Published in final edited form as:

Protein Expr Purif. 2019 June ; 158: 9–14. doi:10.1016/j.pep.2019.02.001.

Rapid and Efficient Purification of *Drosophila* Homeodomain Transcription Factors for Biophysical Characterization

Rachel Orlomoski^{1,2}, Aaron Bogle^{1,2}, Jeanmarie Loss^{1,2}, Rylee Simons^{1,2}, Jacqueline M. Dresch³, Robert A. Drewell^{2,*}, and Donald E. Spratt^{1,*}

¹Gustaf H. Carlson School of Chemistry & Biochemistry, Clark University, 950 Main St., Worcester MA, USA 01610

²Department of Biology, Clark University, 950 Main St., Worcester MA, USA 01610

³Department of Math & Computer Science, Clark University, 950 Main St., Worcester MA, USA 01610

Abstract

Homeodomain transcription factors (HD TFs) are a large class of evolutionarily conserved DNA binding proteins that contain a basic 60-amino acid region required for binding to specific DNA sites. In *Drosophila melanogaster*, many of these HD TFs are expressed in the early embryo and control transcription of target genes in development through interaction with *cis*-regulatory modules. Previous studies where some of the *Drosophila* HD TFs were purified required the use of strong denaturants (i.e. 6M urea) and multiple chromatography columns, making the downstream biochemical examination of the isolated protein difficult. To circumvent these obstacles, we have developed a streamlined expression and purification protocol to produce large yields of *Drosophila* HD TFs. Using the HD TFs FUSHI-TARAZU (FTZ), ANTENNAPEDIA (ANTP), ABDOMINAL-A (ABD-A), ABDOMINAL-B (ABD-B), and ULTRABITHORAX (UBX) as examples, we demonstrate that our 3-day protocol involving the overexpression of His₆-SUMO fusion constructs in *E. coli* followed by a Ni²⁺-IMAC, SUMO-tag cleavage with the SUMO protease Ulp1, and a heparin column purification produces pure, soluble protein in biological buffers around pH 7 in the absence of denaturants. Electrophoretic mobility shift assays (EMSA) confirm that the purified HD proteins are functional and nuclear magnetic resonance (NMR) spectra confirm that the purified HDs are well-folded. These purified HD TFs can be used in future biophysical experiments to structurally and biochemically characterize how and why these HD TFs bind to different DNA sequences and further probe how nucleotide differences contribute to TF-DNA specificity in the HD family.

* Authors to whom correspondence should be addressed: Donald E. Spratt, Gustaf H. Carlson School of Chemistry & Biochemistry, Clark University, 950 Main St., Worcester, MA, USA 01610, Phone: 508-793-7536, Fax: 508-793-7117, dspratt@clarku.edu, Web: <http://wordpress.clarku.edu/dspratt>, Robert A. Drewell, Department of Biology, Clark University, 950 Main St., Worcester, MA, USA 01610, Phone: 508-793-7625, Fax: 508-793-7174, rdrewell@clarku.edu, Web: <http://wordpress.clarku.edu/drewelllab>.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Homeodomain transcription factor; DNA binding; protein-DNA interactions; Electrophoretic Mobility Shift Assay; NMR spectroscopy

Introduction

Gene regulation is a critically important biological process that is required for various cellular functions including homeostasis, cellular differentiation, and intracellular responses to external stimuli [1, 2]. Transcription factors (TFs) are a class of proteins that regulate gene expression by binding to specific DNA sequences within *cis*-regulatory modules (CRMs) and subsequently controlling the recruitment of the basal transcriptional machinery, including RNA polymerase, to specific target promoters in the genome [3–6]. Although the list of well characterized CRMs is relatively short, a number of key molecular features of the system appear to be shared – (i) TFs recognize binding sites of around 7–10 bp in size, (ii) TFs frequently bind a number of different target DNA sequences with varying affinity, (iii) CRMs are hundreds of base pairs in length and often contain multiple DNA binding sequences recognized by a specific TF, and (iv) the functional output from most CRMs is controlled by a complex interacting suite of TFs potentially binding in parallel at many DNA sites [3, 7–14]. As a result, the underlying biochemical and biophysical basis for the formation of these regulatory TF-DNA complexes at CRMs remain poorly characterized.

In this study, we purify five homeodomain (HD)-containing TFs, FUSHI-TARAZU (FTZ), ANTENNAPEDIA (ANTP), ABDOMINAL-A (ABD-A), ABDOMINAL-B (ABD-B), and ULTRABITHORAX (UBX) that are required during embryonic development in the fruit fly *Drosophila melanogaster* [11, 15]. HD proteins are a large class of DNA-binding TFs that contain a highly conserved 60 amino acid sequence encoded by the corresponding 180 bp homeobox in the genes of fungi, plants and animals [16, 17]. *Drosophila* provides an excellent model organism in which to study these evolutionarily conserved TFs due to its extensively studied genome and well-characterized development. The 3D structure of the HD consists of a three alpha-helix bundle. Helix 2 and 3 form a helix-turn-helix (HTH) motif, in which the two helices are anti-parallel to one another, while Helix 1 sits perpendicular to the axes of the other two helices and interacts directly with the major groove of DNA [18–25]. Biological assays, including the bacterial one-hybrid system [26], have been successfully employed to investigate the binding specificity of HD TFs. Based on the results from these studies, HDs are known to bind to a variety of 7–9 bp DNA sequences, but in animal genomes the core of most of these binding sites consist of a canonical 5'-TAAT-3' sequence [11, 27].

To date, the technical challenges and labor-intensive effort associated with the expression, purification and DNA interaction assays of HD TFs have limited the number of detailed biophysical binding studies. However, valuable structural (Nuclear Magnetic Resonance, NMR) and DNA binding (Electrophoretic mobility shift assay, EMSA) data has been obtained for a few *Drosophila* HD TFs, including FTZ, ANTP, ENGRAILED (ENG), and (ABD-B), demonstrating that isolated HDs exhibit significant three-dimensional structural

constraint and that their DNA binding profiles are similar [18–22, 25, 28–33]. A recent structural study of ABD-B in complex with its cofactor extradenticle and four different DNA sequences suggests that the subtle differences between binding affinity may be determined by the initial DNA shape prior to binding rather than the specific DNA sequence itself [34]. However, there still remains many key unanswered biological questions regarding how these structurally similar HD TFs recognize their distinct DNA binding sites, what their binding affinities are at different DNA sites, and how small changes in the DNA sequence affect their affinity.

In this paper we describe a novel and rapid method to overexpress and purify HD TFs using a His₆-SUMO affinity tag system in the absence of denaturants. We demonstrate that the isolated HDs are well-folded by NMR spectroscopy and are functional as judged by their ability to bind specifically to known DNA sequences in EMSA assays. These studies lay the foundation for future biophysical and structural experiments to further probe and decipher the molecular basis for how and why different HD TFs recognize and bind to specific DNA sequences to regulate downstream transcription events.

Materials and Methods

Cloning of HD TF Inserts into a Modified His-SUMO Vector

The HD open reading frames of FTZ (Uniprot: P02835, residues 254–313) and ANTP (P02833, residues 297–356) were *Escherichia coli* (*E. coli*) codon-optimized and synthesized by Synbio Technologies and cloned into a modified T7-inducible HisSUMO vector in between compatible *KpnI* and *XbaI* restriction sites (Figure 1). DNA coding for the HD domains of UBX, ABD-A and ABD-B were generously provided by Dr. Scot Wolfe (University of Massachusetts Medical School, Worcester MA) [26]. The HD domains of UBX (P83949, residues 387–446), AbdA (P29555, residues 398–457), and AbdB (P09087, residues 295–354) were PCR amplified using KOD polymerase (Novagen), digested with *KpnI* and *XbaI*, and ligated into the compatible restriction sites of the modified HisSUMO vector using T4 DNA ligase (Thermo-Fisher). The resulting vectors – HisSUMO-FTZHD, HisSUMO-ANTPHD, HisSUMO-UBXHD, HisSUMOABDAHD, and HisSUMO-ABDBHD – were verified by DNA sequencing (Macrogen, Cambridge, MA).

Expression and purification of HD TF proteins

The HisSUMO-FTZHD, HisSUMO-ANTPHD, HisSUMO-UBXHD, HisSUMO-ABDAHD, and HisSUMO-ABDBHD plasmids were transformed into *E. coli* BL21 (DE3) (Stratagene) and grown in 4 × 1L of Luria Bertani (LB) media supplemented with 100 µg ampicillin. When the OD₆₀₀ of each culture reached 1.0, the temperature was lowered to 30°C and the cells were induced with 1 mM IPTG for 3 hours. Cells were harvested by centrifugation at 6000 × g for 10 minutes using a Sorvall LYNX 4000 superspeed centrifuge with a Fiberlite F10–4×1000 LEX Carbon Fiber rotor (Thermo-Fisher).

Cell pellets were resuspended in 25 mL wash buffer (50 mM Na₂HPO₄ pH 7.5, 300 mM NaCl, 10 mM imidazole) supplemented with 250 µL of ProBlock Gold bacterial protease inhibitor cocktail (100x; Gold Biotechnology, St. Louis, MO) and lysed using an

Emulsiflex-C5 homogenizer (Avestin, Ottawa, ON, Canada). The cell lysate was clarified by ultracentrifugation using an Optima L-80 XP ultracentrifuge with a 70.1 Ti rotor (Beckman-Coulter) at 41,000 rpm for 40 minutes at 4°C. After centrifugation, the clarified supernatant was then applied to 5–10 mL of HisPur Ni²⁺-NTA resin (Thermo-Fisher) pre-equilibrated with wash buffer at a flow rate of 0.5 mL/min. After the flowthrough was collected, the resin was washed with 20–50 column volumes of wash buffer to ensure all of the unbound proteins were removed from the resin. The His-SUMO-HD protein was then eluted with 20 mL of elution buffer (50 mM Na₂HPO₄ pH 7.5, 300 mM NaCl, 250 mM imidazole). The fractions containing eluted protein as judged by SDS-PAGE with Coomassie Brilliant Blue R-250 staining were pooled. The pooled HisSUMO-HD protein was then incubated with 1 mg recombinant Ulp1 (SUMO specific protease) per 25 mL of eluted protein for 1 hour at 25°C to cleave the HisSUMO fusion tag, followed by dialysis against binding buffer (10 mM Na₂HPO₄, pH 7.0, 1 mM DTT) with stirring overnight at 4°C. After dialysis, the Ulp1 cleaved protein sample was centrifuged and syringe filtered through a 0.45 µm filter. The clarified supernatant was then injected on to a 5 mL HiTrap Heparin HP column preequilibrated with binding buffer at a flow rate of 1 mL/min using an AKTA Pure 25L FPLC (GE Healthcare). After washing the column with 10 column volumes of binding buffer, the HD protein was eluted using a linear salt gradient with high salt elution buffer (10 mM Na₂HPO₄, pH 7.0, 2 M NaCl, 1 mM DTT) at a flow rate of 1 mL/min over 60 minutes. Fractions containing pure HD protein as assessed by SDS-PAGE were pooled and quantified by A₂₈₀ using a Nanodrop OneC spectrophotometer (Thermo-Fisher), aliquoted, flash frozen on N₂(l) and stored at –80°C.

Electrophoretic Mobility Shift Assays

The binding of the isolated HDs to double-stranded DNA (dsDNA) was assessed by electrophoretic mobility shift assays (EMSAs) using a positive control 5'-CGGTTTAATTAGCCC-3' (consensus) and a negative control 5'-CGCGCGCCGCGCGCCC-3' (null). These sequences were chosen after analysing aligned binding sequences for FTZ obtained using a bacterial one-hybrid (B1H) system and SOLEXA sequencing [11]. This data set is comprised of 347 total sequences, each of length 7 bp. Since many homeodomain proteins involved in early *Drosophila* development, including ABD-A, ABD-B, ANTP and UBX, are thought to have 8 bp binding sequences with the signature core motif 5'-'TAAT'-3' at nucleotides 3 through 6, we expanded our sequences to 8 bp sequences by adding a nucleotide to the 5' end. The 8 bp sequence underlined in the positive control, TTTAATTA, is the sequence that was found with the highest frequency within the data set, flanked on the 5' end with a 'T', resulted in the highest motif score when all sequences were run through the commonly used bioinformatic algorithm MAST [35]. The 8 bp sequence underlined in the negative control, CGCCGCGC, was chosen due to its lack of representation among the aligned FTZ binding sites and its resulting low MAST motif score. Two oligos were synthesized for the consensus and null sequences by Life Technologies (Carlsbad, CA) with the forward oligo having a 5' fluorescein dye chemically attached to its 5' end (FAM) to aid in visualization. To make FAM-dsDNA, the forward and reverse oligos for each sequence were diluted using a 20 mM HEPES pH 7.4, 8 mM MgCl₂ buffer then mixed in an equimolar ratio. The oligo mixture was heated at 95°C for 5 minutes, followed by 5°C decrease per minute using a T100

Thermal Cycler (Bio-Rad) to facilitate dsDNA annealing. HD-dsDNA complexes were formed by mixing 10 μ M fluorescein-labeled dsDNA with increasing concentrations of isolated HD protein from 0 to 25 μ M (i.e. HD:protein ratios of 0:1 up to 5:1) in a buffer containing 20 mM Tris-HCl pH 7.4, 100 mM KCl, 50 μ g/mL BSA, and 20% glycerol. The HD-dsDNA mixture was then incubated at 25°C for 30 minutes at 25°C, then the glycerol was added to a concentration of 25% to increase the density of the samples in preparation for loading onto a 10% acrylamide/Tris-glycine gel. The gels were pre-run at 130V for 1 hour in 1x Tris-glycine buffer at pH 8.3 to clear the lanes of unpolymerized acrylic acid. After the gels were pre-run, the samples and a tracking dye were loaded and run at 100 V until the tracking dye had migrated two thirds through the gel. The gels were imaged using an iBright FL1000 Imaging System using fluorescent blot mode (Thermo-Fisher).

NMR Spectroscopy

¹⁵N-isotopically labelled HD proteins were expressed in M9 media (2 $\times$ 1L) supplemented with ¹⁵NH₄Cl (1 g/L) and 100 mg/L ampicillin and purified as described above. The NMR samples were prepared in 20 mM MES, 100 mM NaCl, 2 mM TCEP, 5 mM MgCl₂, 10%D₂O/90% H₂O at pH 6.0. The samples were concentrated to a final volume of 600 μ L (¹⁵N-labeled FTZHD at 120 μ M, ¹⁵N-labeled ANTPHD at 120 μ M, ¹⁵N-labeled ABDAHD at 150 μ M) and transferred to a 5 mm O.D. thin walled NMR tube (New-Era, Vineland, NJ). 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS; 1 mM) was added and an internal reference standard and imidazole (1.6 mM) was added as an internal pH indicator to determine the pH of the sample before and after data acquisition [36]. ¹H-¹⁵N HSQC spectra were collected at 25°C using a Varian Inova 600 MHz 4-channel solution-state NMR Spectrometer equipped with a 5-mm PFG triple-resonance probe housed and maintained in the Sackler Sciences Center at Clark University. All data were processed using NMRPipe and NMRDraw [37] and analysed using NMRViewJ [38].

Results and Discussion

HD TFs belong to the second largest class of TFs in metazoans (the first being C2H2 zinc finger proteins) [12, 39]. These HD TF proteins all contain a highly conserved 60 amino acid region called the homeodomain and are highly conserved across a range of different species [40] – the homeobox was found in all metazoan from the least complex sponge to highly organized vertebrates, and it was also found in plants and fungi [41]. Despite this conservation and similarity in binding sequence and structure, HD TFs are able to bind different sites and regulate different genes [12]. This poses several questions about specificity. Are these TFs truly specific in their binding? How do each of these proteins that contain highly conserved DNA binding regions, target a different set of binding sites and have a unique function in terms of the specific genes they regulate? To begin to address these important biological questions, several HD TFs from *Drosophila melanogaster* were chosen for protein overexpression in *E. coli*, protein purification, and use in binding assays in preparation for future biophysical studies.

Construction of HisSUMO-D. melanogaster HD plasmids

FUSHI TARAZU (FTZ) was initially chosen as the protein of interest for determining TFDNA binding affinity due to its importance as an activator for the homeotic genes in *Drosophila* development [13, 42, 43]. Our original goal was to purify and biophysically characterize full-length FTZ. However, it proved to be difficult to obtain adequate amounts of folded, soluble protein. Multiple attempts to purify using a variety of affinity tags (i.e. His₆, His₆-SUMO and GST) in combination with different full-length and *N*- and *C*-terminal truncated FTZ constructs, all failed. We also tested a range of induction and media conditions, varied buffer salts and pH, altered salt and urea concentrations, as well as co-expressing FTZ with its well-characterized cofactor FTZ-F1 [43] using a dual promoter pETDuet vector, hoping that it would act as a chaperone, but all of these strategies were unsuccessful in yielding meaningful quantities of soluble FTZ protein.

After these failed attempts to express and purify full-length and truncated versions of FTZ, due to issues with insolubility, *E. coli* toxicity and degradation, we decided to re-evaluate our approach and focus specifically on the HD of FTZ. Our logic for this decision was based upon the wealth of previous studies that have clearly demonstrated that the HD is responsible for DNA binding and recognition [18–22, 25, 28–33] and homeobox “swap” experiments, where the homeoboxes of different HD TFs have been switched and result in a change in binding specificity and protein activity [44, 45]. For example, when the homeobox of DEFORMED (DFD) was replaced with the homeobox from ULTRABITHORAX (UBX), the chimeric DFD protein with the HD of UBX adopted the functionality of UBX and was able to regulate genes controlled by UBX [45].

To further examine the HD of FTZ as well as other important HD TFs, we created expression plasmids for the overexpression of five *D. melanogaster* HDs. This was achieved by inserting the homeoboxes of ftz, Antp, abd-A, Abd-B, and Ubx between the unique *Kpn*I and *Xba*I in a modified HisSUMO vector with a T7 inducible promoter (Figure 1). These HisSUMO-HD vectors were specifically designed to help improve solubility and to avoid the use of denaturants that have been used previously when studying FTZ [43, 46] and other difficult HD proteins.

Expression and Purification of D. melanogaster HDs

Previous studies have echoed some of the challenges we faced in attempting to isolate HD proteins and have required the use of strong denaturants like urea and multiple columns and ion exchanges to purify the isolated HDs [43, 46]. Since we were focused on purifying our proteins in the absence of any denaturants, as these would potentially interfere with downstream biophysical experiments, we devised a simplified and streamlined expression and purification protocol to produce pure and soluble *Drosophila* HD TFs. After inducing the cells to express our five N-terminal His₆-SUMO-HD constructs (Figure 1), followed by Ni²⁺-IMAC, cleavage with Ulp1, and ion exchange using a heparin column, we were able to reproducibly purify high quantities of the HDs for FTZ, ANTP, ABD-A, ABD-B, and UBX (Figure 2). All of the isolated proteins were purified to homogeneity and deemed to be greater than 95% pure by Coomassie staining. Our typical yields for each protein ranged from 2 to 5 mg/L media. The use of a heparin column was an important polishing step, as it

ensured that no *E. coli* DNA remained bound to our isolated HDs and that we had biochemically pure protein with no DNA contamination for our subsequent experiments. This new method allowed us to improve the solubility and purity of various HDs under non-denaturing conditions within two to three days of growth and expression. We anticipate that this purification technique will also be applicable and appropriate for the isolation of other HDs from other higher organisms.

Electrophoretic Mobility Shift Assays demonstrate that the isolated HDs are functional

Previous bacterial-one hybrid and SOLEXO analysis identified that HDs preferentially bind to dsDNA with a 5'-TAAT-3' core sequence [12]. To examine if our isolated HDs were biochemically functional and would bind to a predicted sequence, we employed a modified EMSA protocol where we used a 5'-fluorescein labelled (5'FAM) dsDNA for tracking the formation of HD-dsDNA complexes in a 10% acrylamide gel. EMSA assays work on the basis that DNA bound by a protein will migrate more slowly through a gel compared to the free DNA due to changes in the overall charge, size and mass of the complex [47]. Since DNA is negatively charged and the isolated HDs are rich in lysines and arginines, making them very positively charged ($pI > 10$), any bound DNA complex will be less negatively charged and have a higher molecular mass. The AT-rich consensus binding site sequence (5'-TTTAATTA-3') used in our EMSA experiments was flanked with GC-rich sequences to ensure the dsDNA complex had a higher melting temperature (see methods for details). The negative control null sequence (5'-CGCCGGCG-3') was chosen as it was not observed in the B1H and SOLEXO study [11], hence, it would not be expected to form a complex with the isolated HDs.

With the consensus DNA sequence, there was an associated increase in the amount of HD-dsDNA complex as increasing amounts of each of the FTZ, ANTP and ABD-A HD proteins were added (Figure 3A–C). This indicates that the isolated HDs recognize and bind to the predicted consensus sequence. In contrast, at protein concentrations up to 10 μ M there was no observed change in migration for the null sequence, confirming that the isolated HDs do not bind to the GC-rich sequence (Figure 3D–F). At higher protein concentrations some weak binding could be observed, likely as a result of non-specific protein-DNA interactions (Figure 3D–F). Taken together, our EMSA assay results clearly demonstrate that the isolated HDs are biochemically functional, as they are capable of binding specifically and in a concentration-dependent manner to the consensus sequence containing the 5'-TAAT-3' core sequence.

NMR Spectroscopy demonstrates that the isolated *D. melanogaster* HDs are folded

The HDs of FTZ, ANTP, and ABD-A were ^{15}N -labeled by overexpressing the proteins in minimal media supplemented with $^{15}\text{NH}_4\text{Cl}$ as the sole nitrogen source. The ^{15}N -labeled proteins were purified using the same protocol with Ni^{2+} -IMAC, Ulp1, and heparin ion exchange. The isolated ^{15}N -labeled HDs of FTZ, ANTP, and ABD-A were then concentrated to $>100 \mu\text{M}$ and analysed by NMR spectroscopy. The ^1H - ^{15}N HSQC spectra collected for these ^{15}N -labelled HD proteins showed good peak dispersion of the backbone amides indicating that the proteins are well-folded (Figure 4). When comparing these three spectra, we observed that many of the amide peaks for each protein were superimposable,

demonstrating that the residues in each of the isolated HD domains are in similar chemical environments as would be expected for proteins having similar sequences. These spectra demonstrate that future NMR structural analysis of these unique HDs is now possible.

Conclusions

We have successfully purified the homeodomains (HDs) from five distinct *Drosophila* TF proteins (FTZ, ANTP, ABD-A, ABD-B, and UBX) using an *N*-terminal HisSUMO fusion tag to increase solubility without the requirement of urea or other solubilizing reagents (Figure 1). This streamlined protocol, which is a significant improvement from prior techniques involving strong denaturing conditions and multiple chromatography columns, produces pure soluble HD protein within three days of growth and harvesting (Figure 2). Analysis of three of the HD proteins (FTZ, ANTP and ABD-B) using EMSAs (Figure 3) and NMR spectroscopy (Figure 4) revealed that the proteins are well folded and functional, as evidenced by selective binding to a consensus DNA sequence (Fig. 3A–C), but not a null sequence (Fig 4D–F). With this new overexpression and purification system in hand, our future studies will focus on investigating the binding affinity and selectivity for a large number of *Drosophila* HD proteins on a broad range of bioinformatically predicted binding sites in direct quantitative biophysical assays. This will allow us to reproducibly address some of the key questions regarding the biochemical and structural basis for HD TFs specifically recognizing and binding different DNA sequences in *Drosophila* and other higher organisms.

Acknowledgments

The authors would like to thank Dr. Guoxing Lin for maintaining the 600 MHz NMR spectrometer housed in the Carlson School of Chemistry and Biochemistry at Clark University. Special thanks for Dr. David Thurlow for his guidance on EMSAs and Steven Beasley for his critical reading of the manuscript. This work was supported by the National Institutes of Health (R15GM126432 to D.E.S., R15GM110571 to R.A.D. and J.M.D.) and start-up funds from Clark University (D.E.S., R.A.D. and J.M.D.).

Abbreviations

TF	transcription factor
HD	homeodomain
dsDNA	double-stranded DNA
EMSA	electrophoretic mobility shift assay
NMR	nuclear magnetic resonance
HSQC	heteronuclear single quantum correlation

References

- [1]. Arnone MI, Davidson EH, The hardwiring of development: Organization and function of genomic regulatory systems. *Development* 124 (1997) 1851–1864. [PubMed: 9169833]
- [2]. Levine M, Davidson EH, Gene regulatory networks for development. *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005) 4936–4942. [PubMed: 15788537]

- [3]. Borok MJ, Tran DA, Ho MC, Drewell RA, Dissecting the regulatory switches of development: lessons from enhancer evolution in *Drosophila*. *Development* 137 (2010) 5–13. [PubMed: 20023155]
- [4]. Kadonaga JT, Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* 116 (2004) 247–257. [PubMed: 14744435]
- [5]. Maeda RK, Karch F, Cis-regulation in the *Drosophila* Bithorax Complex. *Advances in experimental medicine and biology* 689 (2010) 17–40. [PubMed: 20795320]
- [6]. Ptashne M, Gann A, Transcriptional activation by recruitment. *Nature* 6625 (1997) 569–577.
- [7]. Evans NC, Swanson CI, Barolo S, Sparkling insights into enhancer structure, function, and evolution. *Current topics in developmental biology* 98 (2012) 97–120. [PubMed: 22305160]
- [8]. Ho MC, Johnsen H, Goetz SE, Schiller BJ, Bae E, Tran DA, Shur AS, Allen JM, Rau C, Bender W, Fisher WW, Celniker SE, Drewell RA, Functional evolution of cis-regulatory modules at a homeotic gene in *Drosophila*. *PLoS Genetics* 5 (2009) e1000709. [PubMed: 19893611]
- [9]. Kulkarni MM, Arnosti DN, Information display by transcriptional enhancers. *Development* 130 (2003) 6569–6575. [PubMed: 14660545]
- [10]. Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weiszmann R, Celniker SE, Knowles DW, Gingeras T, Speed TP, Eisen MB, Biggin MD, Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biology* 6 (2008) e27. [PubMed: 18271625]
- [11]. Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA, Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* 133 (2008) 1277–1289. [PubMed: 18585360]
- [12]. Noyes MB, Meng X, Wakabayashi A, Sinha S, Brodsky MH, Wolfe SA, A systematic characterization of factors that regulate *Drosophila* segmentation via a bacterial one-hybrid system. *Nucleic acids research* 36 (2008) 2547–2560. [PubMed: 18332042]
- [13]. Starr MO, Ho MC, Gunther EJM, Tu Y-K, Shur AS, Goetz SE, Borok MJ, Kang V, Drewell RA, Molecular dissection of cis-regulatory modules at the *Drosophila* bithorax complex reveals critical transcription factor signature motifs. *Developmental Biology* 359 (2011) 290–302. [PubMed: 21821017]
- [14]. Weingarten-Gabbay S, Segal E, The grammar of transcriptional regulation. *Human genetics* 133 (2014) 701–711. [PubMed: 24390306]
- [15]. Banerjee-Basu S, Baxeavanis AD, Molecular evolution of the homeodomain family of transcription factors. *Nucleic acids research* 29 (2001) 3258–3269. [PubMed: 11470884]
- [16]. Bobola N, Merabet S, Homeodomain proteins in action: similar DNA binding preferences, highly variable connectivity. *Curr Opin Genet Dev* 43 (2017) 1–8. [PubMed: 27768937]
- [17]. Burglin TR, Affolter M, Homeodomain proteins: an update. *Chromosoma* 125 (2016) 497–521. [PubMed: 26464018]
- [18]. Florence B, Handrow R, Laughon A, DNA-binding specificity of the fushi tarazu homeodomain. *Mol Cell Biol* 11 (1991) 3613–3623. [PubMed: 1675428]
- [19]. Muller M, Affolter M, Leupin W, Otting G, Wuthrich K, Gehring WJ, Isolation and sequence-specific DNA binding of the Antennapedia homeodomain. *EMBO J* 7 (1988) 4299–4304. [PubMed: 2907479]
- [20]. Nelson HB, Laughon A, The DNA binding specificity of the *Drosophila* fushi tarazu protein: a possible role for DNA bending in homeodomain recognition. *New Biol* 2 (1990) 171–178. [PubMed: 1982071]
- [21]. Otting G, Qian YQ, Billeter M, Muller M, Affolter M, Gehring WJ, Wuthrich K, Protein–DNA contacts in the structure of a homeodomain–DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J* 9 (1990) 3085–3092. [PubMed: 1976507]
- [22]. Otting G, Qian YQ, Muller M, Affolter M, Gehring W, Wuthrich K, Secondary structure determination for the Antennapedia homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. *EMBO J* 7 (1988) 4305–4309. [PubMed: 2907480]
- [23]. Percival-Smith A, Muller M, Affolter M, Gehring WJ, The interaction with DNA of wild-type and mutant fushi tarazu homeodomains. *EMBO J* 9 (1990) 3967–3974. [PubMed: 1979032]

- [24]. Qian YQ, Furukubo-Tokunaga K, Resendez-Perez D, Muller M, Gehring WJ, Wuthrich K, 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. [PubMed: 7909851]
- [25]. Schier AF, Gehring WJ, Functional specificity of the homeodomain protein fushi tarazu: the role of DNA-binding specificity in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 90 (1993) 1450–1454. [PubMed: 8434005]
- [26]. Meng X, Brodsky MH, Wolfe SA, A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nature biotechnology* 23 (2005) 988–994.
- [27]. Desplan C, Theis J, O'Farrell PH, The sequence specificity of homeodomain-DNA interaction. *Cell* 54 (1988) 1081–1090. [PubMed: 3046753]
- [28]. Affolter M, Percival-Smith A, Muller M, Leupin W, Gehring WJ, DNA binding properties of the purified Antennapedia homeodomain. *Proceedings of the National Academy of Sciences of the United States of America* 87 (1990) 4093–4097. [PubMed: 1971945]
- [29]. Billeter M, Qian Y, Otting G, Muller M, Gehring WJ, Wuthrich K, 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. [PubMed: 2164583]
- [30]. Billeter M, Qian YQ, Otting G, Muller M, Gehring W, Wuthrich K, Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. *J Mol Biol* 234 (1993) 1084–1093. [PubMed: 7903398]
- [31]. Dessain S, Gross CT, Kuziora MA, McGinnis W, Antp-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. *EMBO J* 11 (1992) 991–1002. [PubMed: 1347746]
- [32]. Kissinger CR, Liu BS, Martin-Blanco E, Kornberg TB, Pabo CO, Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* 63 (1990) 579–590. [PubMed: 1977522]
- [33]. Qian YQ, Otting G, Furukubo-Tokunaga K, Affolter M, Gehring WJ, Wuthrich K, NMR structure determination reveals that the homeodomain is connected through a flexible linker to the main body in the *Drosophila* Antennapedia protein. *Proceedings of the National Academy of Sciences of the United States of America* 89 (1992) 10738–10742. [PubMed: 1359544]
- [34]. Zeiske T, Baburajendran N, Kaczynska A, Brasch J, Palmer AG, 3rd, Shapiro L, Honig B, Mann RS, Intrinsic DNA Shape Accounts for Affinity Differences between Hox-Cofactor Binding Sites. *Cell Rep* 24 (2018) 2221–2230. [PubMed: 30157419]
- [35]. Bailey TL, Bodén M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS, MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37 (2009) W202–208. [PubMed: 19458158]
- [36]. Baryshnikova OK, Williams TC, Sykes BD, Internal pH indicators for biomolecular NMR. *J Biomol NMR* 41 (2008) 5–7. [PubMed: 18398685]
- [37]. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6 (1995) 277–293. [PubMed: 8520220]
- [38]. Johnson BA, Blevins RA, NMR View: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4 (1994) 603–614. [PubMed: 22911360]
- [39]. Tupler R, Perini G, Green MR, Expressing the human genome. *Nature* 409 (2001) 832–833. [PubMed: 11237001]
- [40]. McGinnis W, Levine MS, Hafen E, Kuroiwa A, Gehring WJ, A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* 308 (1984) 428–433. [PubMed: 6323992]
- [41]. Mark M, Rijli FM, Chambon P, Homeobox genes in embryogenesis and pathogenesis. *Pediatr Res* 42 (1997) 421–429. [PubMed: 9380431]
- [42]. Drewell RA, Nevarez MJ, Kurata JS, Winkler LN, Li L, Dresch JM, Deciphering the combinatorial architecture of a *Drosophila* homeotic gene enhancer. *Mechanisms of development* 131 (2014) 68–77. [PubMed: 24514265]

- [43]. Yussa M, Lohr U, Su K, Pick L, The nuclear receptor Ftz-F1 and homeodomain protein Ftz interact through evolutionarily conserved protein domains. *Mechanisms of development* 107 (2001) 39–53. [PubMed: 11520662]
- [44]. Hoey T, Warrior R, Manak J, Levine M, DNA-binding activities of the *Drosophila melanogaster* even-skipped protein are mediated by its homeo domain and influenced by protein context. *Mol Cell Biol* 8 (1988) 4598–4607. [PubMed: 2905420]
- [45]. Kuziora MA, McGinnis W, A homeodomain substitution changes the regulatory specificity of the deformed protein in *Drosophila* embryos. *Cell* 59 (1989) 563–571. [PubMed: 2572328]
- [46]. Yu Y, Li W, Su K, Yussa M, Han W, Perrimon N, Pick L, The nuclear hormone receptor Ftz-F1 is a cofactor for the *Drosophila* homeodomain protein Ftz. *Nature* 385 (1997) 552–555. [PubMed: 9020364]
- [47]. Hellman LM, Fried MG, Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc* 2 (2007) 1849–1861. [PubMed: 17703195]
- [48]. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG, Clustal W and Clustal X version 2.0. *Bioinformatics* 23 (2007) 2947–2948. [PubMed: 17846036]
- [49]. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ, Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25 (2009) 1189–1191. [PubMed: 19151095]

Highlights

Overexpression and purification protocol for Homeodomain transcription factors

Functional and well-folded as judged by EMSA and NMR spectroscopy

Future biochemical and biophysical characterization of TF-DNA now possible

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Cloning and sequence alignment of *D. melanogaster* HDs for bacterial overexpression. Vector map (A) and sequence map (B) of HisSUMO-HD constructs. Each of the HDs for FTZ, ABD-A, ABD-B, ANTP, and UBX were cloned into a T7-inducible HisSUMO vector between *KpnI* and *XbaI* restriction sites. (C) Sequence alignment of overexpressed *D. melanogaster* HDs. The 107 member homeodomain family was submitted to ClustalW [48] using default values and analysed using Jalview [49]. Only the five HDs studied in this paper are shown. Highly conserved residues within similar physicochemical properties in all 107 HDs analysed are shown (red – positive, blue – negative, yellow – polar, green – hydrophobic). Absolutely conserved residues and residues with greater than 70% similarity are labelled with a star or cross, respectively.

Figure 2. *D. melanogaster* HD protein purification monitored by reducing 16.5% SDS-PAGE. (A) Purification gel of ANTPHD showing samples taken at each step of the purification process (as described in the Materials & Methods). (B) Gel of purified *D. melanogaster* HDs for FTZ, ABD-A, ABD-B, ANTP, and UBX. Each lane was loaded with 5 ug of purified HD protein. Spectra Multicolor Broad Range Protein Ladder (Thermo-Fisher) was used in both gels to monitor migration and determine molecular weights of the HisSUMO-HD, HisSUMO tag, and isolated HD proteins. Gels were stained with Coomassie Brilliant Blue G-250 and visualized using an iBright FL1000 gel imager.

Figure 3. Electrophoretic mobility shift assays show that the isolated *D. melanogaster* HDs can bind to specific DNA sequences.

EMSA of isolated FTZ, ANTP, and ABD- A HDs binding to the consensus sequence 5'-GCGTTTAATTAGCCC-3' (A-C) and the null sequence 5'-GCGCGCCGGCGGCC-3' (D-F).

Figure 4. Initial NMR spectroscopic characterization of isolated *D. melanogaster* HDs shows that the proteins are well-folded.

^1H - ^{15}N HSQC spectra of ^{15}N -isotopically labelled FTZHD at 120 μM (A), ANTPHD at 120 μM (B), and ABDAHD at 150 μM (C). Samples were prepared in 20 mM MES, 100 mM NaCl, 2 mM TCEP, 5 mM MgCl_2 , 10% D $_2\text{O}$ /90% H $_2\text{O}$ at pH 6.0. All three spectra were acquired for 40 minutes using 16 transients and 64 increments at 25 $^\circ\text{C}$ with DSS used as an internal reference standard.