

Review

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms



journal homepage: www.elsevier.com/locate/bbagrm

Experimental approaches to investigate biophysical interactions between homeodomain transcription factors and DNA



Fadwa Mekkaoui^a, Robert A. Drewell^b, Jacqueline M. Dresch^b, Donald E. Spratt^{a,*}

^a Gustaf H. Carlson School of Chemistry and Biochemistry, Clark University, 950 Main Street, Worcester, MA 01610, United States of America ^b Biology Department, Clark University, 950 Main Street, Worcester, MA 01610, United States of America

ARTICLE INFO

Keywords Homeodomain Transcription factor Drosophila Binding affinity Biophysics Protein-DNA

ABSTRACT

Homeodomain transcription factors (TFs) bind to specific DNA sequences to regulate the expression of target genes. Structural work has provided insight into molecular identities and aided in unraveling structural features of these TFs. However, the detailed affinity and specificity by which these TFs bind to DNA sequences is still largely unknown. Qualitative methods, such as DNA footprinting, Electrophoretic Mobility Shift Assays (EMSAs), Systematic Evolution of Ligands by Exponential Enrichment (SELEX), Bacterial One Hybrid (B1H) systems, Surface Plasmon Resonance (SPR), and Protein Binding Microarrays (PBMs) have been widely used to investigate the biochemical characteristics of TF-DNA binding events. In addition to these qualitative methods, bio-informatic approaches have also assisted in TF binding site discovery. Here we discuss the advantages and limitations of these different approaches, as well as the benefits of utilizing more quantitative approaches, such as Mechanically Induced Trapping of Molecular Interactions (MITOMI), Microscale Thermophoresis (MST) and Isothermal Titration Calorimetry (ITC), in determining the biophysical basis of binding specificity of TF-DNA complexes and improving upon existing computational approaches aimed at affinity predictions.

1. Background

1.1. Homeodomain transcription factors

Eukaryotic gene regulatory networks are controlled by transcription factors (TFs) that bind to specific DNA sequences within the genome. Once bound, TFs contribute to the complex transcriptional regulation of target genes [1]. Many studies have demonstrated that TFs often bind to a range of target DNA sequences with varying affinity [2,3], can form complexes with cofactor proteins [4], and can either activate or repress transcription of a target gene in a context-dependent manner [5,6]. While the importance of TFs in the genetic control of cellular function and embryonic development is well established [7], the biophysical and thermodynamic basis for where, when and how a TF binds to a specific DNA regulatory sequence remains poorly understood.

The biophysical organization of TF proteins has long been recognized as modular [8], with each different protein consisting of a DNA recognition domain and additional structural and functional effector domains [9]. Researchers have classified TF proteins into distinct families based on the shared molecular characteristics of their DNA binding domain [10]. One such DNA binding domain is the homeodomain (HD), which is found in an estimated 15–30 % of all TFs in eukaryotes, depending on the species analyzed [11]. The HD is a 60 amino acid residue motif encoded by a 180 bp homeobox DNA sequence in the corresponding TF protein-coding gene. The homeobox was first discovered in 1984 in *Drosophila melanogaster* [11] and has since been identified in a number of eukaryotic organisms, including yeast, mouse and human [12,13]. Structural studies over the last four decades have revealed that the HD forms a highly conserved helix-turn-helix DNA binding domain consisting of three α -helices and an unstructured *N*-terminal extension (Fig. 1a) [11,14].

The homeobox gene family is large and plays important roles in diverse biological processes. For example, the HD TF proteins regulate cellular differentiation and function, and are critical during embryonic development in metazoans [11,12]. A class of homeobox genes, the HOM-C complex in *Drosophila*, and their mammalian homologs, the Hox genes, are organized in linked chromosomal clusters and exhibit similar expression patterns along the anterior-posterior axis, which determine the basic body plan during embryogenesis [15]. HD TFs regulate a wide range of developmental networks by controlling the temporal, spatial

https://doi.org/10.1016/j.bbagrm.2024.195074

Received 3 July 2024; Received in revised form 26 November 2024; Accepted 1 December 2024 Available online 5 December 2024

1874-9399/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

^{*} Corresponding author. E-mail address: dspratt@clarku.edu (D.E. Spratt).

and cell type-specific expression of target genes, including the nervous system [16] and eye development [17]. Due to their important evolutionarily conserved roles in embryonic development, significant biophysical and structural work has been performed on HD proteins in an effort to expand our understanding of their diverse functions, characterize the molecular activity of these TFs, and improve predictions for TF-DNA binding events.

1.2. Structural characterization of HD TFs

In 1989, the first 3D structure of the HD of a TF was solved for D. melanogaster ANTENNAPEDIA (ANTP), a member of the HOM-C complex, using homonuclear Nuclear Magnetic Resonance (NMR) spectroscopy [18]. This ensemble of structures revealed three α -helices, with helices I and II sitting anti-parallel to one another and connected by a loop while helix III, commonly referred to as the recognition helix, was positioned perpendicular to the other two helices (Fig. 1a) [19]. These studies also uncovered the boundaries of the three α -helices (I – residues 10–21; II – residues 28–38; III – residues 42–52, see Fig. 1b), as well as the hydrophobic core holding the HD together (specifically in ANTP -L16, L26, I34, A35, A37, L38, L40, I45, W48, and F49, see Fig. 1c) [18]. In a subsequent study, Qian et al. determined a very similar structure for FUSHI TARAZU (FTZ) (Fig. 2), a pair-rule TF responsible for establishing the correct number of body segments during Drosophila embryogenesis with 85 % sequence identity to ANTP in the HD [20]. In 2006, the solution NMR structure of a third HD protein, BICOID (BCD) revealed a nearly identical organization to ANTP and FTZ, with some variance in the unstructured N-terminal arm of BCD (residues 2-10; RRTRTTFTS) that can be attributed to greater sequence variability in this region of the protein (Fig. 1b) [21]. These early structural studies provided valuable atomic level detail and indicated a clear rationale for how these HD TFs might bind to DNA to regulate target genes.

The shared structural organization of the HD observed in ANTP, FTZ and BCD has now been characterized in a number of other members of the HD TF family in *Drosophila* (Fig. 2). These studies reveal that all of the *Drosophila* HDs studied to date (listed in Table 1) consist of the three canonical α -helices (Figs. 1a and 2).

Furthermore, NMR spectroscopy and X-ray crystallography have been used to elucidate not only the global fold HD alone for many of these TFs (Fig. 2), but also TF in complex with DNA (Fig. 3) and TF-DNA complexes with other HD cofactors present (Fig. 4). These approaches have also revealed the similar 3D structures for other eukaryotes and prokaryotes homeodomains (summarized in Tables S1 and S2).

1.3. Previous structural studies of HDs bound to DNA

Structural studies have provided valuable information regarding HD-DNA interactions and the DNA binding specificity of HD TFs. ANTP was the first HD-DNA complex structurally examined using both NMR and Xray crystallography [41]. Both methods demonstrated that ANTP achieves specificity through rapidly fluctuating DNA contacts and preferentially binds to the core consensus DNA sequence 5'-TAATGG-3' [42]. As expected, both the X-ray and NMR structures showed that the recognition helix binds in the major groove of DNA. In these structures, I47 was observed to make important contacts with the methyl group of thymine 4 (TAATGG) and the C8 of adenine 3 (TAATGG) [42]. N51 also forms a pair of hydrogen bonds to adenine 3, while Q50 is involved in van der Waals interactions with cytosine 6 and hydrogen bonding with the phosphate backbone (Fig. 3b). Intriguingly, some of the structures in the NMR ensembles also suggest that the HD N-terminal extension engages in the adjacent minor groove (Fig. 3b). In these solution structures, R5 of the HD N-terminal extension forms a hydrogen bond to thymine 1 (TAATGG), fitting into the minor groove (Fig. 3b). These results suggest that HD TFs can potentially experience different conformations in solution, but both the X-ray and NMR structural data demonstrate consistent docking patterns within a HD-DNA complex. In the last two decades, the structure of several Drosophila HD TF-DNA complexes have been characterized in detail (Table 1). All of these structures demonstrate a strikingly similar organization with the recognition helix binding in the major groove of DNA and additional residue contacts from the HD N-terminal region in the adjacent minor groove (Fig. 4).

More recent studies have also shown that HD TFs can bind cooperatively to DNA to potentially modulate site specificities and functional output [43,44]. HD TFs also have the ability to regulate transcription by interacting with each other in homodimeric or heterodimeric configurations (Fig. 4) [43]. For example, BCD monomers interact cooperatively with each other to bind DNA and regulate spatial transcription of various target genes [43]. *Hunchback (hb)* is one such target gene that is activated in a threshold dependent response to the concentration gradient of



Fig. 1. Transcription Factor Homeodomain (TF-HD) Structure. (A) 60 amino acid helix-turn-helix DNA binding domain [163,164]. The HD consists of three alpha helices – Helix I, II and II. Helix III is also referred to as the recognition helix, due to its role in direct binding to DNA [165]. (B) Comparison of TF-HDs BCD, ANTP, and FTZ. Multiple sequence alignment displaying the boundaries of the three α -helices. Residues in the hydrophobic core of the protein are designated with a red box [20,166]. (C) Conserved residues found in the hydrophobic core (shown in pink) and the HD are shown in blue. The HD of ANTP (PDB 1HOM) [25] was visualized using Pymol [167]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

BBA - Gene Regulatory Mechanisms 1868 (2025) 195074



Fig. 2. Solved Transcription Factor-Homeodomain (TF-HD) structures. Overlayed structures of various TF-HDs show the similarities and differences between the homeodomains. All of these structures show the three helices that are characteristic of the HD family, with variations being caused by in the number of extra residues included at the *N*- and *C*-termini, and length of the recognition helix.

the maternally deposited BCD protein in the early embryo. Activation of the hb gene requires the cooperative binding of BCD to multiple DNA binding sites within regulatory regions and the resulting spatial expression profile of hb helps to establish the anterio-posterior axis in the embryo [45].

In summary, structural studies confirmed that the recognition helix (helix III) binds within the major groove of DNA, while residues in the Nterminal arm, located prior to helix I of the HD, appear to wrap around DNA and make contact with the minor groove (Fig. 3) [21]. The reported binding affinities of these HD-DNA complexes are thought to be a function of the base pair sequences within the specific DNA binding site [46] and may also bind cooperatively via protein-protein interactions. Given the overall similarity of the structural organization of HD TF interactions with DNA and the clear conservation of amino acid residues within HDs, a central question for the field is how individual HD TFs exhibit DNA sequence binding specificity and consequently regulate the expression of different target genes. By improving our understanding of the biophysical basis for HD-DNA complex formation, we can further examine binding affinity, investigate how this affinity may be linked to the shape of DNA, improve upon computational predictions, and ultimately elucidate HD-DNA sequence specificity.

1.4. Integrating in silico approaches to characterize TF-DNA interactions

As high quality gene expression data has become more prevalent, the ability of researchers to build accurate mathematical models of the functional activity of TFs within regulatory networks has also increased at an unprecedented rate. We are not only seeing more precise model fits to existing molecular experimental data, but we are developing a wealth of approaches utilizing mathematical tools to unravel the complex regulatory mechanisms involved in gene regulation, including protein-DNA interactions, protein-protein cooperativity, short-range repression, and dual activator/repressor roles of TFs [47–59].

For these mathematical models to accurately predict transcriptional regulation in early Drosophila development via genomic regulatory regions (enhancers), we must first have robust experimental information on the location and affinity of the TF binding sites (TFBSs) present within an enhancer region. For over four decades, computational biologists have worked on developing and refining computational algorithms to predict DNA binding sites. Pioneering studies by Berg and von Hippel in the late 1980s [60-63] established a conceptual understanding of the structural, thermodynamic and statistical rules that impact the binding of TFs to functional sites in DNA. In addition to providing a theoretical framework that enables us to move from qualitative data, in the form of a collection of identified binding sites without affinity measurements, to robust quantitative models that fit the data well [61-63]. One of the most popular bioinformatic approaches to evolve from these early studies for predicting TFBSs is a Position Weight Matrix (PWM)-based approach that scans a DNA sequence for binding sites of a particular TF of interest (Fig. 5) [63-65]. In general, all PWM-based algorithms score sub-sequences (i.e. potential TFBSs) in the DNA region considered and these scores are often interpreted as being

Table 1

| Solve | ed structures | of L | . me | lanogaster | Homeoc | lomain | ı Transcri | ption | Factors. |
|-------|---------------|------|------|------------|--------|--------|------------|-------|----------|
|-------|---------------|------|------|------------|--------|--------|------------|-------|----------|

| | - | - | |
|---------------------|---------------|---------------------------------------|----------------------|
| Homeodomain | PDB/BMRB code | Structural determination method | Citation |
| ABDOMINAL A | 8F37/31057 | NMR | а |
| ABDOMINAL B/ | 5ZJQ, 5ZJR, | X-ray | [22] |
| EXTRADENTICLE | 5ZJS, 5ZJT | | |
| ANTENNAPEDIA | 9ANT, 4XID | X-ray | [23,24] |
| | 1HOM, 1AHD, | NMR | [25-27] |
| | 1SAN/1037, | | |
| | 4104 | | |
| ARISTALESS/CLAWLESS | 3A01, 3A02, | X-ray | [28] |
| | 3LNQ | | |
| | 8F36/31056 | NMR | а |
| BICOID | 1ZQ3/6906 | NMR | [21] |
| EVEN-SKIPPED | 1JGG | X-ray | [19] |
| ENGRAILED | 1HDD, 1ENH, | X-ray | [29-31] |
| | 3HDD | | |
| | 2P81, 2JWT, | NMR | [32,33] ^a |
| | 7YB4/7386, | | |
| | 15536 | | |
| FUSHI TARAZU | 1FTZ | NMR | [20] |
| SEX COMBS REDUCED/ | 2R5Y, 2R5Z | X-ray | [34] |
| EXTRADENTICLE | 1001 101/0 | | FOF 0(1 |
| UBX-EXD-DNA | 4UUS | X-ray | [35,36] |
| VENTRAL NERVOUS | 1NK2, 1NK3, | NMR | [37,38] |
| SYSTEM/NK-2 | 1VND/4141 | | |
| | | | |

Recent advances in machine based learning models, such as AlphaFold2, have also opened new avenues to predict the 3D structures of proteins [39] that can be complementary to conventional structural determination methods like NMR and X-ray crystallography. However, it is important to note the caveat that Alpha-Fold2 is a valuable tool for predicting a protein model but will likely not replace experimentally determined structural approaches that can shed more detailed light on the subtle differences in conformation states experienced by each HD protein [40].

^a To be published on the Protein Data Bank (PDB) https://www.rcsb.org

proportional to the binding affinity of the sequence. However, the accuracy of the quantitative models generated in PWMs may depend on the quality, and perhaps quantity, of the data used to generate them. As a result, conducting detailed experimental binding assays is important to validate this interpretation. Therefore, the field would benefit from a more comprehensive synergistic approach that would not only integrate existing computational and biochemical methods, but also incorporate more advanced quantitative experimental tools to directly assess the biophysical interaction of HD TFs with DNA binding sequences. This system-level approach would undoubtedly make a valuable contribution by expanding our understanding of the intricate relationship of structure and function in HD-DNA complex formation.

2. Established experimental methods to determine protein-DNA binding

Numerous methods have been developed to investigate protein-DNA binding affinity and sequence specificity. In this section, we will discuss the strengths and inherent short-comings of these experimental approaches to examine the binding parameters of a protein-DNA complex.

2.1. DNA footprinting

DNA footprinting was first described in 1978 by Galas and Schmitz [66] and quickly became a commonly used method to determine the sequence specificity of a protein bound to DNA. This approach is based on the principle that any DNA bound by a protein is protected from enzyme cleavage (i.e. nucleases) and/or chemical hydrolysis (i.e. acid/base), while any unprotected DNA will be cleaved (Fig. 6a).

Deoxyribonuclease I (DNase I) is normally used in DNA footprinting experiments because of its large size and ability to function as a singlestranded or double-stranded endonuclease [67]. DNase I binds to the minor groove of DNA and cleaves the phosphodiester backbone producing shorter oligonucleotides. However, DNase I does not cleave uniformly along the DNA [67], due to the inconsistent cleavage rate of the enzyme that is dependent on the local and global fold of the DNA [68] and other factors such as DNA flexibility and sequence differentiation in the minor groove of DNA. When DNase I binds to the oligonucleotide, the DNA becomes dramatically distorted as the minor groove widens and the DNA is bent towards the major groove, eventually allowing for DNA cleavage to occur [69]. Other treatments used in DNA footprinting experiments include hydroxyl radicals, ultraviolet irradiation, and dimethyl sulfate [70]. Hydroxyl radicals are formed by reducing iron(II) with hydrogen peroxide, which then interact with the DNA backbone resulting in random single base deletions [71]. Hydroxyl radicals are not sequence dependent and result in an evenly distributed ladder, but the reaction and digestion time is slow [70]. Ultraviolet irradiation causes the nucleic acid bases to become excited and undergo photoreactions, such as thymine dimerization [72]. Photoreactions are sensitive to the local environment of the DNA and if a protein is bound to the DNA, the environment is altered thereby generating a footprint [72]. Ultraviolet irradiation is a good cleavage agent because it reacts quickly and can penetrate through cell membranes, but the footprinting signal can be unpredictable [72]. Dimethyl sulfate is also a commonly used footprinting reagent that methylates purines in the major groove when not protected by a bound protein. However, this reagent requires extended incubation with the sample that can allow for the dimethyl sulfate to potentially react with the protein, which in turn can result in unwanted DNA footprinting artifacts [72].

Many protein-DNA interaction sites, and the biological function of such interactions, have been determined using this approach. For example, Ekker et al. examined the ULTRABITHORAX (UBX) HD TF, which is responsible for the proper developmental differentiation of wings and halteres in Drosophila [73]. The researchers observed that mutations in the HD of UBX resulted in the transformation of the haltere tissue into the wing tissue [73]. In this study, they were also able to determine that UBX recognizes a consensus sequence of 5'-TTAATGG-3', with the 5'-TAAT'-3' core playing a large role in determining the binding affinity of the protein, while the flanking bases contribute to overall affinity [74]. Although DNase I footprinting was able to experimentally determine this 7 bp binding site for UBX, the authors noted that naturally occurring UBX binding sites in the genome are typically 40-90 bp and consist of multiple TAA tandem repeats [74]. In parallel studies, Han et al. were able to identify nine protein binding sites in the proximal enhancer that can regulate the expression of the ftz gene using DNA footprinting, with each binding site containing the sequence 5'-AGGA-3' [75]. Wang et al. demonstrated that VENTRAL NERVOUS SYSTEM DEFECTIVE (VND)/NK-2 - HD binds to a purified genomic DNA fragment at four distinct sites, with three sites being strongly protected (containing a 5'-AAGTG-3' core sequence) and another site that was weakly protected (5'-CAGAGTTT 3') [76].

Although this approach has uncovered numerous HD binding sites, DNase I footprinting is a very low-throughput method that can only analyze a relatively small region (<1 kb) at a time [77]. Limitations to the DNA footprinting approach include the need for the DNA sequence to be radio-labeled as well as an excess of DNA-binding activity over the amount of DNA fragment used to ensure a clear footprint [78]. Unfortunately, DNA footprinting also cannot distinguish individual components of heterogeneous protein-DNA complexes. As DNase I does not cleave DNA indiscriminately, this can result in the cleaved DNA running unevenly on a polyacrylamide gel that can make it difficult to delineate where the protein provides protection to the DNA (i.e. binding site(s)) [78]. To ensure appropriate digestion, the concentration of DNase I added should be optimized, as DNase I overdigestion can result in faint bands at the top of the probe, more intense bands at the bottom, and an uneven DNA fragment ladder [79]. Another drawback of DNA footprinting is that the yield of the protected fragment may be small and



Fig. 3. ANTP TF-HD bound to double stranded DNA. (A) Top view of double-stranded DNA shows Helix III (aka the "recognition helix") bound to DNA. When rotated 90 degrees counter clockwise, it becomes apparent that the recognition helix inserts into the major groove of DNA. (B) Important residues in the TF-DNA complex. The structure shows that the *N*-terminal arm of the HD makes contact with the minor groove, while the recognition helix binds to the major groove of DNA [165]. Residues R3, R4, R6, T7, T8, F9 on the *N*-terminal arm (shown in pink) make contact with the DNA minor groove [164]. Residues W48, F49, N51, R53, and K/R54 (shown in green) are highly conserved in the TF-HD recognition helix and play a critical role in HD-DNA binding [45]. The ANTP HD-DNA complex (PDB 1ZQ3) [166] was visualized using Pymol [167]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hard to observe if the binding constant for the specific sequence is not larger than the binding constant of the DNA [66]. Due to these limitations, more modern approaches have been implemented to study the interactions between TF HDs and DNA.

2.2. Electrophoretic Mobility Shift Assays (EMSAs)

The Electrophoretic Mobility Shift Assay (EMSA) is a useful method that has been extensively used to examine and verify TF binding to different DNA sequences. EMSA is a qualitative approach that utilizes the fact that protein bound to DNA is less mobile than free DNA when migrating through a gel matrix [80]. The binding of a protein to DNA results in an upward shift on the gel due to changes in size and charge of the protein-DNA complex that is formed (Fig. 6b). The migration of DNA in an EMSA is traditionally monitored through the use of radioisotope-labeled nucleic acids [81].

For example, in 2012 Anderson et al. examined the HD TF MOHAWK (MKX), a key regulator of skeletal muscle and tendon differentiation that specifically binds to an inverted repeat 5'-ACAN₂₅TGT-3' core sequence separated by 25 bases [82]. EMSAs were also performed on *D. melanogaster* HD-TFs FTZ, ANTP, ABDOMINAL-A (ABD-A) and ABDOMINAL-B (ABD-B) in our own earlier studies to confirm that their

isolated HDs were biochemically functional due to their ability to bind specifically to the consensus sequence containing the 5'-TAAT-3' core [83]. Although EMSAs can be adapted to any protein, the disadvantages include the fact that the samples are not at chemical equilibrium during the electrophoresis step, and that the rapid dissociation of the protein-DNA complex can potentially hinder detection [81], and the need for the DNA sequence being tested to be radio-labeled (i.e. ^{32}P) or fluorescently tagged at the 5' and/or 3' end for visualization in classical EMSA assays.

2.3. Systematic evolution of ligands by exponential enrichment (SELEX)

Systematic evolution of ligands by exponential enrichment (SELEX) is a method that has been used to reveal DNA binding profiles of numerous TFs. This method isolates specific nucleotide sequences that bind to TF proteins through a repetitive enrichment process that include the elution of bound oligonucleotides [84], followed by amplification using a DNA polymerase (Fig. 6c).

A prime example of using SELEX to determine TF binding sequences was performed by Slattery et al. in 2011, where eight *Drosophila* TFs (LABIAL (LAB), PROBOSCIPEDIA (PB), DEFORMED (DFD), SEX COMBS REDUCED (SCR), ANTP, ABD-A, ABD-B, and UBX) were examined in

BBA - Gene Regulatory Mechanisms 1868 (2025) 195074



Fig. 4. Solved Transcription Factor-Homeodomain DNA structures. Overlayed structures of various TF-HD (blue) bound to DNA (grey) show the similarities and differences between the homeodomains. The HD recognition helix can be seen binding to the major groove of DNA in each example, with some TFs requiring a cofactor (typically EXTRADENTICLE) to bind to the DNA sequence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

complex with EXD [4]. Intriguingly, all eight HD TFs bound to similar sequences in vitro; however, when the EXD cofactor was present, each TF-HD tested bound to the DNA sequence with improved specificity [4,85]. For example, while both EXD-SCR and EXD-UBX were observed to bind to 5'-TGATTTAT-3', EXD-SCR binds more strongly to 5'-TGA-CAAAT-3' and 5'-TGATTAAC-3', while EXD-UBX binds more strongly to 5'-TGATTTAC-3' [4]. This finding shows that each HD TF has distinct binding preferences that are revealed when the EXD cofactor is present, thereby unlocking latent specificities that are present within the protein-DNA complex formed [4].

While this approach is a high-throughput technique to identify DNA binding sequences, a potential drawback of using SELEX is that it relies on the use of a randomized DNA library. The size of the randomized library defines the libraries diversity, where the smaller the library the lower the chance that the strong binding sites are present in the library [86]. The first two rounds of selection cover a wide range of affinities and the following rounds are biased towards a higher affinity sequence motif, which can lead to unintended sequence selection bias due to multiple rounds of enrichment, elution, and amplification [87]. Confirming SELEX experiments using additional protein-DNA approaches is critically important to insure reproducibility and validate results [4]. Recent advances to this approach have been made by using No Read Left Behind (NRLB) in combination with SELEX [88]. NRLB is a model developed by Rastogi et al. than can be used to predict and identify binding affinities through SELEX data to properly predict protein-DNA binding sites and the input data consists of read out sequences from both "round zero" (R0) and after one cycle (R1) [88]. Rastogi et al. successfully implemented this approach to further investigate the binding specificity and detect binding sites of *D. melanogaster* Hox



Fig. 5. Position Weight Matrix (PWM)-based approaches. Simple schematic of PWM-based approaches (using FTZ protein as an example). All PWM-based approaches require a list of known binding sequences for the protein of interest, shown here in a single FASTA file (seq_FTZ.txt). From this set of sequences, a frequency matrix is constructed that contains the number of times each nucleotide is observed at each position within the binding site. The frequency matrix can then be used to create a graphical visualization of the individual nucleotide preferences at each position, referred to as a sequence logo. In addition, the frequency matrix is converted to a PWM using a logarithmic transformation. Note that this transformation often includes some adjustment for the background frequencies for each of the four nucleotides in the genome and the total number of known binding sequences. The PWM shown has been constructed using log-odds frequencies compared to background frequencies of A:0.3, C:0.2, G:0.2, and T:0.3. This PWM is then used to score individual sequences of interest. An example sequence (5'-GCAATTA-3') is shown below the PWM along with the entries from the PWM used in calculating the sequence's corresponding score, 2.95, highlighted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

proteins in the presence of the cofactor EXD [88]. Although this approach has been used to infer binding affinities of TFs more accurately than SELEX alone and has been employed to assess both Protein Binding Microarrays (PBMs) and Chromatin immunoprecipitation-sequencing (ChIP-seq; please see next section) [89], further investigation into more TF HDs is required.

2.4. Chromatin immunoprecipitation-sequencing (ChIP-seq)

ChIP-seq is another qualitative approach that has been widely used to examine TF-DNA binding [90,91]. This method was initially developed by Albert et al. in 2007 to sequence the DNA of 322,000 individual *Saccharomyces cerevisiae* nucleosomes containing a histone variant [92], and can be used to rapidly decode millions of DNA fragments simultaneously, revealing the specific sequences of the precipitated DNA fragments (Fig. 7a) [93]. The ChIP-seq assay is initiated when formaldehyde is added to crosslink the protein-DNA complex, followed by chromatin shearing into small fragments by sonication. A protein specific antibody is then used to immunoprecipiate the protein-DNA complexes and the crosslinking is reversed using sodium dodecyl sulfate to denature the bound protein. The final step is to extract the isolated DNA fragments, representing the TF-bound sequences, and perform DNA sequencing [94].

ChIP-seq has been employed extensively in the last decade to provide maps for various TF binding sites across the entire genome. For example, Jusiak et al. used ChIP-seq to map the binding regions for the SINE OCULUS (SO), a HD TF that is necessary for *Drosophila* eye development [95]. The authors were able to identify 7566 SO enriched DNA regions that mapped to 5952 genes [95]. Although ChIP-seq can identify the genomic regions bound by a target TF, it requires a high-quality antibody for each TF being studied to ensure the detection of enrichment peaks without considerable background noise [96]. A potential limitation to using ChIP-seq is that it is generally qualitative in terms of the sequences that it extracts and it is not possible to distinguish binding affinities between two different complexes of the same TF [85,97]. Because TFs can bind to a variety of genomic loci with different patterns of chromatin markers and nucleosomal organization, this approach might not just reveal TF DNA binding sites, but also locations of modified histones, and binding between other TFs and nucleosomes [98,99]. Another drawback to ChIP-seq is not having sufficient coverage of the particular genome being sequenced. In the case of Drosophila, even if a ChIP-seq experiment results in >8 million reads, which is generally considered as sufficient coverage for the genome, there is still a high likelihood that some strong TF binding sites could be overlooked [98,100]. Other limitations of this approach include the requirement of a large amount of sample and the importance of having antibodies that offer high specificity and sensitivity needed to obtain a successful result [96,101]. Lastly, with the use of formaldehyde, ChIP-seq is generally considered a non-native approach, but recent advances are shifting towards using more native-like procedures that does not require chemical cross-linking [102] with promising results showing higher immunoprecipitation efficiency [103].

2.5. Bacterial One Hybrid (B1H)

Bacterial One Hybrid (B1H) systems have been used to identify and verify DNA-binding specificity for numerous HD TFs in *Drosophila* [104,105]. This technique involves three major components – a TF expression vector, a bacterial selection strain, and a randomized binding site library in a reporter vector [106] containing a weak promoter [107]. The DNA library undergoes a round of negative selection on plates containing a 5-fluoroorotic acid (5-FOA), allowing for the reduction of self-activating sequences [108]. Two reporter systems can also be used, typically the yeast *HIS3* and/or *URA3* gene, allowing for a combination



Fig. 6. Qualitative methods used to identify transcription factor(TF)-DNA binding sites. (A) DNase I footprinting. After incubating radioactive ³²P-labeled DNA with a TF, the DNA is cleaved by DNase I. Any protein bound to DNA is considered "protected" from nuclease digestion. The "protected" DNA (aka the DNA footprint) indicates the protein binding site in the DNA sequence. Samples are run on a polyacrylamide gel to visualize TF-DNA binding site [78,168]. (B) Electrophoretic Mobility Shift Assay (EMSA). Radio-labeled DNA and the protein of interest are combined and incubated, then run on a polyacrylamide gel. Electrophoretic mobility changes indicate that a protein-DNA complex has formed. A supershift can be observed when a protein specific antibody is also bound to protein-DNA complex [81]. (C) Systematic Evolution of Ligands by Exponential Enrichment (SELEX). An oligonucleotide library consisting of randomly generated sequences are exposed to a target protein. Any unbound sequences or proteins are removed and only bound sequences are amplified by DNA polymerase. The amplified sequences are used for subsequent rounds of selection, which are then increased to identify the tightest binding sequence [87]. Protein structures (PDB 1JGG) [19] were visualized using Pymol [167] and each schematic was created with BioRender.com.



(caption on next page)

Fig. 7. Advanced qualitative methods used to determine transcription factor(TF)-DNA binding sites. (A). Chromatin ImmunoPrecipitation-sequencing (ChIP-seq). Chromosomal DNA crosslinked to a protein is fragmented by sonication. An antibody specific to the protein coupled with a magnetic bead is used to isolate the bound protein-DNA fragments. The chemical crosslinking is then reversed using SDS and the released DNA sequence is analyzed [94]. (B) Bacterial One-Hybrid (B1H). A randomized DNA library representing all potential transcription factor binding sites is cloned immediately upstream from the selectable markers, *HIS3* and *URA3*. If a TF binds to the randomized nucleotide region, the fused RNA polymerase will be recruited to the downstream weak promoter, enabling transcription to occur. Sequences that demonstrate *HIS3* and *URA3* selection are determined to be TF-DNA binding sites [104]. (C) Surface Plasmon Resonance (SPR). An incident light beam is passed through a prism and is reflected off a sensor chip surface containing an immobilized ligand (i.e. DNA fragment). The light is reflected off the surface into a detector. At a certain incident angle (aka resonance angle), light is absorbed by electrons (aka surface plasmons) in the sensor chip causing them to resonate. As molecular binding event occurs, the angular position of light changes [116]. (D) Protein Binding Microarrays (PBMs). A DNA binding protein is expressed and purified containing an epitope tag. The tagged protein is applied to a dsDNA microarray and the microarray is washed to remove any nonspecific protein [120,121]. A fluorescently labeled antibody is then used to quantify the amount of protein bound to DNA. The amount of protein bound to each probe sequence on the array can be quantified [119,121]. Separate microarrays are stained with a dye, SYBR Green I, specific to double-stranded DNA to normalize the PBM data by relative DNA concentrations present in each spot on the microarray [169]. Protein structures (PDB 1JGG) [19] were visual

of positive and negative selection [109]. Each DNA binding domain is expressed as a fusion to an RNA polymerase subunit [109], so that if the DNA-binding domain binds to the target DNA site in the reporter vector, the RNA polymerase is recruited to the promoter and transcription is activated [109] (Fig. 7b).

In 2008, Noyes et al. used a B1H system as a high-throughput method to characterize and determine the specificity of 35 members of the *Drosophila* segmentation network that play important roles in early anterior-posterior patterning [110]. Their study revealed that using a B1H system better defined the recognition motif, 5'-TTTATG-3' for CAUDAL (CAD), when compared to SELEX or DNase I footprinting approaches [110]. The B1H model has also been used to confirm differences between species. For example, B1H was used to determine that the preferred recognition sequence of the murine MKX HD TF was 5'-ACA-3', which was subtly different from the 5-'TnACA'-3' consensus binding site of *Drosophila* MKX [111].

The B1H system is a rapid and streamlined technique used to characterize DNA-binding specificity that can be used to investigate protein-DNA interactions with a DNA library [109]. Although several binding site sequences have been identified for several TFs using the B1H system [109], a drawback to this method is that it only provides qualitative information on a TF-DNA complex. Another limitation is the binding site (s) might not be recognized for TFs with low binding affinity because of competition created by alternative binding sites elsewhere in the bacterial genome limiting the signal generated from a single low affinity binding site upstream of the reporter [108,112].

2.6. Surface Plasmon Resonance (SPR)

Surface Plasmon Resonance (SPR) is an optical technique used to measure molecular interactions in real time [113,114]. SPR examines can be used to measure the rate of complex formation as well as provide thermodynamic information [115] (Fig. 7c). Protein-ligand and proteinprotein interactions have been routinely evaluated using SPR, but protein-DNA interactions have proven to be more difficult using this technique due to higher binding affinities and the contribution of strong electrostatic interactions [116]. For example, Yoo et al. examined the binding of VND, a HD TF that is important for neuroblast formation during the development of the embryonic central nervous sytem in the Drosophila, which recognizes a 5'-T (T/C)AAGT(G/A)G-3' core motif [117,118]. SPR was used to determine the binding affinity and specificity of full length VND (VND-FL) and VND containing a NK2-specific domain (VND-NH), which is located near the C-terminus of the protein. The data showed that VND-FL bound to the DNA sequence with a dissociation constant (K_D) of 1.53 \times 10⁻⁸ M, while VND-NH bound to the same DNA sequence with a K_D of 4.56 \times 10^{-8} M, meaning the binding affinity of VND-NH 3-fold weaker than that of VND-FL [118].

Limitations of SPR include artifacts that can be created by aberrant changes of the refractive index at the sensing surface, which can interfere with the interpretation of binding data [118]. Another issue with using SPR to measure protein-DNA interactions is having to manage the

mass transfer effect, in which the released protein rebinds to another DNA sequence after dissociating [116]. This phenomenon results in a slower dissociation rate measured by SPR, thereby altering the measured binding affinities. Ideally, to determine the binding affinity of protein-DNA using SPR, a previous experimentally determined binding affinity is needed; however, in most cases, an estimated association constant (K_A) value must be used causing longer SPR experimental optimization.

2.7. Protein Binding Microarrays (PBMs)

Protein Binding Microarrays (PBMs) are a high-throughput approach that can used to visualize protein-DNA interactions in vitro [119,120]. This approach allows for the simultaneous characterization and analysis of a protein's binding specificity to thousands of double-stranded DNA (dsDNA) probes affixed to the surface of a microarray [119]. PBMs involve expressing and purifying a DNA binding protein with an epitope tag. The tagged DNA binding protein is then applied to a dsDNA microarray and the microarray is extensively washed to ensure the removal of any nonspecific bound protein [120,121]. A fluorescently labeled antibody is then used to quantify the amount of protein bound to DNA. The amount of protein bound to each probe sequence on the array can be quantified based upon signal intensity [119,121] (Fig. 7d).

Busser et al. defined the DNA sequences bound by eleven TF HDs by using PBMs - SLOUCH (SLOU), MUSCLE SEGMENT HOMEOBOX (MSH), BAGPIPE (BAP), LADYBIRD LATE (LBL), PITUITARY HOMEO-BOX 1 (PTX1), SIX HOMEOBOX 4 (SIX4), TINMAN (TIN), EVEN-SKIPPED (EVE), UBX, and ABD-B [122]. Their PBM results revealed that UBX, ABD-B, SLOU, MSH, EVE and LBL preferentially recognize a binding sequence core of 5'-TAAT-3', which was in good agreement with other TF HD-DNA binding studies [123]. Although these TF HDs bind primarily to a sequences containing a 5'-TAAT-3' core, PBM results showed that SLOU and MSH also recognized a small number of atypical and non-consensus sequences that are unique to the TF HD [122]. Other TF HDs such as TIN and BAP were shown that they could be displaced with DNA sequences that were distinct from the typical 5'-TAAT-3' core sequence altogether [122]. While Busser et al. were able to successfully determine the binding sequences using PBMs, this approach does have some inherent limitations. PBMs are limited by the amount of sequence that can be represented on the microarray as well as the length of the DNA sequences, supporting on average 60 bp long DNA probes [119]. Lastly, the in vitro nature of PBMs can complicate predicting function TF-binding sites in vivo because it is impossible to replicate the in vivo nuclear environment on a microarray surface [119].

2.8. Inherent limitations of existing methods

As described in the previous sections, the various methods used to measure TF-DNA affinities have their individual strengths and drawbacks. For example, although ChIP-seq provides high resolution sequencing and has a wide dynamic range, this technique requires multiple rounds of experimental optimization [91]. Likewise, DNA footprinting can provide a fast and real time analysis, but requires a labeled sample, while SELEX requires multiple rounds of selection [109]. Although performing EMSAs are rapid and can be sensitive, the samples are not in chemical equilibrium and the TF-DNA complex may not be stable resulting in dissociation during electrophoresis, thus causing a miscalculation of binding density [81,91]. For example, Matos et al. used E. coli RNase II as a model system to compare SPR and EMSAs for characterization and interpretation of the stability of RNA-protein complexes [124]. EMSAs are frequently used in the characterization of RNA binding, but they require a radioactive labeled RNA (usually ³²P) and use an indirect method to determine the dissociation constant. To overcome this obstacle, the researchers used SPR to determine the binding ability of wildtype RNase II and sixteen RNase II mutants to RNA [124]. After comparing their EMSA and SPR results, the researchers found that they were not in agreement [124]. EMSA requires a highly stable complex in order to be detected by electrophoresis otherwise a smear is visualized on the gel rather than a clear band. Their study concluded that using SPR was faster and more reliable than running EMSAs that can cause issues with reproducibility and sensitivity [124].

Overall, the techniques discussed in this section provide qualitative protein-DNA binding information, and due to their inherent limitations, many of these techniques have been used in various combinations to assess protein-DNA complex formation. For example, Nitta et al. successfully determined the DNA-binding sites for 92 *Drosophila* TFs [125]. The researchers plotted a dendrogram to compare the similarities between SELEX and B1H motif datasets collected and observed broad agreement in the results, with the same core binding specificities detected with very few exceptions. In some cases, the motifs obtained were not found in both datasets of the dendrogram, which was suggested to be caused by the flanking sequences used for the SELEX not being identical for the B1H dataset [107,125].

Although Matos et al. and Nitta et al. used complementary approaches to examine the binding site specificity, most reported TF-DNA studies only use a single method to determine the DNA binding specificity of a TF. While these individual methods can be useful in assessing TF-DNA binding, they are generally qualitative and may provide inadequate information regarding detailed affinity and sequence specificity. Specificity distinguishes specific binding partners from non-specific binding partners and affinity can be defined as the strength of binding [126], while both are defining characteristics they are not always interrelated. For example, mutations affecting specificity can alter the affinity with their binding partners, but mutations altering the overall binding affinity do not always change the binding specifity [127]. In other cases, the binding affinity and specificity for instant DNA-DNA interactions are anticorrelated as the affinity increases the specificity also increases [128]. One can think of the specifity of any TF as a measure of the variance around the mean affinity; thus, TFs with high specificity have a corresponsing high affinity variance, irrespective of the mean affinity value.

Furthermore, most of the current methods simply determine relative binding affinities for each of the DNA sequences considered. However, the scale of relativity is typically non-linear and any score obtained is itself unitless. As a result, one would need more than a single measurement of a particular sequence to determine absolute affinities for all other sequences investigated. In fact, detailed and accurate parameterization, presumably obtained via robust parameter estimation approaches, would be required to determine exactly how the relative scores relate to absolute binding affinity values for all sequeuences examined.

A major question that still needs to be addressed in the field is how do these TFs recognize and bind to their DNA targets and what are the true binding affinities of these TF-DNA interactions? Here, we call for the adoption of more robust and reproducible biophysical measurements of binding affinities to decipher the important roles that TFs play in cell fate and embryonic development in *Drosophila*.

3. Advanced quantitative approaches to measure TF-DNA affinities

3.1. Computational methods to investigate TF-DNA binding sites

A variety of different computational approaches have been used to further examine the binding affinity of TF-DNA complexes. These approaches have drawn on different areas of computational biology, including sequence alignment, machine learning, and hidden Markov models. To date, the most commonly implemented subset of these approaches are those that involve the construction of PWMs [129,130]. These algorithms begin with a set of known binding sequences for the protein of interest that are typically identified using in vitro approaches such as the B1H system (described in detail above). This list of binding sequences is used to construct a PWM containing the frequency of each of the four possible nucleotides at each position of the binding sequence (Fig. 5). Binding strengths are then inferred for individual proteins and DNA sequences of interest relative to a 'consensus sequence', the sequence found to have the most common nucleotide in each position of the binding sequence (Fig. 5). The underlying assumption of this approach is that sequences with the same physical binding affinity are equally likely to be selected by natural selection, thus those that appear in the set of known binding sequences most often represent the sequences with highest affinity, and the calculated binding energies for protein contacts at each individual base pair are additive (i.e. nucleotide positions within a binding sequence are independent of each other) [63,129–133]. These approaches can quickly score the strength of a large number of potential binding sequences for a protein of interest based on the PWM and nucleotide genomic background frequency [129,130]. For this reason, these methods have been widely used in the discovery of cis-regulatory elements, including enhancers, as well as the identification of TF binding sites used in mathematical modeling studies of transcriptional regulation.

One such study implemented a modified version of the PWM-based program Patser, which allowed the authors to search for potential TF binding sites in the putative cis-regulatory elements identified by ChIPseq analysis [134]. The authors were able to identify potential binding sites for the homeobox TFs SCR, ANTP, and UBX, which are known to regulate the expression of additional genes within the HOM-C complex of Drosophila [134]. A similar study by Ostrin et al. used a PWM-based algorithm to predict potential binding sites for the TF EYELESS (EY) within the genome of *D. melanogaster* [135]. This allowed the authors to identify potential EY binding sites and prioritize a subset of target genes for further functional analysis [135]. More recently, increasingly complex models have been developed that take into account the nonindependence of positions within a binding site [136-138]. Such models allow for the development of PWMs that have been shown in some scenarios to quantitatively outperform a standard PWM in terms of predictive ability [139].

In addition to TF binding site discovery, many mathematical modeling studies would not be possible without the identification of putative TF binding sites through bioinformatic algorithms. Thermodynamic-based models of transcriptional regulation in Drosophila have relied heavily on PWM-based methods over the last two decades [47-59]. These models focus on predicting the expression of a gene based entirely on the sequence of a corresponding enhancer, or cisregulatory module. For example, a recent study by Bhogale et al. employed thermodynamic-based models (an extension of GEMSTAT [50]) to explore the role of indirect binding of TFs during Drosophila embryogenesis and gain insights into how a TF may switch between transcriptional activation and repression roles [59]. The first step in their modeling process was to use a PWM-based algorithm to estimate the binding strengths of all putative sites within each enhancer they analyzed. This was a crucial first step, as the GEMSTAT model relies on these putative binding sites to calculate the predicted target gene expression. They then investigated the role of indirect binding of TFs by

BBA - Gene Regulatory Mechanisms 1868 (2025) 195074

allowing TFs to "piggyback" on other bound TFs to modulate the overall impact on gene expression. The study found that by allowing this indirect binding of TFs to influence the model output, their model was able to more accurately predict gene output during Drosophila mesoderm development. This suggests that TF-TF interactions, where both TFs are not directly bound to the enhancer, may underlie the dual regulatory role observed for some TFs [59]. Although studies such as this demonstrate the usefulness of PWM-based approaches in understanding the regulatory mechanisms underlying Drosophila embryonic development, using PWM-based scores as a proxy for the true binding affinity of a TF for a DNA sequence could be problematic. The assumption made is that there is a direct correlation between binding affinity and the frequency with which we find a particular nucleotide in a given position within the set of known binding sites. To our knowledge, this assumption has never been robustly validated using quantitative biophysical experimental approaches.

Thermodynamic modeling has also been applied to other mammalian systems. For example, Bertolino et al. developed a thermodynamic model which utilizes PWMs to identify binding sites [140]. The researchers were able to identify seven *cis*-regulatory elements neighboring *Cebpa*, a gene encoding for a TF called CCAAT enhancer-binding protein alpha that play a role in the development of myeloid cells [140]. Regardless of the model system, the use of PWMs in combination with experimentally determined binding sites using biophysical approaches can help to decipher where, when, why, and how TFs recognize and bind to different DNA sequences.

3.2. Mechanically Induced Trapping of Molecular Interactions (MITOMI)

Mechanically Induced Trapping of Molecular Interactions (MITOMI) is a newer method that can be used to obtain affinity constants and kinetic rate of protein-protein, protein-ligand, and protein-DNA



Fig. 8. More quantitative methods used to determine affinity and specificity of TF-DNA binding. (A) Mechanically Induced Trapping of Molecular Interactions (MITOMI). During equilibrium binding, the fluorescently labeled ligands bind to a target immobilized on the bottom surface of a membrane [142]. Mechanical trapping occurs when the bottom membrane is actuated, causing the surface-bound complexes to become trapped and expelling the solution phase molecules [142,144]. Unbound molecules are washed away and the trapped material is quantified by signal intensity [142,144]. (B) MicroScale Thermophoresis (MST). MST detects the movement of molecules within a temperature gradient. Binding events are detected by measuring changes in fluorescence intensity with increasing concentration of the ligand. First, the sample is pulled into a capillary tube by surface tension. An Infrared (IR) laser is switched on and unbound/bound molecules move out of the heated spot. A spectral shift is determined by the depletion of fluorescence caused by the slower movement of the bound TF-DNA complex, when compared to the unbound molecules, and then integrated into a titration curve [148]. Protein structures (PDB 1JGG) [19] were visualized using Pymol [167] and each schematic was created with BioRender.com.

interactions [141,142]. This approach allows for the capture of low affinity interactions between TFs and DNA target sequences by measuring the absolute binding affinities [143,144].

This technique employs a microfluidic device to directly measure the binding specifies of TF to numerous DNA sites [144] (Fig. 8A). First, spots of Cy5-labeled dsDNA sequences are printed onto an epoxy-coated microscope glass slide with each DNA sequence spotted with varying concentration to allow for a saturation binding curve and absolute binding affinities to be measured [142]. The DNA arrays are then aligned to a microfluidic chip, containing 768 unit cells [142,143]. Antipenta histidine antibodies are immobilized under the button membrane and the device is loaded with a wheat germ in vitro transcription/ translation (ITT) solution that contains the DNA template coding for the protein [142,143]. The chip is then loaded with the fluorescently labeled protein and binds to the immobilized antibodies. The ITT solution flows into the spotting chamber and the DNA interacts with the protein, the TF-DNA bound complexes are trapped, and any unbound or non-specific DNA is washed away [142,143] (Fig. 8a). A DNA array scanner then visualizes the formed complexes and the amount of protein bound is determined by its fluorescent signal intensity [144]. The detected signals are quantified, plotted, and the binding affinities are determined.

He et al. used PWMs to predict affinity differences of naturally occurring mutations in HUNCHBACK (HB) and BCD binding sites [145], where their PWM scores correctly predicted the affinity change for 21/25 mutations [145]. To evaluate the accuracy of the PWM-based inference, the researchers employed MITOMIs to experimentally measure the binding affinity differences and verified that these combined approaches showed good agreement for naturally occurring mutations in HB and BCD binding sites [145]. Other studies using this approach have examined protein-protein interactions [146] and successfully mapped the binding energy landscape of basic helix-loop-helix TFs [147]. Although MITOMIs allow for the investigation of numerous dynamic molecular interactions [141], when compared to other approaches they are relatively low throughput [142] and requires specialized instrumentation [144].

3.3. Microscale Thermophoresis (MST)

MST is a quantitative approach that detects movement caused by a biomolecular binding event within a temperature gradient using changes in fluorescence intensity (Fig. 8b) [148]. This sensitive technique detects changes in size, charge and the hydration shell of a molecule [149] and can measure a wide range of biochemical interactions, including protein-small molecule, protein-drug, proteinliposome, and protein-DNA [150]. In MST, a fluorescently labeled partner is mixed with varying concentrations of a nonfluorescent partner. The mixture is loaded into a capillary tube and a temperature gradient is applied that allows for the movement of the fluorescent molecule across the temperature gradient to be measured (Fig. 8). Any protein bound to DNA will escape the heated region slower than any unbound protein due to the larger molecular size of the complex [151]. Experiments are run using a low sample volume at low nanomolar concentrations and the binding affinity can be determined in 10 min in free solution [149]. MST was recently used to measure the binding affinity of TRANSCRIPTION TERMINATION FACTOR I (TTF-1), a transcription factor involved in the termination phase of the transcription cycle, with various DNA sequences [148]. A clear strength of MST is the ability to quickly and reproducibly measure protein-DNA interactions in solution. However, the technique is limited in that it does not provide any information on the thermodynamic properties of the interaction (i.e. enthalpy and entropy).

3.4. Isothermal Titration Calorimetry (ITC)

Isothermal Titration Calorimetry (ITC) is a quantitative biophysical approach that can be used to accurately measure real binding affinities for TF-DNA complexes (Fig. 9). ITC is generally considered to be the "gold standard" for biophysical measurements of a biomolecular interaction as it is the only technique that allows for the determination of all thermodynamic parameters (i.e. binding constant, molar ratio, Gibbs free energy, entropy and enthalpy) all in a single experiment [152,153]. The instrument is sensitive enough to characterize biological macromolecules, protein-protein, protein-DNA, and protein-small molecule interactions, and enzymatic activity [152].

ITC is a powerful technique that is now starting to be employed to measure the binding affinity of specific TFs with different nucleotide sequences. To conduct an ITC experiment, the protein of interest (i.e. TF HD) is inserted into the sample cell and the ligand (i.e. DNA sequence being tested) is loaded into the syringe. This is because it is easier to reach higher concentration with the ligand than the protein [154]. Over the course of 20–30 sequential injections, the ligand is titrated into the sample cell until the protein is fully saturated. The incremental heat changes from injection are accurately measured and an isotherm is obtained [155]. The isotherms are then integrated into a titration curve to yield the binding constant (K), stoichiometry (n), and the heat (i.e. enthalpy) of the binding event while also calculating the entropy of the system [155].

For example, the TF SPECIAL AT-RICH SEQUENCE BINDING PROTEIN-1 (SATB1), a HD TF found in humans that regulates over 1000 genes and controls transcription in various biological processes, such as switching of fetal globin genes, metastasis of breast cancer, and differentiation of embryonic stem cells, has been assessed using ITC [156]. SATB1 contains multiple DNA-binding domains including a HD and two CUT-domain repeats, CUTr1 and CUTr2, that work together to modulate DNA binding specificity [156]. SATB1 binds to matrix attachment regions (MARs) in DNA, which are sequences of DNA where the nuclear matrix attaches, and plays an important role in higher-order chromatin organization [157]. ITC was used to demonstrate that CUTr1 of SATB1 has binding specificity for the MAR sequence 5-TAATA-3', while the SATB1 HD has an affinity for a variety of DNA binding sites without specificity [156]. The binding of CUTr1, CUTr2, and the HD to a DNA sequence containing two 5'-TAATA-3' sites was also examined [156]. The change in enthalpy (Δ H) was determined for both the HD and CUTr1 domains to be -20 kJ/mol and -45 kJ/mol, respectively [156], revealing that CUTr1 binds with a tighter affinity to DNA than the HD alone. In contrast, CUTr2 showed no significant heat release, indicating weaker binding than CUTr1 and the HD and overall poor binding to the 5'-TAATA-3' sites [156].

Traditionally, ITC has been considered a low throughput approach due to longer experimental running times and the need for extensive cleaning steps between each run. However, new improvements to ITC instrumentation have been made to make this approach more high throughput and less labor intensive. TA instruments (New Castle, DE, USA) have recently brought to market fully automated versions of their iTC₂₀₀ and Affinity ITC that can be interfaced with a robotic sample changer, injector, and cleaning station [158]. The Affinity ITC Auto uses an industry-proven 96-well plate liquid handling system and the AutoiTC₂₀₀ allows for 30-40 titrations per day, with both systems being able to self-clean and run unattended for extended periods of time [158]. ITC is beginning to emerge as a powerful technique in determining the binding specificity and affinity of TF proteins. With the ability of ITC to experimentally determine all of the thermodynamic parameters of a TF-DNA interaction including stoichiometry and dissociation constant (K_D) ranging from nM to uM range, this approach has the potential to be used to examine many other TFs, including HD TFs, and determine their respective DNA sequence affinities with remarkable reproducibility and accuracy.

3.5. Potential limitations of using in vitro biophysical approaches

Recent in vivo and structural studies examining transcription factors binding to DNA have shown that it is difficult to correlate in vitro



Fig. 9. Isothermal Titration Calorimetry (ITC) is a quantitative biophysical method that can determine affinity and specificity of TF-DNA binding. ITC is a robust method that can experimentally determine binding affinity, specificity, and all of the thermodynamic properties involved in a biomolecular interaction (i.e. enthalpy, entropy, and Gibbs free energy). The instrument contains two cells (reference cell and sample cell) contained within an adiabatic jacket and the system is kept under constant pressure and volume. Initially, the protein of interest is injected into the sample cell, then the ligand (i.e. DNA sequence) is subsequently titrated into the sample cell in discrete injections volumes until the protein is fully saturated. The heat released or absorbed during binding is measured throughout the run. Isotherms are obtained and then integrated to plot a titration curve. Protein structures (PDB 1JGG) [19] were visualized using Pymol [167] and each schematic was created with BioRender.com.

findings with in vivo approaches. For example, researchers have reported that the affinity of some HD TFs can be altered due to chromatin DNA packing, which in turn can also regulating their interactions with other transcriptional regulators in the cell [159,160]. Likewise, Mann and coworkers have reported that HD TF binding is dependent on the 3D shape of the DNA binding site, and that the HD and DNA can undergo conformational changes that can affect the strength of the interaction [22,161,162]. Since chromatin packing and the 3D shape of DNA are also intricately connected to the actual sequence of the DNA, these studies add an additional level of nuance and complexity that need to be considered when it comes to examining HD-DNA complex formation and their respective affinities. It will be important for researchers in this growing field to recognize this caveat because what is observed using an in vivo system.

4. Conclusions

Homeobox genes, and the homeodomain (HD) TF proteins that they encode, are important master regulators of the transcriptional program that controls proper body segmentation formation during early embryonic development. Structural studies using NMR spectroscopy and X-ray crystallography have proven to be instrumental in helping us better understand the molecular structures of isolated HDs, HD-DNA complexes, and HD-HD cofactor-DNA quaternary structures. Numerous qualitative experimental approaches, together with increasingly sophisticated bioinformatic algorithms, have also provided a great deal of insight on TF-DNA binding. A major limitation to these qualitative approaches is that the binding affinity predictions determined from these methods are unitless and can be difficult to reproduce. The modern field of molecular genetics would therefore benefit from the incorporation of more high quality quantitative measurements of HD TF-DNA binding affinities and more robust genome-wide TF binding site predictions.

We therefore recommend that the field shifts towards the adoption of more quantitative approaches, such as ITC, MST, and MITOMI, to analyze the binding affinity and specificity of these HD TF-DNA complexes. These biophysical measurements can then be combined with computational approaches, such as PWM-based approaches, in an iterative manner to improve future predictions on where, when, and how HD TFs are likely to bind in a eukaryotic genome. The use of an integrated system-level approach, that incorporates more biophysical approaches, will be critical to continue to improve and expand our understanding on how these important TFs function in the molecular control of gene expression during embryonic development.

Abbreviations¹

| 5-FOA | 5-fluoroorotic acid |
|----------|---|
| ABD-A | ABDOMINAL-A* |
| ABD-B | ABDOMINAL-B* |
| ANTP | ANTENNAPEDIA* |
| B1H | bacterial one hybrid |
| BAP | BAGPIPE* |
| BCD | BICOID* |
| CAD | CAUDAL* |
| Cebpa | CCAAT enhancer-binding protein alpha |
| ChIP-seq | chromatin immunoprecipitation-sequencing |
| DFD | DEFORMED* |
| DNase I | deoxyribonuclease I |
| KA | association constant |
| KD | dissociation constant |
| EMSA | Electrophoretic Mobility Shift Assays |
| EVE | EVEN-SKIPPED* |
| EXD | EXTRADENTICLE* |
| EY | EYELESS* |
| FTZ | FUSHI TARAZU* |
| hb | hunchback |
| HB | HUNCHBACK* |
| HD | homeodomain |
| ITC | isothermal titration calorimetry |
| ITT | in vitro transcription/translation |
| LAB | LABIAL* |
| LBL | LADYBIRD LATE* |
| MARs | matrix attachment regions |
| MITOMI | mechanically induced trapping of molecular interactions |
| Mkx | Mohawk |
| MKX | MOHAWK* |
| MSH | MUSCLE SPECIFIC HOMEOBOX* |
| MST | Microscale Thermophoresis |
| NMR | nuclear magnetic resonance |
| NRLB | no read left behind |
| PB | PROBOSCIPEDIA* |
| PBM | protein binding microarray |
| PTX1 | PITUITARY HOMEOBOX 1* |
| PWM | position weight matrix |
| SATB1 | SPECIAL AT-RICH SEQUENCE BINDING PROTEIN-1* |
| SCR | SEX COMBS REDUCED* |
| SIX4 | SIX HOMEOBOX 4* |
| SLOU | SLOUCH* |
| SO | SINE OCULUS* |
| SPR | surface plasmon resonance |
| SELEX | Systematic Evolution of Ligands by Exponential Enrichment |
| TFBSs | TF binding sites |
| TFs | transcription factors |
| TIN | TINMAN* |
| TTF-1 | TRANSCRIPTION TERMINATION FACTOR I* |
| UBX | ULTRABITHORAX* |
| VND-FL | full length VND |
| VND-NH | VND containing NK2-specific domain |
| VND/NK | ventral nervous system defective |

CRediT authorship contribution statement

Fadwa Mekkaoui: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Conceptualization. Robert A. Drewell: Writing – review & editing, Writing – original draft, Supervision, Formal analysis, Conceptualization. Jacqueline M. Dresch: Writing – review & editing, Writing – original draft, Supervision, Formal analysis, Conceptualization. Donald E. Spratt: Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Fadwa Mekkaoui reports financial support was provided by Clark University Gustaf H Carlson School of Chemistry and Biochemistry. Robert A. Drewell reports financial support was provided by National Institute of General Medical Sciences. Jacqueline M. Dresch reports financial support was provided by National Institute of General Medical Sciences. Donald E. Spratt reports financial support was provided by National Institute of General Medical Sciences. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2024.195074.

Data availability

No data was used for the research described in the article.

References

- A. Elmas, X. Wang, J.M. Dresch, The folded k-spectrum kernel: a machine learning approach to detecting transcription factor binding sites with gapped nucleotide dependencies, PLoS One 12 (2017) e0185570.
- [2] J.F. Kribelbauer, C. Rastogi, H.J. Bussemaker, R.S. Mann, Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes, Annu. Rev. Cell Dev. Biol. 35 (2019) 357–379.
- [3] M. Schnepf, M. von Reutern, C. Ludwig, C. Jung, U. Gaul, Transcription factor binding affinities and DNA shape readout, iScience 23 (2020) 101694.
- [4] M. Slattery, et al., Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins, Cell 147 (2011) 1270–1282.
- [5] D.S. Latchman, Transcription factors: an overview, Int. J. Biochem. Cell Biol. 29 (1997) 1305–1312.
- [6] Z. Yu, L.-J. Syu, D.M. Mellerick, Contextual interactions determine whether the Drosophila homeodomain protein, Vnd, acts as a repressor or activator, Nucleic Acids Res. 33 (2005) 1–12.
- [7] P. Weidemüller, M. Kholmatov, E. Petsalaki, J.B. Zaugg, Transcription factors: bridge between cell signaling and gene regulation, Proteomics 21 (2021) e2000034.
- [8] A.D. Frankel, P.S. Kim, Modular structure of transcription factors: implications for gene regulation, Cell 65 (1991) 717–719.
- [9] S. Kim, J. Wysocka, Deciphering the multi-scale, quantitative cis-regulatory code, Mol. Cell 83 (2023) 373–392.
- [10] S.A. Lambert, et al., The human transcription factors, Cell 172 (2018) 650–665.[11] S. Banerjee-Basu, A.D. Baxevanis, Molecular evolution of the homeodomain
- family of transcription factors, Nucleic Acids Res. 29 (2001) 3258–3269.
 [12] W.J. Gehring, M. Affolter, T. Bürglin, Homeodomain proteins, Annu. Rev. Biochem. 63 (1994) 487–526.
- [13] T.R. Bürglin, M. Affolter, Homeodomain proteins: an update, Chromosoma 125 (2016) 497–521.
- [14] T. Mitsis, et al., Transcription factors and evolution: an integral part of gene expression (review), World Acad. Sci. J. 2 (2020) 3–8.
- [15] D. Kimelman, B.L. Martin, Anterior-posterior patterning in early development: three strategies, Wiley Interdiscip. Rev. Dev. Biol. 1 (2012) 253–266.
- [16] S.J.H. Wang, et al., Homeodomain-interacting protein kinase (Hipk) plays roles in nervous system and muscle structure and function, PLoS One 15 (2020) e0221006.

¹ *Protein name is presented in all uppercase letters and non-italicized to help distinguish between the protein and gene symbol [170].

- [17] M. Tanaka-Matakatsu, J. Miller, W. Du, The homeodomain of Eyeless regulates cell growth and antagonizes the paired domain-dependent retinal differentiation function, Protein Cell 6 (2015) 68–78.
- [18] Y.Q. Qian, et al., The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors, Cell 59 (1989) 573–580.
- [19] J.A. Hirsch, A.K. Aggarwal, Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity, EMBO J. 14 (1995) 6280–6291.
- [20] Y.Q. Qian, et al., Nuclear magnetic resonance solution structure of the fushi tarazu homeodomain from Drosophila and comparison with the Antennapedia homeodomain, J. Mol. Biol. 238 (1994) 333–345.
- [21] J.M. Baird-Titus, et al., The solution structure of the native K50 Bicoid homeodomain bound to the consensus TAATCC DNA-binding site, J. Mol. Biol. 356 (2006) 1137–1151.
- [22] T. Zeiske, et al., Intrinsic DNA shape accounts for affinity differences between Hox-cofactor binding sites, Cell Rep. 24 (2018) 2221–2230.
- [23] E. Fraenkel, C.O. Pabo, Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex, Nat. Struct. Biol. 5 (1998) 692–697.
- [24] L. Zandarashvili, et al., Entropic enhancement of protein-DNA affinity by oxygento-sulfur substitution in DNA phosphate, Biophys. J. 109 (2015) 1026–1037.
- [25] M. Billeter, et al., Determination of the three-dimensional structure of the Antennapedia homeodomain from Drosophila in solution by 1H nuclear magnetic resonance spectroscopy, J. Mol. Biol. 214 (1990) 183–197.
- [26] M. Billeter, et al., Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex, J. Mol. Biol. 234 (1993) 1084–1093.
- [27] Y.Q. Qian, D. Resendez-Perez, W.J. Gehring, K. Wüthrich, The des(1-6) antennapedia homeodomain: comparison of the NMR solution structure and the DNA-binding affinity with the intact Antennapedia homeodomain, Proc. Natl. Acad. Sci. USA 91 (1994) 4091–4095.
- [28] K. Miyazono, et al., Cooperative DNA-binding and sequence-recognition mechanism of aristaless and clawless, EMBO J. 29 (2010) 1613–1623.
- [29] C.R. Kissinger, B.S. Liu, E. Martin-Blanco, T.B. Kornberg, C.O. Pabo, Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions, Cell 63 (1990) 579–590.
- [30] N.D. Clarke, C.R. Kissinger, J. Desjarlais, G.L. Gilliland, C.O. Pabo, Structural studies of the engrailed homeodomain, Protein Sci. 3 (1994) 1779–1787.
- [31] E. Fraenkel, M.A. Rould, K.A. Chambers, C.O. Pabo, Engrailed homeodomain-DNA complex at 2.2 A resolution: a detailed view of the interface and comparison with other engrailed structures, J. Mol. Biol. 284 (1998) 351–361.
- [32] T.L. Religa, et al., The helix-turn-helix motif as an ultrafast independently folding domain: the pathway of folding of engrailed homeodomain, Proc. Natl. Acad. Sci. USA 104 (2007) 9272–9277.
- [33] T.L. Religa, Comparison of multiple crystal structures with NMR data for engrailed homeodomain, J. Biomol. NMR 40 (2008) 189–202.
- [34] R. Joshi, et al., Functional specificity of a Hox protein mediated by the recognition of minor groove structure, Cell 131 (2007) 530–543.
- [35] N. Foos, et al., A flexible extension of the Drosophila ultrabithorax homeodomain defines a novel Hox/PBC interaction mode, Structure 23 (2015) 270–279.
- [36] J.M. Passner, H.D. Ryoo, L. Shen, R.S. Mann, A.K. Aggarwal, Structure of a DNAbound Ultrabithorax-Extradenticle homeodomain complex, Nature 397 (1999) 714–719.
- [37] J.M. Gruschus, D.H. Tsao, L.H. Wang, M. Nirenberg, J.A. Ferretti, Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance
- spectroscopy: basis of binding specificity, Biochemistry 36 (1997) 5372–5380.
 [38] D.H. Tsao, J.M. Gruschus, L.H. Wang, M. Nirenberg, J.A. Ferretti, The three-dimensional solution structure of the NK-2 homeodomain from Drosophila, J. Mol. Biol. 251 (1995) 297–307.
- [39] Z. Yang, X. Zeng, Y. Zhao, R. Chen, AlphaFold2 and its applications in the fields of biology and medicine, Signal Transduct. Target. Ther. 8 (2023) 115.
- [40] R. Ishima, D.A. Torchia, Protein dynamics from NMR, Nat. Struct. Biol. 7 (2000) 740–743.
- [41] L.K. Robertson, J.W. Mahaffey, Insect homeotic complex genes and development, lessons from drosophila and beyond x, in: Reference Module in Life Sciences, Elsevier, 2017.
- [42] E. Fraenkel, C.O. Pabo, Comparison of X-ray and NMR structures for the Antennapedia homeodomain–DNA complex, Nat. Struct. Biol. 5 (1998) 692–697.
- [43] D.S. Burz, R. Rivera-Pomar, H. Jäckle, S.D. Hanes, Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the Drosophila embryo, EMBO J. 17 (1998) 5998–6009.
- [44] D.S. Wilson, C. Desplan, Homeodomain proteins: cooperating to be different, Curr. Biol. 5 (1995) 32–34.
- [45] P.A. Beachy, et al., Cooperative binding of an Ultrabithorax homeodomain protein to nearby and distant DNA sites, Mol. Cell. Biol. 13 (1993) 6941–6956.
- [46] K.P. Pettie, J.M. Dresch, R.A. Drewell, Spatial distribution of predicted transcription factor binding sites in Drosophila ChIP peaks, Mech. Dev. 141 (2016) 51–61.
- [47] H. Janssens, et al., Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene, Nat. Genet. 38 (2006) 1159–1165.
- [48] E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall, U. Gaul, Predicting expression patterns from regulatory sequence in Drosophila segmentation, Nature 451 (2008) 535–540.
- [49] W.D. Fakhouri, et al., Deciphering a transcriptional regulatory code: modeling short-range repression in the Drosophila embryo, Mol. Syst. Biol. 6 (2010) 341.

BBA - Gene Regulatory Mechanisms 1868 (2025) 195074

- [50] X. He, M.A. Samee, C. Blatti, S. Sinha, Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression, PLoS Comput. Biol. 6 (2010).
- [51] D.S. Parker, M.A. White, A.I. Ramos, B.A. Cohen, S. Barolo, The cis-regulatory logic of hedgehog gradient responses: key roles for gli binding affinity, competition, and cooperativity, Sci. Signal. 4 (2011) ra38.
- [52] M.A. White, D.S. Parker, S. Barolo, B.A. Cohen, A model of spatially restricted transcription in opposing gradients of activators and repressors, Mol. Syst. Biol. 8 (2012) 614.
- [53] A.R. Kim, et al., Rearrangements of 2.5 kilobases of noncoding DNA from the Drosophila even-skipped locus define predictive rules of genomic cis-regulatory logic, PLoS Genet. 9 (2013) e1003243.
- [54] R.A. Drewell, et al., Deciphering the combinatorial architecture of a Drosophila homeotic gene enhancer, Mech. Dev. 131 (2014) 68–77.
- [55] R. Sayal, J.M. Dresch, I. Pushel, B.R. Taylor, D.N. Arnosti, Quantitative perturbation-based analysis of gene expression predicts enhancer activity in early Drosophila embryo, Elife 5 (2016).
- [56] K.A. Barr, J. Reinitz, A sequence level model of an intact locus predicts the location and function of nonadditive enhancers, PLoS One 12 (2017) e0180861.[57] Y. Liu, K. Barr, J. Reinitz, Fully interpretable deep learning model of
- transcriptional control, Bioinformatics 36 (2020) i499–i507.
- [58] P. Dibaeinia, S. Sinha, Deciphering enhancer sequence using thermodynamicsbased models and convolutional neural networks, Nucleic Acids Res. 49 (2021) 10309–10327.
- [59] S. Bhogale, S. Sinha, Thermodynamics-based modeling reveals regulatory effects of indirect transcription factor-DNA binding, iScience 25 (2022) 104152.
- [60] O.G. Berg, P.H. von Hippel, Selection of DNA binding sites by regulatory proteins, Trends Biochem. Sci. 13 (1988) 207–211.
- [61] O.G. Berg, P.H. von Hippel, Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites, J. Mol. Biol. 200 (1988) 709–723.
- [62] P.H. von Hippel, O.G. Berg, On the specificity of DNA-protein interactions, Proc. Natl. Acad. Sci. USA 83 (1986) 1608–1612.
- [63] O.G. Berg, P.H. von Hippel, Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters, J. Mol. Biol. 193 (1987) 723–750.
- [64] R. Staden, Computer methods to locate signals in nucleic acid sequences, Nucleic Acids Res. 12 (1984) 505–519.
- [65] G.D. Stormo, T.D. Schneider, L. Gold, A. Ehrenfeucht, Use of the 'Perceptron' algorithm to distinguish translational initiation sites in *E. coli*, Nucleic Acids Res. 10 (1982) 2997–3011.
- [66] D.J. Galas, DNAse footprinting: a simple method for the detection of protein-DNA binding specificity, Nucleic Acids Res. 5 (9) (1978) 3157–3170.
- [67] T.D. Tullius, DNA footprinting with hydroxyl radical, Nature 332 (1988) 663-664.
- [68] J.E. Herrera, J.B. Chaires, Characterization of preferred deoxyribonuclease I cleavage sites, J. Mol. Biol. 236 (1994) 405–411.
- [69] D. Suck, C. Oefner, Structure of DNase I at 2.0 Å resolution suggests a mechanism for binding to and cutting DNA, Nature 321 (1986) 620–625.
- [70] Ş. Tüzmen, et al., Chapter 14 techniques for nucleic acid engineering: the foundation of gene manipulation, in: D. Barh, V. Azevedo (Eds.), Omics Technologies and Bio-engineering, Academic Press, 2018, pp. 247–315.
- [71] E. Zaychikov, P. Schickor, L. Denissova, H. Heumann, Hydroxyl radical footprinting, Methods Mol. Biol. 148 (2001) 49–61.
- [72] J. Geiselmann, F. Boccard, Ultraviolet-laser footprinting, Methods Mol. Biol. 148 (2001) 161–173.
- [73] S.D. Weatherbee, G. Halder, J. Kim, A. Hudson, S. Carroll, Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the Drosophila haltere, Genes Dev. 12 (1998) 1474–1482.
- [74] S.C. Ekker, K.E. Young, D.P. von Kessler, P.A. Beachy, Optimal DNA sequence recognition by the Ultrabithorax homeodomain of Drosophila, EMBO J. 10 (1991) 1179–1186.
- [75] W. Han, Y. Yu, N. Altan, L. Pick, Multiple proteins interact with the fushi tarazu proximal enhancer, Mol. Cell. Biol. 13 (1993) 5549–5559.
- [76] L.-H. Wang, R. Chmelik, D. Tang, M. Nirenberg, Identification and analysis of vnd/NK-2 homeodomain binding sites in genomic DNA, Proc. Natl. Acad. Sci. 102 (2005) 7097–7102.
- [77] R.A.C. Ferraz, et al., DNA-protein interaction studies: a historical and comparative analysis, Plant Methods 17 (2021) 82.
- [78] B. Leblanc, T. Moss, DNase I footprinting, Methods Mol. Biol. 543 (2009) 37–47.
 [79] M.F. Carey, C.L. Peterson, S.T. Smale, DNase I footprinting, Cold Spring Harb Protoc 2013 (2013) 469–478.
- [80] N.S. Holden, C.E. Tacon, Principles and problems of the electrophoretic mobility shift assay, J. Pharmacol. Toxicol. Methods 63 (2011) 7–14.
- [81] L.M. Hellman, M.G. Fried, Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions, Nat. Protoc. 2 (2007) 1849–1861.
- [82] D.M. Anderson, et al., Characterization of the DNA-binding properties of the Mohawk homeobox transcription factor, J. Biol. Chem. 287 (2012) 35351–35359.
- [83] R. Orlomoski, et al., Rapid and efficient purification of Drosophila homeodomain transcription factors for biophysical characterization, Protein Expr. Purif. 158 (2019) 9–14.
- [84] F. Yang, W.Z. Ouma, W. Li, A.I. Doseff, E. Grotewold, Chapter eleven establishing the architecture of plant gene regulatory networks, in: S.E. O'Connor (Ed.), Methods in Enzymology vol. 576, Academic Press, 2016, pp. 251–304.

- [85] C. Smaczniak, G.C. Angenent, K. Kaufmann, SELEX-Seq: a method to determine DNA binding specificities of plant transcription factors, Methods Mol. Biol. 1629 (2017) 67–82.
- [86] N. Komarova, A. Kuznetsov, Inside the black box: what makes SELEX better? Molecules 24 (2019).
- [87] T.R. Riley, et al., SELEX-seq: a method for characterizing the complete repertoire of binding site preferences for transcription factor complexes, Methods Mol. Biol. 1196 (2014) 255–278.
- [88] C. Rastogi, et al., Accurate and sensitive quantification of protein-DNA binding affinity, Proc. Natl. Acad. Sci. USA 115 (2018) E3692–e3701.
- [89] S. Ruan, S.J. Swamidass, G.D. Stormo, BEESEM: estimation of binding energy models using HT-SELEX data, Bioinformatics 33 (2017) 2288–2295.
- [90] D. Raha, M. Hong, M. Snyder, ChIP-Seq: a method for global identification of regulatory elements in the genome, Curr. Protoc. Mol. Biol. 91 (2010), 21.19.1-21.19.14.
- [91] P. Gade, D.V. Kalvakolanu, Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity, Methods Mol. Biol. 809 (2012) 85–104.
- [92] I. Albert, et al., Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome, Nature 446 (2007) 572–576.
- [93] R. Mundade, H.G. Ozer, H. Wei, L. Prabhu, T. Lu, Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond, Cell Cycle 13 (2014) 2847–2852.
- [94] P.J. Park, ChIP-seq: advantages and challenges of a maturing technology, Nat. Rev. Genet. 10 (2009) 669–680.
- [95] B. Jusiak, et al., Regulation of Drosophila eye development by the transcription factor Sine oculis, PLoS One 9 (2014) e89695.
- [96] B.L. Kidder, G. Hu, K. Zhao, ChIP-Seq: technical considerations for obtaining high-quality data, Nat. Immunol. 12 (2011) 918–922.
- [97] B. Cain, et al., Prediction of cooperative homeodomain DNA binding sites from high-throughput-SELEX data, Nucleic Acids Res. 51 (2023) 6055–6072.
- [98] T.S. Furey, ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions, Nat. Rev. Genet. 13 (2012) 840–852.
 [99] 2 Chromatin patterns at transcription factor binding sites, Nature (2019), https://
- [59] Z chromini patterna a transcription factor binding sites, tradine (2015), https://doi.org/10.1038/nature28171.
 [100] Y.L. Jung, et al., Impact of sequencing depth in ChIP-seq experiments, Nucleic
- Acids Res. 42 (2014) e74. [101] J. Gutin, et al., Fine-resolution mapping of TF binding and chromatin interactions,
- [101] J. Gull Rep. 22 (2018) 2797–2807.
 [102] H. Hou, M. Hudson, M. Liang, Chapter 14 best practices for ChIP-seq and its data
- [102] H. Hot, M. Hudson, M. Lang, Chapter 14 Dest practices for Chir Seq and its data analysis, in: D.F. Duzen, M.H.M. Schmidt (Eds.), Rigor and Reproducibility in Genetics and Genomics, Academic Press, 2024, pp. 319–341.
- [103] T. Ribarska, G.D. Gilfillan, Native chromatin immunoprecipitation-sequencing (ChIP-Seq) from low cell numbers, Methods Mol. Biol. 1689 (2018) 157–166.
- [104] M.B. Noyes, Analysis of specific protein-DNA interactions by bacterial one-hybrid assay, Methods Mol. Biol. 786 (2012) 79–95.
- [105] R.T. Clubb, S. Schumacher, K. Mizuuchi, A.M. Gronenborn, G.M. Clore, Solution structure of the I gamma subdomain of the Mu end DNA-binding domain of phage Mu transposase, J. Mol. Biol. 273 (1997) 19–25.
- [106] X. Meng, M.H. Brodsky, S.A. Wolfe, A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors, Nat. Biotechnol. 23 (2005) 988–994.
- [107] R.G. Christensen, et al., A modified bacterial one-hybrid system yields improved quantitative models of transcription factor specificity, Nucleic Acids Res. 39 (2011) e83.
- [108] M.L. Bulyk, Discovering DNA regulatory elements with bacteria, Nat. Biotechnol. 23 (2005) 942–944.
- [109] X. Meng, M.H. Brodsky, S.A. Wolfe, A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors, Nat. Biotechnol. 23 (2005) 988–994.
- [110] M.B. Noyes, et al., A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system, Nucleic Acids Res. 36 (2008) 2547–2560.
- [111] K. Churion, Y. Liu, H.C. Hsiao, K.S. Matthews, S.E. Bondos, Measuring Hox-DNA binding by electrophoretic mobility shift analysis, Methods Mol. Biol. 1196 (2014) 211–230.
- [112] X. Meng, S.A. Wolfe, Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system, Nat. Protoc. 1 (2006) 30–45.
- [113] R.P. Sparks, J.L. Jenkins, R. Fratti, Use of surface plasmon resonance (SPR) to determine binding affinities and kinetic parameters between components important in fusion machinery, Methods Mol. Biol. 1860 (2019) 199–210.
- [114] M. Seimiya, Y. Kurosawa, Kinetics of binding of Antp homeodomain to DNA analyzed by measurements of surface plasmon resonance, FEBS Lett. 398 (1996) 279–284.
- [115] B. Douzi, Surface plasmon resonance: a sensitive tool to study protein–protein interactions, in: L. Journet, E. Cascales (Eds.), Bacterial Secretion Systems: Methods and Protocols, Springer US, New York, NY, 2024, pp. 363–382.
- [116] S. Wang, G.M. Poon, W.D. Wilson, Quantitative investigation of protein-nucleic acid interactions by biosensor surface plasmon resonance, Methods Mol. Biol. 1334 (2015) 313–332.
- [117] M.S. Yousef, B.W. Matthews, Structural basis of Prospero-DNA interaction: implications for transcription regulation in developing cells, Structure 13 (2005) 601–607.
- [118] S. Yoo, Kinetic analysis of Drosophila Vnd protein containing homeodomain with its target sequence, BMB Rep. 43 (2010) 407–412.

- [119] M.F. Berger, M.L. Bulyk, Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors, Nat. Protoc. 4 (2009) 393–411.
- [120] K.K. Andrilenas, A. Penvose, T. Siggers, Using protein-binding microarrays to study transcription factor specificity: homologs, isoforms and complexes, Brief. Funct. Genomics 14 (2015) 17–29.
- [121] S. Mukherjee, et al., Rapid analysis of the DNA-binding specificities of
- transcription factors with DNA microarrays, Nat. Genet. 36 (2004) 1331–1339.
 [122] B.W. Busser, et al., Molecular mechanism underlying the regulatory specificity of a Drosophila homeodomain protein that specifies myoblast identity, Development 139 (2012) 1164–1174.
- [123] M.B. Noyes, et al., Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites, Cell 133 (2008) 1277–1289.
- [124] R.G. Matos, A. Barbas, C.M. Arraiano, Comparison of EMSA and SPR for the characterization of RNA-RNase II complexes, Protein J. 29 (2010) 394–397.
- [125] K.R. Nitta, et al., Conservation of transcription factor binding specificities across 600 million years of bilateria evolution, Elife 4 (2015).
 [126] X. Du, et al., Insights into protein-ligand interactions: mechanisms, models, and
- methods, Int. J. Mol. Sci. 17 (2016).
- [127] A.M. Bendel, et al., The genetic architecture of protein interaction affinity and specificity, Nat. Commun. 15 (2024) 8868.
- [128] V.V. Demidov, M.D. Frank-Kamenetskii, Two sides of the coin: affinity and specificity of nucleic acid interactions, Trends Biochem. Sci. 29 (2004) 62–71.
 [129] G.Z. Hertz, G.D. Stormo, Identifying DNA and protein patterns with statistically
- [129] G.Z. HETE, G.D. Stormo, identifying DNA and protein patterns with statistically significant alignments of multiple sequences, Bioinformatics 15 (1999) 563–577.
 [130] T.L. Bailey, M. Gribskov, Combining evidence using p-values: application to
- sequence homology searches, Bioinformatics 14 (1998) 48–54.
- [131] G.D. Stormo, DNA binding sites: representation and discovery, Bioinformatics 16 (2000) 16–23.
- [132] G.D. Stormo, D.S. Fields, Specificity, free energy and information content in protein-DNA interactions, Trends Biochem. Sci. 23 (1998) 109–113.
- [133] P.V. Benos, M.L. Bulyk, G.D. Stormo, Additivity in protein-DNA interactions: how good an approximation is it? Nucleic Acids Res. 30 (2002) 4442–4451.
- [134] M.O. Starr, et al., Molecular dissection of cis-regulatory modules at the Drosophila bithorax complex reveals critical transcription factor signature motifs, Dev. Biol. 359 (2011) 290–302.
- [135] E.J. Ostrin, et al., Genome-wide identification of direct targets of the Drosophila retinal determination protein Eyeless, Genome Res. 16 (2006) 466–476.
- [136] M.L. Bulyk, P.L.F. Johnson, G.M. Church, Nucleotides of transcription factor binding sites exert interdependent effects on the binding affinities of transcription factors, Nucleic Acids Res. 30 (2002) 1255–1261.
- [137] T.-K. Man, G.D. Stormo, Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay, Nucleic Acids Res. 29 (2001) 2471–2478.
- [138] R.G. Zellers, R.A. Drewell, J.M. Dresch, MARZ: an algorithm to combinatorially analyze gapped n-mer models of transcription factor binding, BMC Bioinformatics 16 (2015) 30.
- [139] J.M. Dresch, R.G. Zellers, D.K. Bork, R.A. Drewell, Nucleotide interdependency in transcription factor binding sites in the Drosophila genome, Gene Regul. Syst. Biol. 10 (2016) GRSB.S38462.
- [140] E. Bertolino, J. Reinitz, Manu., The analysis of novel distal Cebpa enhancers and silencers using a transcriptional model reveals the complex regulatory logic of hematopoietic lineage specification, Dev. Biol. 413 (2016) 128–144.
- [141] T.A. Duncombe, A.M. Tentori, A.E. Herr, Microfluidics: reframing biological enquiry, Nat. Rev. Mol. Cell Biol. 16 (2015) 554–567.
- [142] J.L. Garcia-Cordero, S.J. Maerkl, Mechanically induced trapping of molecular interactions and its applications, SLAS Technol. 21 (2016) 356–367.
- [143] S. Rockel, M. Geertz, S.J. Maerkl, MITOMI: a microfluidic platform for in vitro characterization of transcription factor-DNA interaction, Methods Mol. Biol. 786 (2012) 97–114.
- [144] G.D. Stormo, Y. Zhao, Determining the specificity of protein-DNA interactions, Nat. Rev. Genet. 11 (2010) 751–760.
- [145] B.Z. He, A.K. Holloway, S.J. Maerkl, M. Kreitman, Does positive selection drive transcription factor binding site turnover? A test with Drosophila Cis-regulatory modules, PLoS Genet. 7 (2011) e1002053.
- [146] L. Shokri, et al., A comprehensive Drosophila melanogaster transcription factor interactome, Cell Rep. 27 (2019) 955–970.e7.
- [147] S.J. Maerkl, S.R. Quake, A systems approach to measuring the binding energy landscapes of transcription factors, Science 315 (2007) 233–237.
- [148] A.M. Mueller, et al., MicroScale thermophoresis: a rapid and precise method to quantify protein-nucleic acid interactions in solution, Methods Mol. Biol. 1654 (2017) 151–164.
- [149] M. Jerabek-Willemsen, C.J. Wienken, D. Braun, P. Baaske, S. Duhr, Molecular interaction studies using microscale thermophoresis, Assay Drug Dev. Technol. 9 (2011) 342–353.
- [150] L. Huang, C. Zhang, Microscale thermophoresis (MST) to detect the interaction between purified protein and small molecule, Methods Mol. Biol. 2213 (2021) 187–193.
- [151] S. El Deeb, et al., Microscale thermophoresis as a powerful growing analytical technique for the investigation of biomolecular interaction and the determination of binding parameters, Methods Appl. Fluoresc. 10 (2022) 042001.
- [152] M.R. Duff Jr., J. Grubbs, E.E. Howell, Isothermal titration calorimetry for measuring macromolecule-ligand affinity, J. Vis. Exp. 2796 (2011).
- [153] A.L. Feig, Applications of isothermal titration calorimetry in RNA biochemistry and biophysics, Biopolymers 87 (2007) 293–301.

- [154] J. Ramirez, Y. Nominé, High-quality data of protein/peptide interaction by isothermal titration calorimetry, Methods Mol. Biol. 1964 (2019) 99–117.
- [155] B.W. Peters, V. Frasca, K.R. Brown, Recent developments in isothermal titration calorimetry label free screening, Comb. Chem. High Throughput Screen. 12 (2009) 772–790.
- [156] K. Yamasaki, T. Yamasaki, The combination of sequence-specific and nonspecific DNA-binding modes of transcription factor SATB1, Biochem. J. 473 (2016) 3321–3339.
- [157] T.-Y. Wang, Z.-M. Han, Y.-R. Chai, J.-H. Zhang, A mini review of MAR-binding proteins, Mol. Biol. Rep. 37 (2010) 3553–3560.
- [158] F.E. Torres, M.I. Recht, J.E. Coyle, R.H. Bruce, G. Williams, Higher throughput calorimetry: opportunities, approaches and challenges, Curr. Opin. Struct. Biol. 20 (2010) 598–605.
- [159] D. Porcelli, B. Fischer, S. Russell, R. White, Chromatin accessibility plays a key role in selective targeting of Hox proteins, Genome Biol. 20 (2019) 115.
- [160] L. Bridoux, et al., HOX paralogs selectively convert binding of ubiquitous transcription factors into tissue-specific patterns of enhancer activation, PLoS Genet. 16 (2020) e1009162.
- [161] J.F. Kribelbauer, et al., Context-dependent gene regulation by homeodomain transcription factor complexes revealed by shape-readout deficient proteins, Mol. Cell 78 (2020) 152–167.e11.

- BBA Gene Regulatory Mechanisms 1868 (2025) 195074
- [162] N. Abe, et al., Deconvolving the recognition of DNA shape from sequence, Cell 161 (2015) 307–318.
- [163] S.K. Menon, C.M. Lawrence, Helix-turn-helix motif, in: S. Maloy, K. Hughes (Eds.), Brenner's Encyclopedia of Genetics, second edition, Academic Press, San Diego, 2013, pp. 412–415.
- [164] S.W. Chu, et al., Exploring the DNA-recognition potential of homeodomains, Genome Res. 22 (2012) 1889–1898.
- [165] A.I. Dragan, et al., Forces driving the binding of homeodomains to DNA, Biochemistry 45 (2006) 141–151.
- [166] J.M. Baird-Titus, et al., The solution structure of the native K50 Bicoid homeodomain bound to the consensus TAATCC DNA-binding site, J. Mol. Biol. 356 (2006) 1137–1151.
- [167] L.a.W.D. Schrödinger, The PyMOL Molecular Graphics System, 2.0 edn, 2015 (New York, NY, USA).
- [168] C.M. Bergman, J.W. Carlson, S.E. Celniker, Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, *Drosophila melanogaster*, Bioinformatics 21 (8) (2005) 1747–1749.
- [169] M.F. Berger, M.L. Bulyk, Protein binding microarrays (PBMs) for rapid, highthroughput characterization of the sequence specificities of DNA binding proteins, Methods Mol. Biol. 338 (2006) 245–260.
- [170] P.W.H. Holland, H.A.F. Booth, E.A. Bruford, Classification and nomenclature of all human homeobox genes, BMC Biol. 5 (2007) 47.