disordered regions (IDRs), most of which are poorly characterized. While not generally considered as determinants of TF binding specificity, IDRs guide protein–protein interactions within transcriptional condensates, and multiple examples exist in which short IDRs flanking the DBD contribute to binding specificity via direct contact with the DNA. We recently reported that long IDRs, present away from the DBD, act as major specificity determinants at the genomic scale. Here, we discuss mechanisms through which IDRs contribute to DNA binding specificity, highlighting the role of long IDRs in dictating the *in vivo* binding site selection.

Addresses

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, 76100, Israel

Corresponding author: Barkai, Naama (naama.barkai@weizmann.ac.il)

[a] These authors contributed equally.

Current Opinion in Structural Biology 2021, **71**:110–115

This review comes from a themed issue on **Protein Nucleic Acid Interactions**

Edited by **Gordon Hager** & **Hitoshi Kurumizaka**

For a complete overview see the Issue and the Editorial

https://doi.org/10.1016/j.sbi.2021.06.011

Keywords

Transcription factors, DNA binding, Intrinsically disordered regions.

The regulation of gene expression depends on the binding of transcription factors (TFs) to upstream gene regulatory regions. When viewed across the genome, TFs localize preferentially to specific DNA sequences, called *cis*-regulatory motifs. These motif preferences are defined by the inherent binding affinities of the TF globular DNA-binding domains (DBDs). Yet, the DBD preferences only partially explain the *in vivo* TF-binding profiles because only a small fraction of genomic sites containing these short sequence motifs is TF occupied. How TFs distinguish the relevant motif-containing sites, to which they bind, from nonrelevant sites that remain unbound is a fundamental question pertinent to gene expression in all eukaryotic cells.

At the molecular level, much of what we know about TF binding to DNA comes from three-dimensional structures of DBD–DNA complexes. DNA binding involves nonspecific interactions, through which the DBD attaches to the DNA regardless of its sequence, and specific interactions, where residues within the DBD interact with a particular nucleotide sequence. Nucleotide identities are revealed within the DNA major groove, where each base has a unique hydrogen-bonding signature [1,2]. Accordingly, specific DBD–DNA interactions are commonly localized to the major groove and involve hydrogen bonding between amino acid side chains and individual base pairs [3].

Missing from crystal structures are flexible regions that do not adopt a stable fold. These so-called intrinsically disordered regions (IDRs) may transit rapidly between folds, or remain extended. IDRs are recognized by sequence analysis as low-complexity regions lacking hydrophobic residues and are particularly enriched within TFs [4–10]; analysis of 1121 species across the tree of life, for example, revealed that in all eukaryotes (but not in bacteria or archaea), DNA-binding proteins are abundant with such sequences [8]. In fact, in many TFs, IDRs span over hundreds of amino acids and encompass a major fraction of the TF sequence.

The enrichment of IDRs within TFs suggests a general contribution to gene regulation. Activation domains (ADs), for example, are often disordered, which may increase the range and flexibility of their interaction with the general transcriptional machinery. Still, most IDRs within TFs remain uncharacterized.

In this review, we consider the role of IDRs in TF binding specificity. We begin with a brief discussion about transcriptional condensates, an emerging concept referring to the formation of large concentrates of transcription apparatus at particular genomic loci. In most described cases, incorporation of TFs into such condensates depends on multivalent interactions between disordered ADs and components of the general transcriptional machinery, suggesting little specificity. As a

more direct contribution of IDRs to binding site selection, we describe cases where IDRs directly interact with the DNA. For this, we first focus on short flexible segments flanking defined DBDs. In some cases, these flexible regions contribute to DNA binding through base-dependent, minor-groove interactions. We next extend the discussion to long IDRs located away from the DBD. In a large fraction of TFs, IDRs can span over hundreds of amino acids and constitute a major fraction of the TF sequence. Finally, we review our recent results, in which such long IDRs direct TFs to their *in vivo* binding sites through multiple, weak, and partially redundant determinants distributed throughout the entire TF. These IDRs may function through direct DNA interaction or through multivalent protein–protein interactions of the type leading, perhaps, to transcriptional hubs or condensates.

Disordered ADs incorporate TFs into transcriptional condensates

Condensates are membraneless assemblies of biomolecules [11–15]. Transcriptional condensates are composed of a large number of transcription components, including specific TFs, cofactors, and RNA polymerases, as well as signaling factors [16] and transcribed RNAs [17]. Condensates appear to form primarily at super-enhancers and may promote rapid (burst-like) transcription and chromosome conformational changes by bringing together enhancers that share binding of specific TFs or cofactors [18]. Recently, condensates were linked to human diseases, by a study showing that disease-associated expansion of IDR repeats within TFs alters condensate formation and composition [19].

Condensate formation depends on multivalent interaction characteristics of IDRs. Consistently, protein regions required for condensate formation are commonly disordered, and artificial IDRs composed of naturally occurring intrinsically disordered repeats can be designed to generate condensates of defined properties [20]. Cell microscopy reveals clustering of general transcriptional components of various sizes and dynamics, ranging from small and transient to large and stable [15], some of which display properties of phase-separated liquid droplets. Condensates involving TFs of the FET family, for example, were transient and appeared as phase-separated liquid droplets. However, these droplets appeared only at concentrations exceeding those required for transcription [21].

Transcriptional condensates may present a general mean through which TFs activate transcription. Whether they also contribute to TF binding specificity, however, is less clear. In principle, different condensates, localized at specific genomic loci, could incorporate different subsets of specific TFs. Condensates of different compositions are indeed possible: RNA polymerase II, for example, transits between at least two types of condensates involved in transcription and RNA processing, depending on its phosphorylation pattern [22,23]. Furthermore, the polycomb group of repressors may also form unique condensates that help organizing polycomb group–bound chromatin [24].

Yet, in the context of specific TFs, transcriptional condensates appear to depend primarily on interactions between TF ADs and components of the general transcriptional machinery, suggesting little specificity. The mediator complex, for example, emerged as a major element driving condensate formation and as a key for the incorporation of specific TFs, such as Gcn4 and Oct4, into such condensates [13]. Other general coactivators that form phase condensates through interactions with ADs of TFs include the histone acetyltransferase p300 [16] and the TATA-binding protein-associated factor TAF15 whose condensate-forming ability is enhanced by its interactions with the C-terminal domain of RNA polymerase II [*25].

In yet other examples, condensate-forming IDRs could be replaced by IDRs of general factors. NELF, for example, a factor that represses transcription by impairing RNA polymerase II elongation, rapidly forms nuclear condensates upon stress in human cells. The formation of these condensates depends on an IDR within NELF, yet this region can be functionally replaced by IDRs of the FUS or EWSR1 proteins [**26]. Similarly, the pioneering activity of the EBF1 TF depends on a prion-like domain which allows its association with FUS-generated condensates. Yet, this essential function was replaceable by heterologous prion-like domains [27]. Therefore, although in principle, the incorporation of TFs into specific transcriptional condensates could guide their binding specificity, this is not yet well supported by existing evidence.

Direct DNA binding by short disordered segments flanking the DBD

Most known DBDs have a characteristic fold. Yet, exceptions, such as the disordered basic and AT-hook domains, exist. These exceptions best exemplify the ability of IDRs to guide specific DNA binding [4]: the disordered AT-hook, for example, binds DNA using a conserved peptide motif P–R–G–R–P that specifically recognizes minor-groove A/T-rich DNA [4,28].

IDRs are quite frequent in regions flanking TF DBDs [29] and may contribute to DNA binding. Early observations supporting this came from studies of the phage λ repressor. Similar to other DBDs, the DNA binding of the λ repressor involves an alpha helix that is inserted into the major groove of the DNA. *In vivo* and *in vitro* studies, however, revealed an additional N-terminal

'arm', consisting of five predominantly basic residues that extend along the major groove to the opposite side of the DNA. This extended region contributes to both the specificity and the affinity of DNA binding [30,31]. Similar contributions of such flexible regions to DNA binding were characterized in additional cases, including the *Caenorhabditis elegans* SKN-1 protein [32,33], the human estrogen-related receptors, hERR [34,35], and the *Drosophila* Hox TFs, where it endows related Hox paralogs with distinct binding preferences [36–41].

DNA interactions by short IDRs found in proximity to the DBD may therefore present a common theme. In most described cases, these interactions localized to the minor groove [32–41]. Unlike the major groove, the minor groove exposes little specifics of the nucleotide sequence, at least with respect to hydrogen bonding. Minor groove recognition may therefore depend mostly on the shape and charge of the DNA. This is often exemplified by localization of positively charged arginine residues to narrow (A/T-containing) and negatively charged minor grooves [3,37,42]. Of note, while shape readout appears most common, in at least one example—the HapB TF—sequence-specific binding within the minor groove was demonstrated [43]. Here, the CCAAT motif was bound by a combination of a helix and a disordered anchor, the latter accelerating the association rate by ~300 folds.

Thermodynamics analysis provided further insight into the energetics of DNA binding by IDRs [44]. In the case of homeodomains, the disordered tail significantly contributed to binding through nonelectrostatic (and therefore specific) interactions. By contrast, the disordered extensions of two TFs of the HMG family contributed to binding affinity largely through electrostatic interactions and were therefore judged as nonspecific.

Do long IDRs located away from the DBD play a role in TF binding specificity?

The ability to crystalize short IDRs flanking the DBD while DNA bound allows for direct visualization of their interaction with the DNA. These short and disordered segments, however, represent only a small fraction of IDRs within TFs. In fact, TF sequences are particularly enriched in long IDRs that are present away from the DBDs. For example, over 25% of human TFs are predicted to be mostly disordered (>50% of the protein sequence) [6]. These disordered regions can become exceedingly long, reaching hundreds of amino acids. Such IDRs are more challenging to study, as they are difficult to purify and are absent from crystal structures. Furthermore, these sequences are poorly conserved when analyzed using existing (alignment-based) sequence analysis tools, making it hard to predict functionally important regions. The contribution of such long IDRs to DNA binding remained relatively unexplored.

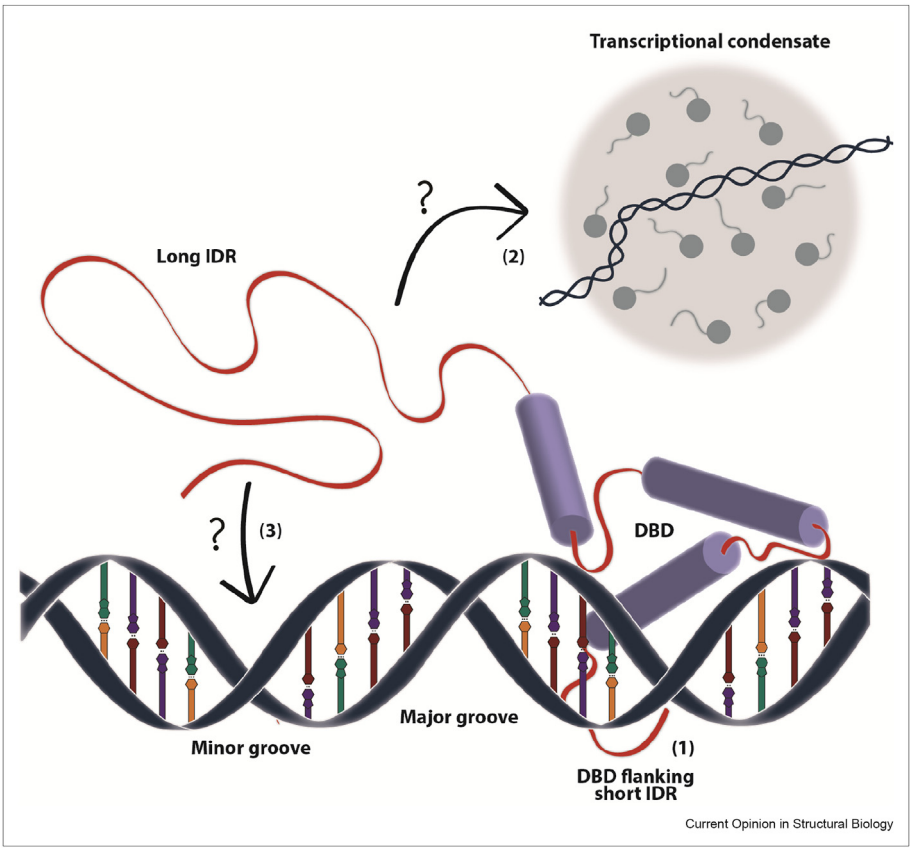
A role of an intermediate length IDR in TF binding specificity was described in the case of the herpes virus TF ICP4 [45]. Here, specific interactions between the DNA and a ~30–amino acid IDR, located away from the globular DBD, were revealed *in vitro* through a combination of crystal structures and solution-based methods, such as nuclear magnetic resonance (NMR). It was found that the known ICP4 consensus motif (RTCGTCNNYNYSG) combines the binding preferences of a globular domain, which contacts the first four nucleotides (RTCG), with those of the relatively distant IDR, which binds the downstream fuzzier sequence (*YNYSG*).

An example supporting the role of long IDRs in directing TF binding *in vivo* came from a recent single-molecule analysis of the glucocorticoid receptor (GR) TF [**46]. This analysis uncovered two distinct states of limited GR mobility within the nucleus. One of these, of smaller confinement but longer residence time, was lost upon deletion of its 400–amino acid long IDR, whereas the second was dependent on the DBD. Of note, this IDR-based confinement appears to affect DNA binding, as this same IDR deletion resulted in the loss of most of the GR *in vivo* binding sites, as assayed by ChIP-seq, although the truncated TF still localized to sites containing the DBD-preferred motif. Therefore, this long IDR is instrumental for the navigation of the GR TF within the nuclear environment and helps detect or stabilize the binding at a large fraction of its *in vivo* binding sites.

In a recent study, we took a broader *in vivo* approach to examine whether long IDRs, composed of hundreds of amino acids, contribute to TF binding specificity at the genomic scale [**47]. As models, we considered two well-characterized budding yeast TFs, Msn2 and Yap1, that contain long (>500 aa) and almost exclusively disordered sequences outside their DBDs. Msn2 is a zinc-finger TF that acts as a master regulator of the general stress response, activating the expression of several dozens of genes in response to a variety of environmental stresses, whereas Yap1 is a member of the basic leucine zipper (bZip) family which activates genes under oxidative stress.

As described previously, *in vitro* analysis enabled to assess the contribution of short IDR segments flanking the DBD in extending the DBD-preferred motif. In the case of *in vivo* binding, the question of specificity is approached more broadly: rather than focusing on motif preference, the key question is how TFs select the subset of motif-containing sites to which they bind. In the case of the Msn2 and Yap1, these *in vitro* motif preferences are well known. We therefore asked whether their long IDRs contribute to the choice of motif-containing sites bound by the TFs.

Figure 1



Possible mechanisms by which IDRs contribute to TF binding specificity: Short IDRs flanking the DBD often stabilize TF binding and contribute to binding specificity by interaction with the opposite DNA minor groove **(1)**. We recently showed that long IDRs, located away from the DBD, play a major role in TF promoter selection. Possible mechanisms generating this IDR-dependent binding septicity include incorporation of TFs into condensates localized at particular genomic regions **(2)**. Alternatively, promoter recognition might be achieved by direct interaction of long TF IDRs with DNA **(3)**.

We examined the role of different regions within the TFs in directing *in vivo* binding to their selected genomic binding sites. For this, we used spatially resolved profiling of binding locations, coupled with extensive manipulations of the TF sequences. This analysis revealed that the long IDRs, of both Msn2 and Yap1, act as major specificity determinants and are in fact required and sufficient for the localization of the TFs to most of their target promoters. When testing the binding profile of a TF variant that contains only the DBD (lacking the extended IDR), we found that it localizes to sites containing the preferred motif, as expected, but these selected sites were different from the ones bound by the intact TF. A variant that lacks the DBD, on the other hand, does not localize at the same preferred motif, but does recognize most promoters bound by the intact TF. Furthermore, we found that these IDRs direct promoter binding through a new paradigm, which we term 'distributed specificity': multiple, weak, and partially redundant determinants are distributed throughout the disordered sequence and contribute to promoter selection in a cumulative manner. Interestingly, these specificity determinants are conserved over long evolutionary distances and among orthologs that show little (alignment-based) sequence similarity. Defining sequence analysis tools compatible with IDRs is a contemporary challenge that will greatly benefit the search for the molecular grammar underling IDR-based promoter recognition [*48].

Conclusion

In this review, we presented the hypothesis that long IDRs, which are prevalent among TF sequences, play a role in directing TFs to their preferred binding sites along the genome. What could be the mechanistic basis for this IDR-promoter recognition? Different models can be envisioned (Figure 1). First, IDRs allow multivalent interactions that could support diverse protein–protein interactions. Interactions of disordered ADs with components of the general transcriptional machinery, for example, incorporate TFs into transcriptional condensates localized at particular genomic

regions. Yet, current evidence provides little support for the role of such condensates in TF binding specificity. IDRs could also be involved in other, more specific interactions with DNA-binding proteins or with chromatin features. Alternatively, IDRs could directly interact with DNA, a possibility we focused on in this review. *In vitro* analysis was highly instrumental in revealing specific DNA binding of short IDRs flanking DBDs, and there is also scattered evidence suggesting specific interactions between moderate-sized IDRs and DNA. Such data are not yet available for long IDRs containing hundreds of residues, at least in part because of the difficulty of purifying and handling such regions. Computational analysis suggests an enrichment of TF IDRs with molecular recognition features, most notably features predicted to form a helical structure upon binding to nucleic acids or protein partners [4,49]. Additional studies are required to distinguish these possibilities and to examine the general use of IDRs as a mediator of TF binding specificity.

Conflict of interest statement

Nothing declared.

Acknowledgment

This project was supported by the ERC, Minerva and BSF-NSF.

References

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

* of special interest
** of outstanding interest

1. Seeman NC, *et al.*: **Sequence-specific recognition of double helical nucleic acids by proteins**. *Proc Natl Acad Sci USA* 1976, **73**:804–808.
2. Garvie CW, Wolberger C: **Recognition of specific DNA sequences**. *Mol Cell* 2001, **8**:937–946.
3. Rohs R, *et al.*: **Origins of specificity in protein-DNA recognition**. *Annu Rev Biochem* 2010, **79**:233–269.
4. Liu J, *et al.*: **Intrinsic disorder in transcription factors**. *Biochemistry* 2006, **45**:6873–6888.
5. Ward JJ, *et al.*: **Prediction and functional analysis of native disorder in proteins from the three kingdoms of life**. *J Mol Biol* 2004, **337**:635–645.
6. Minezaki Y, *et al.*: **Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation**. *J Mol Biol* 2006, **359**:1137–1149.
7. Peng Z, *et al.*: **Genome-scale prediction of proteins with long intrinsically disordered regions**. *Proteins: Struct, Funct, Bioinf* 2014, **82**:145–158.
8. Wang C, *et al.*: **Disordered nucleiome: abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea**. *Proteomics* 2016, **16**:1486–1498.
9. Skupien-Rabian B, *et al.*: **Proteomic and bioinformatic analysis of a nuclear intrinsically disordered proteome**. *J Proteomics* 2016, **130**:76–84.
10. van der Lee R, *et al.*: **Classification of intrinsically disordered regions and proteins**. *Chem Rev* 2014, **114**:6589–6631.
11. Banani SF, *et al.*: **Biomolecular condensates: organizers of cellular biochemistry**. *Nat Rev Mol Cell Biol* 2017, **18**:285–298.
12. Shin Y, Brangwynne CP: **Liquid phase condensation in cell physiology and disease**. *Science* 2017, **357**.
13. Boija A, *et al.*: **Transcription factors activate genes through the phase-separation capacity of their activation domains**. *Cell* 2018, **175**. 1842–1855.e16.
14. Sabari BR, *et al.*: **Coactivator condensation at super-enhancers links phase separation and gene control**. *Science (New York, NY)* 2018, **361**, eaar3958.
15. Cho W-K, *et al.*: **Mediator and RNA polymerase II clusters associate in transcription-dependent condensates**. *Science* 2018, **361**:412–415.
16. Ma L, *et al.*: **Co-condensation between transcription factor and coactivator p300 modulates transcriptional bursting kinetics**. *Mol Cell* 2021, **81**. 1682–1697.e7.
17. Henninger JE, *et al.*: **RNA-mediated feedback control of transcriptional condensates**. *Cell* 2021, **184**. 207–225.e24.
18. Kim S, Shendure J: **Mechanisms of interplay between transcription factors and the 3D genome**. *Mol Cell* 2019, **76**: 306–319.
19. Basu S, *et al.*: **Unblending of transcriptional condensates in human repeat expansion disease**. *Cell* 2020, **181**. 1062–1079.e30.
20. Dzuricky M, *et al.*: **De novo engineering of intracellular condensates using artificial disordered proteins**. *Nat Chem* 2020, **12**:814–825.
21. Chong S, *et al.*: **Imaging dynamic and selective low-complexity domain interactions that control gene transcription**. *Science* 2018, **361**, eaar2555.
22. Portz B, Shorter J: **Switching condensates: the CTD code goes liquid**. *Trends Biochem Sci* 2020, **45**:1–3.
23. Guo YE, *et al.*: **Pol II phosphorylation regulates a switch between transcriptional and splicing condensates**. *Nature* 2019, **572**:543–548.
24. Tatavosian R, *et al.*: **Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation**. *J Biol Chem* 2019, **294**:1451–1463.
25. * Wei MT, *et al.*: **Nucleated transcriptional condensates amplify gene expression**. *Nat Cell Biol* 2020, **22**:1187–1196.
Using optogenetic, this study reported that TFs of the FET-family phase separate within living cells and that this process may drive localized transcription
26. ** Rawat P, *et al.*: **Stress-induced nuclear condensation of NELF drives transcriptional downregulation**. *Mol Cell* 2021, **81**. 1013–1026.e11.
This study desmonsteates that NELF undergoes sterss-indeuce condensation. An IDR within NELF is required for this condensation, but can be functionally replaced by IDRs of the FUS or EWSR1 proteins
27. Wang Y, *et al.*: **A prion-like domain in transcription factor EBF1 promotes phase separation and enables B cell programming of progenitor chromatin**. *Immunity* 2020, **53**. 1151–1167.e6.
28. Bewley CA, *et al.*: **Minor groove-binding architectural proteins: structure, function, and DNA recognition**. *Annu Rev Biophys Biomol Struct* 1998, **27**:105–131.
29. Guo X, *et al.*: **Intrinsic disorder within and flanking the DNA-binding domains of human transcription factors. Pacific Symposium on Biocomputing**. *Pac Symp Biocomput* 2012: 104–115.
30. Eliason JL, *et al.*: **NH2-terminal arm of phage lambda repressor contributes energy and specificity to repressor binding and determines the effects of operator mutations**. *Proc Natl Acad Sci USA* 1985, **82**:2339–2343.
31. Jordan, Pabo C: **Structure of the lambda complex at 2.5 A resolution: details of the repressor-operator interactions**. *Science* 1988, **242**:893–899.

32. Rupert PB, *et al.*: **A new DNA-binding motif in the Skn-1 binding domain–DNA complex**. *Nat Struct Biol* 1998, **5**: 484–491.

33. Kophengnavong T, *et al.*: **The SKN-1 amino-terminal arm is a DNA specificity segment**. *Mol Cell Biol* 1999, **19**:3039–3050.

34. Gearhart MD, *et al.*: **Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition**. *J Mol Biol* 2003, **327**:819–832.

35. Meinke G, Sigler PB: **DNA-binding mechanism of the monomeric orphan nuclear receptor NGFI-B**. *Nat Struct Biol* 1999, **6**: 471–477.

36. Passner JM, *et al.*: **Structure of a DNA-bound ultrabithorax–extradenticle homeodomain complex**. *Nature* 1999, **397**:714–719.

37. Joshi R, *et al.*: **Functional specificity of a Hox protein mediated by the recognition of minor groove structure**. *Cell* 2007, **131**: 530–543.

38. Slattery M, *et al.*: **Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins**. *Cell* 2011, **147**: 1270–1282.

39. Merabet S, Mann RS: **To be specific or not: the critical relationship between Hox and TALE proteins**. *Trends Genet* 2016, **32**:334–347.

40. Abe N, *et al.*: **Deconvolving the recognition of DNA shape from sequence**. *Cell* 2015, **161**:307–318.

41. Zeiske T, *et al.*: **Intrinsic DNA shape accounts for affinity differences between Hox-cofactor binding sites**. *Cell Rep* 2018, **24**:2221–2230.

42. Rohs R, *et al.*: **The role of DNA shape in protein–DNA recognition**. *Nature* 2009, **461**:1248–1253.

43. Huber EM, *et al.*: **DNA minor groove sensing and widening by the CCAAT-binding complex**. *Structure* 2012, **20**:1757–1768.

44. Crane-Robinson C, *et al.*: **The extended arms of DNA-binding domains: a tale of tails**. *Trends Biochem Sci* 2006, **31**:547–552.

45. Tunnicliffe RB, *et al.*: **The herpes viral transcription factor ICP4 forms a novel DNA recognition complex**. *Nucleic Acids Res* 2017, **45**:8064–8078.

46. ** Garcia DA, *et al.*: **An intrinsically disordered region-mediated confinement state contributes to the dynamics and function of transcription factors**. *Mol Cell* 1 April 2021, **81**:1484–1498.e6.
Through single-molecule analysis, this study discovered a role of an IDR of the TF GR in its dynamics wtihin the nucleus. IDR deletion led to a loss of a stablized-diffusion mode which correlated with a loss of TF binding from multiple promoters.

47. ** Brodsky S, *et al.*: **Intrinsically disordered regions direct transcription factor in vivo binding specificity**. *Mol Cell* 2020, https://doi.org/10.1016/j.molcel.2020.05.032.
This study shows that Msn2 and Yap1, two budding yeast TFs, are guided to their target promoters through multiple weak and partially redudant specificity determinants distributed wtihin their long (>500 residues) IDRs.

48. * Chong S, Mir M: **Towards decoding the sequence-based grammar governing the functions of intrinsically disordered protein regions**. *J Mol Biol* 2020:166724.
This review paper describes emerging approaches for revealing the molecular grammar defining IDR functionality

49. Oldfield CJ, *et al.*: **Coupled folding and binding with alpha-helix-forming molecular recognition elements**. *Biochemistry* 2005, **44**:12454–12470.