Spatiotemporal dynamics of enhancers activity in the early *Drosophila* embryo

Stefano Ceolin

München 2020

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig–Maximilians–Universität München

Spatiotemporal dynamics of enhancers activity in the early *Drosophila* embryo

Stefano Ceolin

Stefano Ceolin
aus San Vito al Tagliamento, Italien
2020

Erklärung:
Diese Dissertation wurde im Sinne von 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Don Lamb betreut.

Eidesstattliche Versicherung:
Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 23.04.2021

Stefano Ceolin

Dissertation eingereicht am: 13.10.2020
1. Gutachter: Prof. Dr. Don Lamb
2. Gutachter: Prof. Dr. Nicolas Gompel
Mündliche Prüfung am: 19.11.2020

Summary

The precise control of gene expression by transcriptional enhancers is key to a variety of biological processes. Each enhancer sequence contains a complex constellation of binding sites for multiple transcription factors. When these factors are bound to their cognate sites, they can either promote or repress enhancer activity, thus determining the expression of target genes. The network of enhancers governing the anterior-posterior axis segmentation in the earliest stages of *D. melanogaster* development is a well-established paradigm to study transcription in developmental biology, and notably how the sequence of an enhancer determines its transcriptional output.

While much is known about the regulatory logic of segmentation enhancers, detailed quantitative models of enhancer activity are still missing. To develop predictive sequence-to-expression models, one needs to precisely measure transcriptional enhancer activity in space and time, both for weak and strong enhancers. However, the rapid changes in gene expression during early stages of development hinder the accurate measurement of the dynamics of enhancers activity with standard imaging techniques. In fact, the few existing methods available to measure transcription have significant limitations in either sensitivity, resolution or throughput.

In this thesis, in order to address the need for new experimental strategies to track enhancer activity, I introduce a new method that overcomes some of these limitations. I optimized the bright and fast-maturing fluorescent protein mNeonGreen as a real-time, quantitative reporter of enhancer expression and derived enhancer activity from the reporter fluorescence dynamics with high spatial and temporal resolution, using a robust reconstruction algorithm. By comparing the results obtained using the new reporter with data obtained with the established MS2-MCP system, I demonstrated the reliability and the higher detection sensitivity of the new reporter. Moreover, I demonstrated the usefulness of the new reporter for investigating fundamental questions regarding transcriptional regulation by quantifying the activity of variants of a simple synthetic enhancer. In this setting, I observed how different features of the enhancer, such as the reduction of enhancer-promoter distance or the addition of binding sites for the pioneer transcription

factor Zelda, influence activity.

I then moved on to apply our methodology to measure the activity of 20 synthetic enhancers designed to gain new insights into the effect of binding sites for the transcription factor Hunchback on enhancer activity. Hunchback is considered to be a context-dependent transcription factor able to either activate or repress different enhancers during *D. melanogaster* embryo segmentation. The mechanism driving the context dependent activity of Hunchback is, however, not yet understood. To address this question, I generated multiple synthetic enhancers combining a sequence of 3 Hunchback binding sites with binding sites for two groups of activators, varying the binding sites spacing, strength and orientation, and I measured the spatiotemporal dynamics of their activity in-vivo. These data revealed a dual role of Hunchback binding sites in segmentation enhancers: on the one hand, Hunchback acts as a typical short range repressor by binding to its cognate sequences; on the other hand, I report a novel effect of a sequence containing multiple Hunchback binding sites, that is able to increase enhancer activity independently from Hunchback binding.

I conclude that the mNeonGreen reporter system is a powerful new approach that will be useful for systematically and comprehensively characterizing enhancers activity. In particular, the new reporter provided substantial advances in throughput and sensitivity that make it an ideal tool to characterize the activity of a larger number of synthetic enhancers. This achievement is an important step in the ongoing effort to further advance our understanding of transcriptional regulation.

Contents

List of Abbreviations

AP	Anterior-Posterior axis
Bcd	Bicoid (Transcription factor)
bp	base pairs
CRE	cis regulatory element
DSCP	Drosophila synthetic core promoter
DV	Dorsal-Ventral axis
Dl	Dorsal
EL	Embryo length
GFP	Green Fluorescent Protein
Gt	Giant (Transcription factor)
Hb	Hunchback (Transcription factor)
hb_ant	hunchback anterior enhancer
Kni	Knirps (Transcription factor)
Kr	Kruppel (Transcription factor)
mNeonRep	full sequence of the mNeonGreen reporter including the 3'UTR and 5'UTR
NLS	Nuclear localization signal
nc	nuclear cycle
ODE	Ordinary differential equation
PWM	Positional Weight Matrix
SNR	Signal to noise ratio
smFISH	single molecule fluorescence *in-situ* hybridization
TF	Transcription Factor
TSS	Transcription start site
Twi	Twist
UTR	Untranslated region
Zld	Zelda (Transcription factor)

Chapter 1

Introduction

It took centuries for philosophers and scientists to reach a unified view in which the shapes of non-living matter, from the regular trajectories drawn by the periodic movements of planets in the sky to the orderly structure of crystals, are understood as a result of mechanical forces at play among their material constituents. Including the even more complex and fascinating forms observed in living organisms into this scheme took even longer: only around the beginning of the XX century, have researchers of natural history, or biology, started to describe their findings with a mathematical language and interpret them in light of physical laws. In a seminal and influential work, D'Arcy Wentworth Thompson was probably among the firsts to describe the influence of physical phenomena on the growth and forms of living things. For example, describing the effect of surface tensions on the form of a cell, or temperature on the growth of an organism (Thompson, 1917). By the mid XX century, many more ideas were developed on how simple physical principles can influence the forms of living things in less obvious ways, for example, through biochemical processes that fuel pattern formation in reaction-diffusion systems(Turing, 1952). Finally, the extraordinary discoveries of genetics in the last decades, provided a unified picture in which the effects of physical forces and evolution shape the development of the forms of living things.

Modern developmental biology identified one of the main molecular mechanisms driving organism development in the precise control of gene expression by transcriptional enhancers. A mechanism that has been proven to be the key in a variety of biological processes ranging from animal development to cancer biology (Sur and Taipale, 2016), (Peter and Davidson, 2011). A striking example of precise transcriptional regulation at work is found in the early stages of *Drosophila melanogaster* (*D. melanogaster*) embryonic development. There, a network of transcriptional enhancers reads gradients of transcriptional activators and repressors (Jaeger, 2011) (Nusslein-Volhard, 1991), which

are provided maternally in the egg. The resulting activation of these enhancers taking place only in certain groups of syncitial nuclei, that are uniformly distributed beneath the cortex of embryo, lays down the blueprint of the future fly body structure with stunning precision (Petkova et al., 2019).

Thanks to the high reproducibibility of its developmental processes, *D. melanogaster* embryos are a natural laboratory for studying transcriptional responses simultaneously under a large number of conditions imposed by the maternal gradients(Gregor et al., 2014). Therefore, this system has become a paradigm to decipher the function of enhancers, and notably how the sequence of an enhancer determines its transcriptional output (Pennacchio et al., 2013) (Spitz and Furlong, 2012). Ultimately, the functioning of enhancers should be determined by their sequence, which typically contains a complex architecture of clusters of binding sites for transcription factors (TFs). How this architecture of binding sites determines expression is the subject of intensive investigation. The occupancy of these sites can be influenced by cooperativity (Lebrecht et al., 2005), competition (Small et al., 1996) and by the presence of nucleosomes. Gene expression is, in turn, a complex function of site occupancy, integrating opposing effects of different factors that can either promote or inhibit transcription (Levine, 2010). The ultimate goal of researchers investigating transcriptional enhancers is to quantitatively understand how all these effects intertwine and to develop predictive models that can link enhancer sequences to expression (Segal et al., 2008). A fundamental step towards this goal is to precisely measure the transcriptional enhancer activity in space and time. Multiple pieces of information, like the concentrations of input TFs, the enhancer sequence and its activity, are the bases for quantitatively understanding how gene expression is 'computed' by the combinatorial occupancy of TFs binding sites. In this context, methods to collect quantitative and time-resolved activity data for large numbers of enhancer sequences are of primary importance and can open the door to a wider study of enhancer architecture based on synthetic enhancer sequences designed to test specific hypothesis on transcriptional regulation (Crocker et al., 2017) (Fakhouri et al., 2010).

1.1 Transcriptional regulation

The control of gene expression is a critical process for development and cellular differentiation. In fact, even if all the cells of an organism share the same genome, they are both structurally and functionally very different. These differences are the result of a differential regulation of the expression of the genes encoded in the genome of all cells, ensuring that only the correct genes are expressed in different cell types. Transcriptional

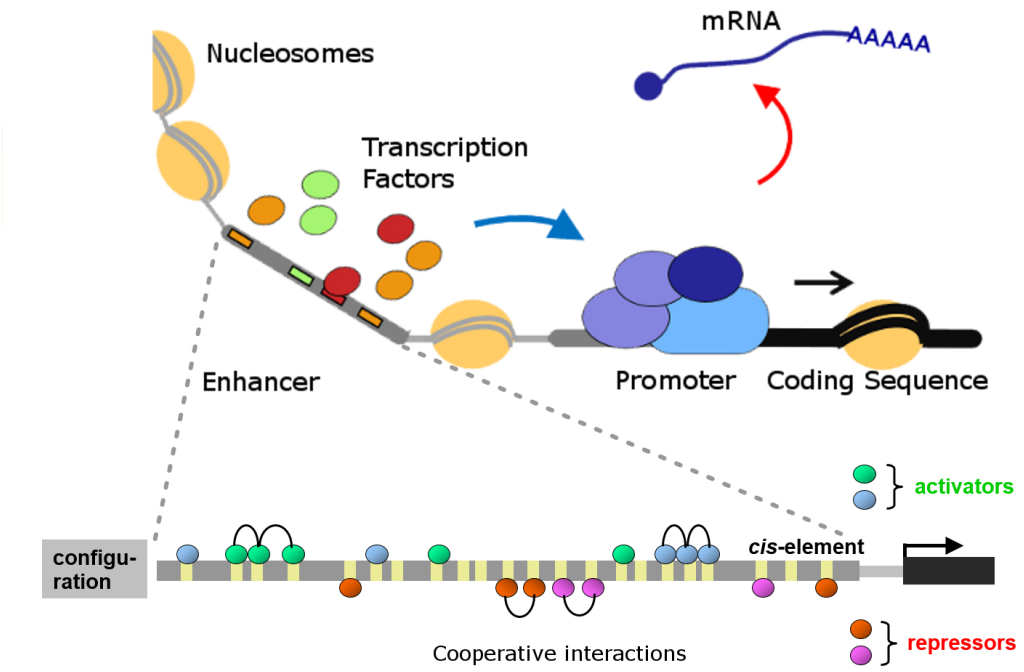

Figure 1.1: Sketch of transcriptional regulation of gene expression. The coding sequence of a gene (in black) is transcribed by RNA PolII (in blue), which is recruited to the basal promoter of the gene. The recruitment of RNA PolII depends on the activity of transcription factors, which recognize and bind to specific DNA motifs contained in enhancer sequences. The enhancer carries clusters of binding sites for various transcription factors, which could either promote or disfavor the recruitment of PolII to the basal promoter and thus control gene expression. The arrangement of the binding sites along the enhancer sequence can influence the binding of transcription factors, for example by allowing cooperative interactions among factors. Different regions of enhancer, promoter and intervening sequences are differentially occupied by nucleosomes (in yellow) which can reduce the accessibility of the DNA sequence to TFs.

regulation is the first fundamental layer of control of gene expression and takes place at the level of transcription, regulating the rate at which mRNA is produced from a gene. The necessary instructions for transcriptional regulation are also found in the genome, often in close proximity to the coding region of each gene, and are encoded in the DNA sequence of regulatory elements which can be divided in two classes: promoters and enhancers. Promoters are always located closely upstream of the coding sequence of a gene, they are the site of RNA PolII recruitment and therefore the start site for transcription.

Certain classes of promoters are not constitutively active and their activity has to be induced by enhancer elements, also called cis-regulatory elements (CREs). Enhancers have a modular structure, which means that the expression of a single gene can be controlled, often independently, by multiple enhancer modules. Often, each of these modules is responsible for the expression of its target gene in response to different external conditions (Maston et al., 2006).

The information encoded in an enhancer module has a fundamental unit: the binding site for specific TFs. TFs are DNA binding proteins that are able to regulate transcription of genes. Each TF is able to recognize and bind a specific DNA sequence, typically between 8 base pairs(bp) and 16 bp long, using various chemical interactions that are sensitive to both the sequence of DNA bases and its conformation. In eukaryotes, TFs can be classified into two main groups: the first group consists of general TFs that are necessary for the expression of all genes. General TFs directly or indirectly bind to promoters forming a large protein complex called the transcription pre-initiation complex. This complex has multiple functions: it helps the loading of RNA polymerase II at the transcription start site, denatures DNA and starts transcription (Reese, 2003). The second group consists of TFs responsible for the differential regulation of genes. These TFs typically bind DNA at enhancers and influence gene expression by protein-protein interactions with transcriptional coregulators, like chromatin remodelers or the mediator complex (Maston et al., 2006). Once bound, activator TFs promote polymerase recruitment at the core promoter of a nearby gene whereas repressor TFs disfavor it, either directly or by preventing activator TFs from binding (Levine, 2010). The role of a specific TF as activator or repressor in a specific gene regulatory network is usually well determined. However, it has been known for a while that the activity of a TF is in some cases context dependent and can be influenced by interactions with other TFs bound to the same enhancer(Pan and Courey, 1992).

TFs binding sites are the building blocks of enhancers but single sites are usually not able to act autonomously. Since TFs have a certain level of flexibility in recognizing their target sites and binding sites already have an intrinsically low specificity because of their small size, there is a widespread distribution of potential binding sites throughout the genome. In most cases, however, these sites are not functional. Functional enhancers typically contain clusters of binding sites concentrated over few hundreds base pairs. The search for such clusters is a successful means of identify new enhancer modules(Berman et al., 2002)(Schroeder et al., 2004). However, not all clusters of binding sites found in the genome constitute active enhancers since the background sequence in which the binding sites are embedded and the distances and orientation of the binding sites can influence the enhancer activity of a sequence (Grossman et al., 2017). Moreover, like most of the

genome, also the enhancer DNA is typically wrapped around nucleosomes. The resulting competition between nucleosomes and TFs has been found to be a major determinant of enhancer activity(Grossman et al., 2017). As a consequence, some binding sites clusters are not functional enhancers because they are not accessible to TFs.

Since an enhancer carries a complex constellation of binding sites, its regulation is highly combinatorial in nature. In an over simplified model, we can imagine that an enhancer will only drive gene expression in the presence of activators that can bind to it and in the absence of repressors that can bind to it. In reality, enhancers respond to the input TFs in a concentration dependent manner and the interpretation of the regulatory input depends on the number, affinity and organization of the binding sites. Different enhancers may depend on different mechanisms for interpreting their regulatory input and achieving a precise, spatially and temporally modulated, transcriptional output. One mechanism is the interplay of various TFs, each expressed in partially overlapping temporal and/or spatial windows. Another mechanism is the presence of cooperative processes that can give rise to a binary response of the enhancer activity when the input concentration of TFs reaches precise thresholds. For example, the transcriptional activator Bicoid (Bcd), which is active in *D. melanogaster* development, is able to bind cooperatively to enhancers in which multiple, closely positioned, binding sites are present (Hanes et al., 1994). This process, which likely relies on protein-protein interactions among Bcd molecules reinforcing DNA-binding, provides a mechanism for threshold dependent gene activation in the embryo(Lebrecht et al., 2005). Since cooperative binding turns out to be dependent on the spacing, orientation and number of binding sites for Bcd, different enhancers architectures can respond to different concentration thresholds(Burz et al., 1998).

Even though cooperative processes are often associated with protein-protein interactions, other modes of cooperativity among TFs also exist. One example is given by synergistic interactions of TFs with coactivators and components of the transcriptional machinery, which can be recruited by the cooperative interaction with two or more TFs bound to the enhancer(Merika et al., 1998). Another example of cooperativity that does not require direct protein-protein interactions among the factors is given by the influence that a bound TF has on the local bending of DNA. In fact, a change of DNA bending can propagate to nearby binding sites for additional factors, influencing their occupancy(Falvo et al., 1995). Unexpectedly, positive cooperative effects might even occur when two factors are competing for shared binding sites, as the alternating binding of the two factors can more effectively compete with nucleosome binding and thus maintain the accessibility of the enhancer (Voss et al., 2011).

This variety of cooperative interactions among TFs is also reflected on the structure of enhancer sequences. Some enhancers have a rigid structure of binding sites in fixed orien-

tations at fixed distances that cannot be altered without disrupting the enhancer activity. These enhancers, termed enhanceosomes, are strongly dependent on protein-protein interactions among the factors and their activity requires the presence of all factors(Merika and Thanos, 2001). Other enhancers, named 'billboard' enhancers, seem to have a more flexible organization in which only a subset of factors interact cooperatively. For the billboard enhancers, changes in distance and composition of the binding sites are partially tolerated and the enhancer can be active even if only a subset of the input TFs are present (Arnosti and Kulkarni, 2005). In one more class of enhancers, named TF collectives, strong cooperative interactions with co-activators create a scenario in which the composition and architecture of binding sites is relatively flexible, but all factors are strictly required for the enhancer to be active(Junion et al., 2012). Recent experimental findings (Shrinivas et al., 2019) and hints from theoretical modeling(Bialek William, Gregor Thomas, 2019) are just starting to unveil a more general model of enhancer activity. In light of these recent findings, the various enhancer models could be seen as different cases of a common scenario in which the activation of the enhancer corresponds to the formation of a transcriptional condensate, a phase-transition that occurs at enhancers when enough DNA-protein and protein-protein interactions take place(Shrinivas et al., 2019).

An interesting question is whether, and to what extent, the binding of a TF and its transcriptional activity are regulated by the same features of an enhancer sequence. A large study, focused on enhancers responding to the human transcriptional activator PPARγ, looked at more than 30000 natural and synthetic enhancers and investigated how TF binding and enhancer activity are influenced by various enhancer features (Grossman et al., 2017). The results of this study showed that TFs occupancy at binding sites on transfected plasmids is linearly dependent on the affinity of the core binding motif with little or no influence from the flanking sequences. However, this is no longer true when looking at binding sites in the genome: the vast majority of binding sites that are actually occupied lie in regions where chromatin is open and, furthermore, TF binding in these regions correlates with the level of histone modifications corresponding to active chromatin. Therefore, the core binding sites and the chromatin landscapes seem to be the major determinants of TFs binding. The occupancy of activator TFs is, however, known to correlate only weakly with the expression of neighboring genes(Sandmann et al., 2006)(Vokes et al., 2008). Although the disruption of activator binding sites is sufficient to reduce or even completely suppress enhancer activity, the core binding sites motifs are not sufficient to drive expression, which strongly depends on the flanking sequences(Grossman et al., 2017). The influence of flanking sequences on expression is due to the presence of specific motifs recognized by additional TFs, which can be consistently classified in distinct

functional groups as additive or synergistic co-activators or inhibitors(Grossman et al., 2017). An additional layer of complexity in the relation between TF binding and activity comes from the fact that the binding of one TF can label enhancers that will be only active at future times in response to different TFs. This behavior is typically observed for pioneer TFs, which are not by themselves able to activate an enhancer but initiate processes that lead to activation at later stages in development. One example is given by the pioneer TF Zelda(Liang et al., 2008), which is a master regulator of the maternal to zygotic transition in *D. melanogaster* development. Binding of Zelda to its cognate sites promotes deposition of histone modifications and increases DNA accessibility(Li et al., 2014), thus increasing binding sites occupancy of other TFs.

As the interaction of multiple TFs shapes the activity of an individual enhancer module, multiple enhancer modules control the expression of the same gene. It has been known for a long time that the activity of each of these modules is mostly independent from each other. This fact is clearly demonstrated by reporter assays in which an enhancer is coupled to an heterologous promoter and drives the expression of a reporter gene(S. Small and Levine, 1993)(I. Gray and Levine, 1994). These artificial constructs are then integrated in the genome in regions that can be far away from the original enhancer locus but they, nevertheless, recapitulate the native expression patterns in space and time. However, more recent studies have reported that the activity of multiple enhancers is not completely independent. In particular, among the many enhancers that control the expression of developmental genes, some have overlapping, or even completely identical, expression patterns. These secondary, or shadow, enhancers could improve the robustness of the expression patterns of genes from environmental perturbations(Perry et al., 2010). The fact that they simultaneously engage with the same promoter may be the reason that their activity is often not additive. For example, a study that focused on the *D. melanogaster* developmental genes Hunchback and Knirps, found that the activity of weak enhancers acting on the Knirps gene is additive or even super-additive. By contrast, the stronger enhancers controlling the expression of the Hunchback gene are sub-additive, probably because two or more strong enhancers can impede one another by frequently contacting the same promoter(Bothma et al., 2015).

Despite the extensive amount of available information, transcriptional regulation is not yet fully understood and many central questions are still unanswered (Pennacchio et al., 2013). For example, how are enhancers able to interact with promoters in the complex three-dimensional structure of the genome? Or how do enhancer mutations affect their function and what influence does this have on human disease? As we have seen throughout this section, various mechanisms regarding the activity of specific enhancers or TFs have been investigated in detail. However, a more general, quantitative understanding of how

enhancer elements 'compute' expression from the concentration of input TFs and various features of their DNA sequence is still missing.

1.2 The patterning of *D. melanogaster* embryos

As we have seen in the previous section, transcriptional regulation has been studied in great detail in the past decades. A central paradigm in which many mechanisms of transcriptional regulation were discovered, and a striking example of transcriptional regulation at work, is the segmentation of *D. melanogaster* embryos. In this section we will summarize the most relevant information about the development of *D. melanogaster* embryos in general and we will then describe the genetic networks responsible for the embryo segmentation in greater detail.

1.2.1 The early stages of *D. melanogaster* development

The first stages of development of *D. melanogaster* are a tightly regulated and extremely reproducible series of events that starts when the embryo is fertilized and layed. Structurally, the embryo is roughly an ellipsoid of average length 500 μm and diameter 180 μm, with a flattened dorsal side and a more convex ventral side. The embryo is protected by an eggshell composed of a thick and opaque outer layer, the chorion, and an inner transparent layer, the vitelline membrane (Margaritis et al., 1980). The eggshell encloses the egg cytoplasm, the first cell nucleus located around the middle of the embryo and a large number of yolk particles.

The first three hours of embryo development are marked by 13 rounds of synchronized mitotic divisions, thus corresponding to 14 cell cycles. During the first 6 of these cell cycles, no transcription of zygotic genes takes place. Therefore, during this time period, the first and only source of regulation comes from maternally-deposited mRNA and protein molecules. These factors regulate the timing of mitotic divisions (Edgar and O'Farrell, 1989) and establish the initial chromatin organization. Some of the maternally deposited factors, often called morphogenes, form gradients inside the embryo and therefore provide the first clue that will determine the spatial organization of the new fly(Nusslein-Volhard et al., 1987). Starting from nuclear cycle (nc) 5, the cell nuclei migrate from the middle of the embryo towards the surface where they form a regularly arranged layer by nc 8.

Transcription of zygotic genes is first observed during the seventh cell cycle but it is only after the thirteenth nuclear division that a more widespread zygotic genome activation takes place (De Renzis et al., 2007). At this point of embryo development, the full activation of various gene regulatory networks rapidly establishes spatial patterns of expression of multiple genes that (see Figure 1.2), taken together, specify cell identities and constitute a precise blueprint of the future body structure of the adult fly (Karaiskos et al., 2017). Cell membranes start to invaginate from the embryo cortex and separate

stage	cell cycles	time from fertilization	developmental activity
1	1-2	0-25 min	first cleavage cycles
2	3-8	25-65 min	formation of perivitelline space
			migration of the nuclei towards the cortex
3	9-10	65-80 min	formation of the polar buds
4	11-13	80-130 min	syncitial blastoderm
5	14	130-170 min	cellularization
6	14	170-180 min	gastrulation

Table 1.1: The earliest stages of development of a *D. melanogaster* embryos. The table reports the timeline of the first 6 stages of *D. melanogaster* embryo development together with the corresponding mitotic cleavages, developmental time at 25°C and the main developmental activity defining each stage.

the nuclei into distinct cells right before gastrulation takes place.

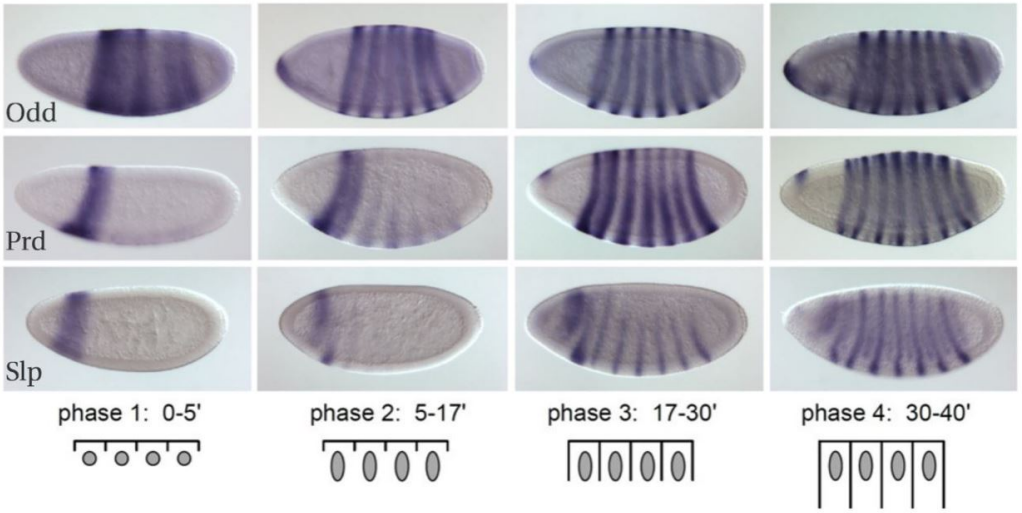

Figure 1.2: Highly dynamic gene expression patterns form the blueprint of the future adult fly. *In-situ* hybridization stainings of the expression patterns of three representative pair-rule genes (from top to bottom: *odd skipped*, *paired* and *sloppy paired*) at four consecutive time points during stage 5 of embryo development. The progressive invagination of cellular membranes separating cell nuclei is represented in the sketches in the last row and can be used to stage images of fixed embryos. Figure kindly provided by Prof. Ulrike Gaul.

1.2.2 The segmentation network

The process of cell differentiation during *D. melanogaster* development that starts with the transcription of the first zygotic genes is orchestrated by two genetic networks which pattern the embryo along the dorsal-ventral (DV) or anterior-posterior (AP) axes. In contrast to other genetic networks for which post-transcriptional and post-translational regulation are critical, genes in these two networks encode mostly transcription factors and their expression is regulated mostly at the level of transcription.

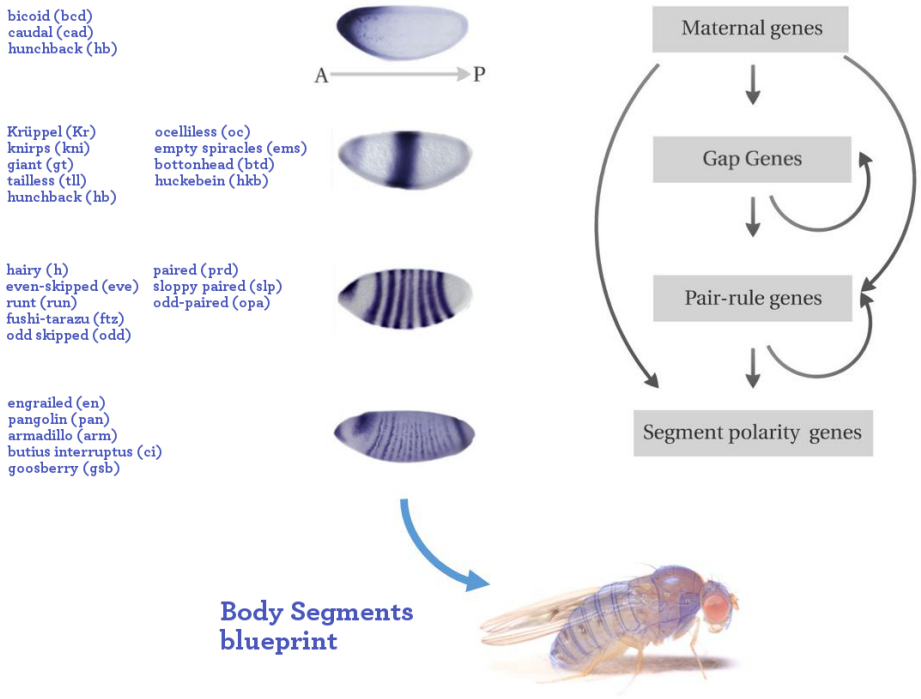

Figure 1.3: Outline of the *D. melanogaster* segmentation network with expression patterns of representative genes. The diagram on the right illustrates the hierarchic structure of the segmentation network and the genetic interactions between genes belonging to different classes. On the left, a list of genes from each class of the network, with embryo images showing a representative expression pattern for each class. RNA in situ hybridization images represent: bicoid, knirps, eveen-skipped and engrailed (from Berkeley Drosophila Genome Project; (Tomancak et al., 2002)). The expression patterns of polarity genes can be considered a blueprint of the future body structure of the adult fly.

Both the AP and DV networks process the information of maternally deposited RNA and protein molecules and produce increasingly more complex and refined patterns of

zygotic regulators. The AP axis patterning network, which is the focus of this thesis, defines the body segments of the future adult fly and is, therefore, often referred to as "the segmentation network" (Nasiadka et al., 2002).

The segmentation network has a hierarchical structure that is summarized in Fig. 1.3. The genes of the segmentation network are grouped in four classes based on the similar phenotypes of mutant flies in which their expression is removed, and on their roles in the regulatory logic of the network. The spatial patterns of expression become more and more complex as we move down the cascade of genes in the network: At the top of the network we find the maternally deposited regulators that are distributed in broad gradients throughout the embryo. The maternal regulators activate the expression of zygotic "gap" genes that are typically expressed in one or two stripes in the embryo. The removal of gap genes cause the loss of multiple contiguous body segments in the larva, forming a gap in its body plan. Pair-rule genes are a critical turning point in the network: the combined action of maternal and gap-factors controls their expression in narrow periodic stripes that identify alternating para-segments of the fly body structure. Finally, pair-rule genes control the expression of segment polarity genes which, differently from all genes upstream in the network, do not exclusively encode transcription factors but also other proteins involved in intercellular signaling. The names of segmentation network genes describe some features of the phenotype of mutant fruit flies in which their expression have been knocked-out. For example, the gene *Krüppel* (Kr) is named after the crippled appearance of mutant embryos lacking its expression and loss of functions mutations of the gene *even-skipped* (eve) remove denticle bands only in alternating segments.

Most TFs in the segmentation network have a clear role as transcriptional repressors or activators on all enhancers. Importantly, we should note that the maternal factors in the network are mostly transcriptional activators (Bicoid, Caudal, Hunchback). The concentration of these maternal factors is not uniform across the embryo: they exhibit broad concentration gradients along the embryo AP axis, thus providing the necessary positional information to guide the embryo development. An additional activating input comes from ubiquitously expressed factors like Zelda (Liang et al., 2008) and D-STAT (R. Yan S. Small and Darnell, 1996) which are both zygotically and maternally expressed. By contrast, the downstream genes mostly encode transcriptional repressors. For example, many gap genes, like *Knirps*, *Krüppel* and *Giant*, are short-range repressors acting locally, in proximity of the repressor binding sites in a single enhancer, to quench the activity or block the binding of activators targeting nearby sites. Pair-rule TFs, like *Hairy*s, act instead as long range repressors silencing an entire locus (Cai et al., 1996)(Martinez and Arnosti, 2008). Overall, the network mainly works by carving out enhancer activity through the localized expression of transcriptional repressors which counteract the activity

of more broadly distributed transcriptional activators (Small et al., 1996).

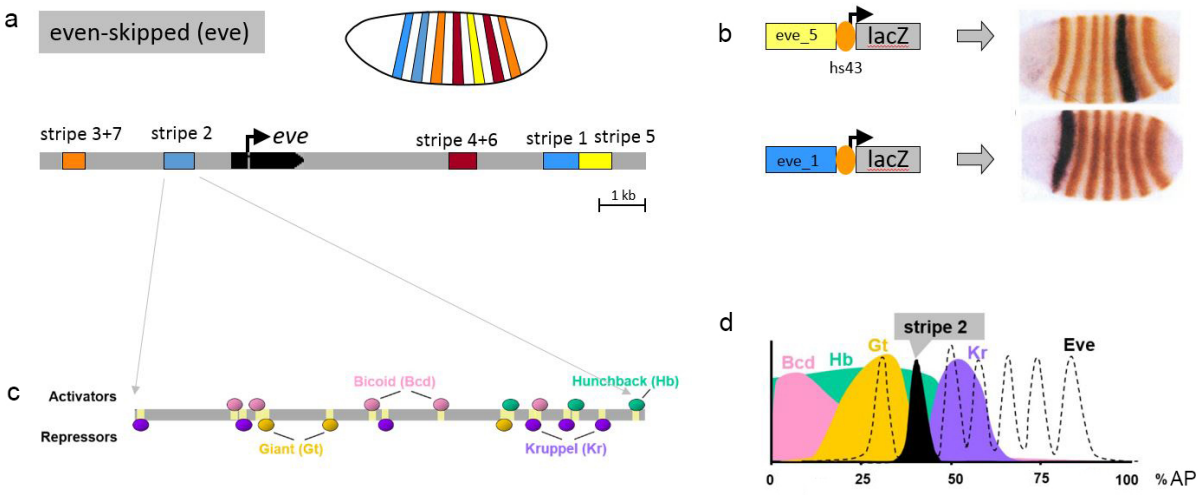

Figure 1.4: Structure of the *even-skipped* gene locus. a) The pair-rule gene *even-skipped* is expressed in seven stripes along the embryo AP axis. Expression in these stripes has been linked to the activity of five autonomous enhancers that are all located in close proximity to the target gene. b) Enhancer elements work autonomously and each of them controls expression in one or two stripes, as shown here for the stripe 5 enhancer (eve_5) and the stripe 1 enhancer (eve_1) c) Enhancer element contains a complex architecture of binding sites for multiple activators (e.g. Bicoid and Hunchback) and repressors (e.g. Giant and Krüppel), as illustrated here using the enhancer responsible for the expression of stripe 2 (eve_2). d) Sketch of the concentration profile of input TFs for the eve_2 enhancer along the AP axis. Expression of the eve_2 enhancer (in black) and of all eve stripes (dashed line). The eve_2 enhancer is active in a region of the embryo where a net activating input is present. Figure kindly provided by Prof. Ulrike Gaul.

The expression of each gene in the segmentation network is controlled by multiple transcriptional enhancers that are responsible for the expression of the gene in different domains or at different developmental stages (Fig. 1.4 a-b). These enhancers are typically located in close proximity to the gene body and are able to act autonomously without significant cross-talk (I. Gray and Levine, 1994)(S. Small and Levine, 1993). Most enhancers have been discovered by a systematic dissection of regions proximal to promoters and tested by means of enhancer-reporter assays (Fujioka et al., 1999)(Hoch et al., 1990)(Small et al., 1992). Each enhancer receives input from multiple TFs and will only be active in presence of a net activating input. Since the concentration of these factors

follows different profiles along the embryo, each enhancer will only be active in regions where this net activating input is achieved (Fig. 1.4 c-d). Which TFs actually influence the activity of a specific enhancer depends on the binding sites present in its sequence, which have been characterized with multiple approaches: predicted candidate binding sites have been identified using in vitro derived TFs binding preferences in the form of positional weight matrices (PWM) (Berman et al., 2002)(Schroeder et al., 2004)(Schroeder et al., 2011), functional binding sites have been confirmed by CHIP experiments (Chen et al., 2012)(Perry et al., 2011) and, finally, the relevance of few specific binding sites has been investigated by means of mutational analysis(Berman et al., 2004).

On the whole, the segmentation network is able to process the positional information derived from maternal inputs with stunning precision. The noise levels in the concentration of segmentation TFs is as low as it is physically possible given the stochastic fluctuations induced by diffusion(Gregor et al., 2007a). Moreover, the precision and reproducibility of embryo segmentation can be directly related to the reproducibility of maternal inputs(Petkova et al., 2014). This means that the positional information carried by the concentration of maternal factors is read out optimally by the network, allowing each cell in the blastoderm to know its position along the embryo longitudinal axis with 1% precision(Petkova et al., 2019). The high robustness and precision of the segmentation network are achieved by multiple levels of control, for example, by the regulation of the same gene by multiple enhancers with overlapping expression patterns, which produce sharper expression patterns(Perry et al., 2011).

1.3 Experimental methods to measure enhancer's activity in Drosophila embryos

In order to answer the many open questions about transcriptional regulation in general and the *D. melanogaster* segmentation network in particular, we need quantitative information on the spatiotemporal evolution of each element of the network. Ideally, this means to measure the concentration and production rate of each protein and mRNA molecules involved in the process at all times and positions in the embryo. For both mRNA and protein quantification, two complementary approaches with different sets of advantages and limitations can be used. One approach is based on the labeling of a target molecule in chemically fixed embryos, thus providing a static picture of the system. The second approach is instead based on genetically encoded fluorescent reporters and therefore allows one to follow developmental dynamics in real time(Gregor et al., 2014).

For protein quantification in fixed embryos, immunohistochemistry has been used ex-

tensively to observe protein localization and relative protein levels in *D. melanogaster* (Warn et al., 1979). However, the quantification of absolute protein levels with this approach is only possible when reference samples can be measured together with the fixed tissues. Moreover, obtaining a full-time course of protein expression profiles, although possible through a careful analysis of morphological features of the embryos(Fowlkes et al., 2008), is very laborious and not extremely precise.

The discovery of fluorescent proteins provided a new tool to follow protein dynamics in general, and transcription factor dynamics in particular, using live imaging techniques. Genetically encoded protein fusions of TFs and fluorescent proteins allowed us to measure absolute protein levels and investigate additional aspects of TFs dynamics, like their diffusion or degradation(Gregor et al., 2007b). This approach, however, has been limited to the study of maternal factors of the segmentation network. This limitation comes from the fast dynamics of gene expression at the level of zygotically expressed genes, which introduce additional challenges. Changes in gene expression happen on a timescale of a few minutes, which is comparable to the maturation time of fluorescent proteins(Cubitt et al., 1995). Therefore, the resulting delay between the translation of fluorescent proteins and the onset of a fluorescent signal renders protein fluorescence difficult to interpret(Ludwig et al., 2011b). More recently, Bothma et al. developed a new approach to quantify TFs concentration in *D. melanogaster* embryos. In this approach, eGFP is provided maternally and is therefore already fluorescent and uniformly distributed in the entire embyro. Crispr/Cas9 is used to tag the specific TF of interest with a nanobody that is able to bind to eGFP. Therefore, upon expression of the TF and its localization in the nucleus, eGFP is also translocated to the nucleus, resulting in a local increase of eGFP fluorescence (Bothma et al., 2018).

For mRNA quantification, fluorescent *in-situ* hybridization has been historically the method of choice in fixed embryos. This technique is based on a primary antibody recognizing an epitope incorporated in the DNA probe and a secondary fluorescently labeled- or enzyme conjugated- antibody for the readout. These methods provide high sensitivity and make it possible to characterize the spatial modulation of gene expression levels driven by many enhancers (Edgar et al., 1986b)(Weil et al., 2010), but can only offer a qualitative or semi-quantitative readout of mRNA concentration, because of the non-linear nature of enzymatic labeling(Hughes and Krause, 1998). However, more laborious protocols combined with advanced imaging techniques can be used to obtain quantitative data on mRNA expression (Fowlkes et al., 2008). More recently, single molecule fluorescent *in-situ* hybridization (smFISH) has reached the ultimate single mRNA molecule sensitivity. This approach is based on short nucleic acid probes complementary to a gene of interest, which are directly linked with a fluorophore. The resulting high contrast la-

beling allows one to detect both single mRNA molecules in the cytoplasm and foci of transcription in the nucleus, where multiple mRNA molecules cannot be resolved from each other. However, the fluorescence of single mRNA molecules in the cytoplasm can be used as a reference to quantify the amount of mRNA being produced at foci of active transcription in the cell nucleus, thus quantitatively measuring both cytoplasmic and nascent transcripts(Little and Gregor, 2018).

Also in this case, methods based on fixed-tissue staining have a range of precision, in particular in embryo staging, that is insufficient to study the molecular dynamics of transcriptional processes. To overcome this limit, researchers have once more turned to *in-vivo* labeling and live imaging. The most successful approach uses RNA binding proteins that can specifically recognize and bind certain mRNA sequences. An example of this approach is the MS2-MCP bacteriophage system(Bertrand et al., 1998), which has been used to directly capture the temporal dynamics of transcription in living cells. It is based on the co-expression of two components: an MCP-GFP fusion protein and a series of multiple RNA hairpins (MS2 loops) integrated in the sequence of the transcribed gene of interest or reporter gene.

To apply this method to *D. melanogaster* embryos two transgenic lines of flies are required (Fig. 1.5 a) (Garcia et al., 2013). In one line, MCP-GFP is expressed under the control of a maternal driver (e.g. the "nanos" maternal driver) and MCP-GFP molecules are deposited in the embryo before fertilization. This provides plenty of time for the maturation of MCP-GFP molecules to occur before embryo development starts. In the second fly line, an MS2-*yellow* or MS2-*lacZ* reporter gene, consisting of 24 repeats of the MS2 stem loop upstream the *yellow* or *LacZ* gene coding sequence (6.4kb or 5.7kb), is expressed under the control of the an enhancer of interest. The two fly lines are crossed and embryos are imaged with live confocal, light-sheet or two-photon fluorescence microscopy. In the developing embryo the reporter MS2-*yellow* gene is transcribed and nascent mRNA forms MS2 hairpins (Fig. 1.5 b). As they form, the hairpins are recognized and bound by MCP-GFP molecules (Fig. 1.5 c). The resulting local accumulation of MCP-GFP molecules is detected as a diffraction limited fluorescent spot over a fluorescent background of the unbound MCP-GFP molecules. Each fluorescent spot is detected and segmented in 3D (Fig. 1.5 d), and the ratio between the intensity of the spot and its local background is used as an estimator of the amount of mRNA being transcribed.

Importantly, the inherent fluorescence background from unbound MCP-GFP molecules limits the sensitivity of detection. Tens of accumulated GFP molecules are necessary to lead to a measurable signal. Therefore, in order to reach a high signal to noise ratio, it's important that the reporter construct includes multiple copies of the MS2 stem loop upstream of a long reporter gene. Indeed, since the MS2-reporter gene construct is tran-

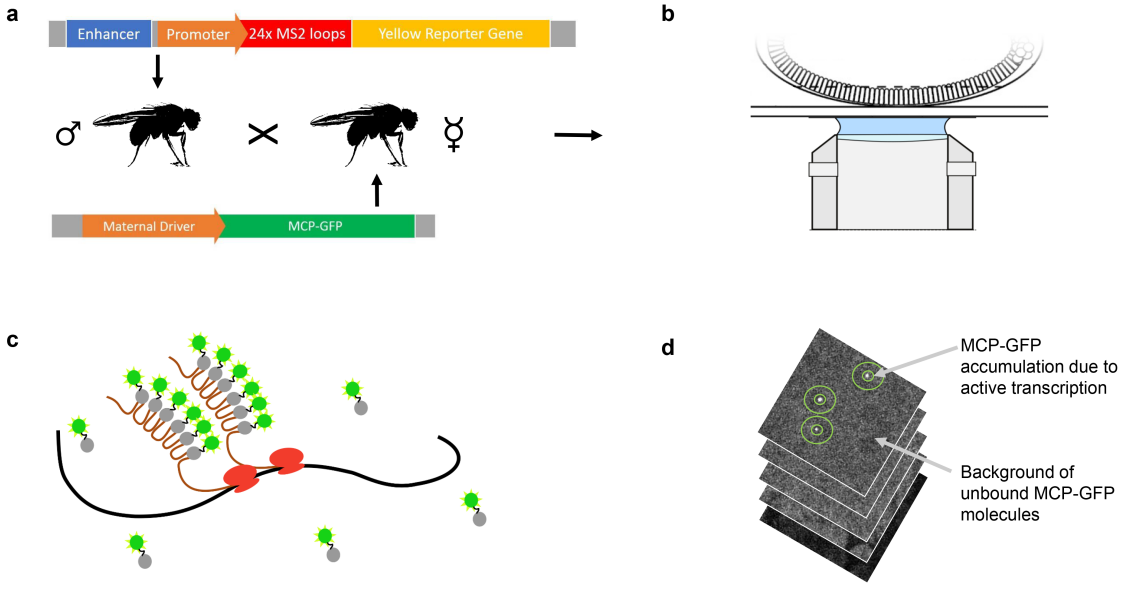

Figure 1.5: Live imaging of transcription in *D. melanogaster* embryos with the MS2-MCP system. a) Cross of transgenic fly stocks required for imaging transcription with the MS2-MCP system in *D. melanogaster* embryos. Male flies, carrying an MS2-yellow reporter gene expressed under the control of an enhancer of interest, are crossed with female virgins in which MCP-GFP is expressed under the control of a maternal driver. b) Embryos are imaged by means of live fluorescence microscopy. Confocal, light-sheet or two photon microscopy are required for their sectioning capabilities to achieve high contrast in a thick specimen like *D. melanogaster* embryos (Mavrakis et al., 2008). c) As soon as the MS2-yellow reporter construct is transcribed, MS2 hairpins are formed. The hairpins are recognized and bound by MCP-GFP molecules. The simultaneous transcription of the reporter gene by multiple RNA Pol II molecules causes further accumulation of MCP-GFP molecules at the site of transcription. d) Fluorescent spots resulting from the local accumulation of MCP-GFP are detected in 3D. The ratio between the intensity of the spot and its local background is used as a proxy for the amount of mRNA being transcribed.

scribed by multiple polymerase molecules sequentially and the elongation time depends on the length of the transcribed gene, using a longer reporter gene increases the local accumulation of MCP-GFP molecules and the signal-to-noise ratio.

In cell culture, despite the fluorescent background discussed above, this method can

reach single mRNA molecule sensitivity. However, this is not the case in *D. melanogaster* embryos where the detection threshold is higher due to more difficult imaging conditions (light scattering, higher autofluorescence background, etc.), thus limiting MS2-MCP sensitivity in comparison to smFISH(Gregor et al., 2014), for instance. Nevertheless, the approach of the MS2-MCP system to measure transcription in *D. melanogaster* embryos has made it possible to unravel new aspects of enhancer activity(Garcia et al., 2013) (Lucas et al., 2013), for example, the role of enhancers in controlling transcriptional bursting(Fukaya et al., 2016).

1.4 Research questions

The work presented in this thesis has two distinct research goals: the first goal is to develop a new experimental strategy to track enhancer activity *in-vivo*, overcoming some of the limitations of existing techniques and allowing us to track the activity of both native and synthetic enhancers. The second goal is to exploit the advantages of a new reporter to obtain quantitative data on the activity of synthetic enhancers that have been designed to study specific questions about the segmentation network. In particular, our aim is to use synthetic enhancers to clarify the role of the bifunctional transcription factor Hb and determine how different features of the enhancers sequence can coordinate its different behaviors.

1.4.1 A new method to measure enhancer's activity

Measuring the spatiotemporal dynamics of enhancer activity during the rapid development of *D. melanogaster* embryos requires a quantitative, sensitive and scalable method. Changes in gene expression in fly embryos indeed happen over minutes and requires a sensitive system offering high enough spatial and temporal resolution. As we have seen, the few existing methods available to measure enhancer activity have all significant limitations: *in-situ* hybridization has been historically the method of choice for direct, spatially resolved, mRNA quantification and, more recently, smFISH has reached the ultimate single mRNA molecule resolution. Unfortunately, both methods rely on the staining of fixed embryos, and therefore cannot measure transcriptional dynamics. On the other hand, the MS2-MCP system relies on the detection of nascent transcripts as fluorescent spots over a fluorescent background of unbound MCP-GFP molecules, which could significantly impair signal-to-noise ratios and, as a consequence, limits the sensitivity of detection that can be reached with this system.

The first goal of this thesis is to explore new methods that could overcome these various limitations. In particular, we aim for a quantitative method that could offer experimental advantages in measuring the activity of medium to large numbers of native or synthetic enhancers, with respect to the existing techniques. In this thesis, we explore the idea that bright and fast-maturing fluorescent proteins could be used as a real-time, quantitative reporter of enhancer activity. An obvious potential issue of protein reporters is that they only offer an indirect readout of transcription. However, enhancer activity and mRNA concentration could be, in principle, inferred from the time course of reporter fluorescence. On the other hand, fluorescent protein reporters can potentially offer several advantages that we want to exploit: they make it possible to perform live imaging and

thus to monitor enhancer dynamics in real time. They have no intrinsic background, they thus limit the background to the specimen auto-fluorescence, which would lead to higher sensitivity. They require simpler crossing schemes of genetically engineered fly stocks compared to the MS2-MCP system, thus improving experimental throughput.

1.4.2 The role of Hunchback binding sites in segmentation enhancers

Considering all the complex biochemical processes underlying transcriptional regulation that we discussed in the previous sections, it comes as no surprise that also the effect of some TFs involved in *D. melanogaster* segmentation is context dependent. One example is given by Hb, a key regulator of many gap and segment polarity genes of the segmentation network. It is expressed, both zygotically and maternally, in the anterior half of the embryo. Hb has been reported to act as an activator or a repressor on different enhancers (Zuo et al., 1991)(Staller et al., 2015). Early reports postulated that Hb could switch from a repressive to an activating factor when bound to the enhancer in close proximity to Bicoid (Simpson-Brose et al., 1994). Other studies have instead suggested that different binding modalities of Hb (e.g. as dimers vs monomers) could be a key determinant of its effect on gene expression (Papatsenko and Levine, 2008). Even when Hb is clearly acting as a repressor (for example in setting the anterior boundary of activity of various gap or segment-polarity enhancers like those controlling the genes *Krüppel*, *Knirps* and *Giant*) its repressive activity can show different context dependent behavior. While, for some of these enhancers, Hb is sufficient for repression and works in a simple concentration dependent manner, for other enhancers Hb creates a permissive environment for the action of additional repressive factors (Yu and Small, 2008). Despite this relatively large set of observations, we haven't yet reached a satisfactory understanding of the role of the Hb binding site in segmentation enhancers and, in particular, how different features of the enhancers sequence can coordinate the different behaviors of Hb.

The second goal of this thesis is to elucidate how Hb bifunctionality works and which features of an enhancer sequence controls its behavior. We believe that combining synthetic biology with a quantitative reporter of enhancer activity could offer a mean to clarify the role of Hb binding sites in enhancers. In particular, Hb bifunctionality could be studied by measuring the activity of synthetic sequences carrying different arrangements of binding sites for Hb and other activator TFs. In this thesis, we focus on two sets of synthetic enhancers. First, following a similar experimental strategy as introduced by Fakhouri et al. (Fakhouri et al., 2010), we will combine Hb binding sites with binding sites for two activating transcription factors of the orthogonal D-V patterning system: Twist

(Twi) and Dorsal (Dl). Importantly, this design allows one to simultaneously monitor the enhancer activity both in the presence or absence of the Hb protein, all in a single embryo. Then, in a second set of experiments, we will exchange the Twi and Dl binding sites with binding sites for a different TF, the anterior activator Bcd. This experimental design will allow us to understand if Hb acts consistently on different activators or if its activity depends on specific protein-protein interactions. Moreover, this second set of experiments will also allow us to test the role of Hunchback in a context that is closer to that of native enhancers, since Hb and Bcd often regulate the same enhancers.

1.5 Summary of the results

In the first part of this thesis (Chap. 2), we demonstrate that bright and fast-maturing fluorescent proteins can be used as a real-time, quantitative reporter of enhancer activity in living *D. melanogaster* embryos. We present a new reporter based on the expression of an optimized fluorescent protein. The reporter consists of the bright and fast maturing fluorescent protein mNeonGreen (Shaner et al., 2013) fused to multiple localization signals, and coupled to translational enhancers sequences. We monitor the protein expression level by confocal fluorescence microscopy. Since protein concentration is only an indirect readout of transcription, we infer mRNA levels by analyzing the time course of the mNeonGreen fluorescence intensity with a reconstruction algorithm based on a model of ordinary differential equations. We validate our approach by comparing our data with those we obtain with the MS2-MCP system for the activity of the well-studied hunchback anterior (hb_ant) enhancer. In addition, we challenge the sensitivity of our reporter system by measuring the weak expression patterns driven by a short synthetic enhancer carrying three binding sites for the transcription factor Bcd. The weak activity driven by this short enhancer turns out to be undetectable using the MS2-MCP system, but it is well captured by the new reporter, thus proving the superior sensitivity of this approach. Finally, we use our reporter in a proof-of-principle study of synthetic enhancers. We measure quantitative differences in the activity of three synthetic enhancers, which carry different combinations of binding sites for Bcd and Zelda (Zld). Our data show how the distance between an enhancer and its target promoter affects transcription , and how Zld influences both the intensity and dynamics of Bcd-dependent transcription. The work described in Chap. 2 of this thesis has been accepted for publication in *Communications Biology*.

In the second part of this thesis (Chap. 3), we apply our new reporter system to a larger set of synthetic enhancers, designed to elucidate a specific aspect about *D. melanogaster* segmentation: the role of Hb binding sites. We obtain quantitative data on the spatiotemporal dynamics of expression of a set of 20 synthetic enhancers. Our data reveal a novel dual role of Hb binding sites in shaping segmentation enhancers activity: on the one hand, Hb can act as a typical short range repressor by binding to its cognate sequences; on the other hand, a sequence containing multiple Hb binding sites increases enhancer activity independently from Hb binding. This sequence is able to promote a permissive environment for the enhancer activity driven by different activators, possibly by disfavoring nucleosome occupancy due to the fact that Hb binding sites coincide with Poly-dA stretches. The distance dependencies of Hb repression and Hb binding sites activation are different, thus creating a strong non-linear behavior of enhancer's activity as a function

of the enhancer's architecture. The work described in Chap. 3 of this thesis has been prepared for publication and will be submitted to a peer-reviewed journal.

The relevant methods required are described within the respective chapters.

Chapter 2

A new reporter system to measure enhancer activity

2.1 Introduction

The precise control of gene expression by transcriptional enhancers, or its deregulation, are key to a variety of biological processes ranging from animal development to cancer biology (Sur and Taipale, 2016), (Peter and Davidson, 2011). As we have seen in the previous chapter, a striking example of precise transcriptional regulation at work is found in the early stages of *D. melanogaster* embryonic development. There, a network of transcriptional enhancers reads gradients of transcriptional activators and repressors (Jaeger, 2011)(Nusslein-Volhard, 1991), provided maternally. The resulting activation of these enhancers lays down the blueprint of the future fly body structure with stunning precision (Petkova et al., 2019). This process has been a fertile platform for interdisciplinary research devoted to decipher the function of enhancers, and notably how the sequence of an enhancer determines its transcriptional output (Pennacchio et al., 2013), (Spitz and Furlong, 2012). This is a two-prone problem. On the one hand, one needs to understand the logic of a complex architecture of clusters of binding sites for transcription factors (TFs), which are encoded in an enhancer's sequence. On the other hand, it is necessary to precisely measure the transcriptional enhancer activity in space and time. These two pieces of information are prerequisites for quantitatively understanding how gene expression is 'computed' by the combinatorial occupancy of TFs, and for developing predictive models that can link enhancer sequences to gene expression (Segal et al., 2008)(Fakhouri et al., 2010)(Crocker et al., 2017). In this context, methods capable of collecting quantitative and time-resolved activity data for large numbers of natural or synthetic enhancer sequences are of primary importance.

As we have seen in section 1.3, techniques to measure transcription have steadily improved over the past decades. *In-situ* hybridization has been historically the method of choice for direct mRNA quantification in fixed embryos. More recently, smFISH has reached the ultimate single mRNA molecule sensitivity, measuring both cytoplasmic mRNA molecules and nascent transcripts at the site of transcription (Little and Gregor, 2018). However, as sensitive as these labeling methods are, they all rely on the staining of fixed embryos. As a consequence, measuring expression at multiple time points is only possible through labor-intensive procedures and, in any case, these methods are incapable of elucidating the molecular dynamics of transcriptional processes. Live imaging of standard fluorescent protein reporters (e.g. GFP) is extensively used to study gene expression dynamics with reporter constructs. However, their application to *D. melanogaster* embryos has been limited by the slow maturation time of most fluorescent proteins compared to the rapidity of embryo development. In *D. melanogaster* embryos, these methods could provide a delayed and possibly inaccurate representation of the dynamics of enhancer activity (Ludwig et al., 2011a). To overcome this limit, researchers have turned to *in-vivo* mRNA labeling techniques such as the MS2-MCP system, which captures directly the temporal dynamics of the enhancers' activity in living cells. However, unlike experiments in cell culture where this method can reach single molecule sensitivity, the detection threshold is high in fly embryos due to the more difficult imaging conditions. Indeed, the *D. melanogaster* embryo is a thick specimen for optical fluorescence microscopy, which causes several problems like light scattering, absorption and a high background signal (Mavrakis et al., 2008). In turn, this results in a lower sensitivity of the MS2-MCP system in comparison to, for example, smFISH(Gregor et al., 2014).

In this chapter, we introduce a new method that overcomes various limitations of the existing approaches. We develop an optimized version of the fluorescent protein mNeonGreen and use it as a real-time, quantitative reporter of gene expression. mNeonGreen is the brightest among all monomeric fluorescent proteins and is one of the fastest maturing, with a maturation time shorter than 10 minutes (Shaner et al., 2013). We derive mRNA production rate and mRNA concentration from the dynamics of the reporter fluorescence with high spatial and temporal resolution. In comparison with the MS2-MCP system, our results demonstrate a higher detection sensitivity.

To further illustrate the potential of this technique, we apply our methodology to study the influence of (i) enhancer-promoter distance, and (ii) the presence of binding sites for a pioneer transcription factor Zld on Bcd-dependent transcription. We observed an increase in enhancer activity upon reduction of the enhancer-TSS distance, a phenomenon that was already reported previously in other systems. Moreover, whereas we found, not surprisingly, that the binding of Zld results in an increased enhancer activity (presumably

due to an increased accessibility of activator binding sites), we also observed that Zld alters transcriptional dynamics, increasing the transcriptional rate over time. To our knowledge, this phenomenon was never reported before.

Overall, we believe that the mNeonGreen reporter system is a powerful new approach that will be useful to systematically and comprehensively characterize enhancer activities in *D. melanogaster* embryos, and easily extendable to other systems as well.

2.2 Results

2.2.1 A mNeonGreen reporter system to measure enhancer activity in living *D. melanogaster* embryos.

Measuring the spatiotemporal dynamics of enhancer activity during the rapid development of *D. melanogaster* embryos requires a quantitative, sensitive and scalable method. We reasoned that engineering a fast-maturing fluorescent protein could provide us with a sensitive and versatile candidate reporter, as very low protein levels can be readily detected using a state-of-the-art fluorescence microscope. Given the rapidity at which gene expression changes during development and the difficult imaging conditions in *D. melanogaster* embryos, maturation time and brightness of the reporter protein are pivotal parameters. Hence, we considered various fast-maturing fluorescent proteins and selected the recently discovered fluorescent protein mNeonGreen (Shaner et al., 2013), which has the highest known ratio of molecular brightness over maturation time (Lambert, 2019) and has been established as a superior fluorescent tag and expression reporter (Hostettler et al., 2017).

To prevent diffusion of the reporter protein in the syncytial blastoderm to regions far from where it is synthesized, we created an mNeonGreen protein fusion with multiple nuclear localization signals (NLS) at both the C- and N-terminus (Fig. 2.1 a). The nuclear localization of the reporter protein has the additional advantage of increasing the signal intensity over the embryo autofluorescence background, since the background is lower in the nuclei. In preliminary experiments (see Appendix C.1), we tested different arrangements of NLSs of various classes (Kosugi et al., 2009) in *D. melanogaster* S2 cells, and we selected the one achieving the highest ratio of nuclear to cytoplasmic mNeonGreen fluorescence (two NLS sequences downstream and one upstream of the mNeonGreen coding sequence; see Fig. 2.1 a). Finally, to further boost the signal produced by enhancer activation, we codon-optimized the coding sequence of the reporter, coupled the enhancer to a strong synthetic promoter, the Drosophila Synthetic Core Promoter (DSCP) (Pfeiffer et al., 2010)), and included a translational enhancer at the 5'UTR (Fig. 2.1 a), which has been proven to increase transgene translation in *D. melanogaster* (Pfeiffer et al., 2012). We refer to the full sequence of the reporter including the 3'UTR and 5'UTR as mNeonRep.

While a protein reporter provides us with an indirect readout of transcription, the mRNA concentration can be reconstructed from the time course of protein concentration, provided that kinetic parameters for protein and mRNA degradation and protein maturation are known (Fig. 2.1 b). We inferred the mRNA concentration from the temporal

dynamics of the reporter fluorescence by adapting an approach used to analyze protein expression in bacterial cell cultures(Zulkower et al., 2015). We modeled expression and maturation of the mNeonGreen reporter with a set of linear ordinary differential equations:

$$\frac{\mathrm{d}M(t)}{\mathrm{d}t} = \mathrm{mRNA_p}(t) - k_{dm} * M(t) \tag{2.1}$$

$$\frac{\mathrm{d}D(t)}{\mathrm{d}t} = k_p * M(t) - (k_{dp} + k_{mat}) * D(t) \tag{2.2}$$

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t} = k_{mat} * D(t) - k_{dp} * F(t) \tag{2.3}$$

The model includes the concentration of reporter mRNA $M(t)$, the concentration of non-mature dark reporter protein $D(t)$, and the concentration of the mature fluorescent protein $F(t)$. Maturation of the protein and degradation of both protein and mRNA are accounted for by the rates of mRNA degradation k_{dm}, protein degradation k_{dp}, and protein maturation k_{mat}. We directly measured the kinetic parameters in a dedicated set of experiments using alpha-amanitin- or cycloheximide-treated embryos in which either transcription or translation of our reporter gene is blocked (see Fig. 2.12 in the methods section). We found an average maturation time $1/k_{mat} = 6.6$ min, an average mRNA lifetime $1/k_{dm} = 35$ min and a negligible rate of protein degradation over the observed time scale.

By means of linear inversion and a regularized non-negative least square algorithm, we used the ODE model to reconstruct the mRNA concentration, the rate of instantaneous mRNA production and the cumulative or total mRNA production underlying the time course of protein fluorescence. To verify the robustness of this data analysis pipeline, we implemented a bootstrapping algorithm (see Fig. 2.8 and Methods) that estimates the confidence interval of the inferred RNA levels. The interested reader can find all the details about the model fitting and the bootstrapping procedure in section 2.4.7.

We tested the performance and reliability of the reconstruction algorithm with different sets of simulated data to determine the effects of different signal-to-noise ratios (SNR) in the range that we typically observed with our mNeonGreen reporter. In addition, we tested the ability of the algorithm to discriminate between different dynamics of enhancer activity (see Figures A.1 and A.2 in Appendix A). We examined the systematic errors that can be introduced if the most important kinetic parameters are under- or over-estimated (see Figures A.3 and A.4). Overall, this analysis shows that the reconstruction algorithm allows for a robust determination of the mRNA concentration and rate of mRNA production.

Next, we characterized our reporter system by measuring the expression driven by

the well-known hb_ant enhancer (Segal et al., 2008) (Fig. 2.1). As for every enhancer-reporter construct in this thesis, we used live confocal microscopy with a time resolution of 1 minute to image, in triplicates, embryos laid from hb_ant>mNeonRep homozygous parents. Starting from nuclear cycle 13, we observed a strong signal localized in the nuclei, which recapitulates the known (Garcia et al., 2013)(Segal et al., 2008) spatiotemporal activity of the hb_ant enhancer (Fig. 2.1 c-e). During the post-processing procedure we analyzed the fluorescence signal recorded from the cortical region of the embryo. We divided the cortical region in bins corresponding to of 2% of embryo length (EL) defined as the distance along a line connecting the embryo anterior and posterior tips (AP axis, 0% anterior tip, 100% posterior tip, all expression plots represents the dorsal side of the embryo) (Fig. 2.1 e-f-g and Methods). For each bin, we computed the spatial average of fluorescence signal and we analyzed its time course to reconstruct the cumulative mRNA production (Fig. 2.1 h-i and Fig. B.1 a-b) and the instantaneous mRNA production (Fig. B.1 c-d). We achieved temporal registration between different datasets by visual inspection of the embryo images simultaneously acquired by DIC microscopy. We defined time zero as the onset of nuclear cycle 14 (nc14). The spatiotemporal dynamics of the reconstructed mRNA levels matched the known dynamics of the hb_ant enhancer activity (Fig. 2.1 i): the expression domain was first established as a broad gradient with stronger expression in the anterior-most part of the embryo. Later in nc14, expression expanded towards the middle section of the embryo, thus creating a larger domain with a sharp boundary centered around 40% EL (Fig. 2.1 i). The results of mRNA reconstruction for the three biological replicates show a very reproducible pattern with a global correlation coefficient of r=0.99 (Fig. B.1 b).

To validate our approach, we carried out expression measurements using the MS2-MCP tagging technique. We performed live imaging of an MS2-yellow reporter gene expressed under control of the hb_ant enhancer. In order to perform a fair comparison between the two approaches we coupled the enhancer to the same promoter (DSCP) used for the mNeon reporter. The MS2-MCP system requires different imaging conditions than the mNeonGreen reporter to achieve the best SNR. In particular, for the mNeonGreen reporter the detection sensitivity is the most important parameter while spatial resolution is not critical. By contrast, the MS2-MCP system requires high spatial resolution to reliably detect the diffraction limited MCP-GFP spots. Therefore, for these experiments we applied imaging conditions similar to those reported in previous studies that used the MS2-MCP system in *D. melanogaster* embryos (Garcia et al., 2013) (Section 2.4.3). We observed the typical MS2-MCP fluorescent spots in the anterior half of the embryo (Fig. 2.2 c). The strong magnification of the high numerical aperture microscope objectives that are necessary to detect these localized fluorescent signals limits the field of view to

about 30% EL, as for previous works (Garcia et al., 2013)(Lucas et al., 2013). However, to characterize the enhancer activity domain, it is necessary to examine larger portions of the embryos and the acquisition speed of our confocal microscope does not allow tiling without compromising on time resolution. To overcome this limitation, we collected data at different positions in multiple embryos. Our results for the activity of hb_ant agree with those reported in literature using the MS2-MCP system (Garcia et al., 2013). After merging the datasets obtained from several nuclei, we observed an expression pattern similar to that obtained with our mNeonGreen reporter (Fig. 2.2 b and d). We then compared the cumulative mRNA levels measured with both techniques at all points in time and space (Fig. 2.2 e), and we found a high correlation between the two datasets (r=0.95).

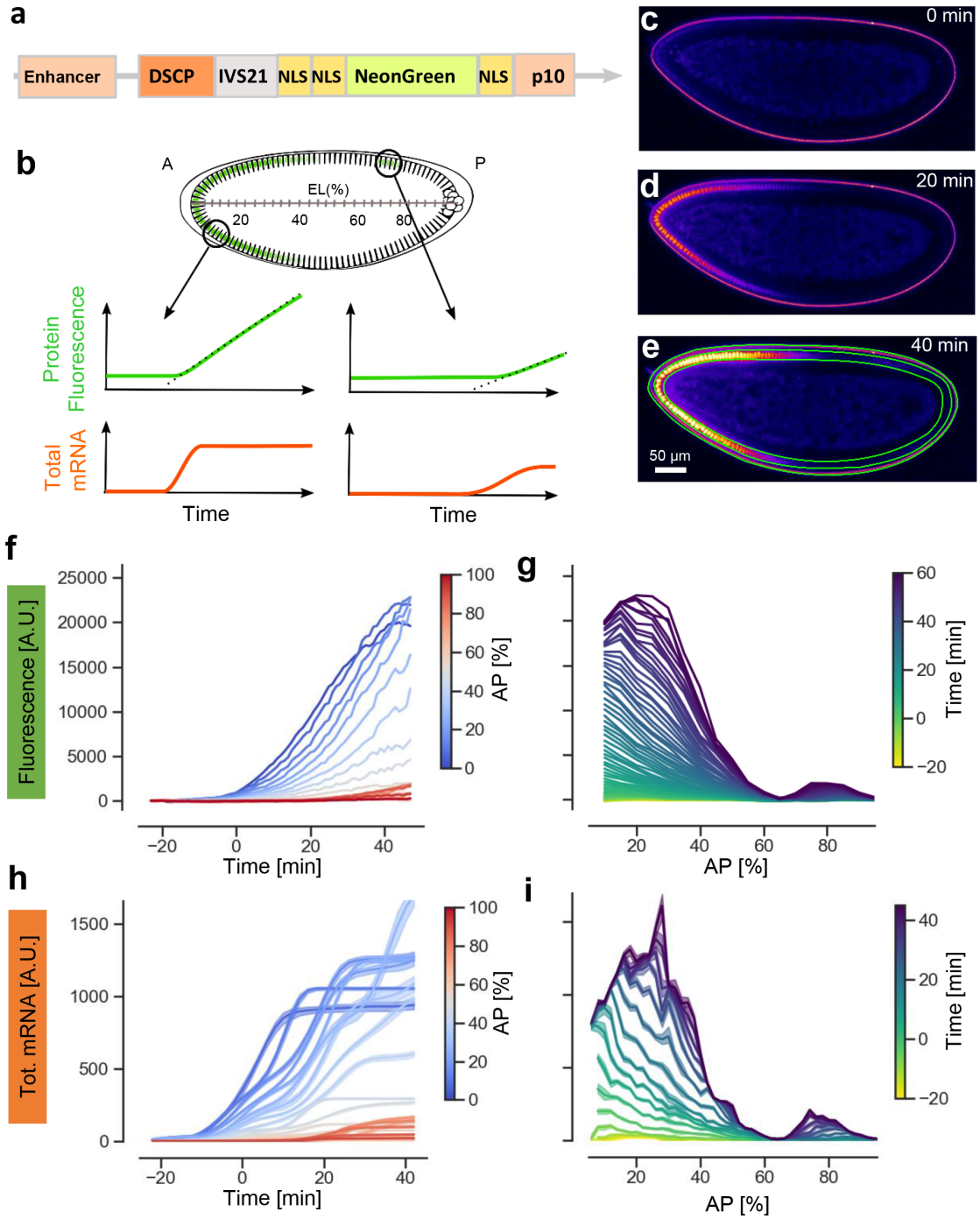

Figure 2.1: (Continued on the following page.)

Figure 2.1: A mNeonGreen reporter system to measure enhancer expression in living *D. melanogaster* embryos. a) The composition of the mNeonGreen reporter construct including: Enhancer, basal DSCP promoter, the translation enhancer sequence IVS21 at the 5'UTR, the mNeonGreen fused to three nuclear localization signals and the terminator sequence p10. b) Illustration of the use of a fluorescent protein as a transcriptional reporter. The time course of protein fluorescence (in green) is different in different portions of the embryo and carries information on the underlying dynamics of the total mRNA production (in orange). Representative confocal slices of embryos carrying the hb_ant-mNeonRep construct at three different time points during embryo development, showing mNeonGreen fluorescence in false colors. f)Traces of the time course of the average fluorescence signal after spatial binning in bins corresponding to 2% of the embryo length. The traces are color coded by their position along the axis. g) Fluorescence expression patterns along the AP axis. Each track corresponds to a different time point of embryo development. h) Time course of cumulative mRNA production, color coded by their position along the axis with a spatial resolution of 2% of the embryo length. i) Total mRNA production patterns at selected time points of embryo development.

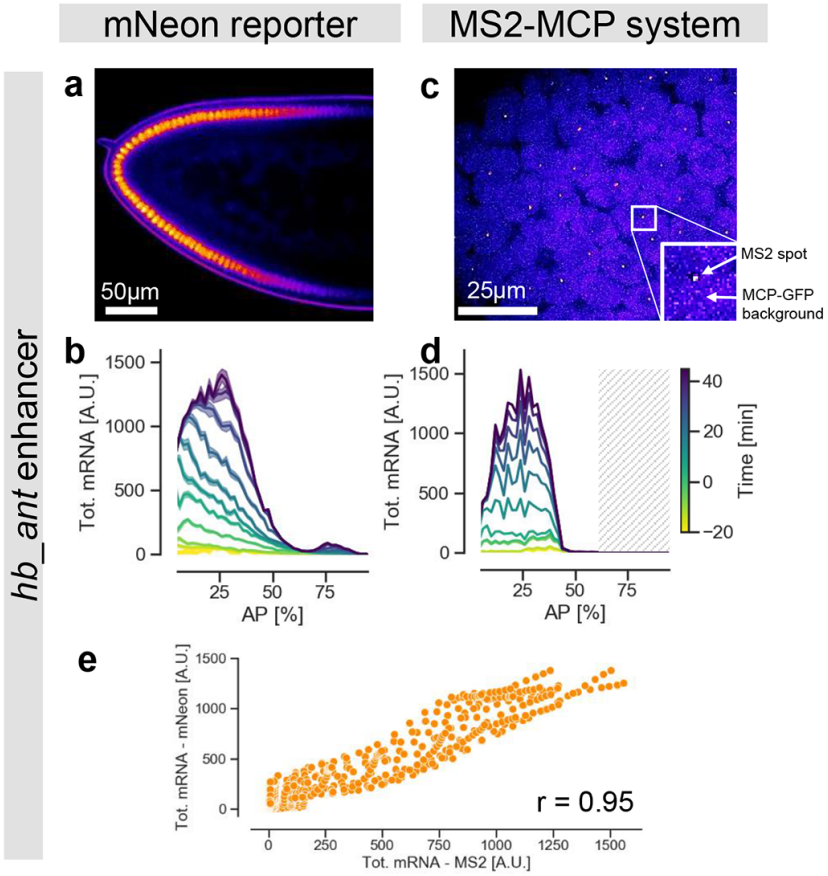

Figure 2.2: mNeonGreen and MS2 system measurements of hb_ant enhancer activity are compatible. a) A confocal slice of the anterior region of an hb_ant-mNeonRep embryo showing mNeonGreen fluorescence in false colors. b) Cumulative mRNA production patterns at different times of embryo development measured with the mNeonGreen reporter. c) A confocal slice of the anterior region of a hb_ant-DSCP-MS2-yellow embryo, MCP-GFP fluorescence represented in false colors. d) Cumulative mRNA production patterns for the hb_ant enhancer as a function of AP position, measured with the MS2-MCP system. Since the field of view is limited to about 30% of the embryo, data are pooled from two independent experiments, focusing on the regions 0%-30% and 30%-60% AP, respectively. The shaded area represents the region of the embryo for which no data was collected. e) Comparison of the cumulative mRNA production patterns measured with the mNeonRep and MS2-MCP reporter systems, at all times and positions. r = Pearson's correlation coefficient.

2.2.2 The mNeonGreen reporter detects weaker expression patterns than the MS2 system.

We then investigated the ability of our reporter system to detect lower mRNA production levels. To this end, we used both techniques to measure the activity of a much weaker synthetic enhancer, consisting of three binding sites for Bcd (Bcd3 enhancer) (Ronchi et al., 1993). Previous studies used *in-situ* hybridization staining to characterize the expression of Bcd3 and observed a weak signal in the anterior part of the embryo (Ronchi et al., 1993)(Simpson-Brose et al., 1994). Using our mNeonGreen reporter assay as above, and in contrast to the hb_ant enhancer for which we observed strong activity characterized by a sharp boundary (Fig. 2.2 a-b), the Bcd3 enhancer produced a weak but reproducible (r=0.83 and Fig. B.2) gradient of reporter expression. The mRNA concentration produced by the Bcd3 enhancer has a peak at around 10% AP and slowly decreases towards the middle of the embryo, where it becomes indistinguishable from background noise at 40% AP (Fig. 2.3a-b). The inferred cumulative mRNA production level at its maximum position (arrow in Fig. 2.3 b) was 30 times weaker for the synthetic Bcd3 enhancer than for the native hb_ant enhancer coupled to the same heterologous promoter. In addition, the dynamics of expression driven by the two enhancers also showed important differences: the cumulative mRNA production from Bcd3 tends to saturate during nc14. This implies that Bcd3 activity progressively declined during nc14, whereas, for the hb_ant enhancer, most of the activity took place around mid nc14 (compare Fig. 2.2 b and Fig. 2.3 b).

Strikingly, we did not observe any clear activity domain of the Bcd3 enhancer when we applied the MS2-MCP technique. We could only detect a limited number of very weak fluorescent dots, hardly distinguishable from the background noise (Fig. 2.3 c-d, and Fig. 2.4). To substantiate this observation, we examined the histogram of the intensity of detected spots (Fig. 2.4). We found that spots intensities show a similar distribution between the Bcd3 enhancer and embryonic regions from the previous experiment where the hb_ant enhancer is not expected to produce any expression. Thus, the detected MS2-MCP spots for the Bcd3 enhancer exhibit a fluorescence signal comparable to noise level. We concluded that the mNeonGreen reporter system has a lower detection limit for mRNA production, making it suitable to study enhancers with low activity levels, such as Bcd3.

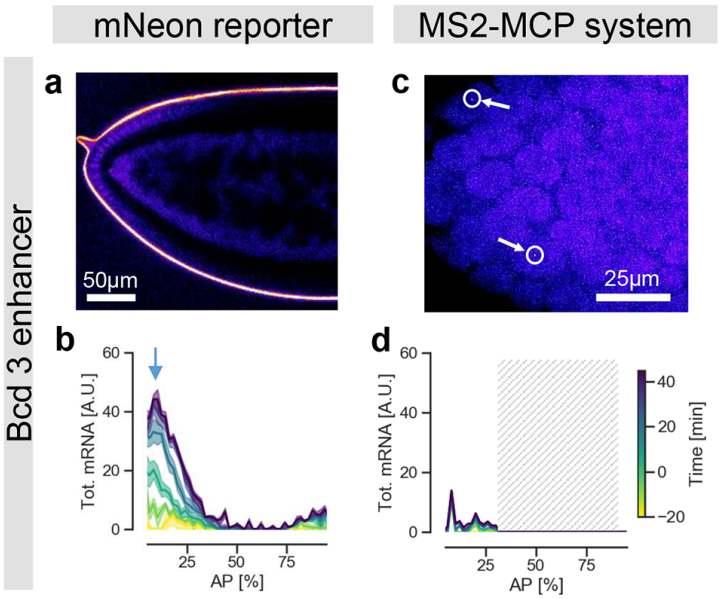

Figure 2.3: The mNeonGreen reporter detects weaker expression patterns than the MS2-MCP system. a) A confocal slice of the anterior region of a Bcd3-mNeonRep embryo, showing mNeonGreen fluorescence in false colors. b) Cumulative mRNA production patterns at different times of embryo development measured with the mNeonGreen reporter. c) A confocal slice of the anterior tip of a Bcd3-DSCP-MS2-yellow embryo with MCP-GFP fluorescence represented in false colors. Circles and arrows indicate the only two potential MS2 spots in the image. d) Cumulative mRNA production patterns for the Bcd3 enhancer as a function of the AP position, measured with the MS2-MCP system.

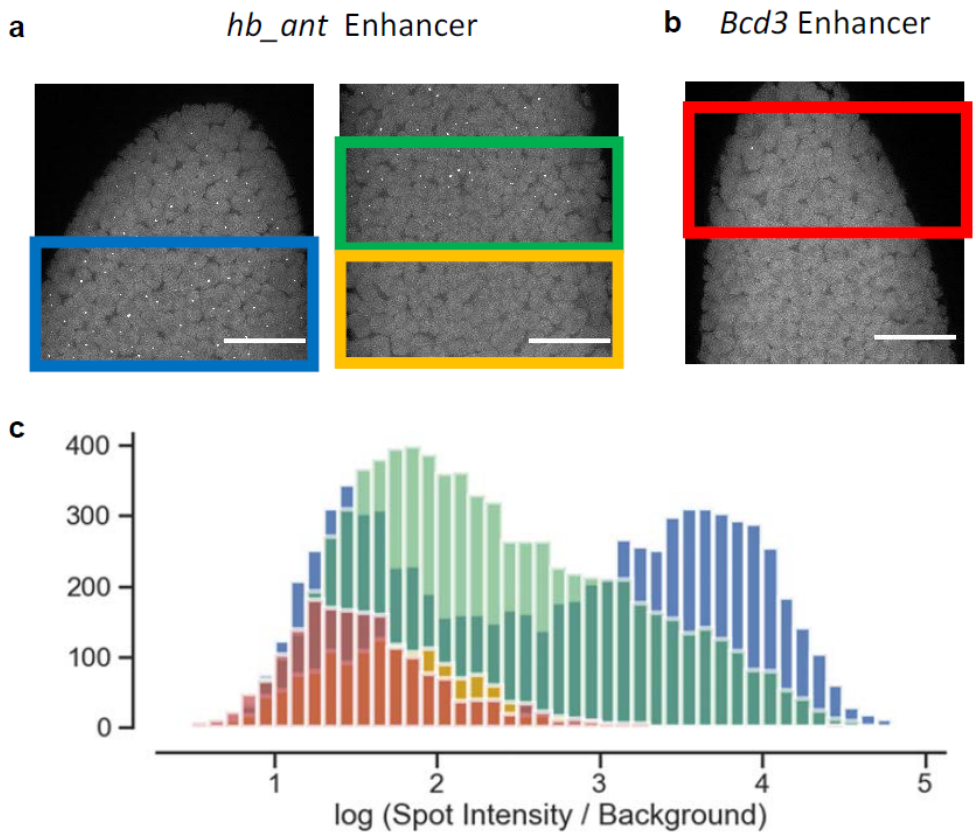

Figure 2.4: Transcriptional dynamics of the Hunchback Anterior and Bcd3 enhancers measured with the MS2 reporter system. a) Snapshots of anterior and middle region of embryos expressing MS2 stem loops under the control of the hb_ant enhancer and b) the Bcd3 enhancer. c) Histograms of the intensity of MS2-MCP fluorescent spots detected in different regions of embryos for the hb_ant and Bcd3 enhancers. Data represented in blue, green and yellow in the histogram corresponds to the intensities of the MS2 dots produced by the activity of the hb_ant enhancer and detected inside the blue, green or yellow boxes in panel a. Data represented in red corresponds to the intensities of the MS2 dots produced by the activity of the Bcd3 enhancer inside the red box in panel c. The MS2-MCP spots for the Bcd3 enhancer exhibit a fluorescence signal comparable to that of the hb_ant enhancer in a region where no expression is expected (yellow box) and their fluorescence intensity is therefore comparable to noise level.

2.2.3 Applying the mNeonGreen reporter to quantify enhancer activity.

To better demonstrate the capacity of our method to quantify the activity and dynamics of enhancers, we first used the mNeon reporter to image additional native *D. melanogaster* enhancers, located in the Krüppel cis-regulatory region. They belong to the two separate cis-acting control units CD1 and CD2: B_element, which is a short (142bp) core part of the Kr_CD1 enhancer, and Kr_CD2 (Fig. 2.5). The three enhancers exhibit very different segmentation patterns and dynamics of total mRNA production (Figure 2.5 a-c, average total mRNA for the three replicates of each enhancer). Whereas hb_ant shows its characteristic expression gradient along the AP axis (Fig. 2.5 a, upper and middle panels), the B_element shows a broad, weak stripe of expression in the anterior half of the embryo (Fig. 2.5 b, upper and middle panels), and Kr_CD2 contrasts with a well defined expression of two strong, narrow stripes at 20% and 50% of the AP, respectively (Fig. 2.5 c, upper and middle panels).

These results are in agreement with what was reported before using either *in-situ* staining (Hoch et al., 1991) or the MS2-MCP system (El-Sherif and Levine, 2016)(Scholes et al., 2019). Previous work on the Kr_CD2 enhancers focused particularly on the central stripe of expression and provided quantitative data on its dynamics using the MS2-MCP system. The MS2-MCP data revealed a dynamic shift of the stripe peak towards the embryo anterior, covering 4% of the embryo length (El-Sherif and Levine, 2016), in perfect agreement with the shift observed in our data (Fig. 2.5 c, inset in the middle panel). The inspection of the temporal dynamics reveals differences between the three different cis-elements (Fig. 2.5 a-c, lower panels and Fig. 2.5 d): the total mRNA levels in the anterior (dots in Fig. 2.5 d) show similar dynamics with a gradual increase already before nc13 (our t=0), however with one noticeable difference: whereas one observes a continual and similar increase of the total mRNA for hb_ant and Kr_CD2, the total mRNA reaches a plateau at low expression levels after t= 10mins for the B_element. Interestingly, expression in the posterior stripes of hb_ant and Kr_CD2 starts later ($t > 20$ mins; crosses in Figure 2.5 d).

To further illustrate the potential of our method to quantify the activity of synthetic enhancers, we tested two variants of the Bcd3 enhancer. We modified two independent features of the enhancer's structure that are believed to influence activity(Amit et al., 2011)(Dufourt et al., 2018) and examined whether our experimental approach could capture any resulting differences. The first feature is the distance between the enhancer and the transcriptional start site (TSS), the second feature, the enhancer accessibility (i.e., its chromatin state). To change the enhancer-TSS distance, we removed a linker sequence

between the Bcd3 enhancer and the DSCP promoter, thereby shortening the distance between the TSS and the most proximal Bcd binding site from 146 bp to 73 bp (enhancer termed Bcd3-proximal). A comparison of the average expression profiles shows that the activity of the Bcd3-proximal enhancer is indeed consistently 2.5 fold stronger than that of Bcd3 (Fig. 2.6 a,b).

To test the influence of enhancer accessibility, we added binding sites for the transcription factor Zld. Zld is a pioneer transcription factor: it is not able to activate an enhancer by itself but initiates processes that lead to activation at later stages in development. In particular, Zld is known to promote chromatin decompaction. To test the effect of Zld on Bcd dependent transcription, we embedded three Zld consensus sites in a 56 bp stretch of a neutral DNA sequence(Crocker et al., 2017) upstream of the Bcd3 enhancer (Fig. 2.6 c and Fig. B.2). Remarkably, the resulting enhancer, termed Zld3-Bcd3, shows a pronounced increase of reporter expression (6 fold stronger on average) compared to the Bcd3 enhancer (Fig. 2.6 b-c). Again, this finding confirms the role of Zld in increasing enhancer activity(Crocker et al., 2017)(Dufourt et al., 2018), without altering the spatial pattern.

To compare the activity profiles of the three Bcd3 enhancer variants more directly, we plotted the ratios of cumulative mRNA production levels between Bcd3-proximal and Bcd3 (Fig. 2.6 d), and between Zld3-Bcd3 and Bcd3 (Fig. 2.6 e). Interestingly, while in the first case the ratio remains constant throughout the blastoderm development, suggesting that decreasing the enhancer-TSS distance simply leads to a rescaling of the activity, the Zld3-Bcd3/Bcd3 average ratio grows over time with values increasing from 5 to 7 during nc14. Thus, Zelda not only increases the rate of the enhancer activity but also alters its dynamics. The cumulative mRNA profiles for the two enhancers indeed show that the saturation observed early in nc14 for Bcd3 does not occur in the Zld3-Bcd3 variant, which remains active throughout the period measured. We conclude that features that are expected to modulate enhancer activity in similar ways, in this case by increasing the transcription rate, may result in quite different dynamics. Such subtle differences are not immediately obvious or accessible with classical reporter assays, and our experiments demonstrate the power of our mNeonRep system to capture and quantify them.

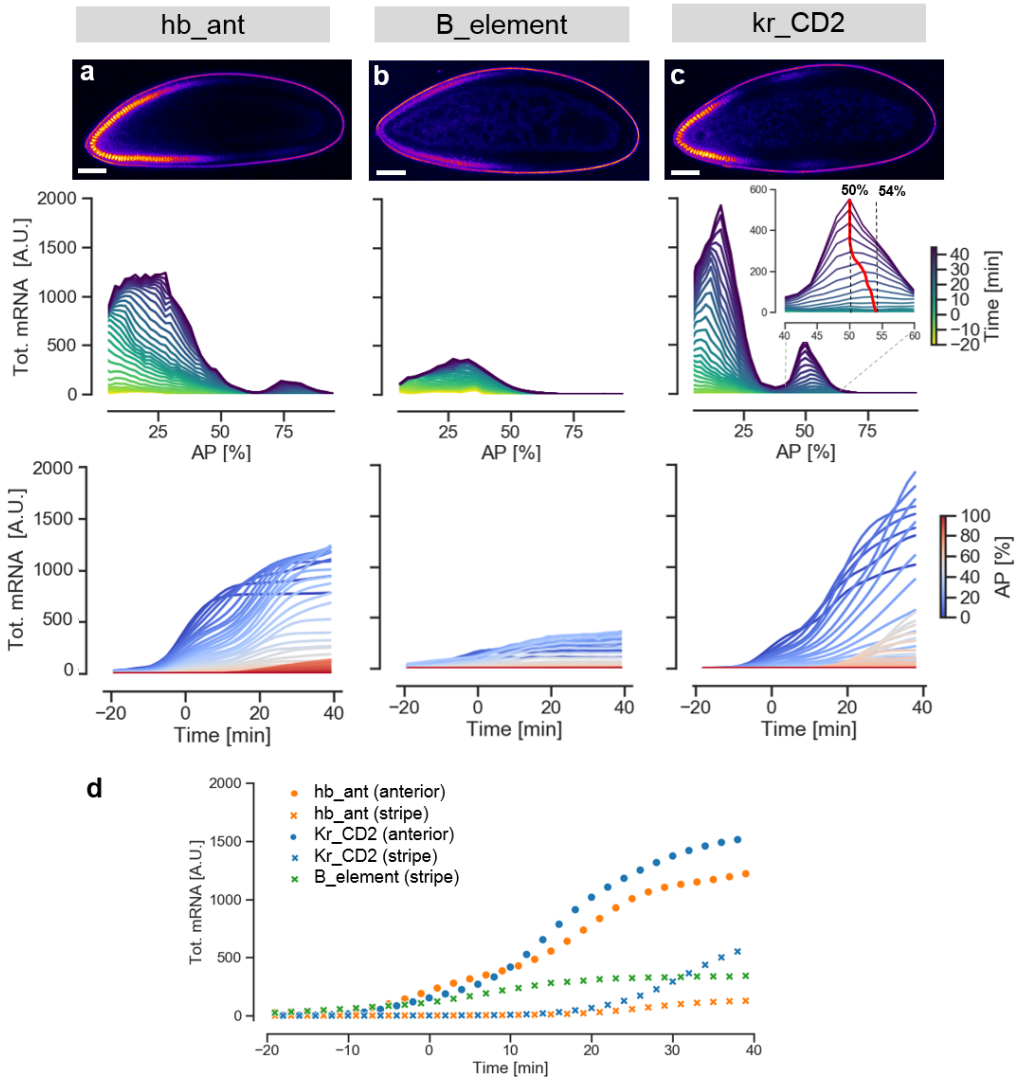

Figure 2.5: Activity and dynamics of embryos carrying native *D. melanogaster* enhancers using the mNeon reporter system. a-c) upper panel: Confocal fluorescence sections of a) hb_ant, b) B_element and c) kr_CD2 embryos just before gastrulation. Middle panels: total mRNA levels along the AP axis and as a function of time. The time is color-coded (see inserts). The data are averages of three biological replicates of each enhancer. Lower panel: Total mRNA levels as a function of time. The AP position is color-coded (see inserts). Averages for three replicates of each enhancer. Whereas hb_ant shows its characteristic expression gradient along the AP axis, B_element shows a broad stripe of weak expression in the anterior half of the embryo, and kr_CD2 contrasts with a well defined expression of two strong, narrow stripes at 20% and 50% of the AP, respectively. d) Temporal dynamics for the anterior region (dots) and the posterior stripe (crosses).

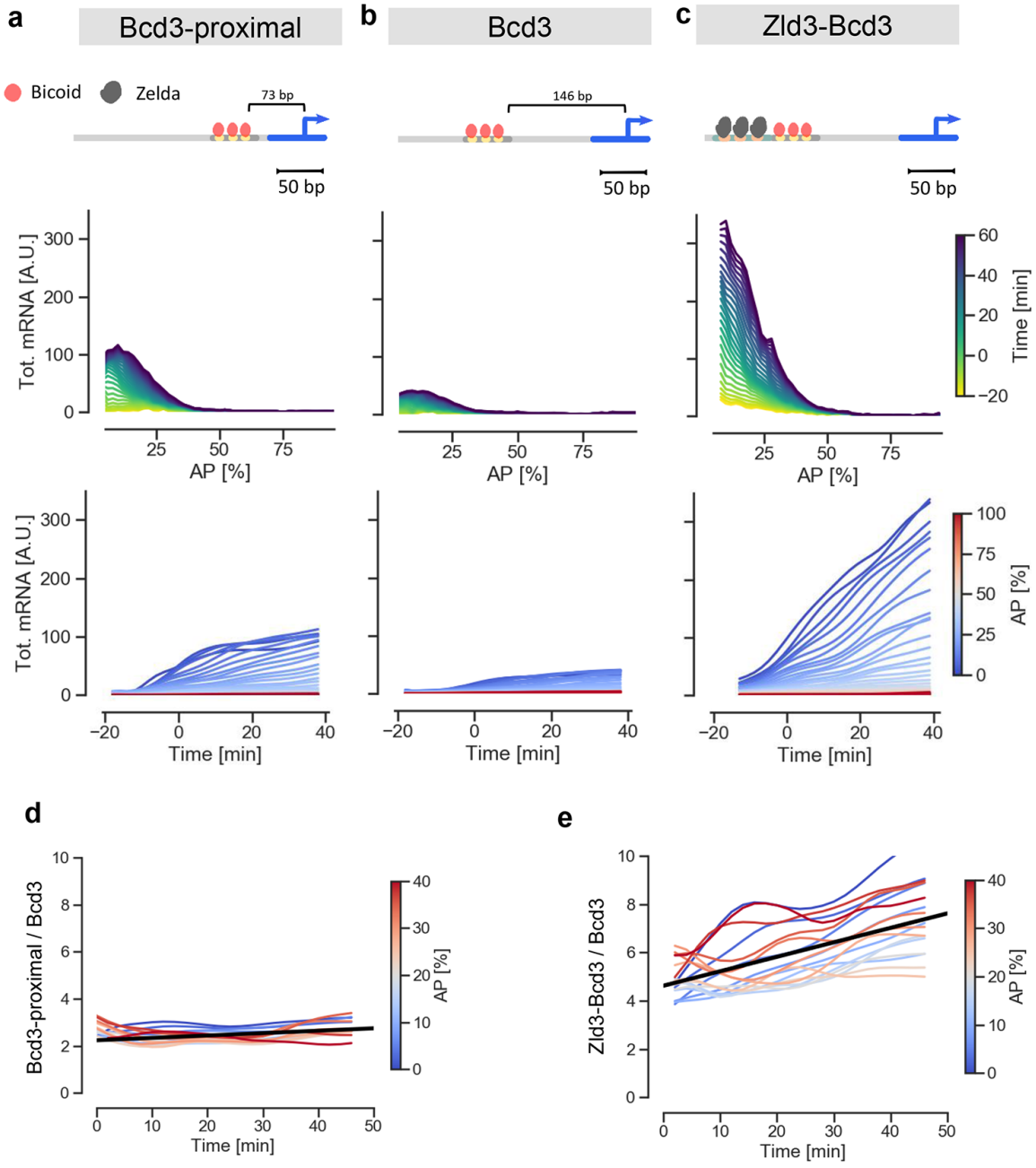

Figure 2.6: (Continued on the following page.)

Figure 2.6: Applying the mNeonGreen reporter to measure transcriptional dynamics of synthetic enhancers. a - b - c) Dynamics of the activity of the Bcd3-proximal, Bcd3 and Zld3Bcd3 enhancers, respectively. From top to bottom: sketch of enhancer architectures, illustrating arrangement of binding sites and distance from the TSS. Cumulative mRNA production patterns along the AP axis, color coded by the time of embryo development (average of 3 biological replicates). Time course of cumulative mRNA production, color coded by position along the AP axis (average of 3 biological replicates). d) Ratio of the cumulative mRNA production for Bcd3-proximal and Bcd3 enhancers as a function of time, color coded by the position along the AP axis. e) Ratio of the cumulative mRNA production for Zld3Bcd3 and Bcd3 enhancers as a function of time, color coded by the position along the AP axis.

2.3 Discussion

2.3.1 Advantages and limitations of the mNeon reporter system.

The mNeonGreen reporter system and the data analysis pipeline presented in this thesis constitute a valuable method for measuring transcriptional dynamics in vivo. The optimized protein reporter presents several advantages. First, it relies on the expression of a single transgenic component, which avoids complex crossing schemes of transgenic organisms and accelerates experimental work. Second, it provides higher detection sensitivity in comparison to the MS2-MCP system, the main and widely used method for studying enhancer dynamics in vivo. This superior sensitivity is most probably achieved thanks to the absence of the fluorescence background that arises from unbound MCP-GFP molecules. Third, the bright mNeonGreen signal makes it possible to image through low numerical aperture/low magnification objectives, which can capture large fields of view and thereby whole embryos at once, with enough spatial resolution to clearly distinguish single cells. Altogether, these advantages significantly increase detection sensitivity and throughput to measure transcriptional dynamics in organisms.

The application of the mNeonGreen reporter system to the *D. melanogaster* blastoderm system nevertheless poses specific challenges that limits its spatial resolution in comparison to other techniques (sub-nuclear resolution for the MS2–MCP system and smFISH). In this system, the absence of cellular membranes allows for the diffusion of both the reporter protein and RNA. Nevertheless, this process is likely limited because of the strong nuclear localization of the reporter protein. This is supported by the following observations. First, the observed expression patterns are compatible with those measured with other techniques, which would not be the case in the presence of rapid diffusion of either protein or RNA. In particular, we found a quantitative agreement of our data on the kr_CD2 central stripe with published data. Since this domain of expression is very narrow, it constitutes a challenging test to rule out a strong impact of the reporter diffusion on our data. The agreement between our data with published data thus demonstrates the spatial localization of our reporter, which was able to detect the dynamic shift in the position of the stripe peak, covering just 4% of the embryo length. Second, a study demonstrated that diffusion of mRNAs between neighboring nuclei is limited to only 1-2 nuclei in the *D. melanogaster* blastoderm(Bothma et al., 2018). Note that in the absence of cell membranes, the possible cross-talk of the mNeonGreen signal between neighboring nuclei, which we could not characterize directly, does not make our reporter system suitable for studying single cell dynamics, in particular phenomena like transcriptional bursting, for which the MS2-MCP system is more useful. However, dynamical informa-

tion can still be extracted with the mNeonGreen reporter system when considering it to be a spatially averaged signal over 2-3 nuclei in *D. melanogaster* blastoderm, which is therefore the spatial resolution of our reporter. Moreover, this limit is strictly related to the syncitial nature of the *D. melanogaster* blastoderm, while single cell resolution can certainly be reached in other organisms or cell cultures.

Another important aspect of our reporter system is its relatively high temporal resolution, which depends on the signal-to-noise ratio of the protein fluorescence signal. We estimated that the reconstruction algorithm is able to resolve fast changes in enhancer activity (e.g., a short pulse of RNA production) with a resolution of about 7 minutes under our imaging conditions (as assessed by numerical simulations; see Appendix A, Fig. A.2). This value is modestly lower than that achievable with the MS2-MCP system (around 2-3 minutes), presumably because of a longer delay between RNA transcription and signal accumulation and the needed reconstruction step. However, in experiments with a higher signal-to-noise ratio of the fluorescence signal, the temporal resolution of our system can be readily improved by tuning a single regularization parameter of the reconstruction algorithm (λ; see Methods section, Fig. 2.8 and Fig. A.2 d).

2.3.2 The effect of enhancer-promoter distance and Zelda on Bicoid dependent enhancers.

To validate and illustrate the technical enhancements discussed in the previous section, we applied our method to a particular biological question: How structural or architectural features of an enhancer affect its activity? The three synthetic enhancers we used to this end; Bcd3, Bcd3-proximal and Zld3-Bcd3, show unambiguous differences in their spatiotemporal dynamics. In that sense, they provide a convincing illustration of the power of our system. They also reveal interesting biological properties. First, although enhancer activity was originally thought to be independent from the orientation and distance of the enhancer respect to the promoter (Banerji et al., 1981), subtle effects of locus architecture on enhancer function have been reported (Tara Lydiard-Martin Meghan Bragdon, 2014)(Symmons et al., 2016). In the simpler case of bacterial enhancers, a systematic analysis of synthetic enhancers positioned at close distances from the promoter (20-500 bp) revealed a non-linear dependency on the distance with a peak activity at around 70 bp (Amit et al., 2011). In *D. melanogaster*, the enhancer-promoter distance has also been shown to influence absolute levels of enhancer activity (Tara Lydiard-Martin Meghan Bragdon, 2014). The increased enhancer activity that we observed upon reduction of the enhancer-TSS distance (Bcd3-proximal) is in agreement with these notions. At the molecular level, it may reflect the frequency of contacts between enhancer and core promoter.

Interestingly, changing the enhancer-promoter distance tuned the level of expression without altering enhancer dynamics: both Bcd3 and Bcd3-proximal exhibit a rapid reduction of transcription after the beginning of nc14. The reduction of Bcd-dependent transcription for native and synthetic enhancers have been studied in previous works and is thought to depend on Bcd sumoylation (J. and Ma, 2012), (Li et al., 2014). The fact that we observe similar dynamics for both enhancers may imply that increasing the chances of enhancer-promoter communication does not interfere with TF binding to the enhancer.

We also used our new methodology to examine another enhancer feature, the chromatin state. We focused on binding sites for the pioneer transcription factor Zld (Liang et al., 2008), whose role in enhancers has been extensively studied. Binding of Zld to its cognate sites promotes the deposition of histone modifications and increases DNA accessibility (Liu and Ma, 2015). Furthermore, single molecule studies have shown that Zld creates local hubs of increased Bcd concentration in the vicinity of enhancers, thus increasing the occupancy of the binding sites (Mir et al., 2017). Artificially inserting binding sites for Zld in native or synthetic enhancers has been reported to increase the enhancer activity and accelerate enhancer activation after nuclear divisions (Crocker et al., 2017)(Dufourt et al., 2018). The increased activity we measured for Zld3-Bcd3 compared to Bcd3 is therefore not a surprise. Our method, however, also reveals an additional property of these added Zld sites, relating to the temporal dynamics of enhancer activity: Zld binding sites prolong enhancer activity. Whereas Bcd3 and Bcd3-proximal enhancers' activities are substantially reduced already 10 minutes after the beginning of nc14, Zld3-Bcd3 activity remains sustained throughout the entire nc14 (Fig 2.6 b-c). It was recently observed that Zld responsive enhancers require continuous Zld activity throughout the entire phase of zygotic genome activation that takes place during the entire nc14 (McDaniel et al., 2019). However, our observation would also suggest that the introduction of Zld sites interferes with the deactivation of Bcd, which is also taking place during nc14. We speculate that this effect could be mediated by different mechanisms: a direct interaction with the sumoylating enzymes, or the sequestration of Bcd in a local enhancer microenvironment where sumoylating enzymes are not present. A last possibility is that secondary co-factors which read out Bcd sumoylation could also be influenced by the presence of Zld or Zld-induced histone modifications.

In summary, we have shown that our mNeonGreen reporter system is a powerful tool to study transcriptional dynamics and is particularly suited for studies that aim to quantify expression dynamics of larger numbers of native or synthetic enhancers including constructs with weak expression levels. Our approach offers significant advantages in terms of sensitivity and throughput compared to existing methods and can, in principle, be applied to other organisms or cell cultures, provided the necessary optimizations of

the construct to a new organism (e.g. using a different promoter and a codon optimized sequence adapted for the new organism) and the characterization of the rates that are required for the reconstruction algorithm in may vary in a different system.

2.4 Methods

2.4.1 Cloning of transgenes

The mNeonGreen reporter construct was generated by C- and N-terminal fusion of a codon optimized mNeonGreen (Shaner et al., 2013) coding sequence (optimized with the Eurofins genomics GENEius software package - Munich, Germany , and obtained by gene synthesis) to three different nuclear localization signals: the Bipartite-N-term NLS (Magico and Bell, 2011), the SV40 NLS and a Class3 C-term NLS (Kosugi et al., 2009). All enhancers were coupled to a basal Drosophila synthetic core promoter (DSCP), a particularyl strong and inducible synthetic promoter that has been widely used to characterize enhancer activity in *D. melanogaster* (Pfeiffer et al., 2010). A complete list of all sequences is provided in Appendix C.1.

The sequences of hb_ant and Kr_CD2 enhancers were amplified from genomic DNA. The sequence of the Bcd3 enhancer (Ronchi et al., 1993) was generated by annealing of single-stranded oligonucleotides corresponding to the forward and reverse strands. The sequence containing 3 Zld binding sites was designed by inserting three consensus binding sites for Zld (Jung et al., 2018) into a neutral background described previously (Crocker et al., 2017), and was generated by oligo annealing. In all constructs, except for Bcd3-proximal, a 73-bp linker separated the enhancer from the basal promoter. This sequence does not contain any predicted binding site for transcription factors of the segmentation network, based on available positional weight matrices (PWM) (Jung et al., 2018). To optimize expression, we included the IVS+Syn21 translational enhancer sequences (Pfeiffer et al., 2012) at the 5'UTR and the p10 terminator sequence (Pfeiffer et al., 2012) at the 3'UTR. We refer to the full sequence of the reporter including 3'UTR and 5'UTR as mNeonRep. All elements were cloned into an expression construct based on the pBDP backbone (a gift from Gerald Rubin; Addgene plasmid #17566) as described in (Bozek et al., 2019) with only one difference: the insertion of an additional 340bp long neutral spacer (Sayal et al., 2011) upstream of the enhancer. The rational for this additional step is given by the fact that a preliminary analysis revealed the presence of significant binding sites for segmentation factors in the backbone of the reporter plasmid, which could have potentially interfered with the architecture of our synthetic enhancers (see predicted bindings sites in Fig. D.1 and Fig. D.2 in Appendix D).

For the generation of the MS2 reporter construct, the 24XMS2 tag (a gift from Robert Singer, Addgene plasmid #31865) was fused upstream of the yellow reporter gene21 coding sequence (a gift from Liqun Luo, Addgene plasmid #24350). The 24xMS2- yellow sequence was then cloned immediately downstream the enhancer-linker-DSCP sequence

into the same pBDP backbone used to clone the mNeonGreen reporter construct.

2.4.2 Fly Stock generation

All reporter plasmids for both the mNeonGreen and MS2 reporters were integrated in the same attP2 docking site using PhiC31 integrase(Pfeiffer et al., 2010). Homozygous fly stocks were generated by crossing a single male with a single homozygous virgin female, and the insertion of the correct construct was verified by single-fly PCR of both parents and sequencing of the PCR products.

2.4.3 Live Imaging

mNeonGreen:
Enhancer-mNeonRep embryos were collected, dechorionated in 50% bleach and mounted between a semipermeable membrane and a microscope cover glass, immersed in halocarbon oil (Sigma). Imaging was performed at 24±1°C on a Zeiss LSM710 confocal microscope using a 40x 1.2NA water immersion objective. The pixel size was set to 1.1 μm. Since the field of view of the 40x lens is not large enough to cover the entire embryo, we acquired two tiled z-stacks, each consisting of 3 images separated by 7.5 μm in z, were acquired at each time point. The resulting field of view of 250 μm x 580 μm allows for imaging of an entire embryo in a single movie with a final time resolution of 60 s per z-stack. The laser power was optimized to obtain the maximum signal while avoiding significant photobleaching (see Fig. 2.7). The optimal laser power was found to be 8μW, measured in the back focal plane of the objective

MS2:
yw;Histone-RFP;MCP-GFP (Bloomington Drosophila Stock center #60340) virgins were crossed with males carrying either the hb_ant-DSCP-MS2-Yellow or Bcd3- DSCP-MS2-Yellow reporter genes. Embryos were collected and mounted as previously described. Imaging was performed on a Zeiss LSM710 confocal microscope using a 63x 1.4NA oil immersion objective. The pixel size was set to 0.33 μm and the image field of view to 169 μm x169 μm. A stack of 15 images separated by 1.3 μm in z was acquired at each time point. The final time resolution was 60 s per z-stack. At the end of each movie, a single tiled z-stack with a much larger field of view of 500 μm x 840 μm was acquired to precisely locate the imaged area relative to the embryo tips.

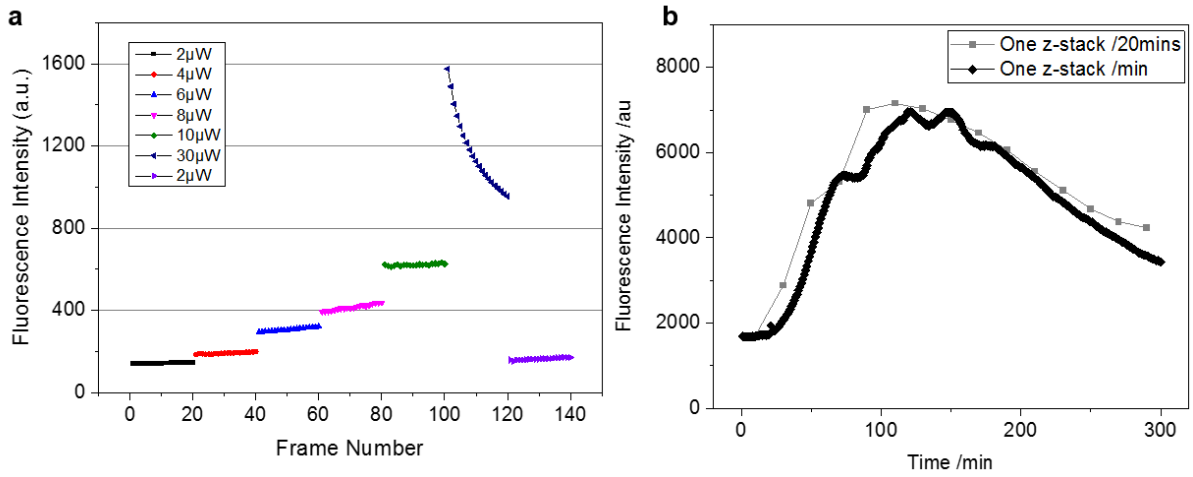

Figure 2.7: Optimization of the laser excitation power for the mNeon reporter sytem. a) In order to optimize the laser excitation power and check for photobleaching, we looked at the average intensities of consecutive confocal fluorescence time series of the anterior part of a developing hb_ant embryo. Since we have constant expression of the reporter and negligible degradation, we expect that over a short time frame the fluorescence signal rises with a constant rate. We changed the laser power every 20 frames from 2μW to 30μW (measured at the microscope objective back focal plane), as indicated in the insert. The laser power was set at the same value of 2μW for the first and the last cycle of confocal images to check that the rate of expression of the mNeonGreen reporter did not substantially changed during the measurements. Since the reporter concentration is increasing at a constant rate, we expect the slope of the signal to increase linearly with the laser power until photobleaching occurs. The optimal laser power provides the highest slope, for which the increase of the fluorescence signal is maximal while photobleaching is negligible. Here, the optimal laser power was found to be 8μW (pink points). b) As an additional control that photobleaching is negligible at the selected laser power of 8μW, we looked at the average fluorescence time course of two embryos carrying the hb_ant enhancer and imaged over 5 hrs, with one frame per minute (our usual condition) and one frame every 20 mins, respectively. The fluorescence time courses of the two datasets overlap very well, demonstrating that photobleaching is very low.

2.4.4 Alfa-Amanitin and Cycloheximide injections

To measure the degradation rate of reporter RNA in the embryo, transcription was blocked by injecting alpha-amanitin (Sigma-Aldrich: A2263) using a needle, at a concentration

of 0.4 mg/ml (Edgar et al., 1986a). To measure the maturation rate of fluorescence of the reporter protein in the embryo, translation was blocked by injecting cycloheximide (Sigma-Aldrich: 01810) using a needle, at a concentration of 0.9 mg/ml (McCleland and O'Farrell, 2008). In both cases, hb_ant-mNeonRep embryos were collected, dechorionated in 50% bleach, glued on a coverslip and dried at 18°C for 20 minutes prior to the injections. Injected embryos were imaged as described above.

2.4.5 Image segmentation

The region of interest to read out the reporter signal is the cortical region of the embryo which contains the nuclei and is located just beneath the vitelline membrane. In order to automatically detect this area and read out the reporter fluorescence while taking into account the variability in shape and size of the embryos we implemented an image analysis pipeline (Fig. 2.8 a,b and c). Confocal stacks of embryos were processed as follows using the *Definiens XD 2.0* software package (Munich, Germany). In brief, for each stack the external contour of each embryo was identified using the strong autofluorescence signal arising from their vitelline membrane. The external contour of the embryo was then shrunk twice by a few pixels towards the interior of the embryo (region delimited by the blue lines in Fig. 2.8 c). We then applied a watershed segmentation to produce small segmented patterns, randomly distributed and of various sizes in the segmented cortical region (typically 6-10 pixels; each individual size corresponded typically to less than 1% of the embryo length). The mean mNeonGreen fluorescence signal was then computed for each element of the small segmented pattern and the areas of the image with a particularly strong fluorescence compared to their surroundings, such as those arising from the presence of yolk particles in the embryo cortex during the earlier stages of development, were removed from the analysis. Finally, the segmented elements from all stacks were binned together based on their position along the AP axis of the embryo with a resolution of 2% of the embryo length to compute their average mNeonGreen fluorescence intensity.

2.4.6 Temporal registration

In order to compare expression patterns between different embryos, it is important to precisely register the time relative to a common reference point. We chose the common reference point (time zero, t=0) as the instant at which membranes reappear throughout the whole embryo after the mitotic division that follows nc13 and we used differential interference contrast (DIC) images for the calibration procedure of each movie. Time

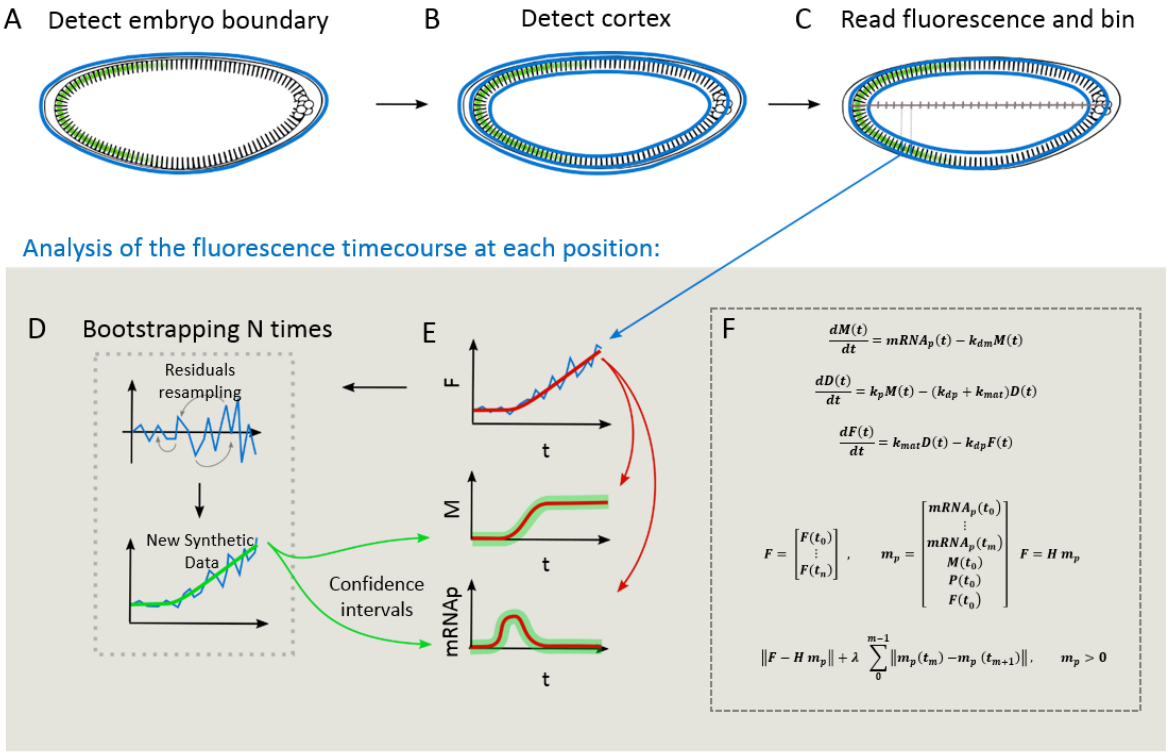

Figure 2.8: Data analysis pipeline for the mNeonGreen reporter. a-b-c) Sketch of the image analysis pipeline: in order to read out the fluorescence signal of the reporter protein we first detect the embryo boundary. By shrinking twice the contour that describes the boundary of the embryo, we generate two new contours which delimit the cortical region of embryo. We read out the fluorescence in this region and bin the data based on the position along the AP axis of the embryo using bins of 2% of the embryo length. d-e-f) Sketch of the mRNA reconstruction analysis. The time course of the signal in each bin is analyzed separately by fitting the model expressed by equations in f to the fluorescence data. The model is fitted to the data minimizing a regularized sum of squared residuals. After the first fit of the model to the data, we apply a bootstrapping algorithm based on the resampling of the residuals. The residuals are reshuffled and added again to the fluorescence time course predicted by the model to obtain a new synthetic dataset. This new dataset is reanalyzed and the bootstrapping procedure is repeated N times to construct confidence intervals around the predicted mRNA and mRNA production levels. The reshuffling of the residuals is performed taking into account that the residuals are uncorrelated but their amplitude depends on the signal intensity.

zero was determined with a precision of ±1min, which also matched the time resolution of our measurements.

2.4.7 Analysis of Reporter time course and mRNA reconstruction

Protein levels offer an indirect readout of transcription. To overcome this limitation, we followed the approach presented in (Zulkower et al., 2015) and we modeled the expression of our reporter with a set of ordinary differential equations:

$$\frac{\mathrm{d}M(t)}{\mathrm{d}t} = \mathrm{mRNA_P}(t) - k_{dm} * M(t) \tag{2.4}$$

$$\frac{\mathrm{d}D(t)}{\mathrm{d}t} = k_p * M(t) - (k_{dp} + k_{mat}) * D(t) \tag{2.5}$$

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t} = k_{matT} * D(t) - k_{dp} * F(t) \tag{2.6}$$

Using the linearity of this model and discretizing the time, this system of differential equations can be rewritten as a linear model. The linear model links, through a model matrix H, the array of observed fluorescence values $F(t_i)$ to the mRNA production rate $m_p(t_j)$ and to the initial concentrations of mRNA $m(0)$, to the dark and mature fluorescent protein concentration $P(t_0)$ and $F(t_0)$.

$$\boldsymbol{F} = \begin{bmatrix} F(t_0) \\ F(t_1) \\ \vdots \\ F(t_n) \end{bmatrix}, \quad \boldsymbol{m_p} = \begin{bmatrix} \mathrm{mRNA_P}(t_0) \\ \mathrm{mRNA_P}(t_1) \\ \vdots \\ \mathrm{mRNA_P}(t_m) \\ \mathrm{P}(t_0) \\ \mathrm{F}(t_0) \end{bmatrix}, \qquad \boldsymbol{F} = \mathbf{H}\,\boldsymbol{m_p} \tag{2.7}$$

The matrix **H** in the equation above depends only on the structure of the original ODE system and on the rates of mRNA degradation, k_m, protein degradation, k_{dp}, and protein maturation, k_{dm}. In particular, each column of this matrix represents the solution of the ODE system for F at all times t_i, when only one burst of mRNA production of unitary amplitude takes place at time t_j, or, for the last two columns of H, when no mRNA production takes place but a unitary amount of either non-fluorescent (P) or fluorescent (F) protein is present at t_0. The matrix can therefore be generated by solving, either numerically or analytically, the ODE system for all these conditions.

The linear model can be fit to the time course of protein fluorescence $F(t_i)$ using a regularized non-negative least square algorithm which determines the rate of RNA

production minimizing:

$$\|\boldsymbol{F} - \mathbf{H}\,\boldsymbol{m_p}\|^2 + \lambda \sum_{i=0}^{m-1} \|\mathrm{mRNA_P}(t_i) - \mathrm{mRNA_P}(t_{i+1})\|, \quad \mathrm{mRNA_P}(t_i) > 0 \tag{2.8}$$

which includes the assumptions that the mRNA production rate and all concentrations must be positive definite. In order to implement the regularization numerically, it is convenient to use a trick and define an extended model matrix $\mathbf{E}$ that already includes the regularization and an extended arrays of observed fluorescence $\boldsymbol{F^*}$:

$$\mathbf{E} = \left[\begin{array}{ccccc} \multicolumn{5}{c}{\mathbf{H}} \\ \hline \lambda & -\lambda & 0 & \dots & \dots \\ 0 & \lambda & -\lambda & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \dots \\ \hline 0 & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & 0 \end{array}\right], \quad \boldsymbol{F^*} = \begin{bmatrix} F(t_0) \\ F(t_1) \\ \vdots \\ F(t_n) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.9}$$

The sum of squared residuals for this extended model coincides with the regularized sum of squared residuals for the original model:

$$\|\boldsymbol{F^*} - \mathbf{E}\,\boldsymbol{m_p}\|^2 = \|\boldsymbol{F} - \mathbf{H}\,\boldsymbol{m_p}\|^2 + \lambda \sum_{i=0}^{m-1} \|\mathrm{mRNA_P}(t_i) - \mathrm{mRNA_P}(t_{i+1})\| \tag{2.10}$$

Having included the regularization in the definition of the model allows us to use all available methods for the minimization of the squared residuals of linear models. If we wouldn't have to include any additional constraints to the model, this would make it possible to find the solution that minimizes the squared residuals exactly. Instead, we want to include the additional constraint that all $\mathrm{mRNA_P}(t_i)$, $\mathrm{P}(t_0)$ and $\mathrm{F}(t_0)$ are positive definite, but we can still implement these additional constraints by using an efficient non-negative least square minimization algorithm such as, for example, *nnls* in the Python SciPy optimization library(Lawson and Hanson, 1995).

After fitting, we calculate the cumulative mRNA production as:

$$mRNA_{Tot}(t_i) = \sum_{j=0}^{i} \mathrm{mRNA_P}(t_j) \tag{2.11}$$

In order to estimate the sensitivity of the reconstruction to noise, we implemented a bootstrapping of the residuals algorithm (Kreiss and Lahiri, 2012) (Fig. 2.8 d,e and f).

We first checked that the residuals of the fit showed no correlation at different times (Fig. 2.9 a), and that the standard deviation of the residuals scaled linearly with the intensity of the signal (Fig. 2.9 b). Then, we implemented bootstrapping by rescaling the residuals with the average signal, reshuffling the normalized residuals and rescaling them back again (Fig. 2.8). Using this protocol, we build a "synthetic" time course of fluorescence and, by fitting the data, we obtain a set of predicted mRNA production rates. Since the procedure can be repeated N times, we can use it to obtain a population of predicted mRNA production rates at each time, from which we derive confidence intervals.

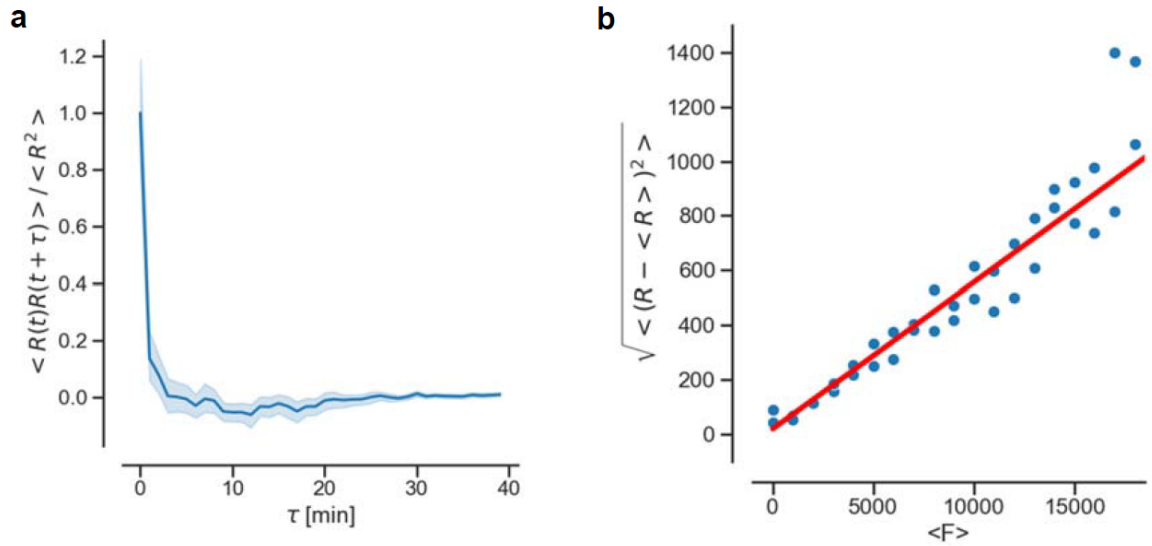

Figure 2.9: Statistical analysis of fit residuals. a) An autocorrelation function of the residuals of the mRNA reconstruction fitting procedure for hb_ant Enhancer. No time correlation of the residuals is observed. <> denotes the average of all the data at different positions and times and R denotes the residuals of the fit. b) The amplitude of the residuals of the mRNA reconstruction fitting procedure depends linearly on the intensity of the fluorescence signal. A linear fit of the data gives: $26 + 0.018 < F >$

Since the RNA reconstruction algorithm includes a regularization parameter λ, we explored the impact of this parameter with numerical simulations of the data analysis pipeline. A summary of the results is presented in Appendix A. Decreasing the value of this parameter increases the time resolution (Fig. A.2 d). However, too small of values for λ give rise to unstable reconstructions creating artifacts and large errors in the estimates (Fig. A.2 e). In this respect, bootstrapping also offers an internal control that guides the choice of λ. The minimum value of λ that gives reproducible reconstructions depends

on the signal-to-noise level of the measured protein fluorescence. A higher signal-to-noise ratio allows for obtaining reproducible results with lower values of λ and increased time resolution.

Both the regularization parameter λ and the number of bootstraps N are hyper-parameters of the reconstruction algorithm. To systematically fix their values, we relied on numerical simulations of the data analysis pipeline. In particular, we considered three different dynamics of mRNA production, similar to those represented in Fig. A.1 where we simulated fluorescence data and added noise to the simulated fluorescence, in order to reproduce the same amplitude and statistical properties of the noise observed in our actual measurements. We quantified the precision of the reconstruction by looking at the average correlation between the simulated ground-truth mRNA production timecourse and the result of the reconstruction algorithm. We then looked at how the regularization parameter λ and the number of bootstraps influence the precision of the reconstruction (Fig. A.5). We found that the average correlation reaches a maximum for $\lambda = 7$ and that no further significant improvement is achieved for $N > 40$.

2.4.8 Signal linearity, background subtraction and correction for photobleaching

An important assumption for quantitative fluorescence live imaging is that the recorded fluorescence signal is directly proportional to the fluorophore concentration. We checked the validity of this assumptions under our imaging conditions using a serial dilution of Rhodamine6G in a 96 well plate. We found that the response of our setup is non-linear only at very low concentration (Fig. 2.10). The observed non–linearity can be described by a polynomial fit, which we used to correct all raw data.

The fluorescence signal measured in the embryo cortex arises from two independent contributions: the reporter protein fluorescence that we aim to isolate and the embryo auto-fluorescence. Moreover, the background auto-fluorescence is not homogeneous and not constant in time but, under our imaging conditions, is subjected to photobleaching. Under our imaging conditions the mNeonGreen signal is minimally affected by photobleaching as we have already shown. In order to estimate the spatial pattern of the background auto-fluorescence, we computed an average background profile by averaging the signal of 4 frames at the beginning of each movie, when no mNeonGreen fluorescence is detectable. Moreover, in order to compensate for the photobleaching of the background, the auto-fluorescence background profile was then rescaled at each time point. An estimation of the rate of photobleaching was calculated for each experiment by fitting the timecourse of the signal in a region of the embryo where no expression of the reporter pro-

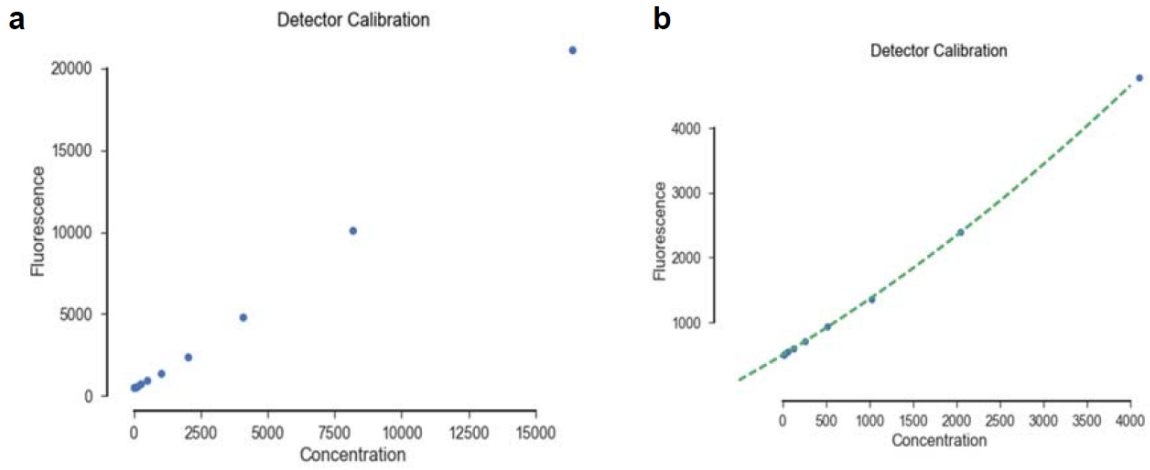

Figure 2.10: Linearity of the imaging setup. a)The plot represent the fluorescence signal from a serial dilution of Rhodamine6G in a 96 well plate as a function of the dye concentration under the same imaging conditions used to image the mNeonReporter. As expected the signal depends linearly on the dye concentration. However, when zooming in at low concentrations (b) a slightly non-linear behavior can be observed, potentially due to dark counts or a non-linearity of the detector itself. However, this non linear trend can been interpolated with a polynomial fit, which we used to correct the data obtained from expression measurements in living embryos.

tein was observed nor expected. Finally, the average background was rescaled at each time point, assuming the embryo autofluorescence to be equally affected by photobleaching at all positions in the embryo.

As a final negative control for the entire image and data analysis pipeline we measured a wild-type embryo that does not express mNeonGreen (Fig. 2.11). Even though a relatively strong autofluorescence can be observed in the yolk, the background signal is extremely weak in the cortical region of the embryo where the nuclei are located. We analyzed confocal images from the wild wild-type embryo using the same analysis pipeline applied to embryos expressing the mNeon reporter, including image segmentation, background correction and mRNA reconstruction. The final result doesn‘t show any evident artifact or systematic error introduced by the data analysis pipeline. Moreover, thanks to these data we could also characterize the noise level of our measurements which corresponds to 5 A.U. of total mRNA. This level is 10 times lower than the expression level of the weakest enhancer presented in this this thesis.

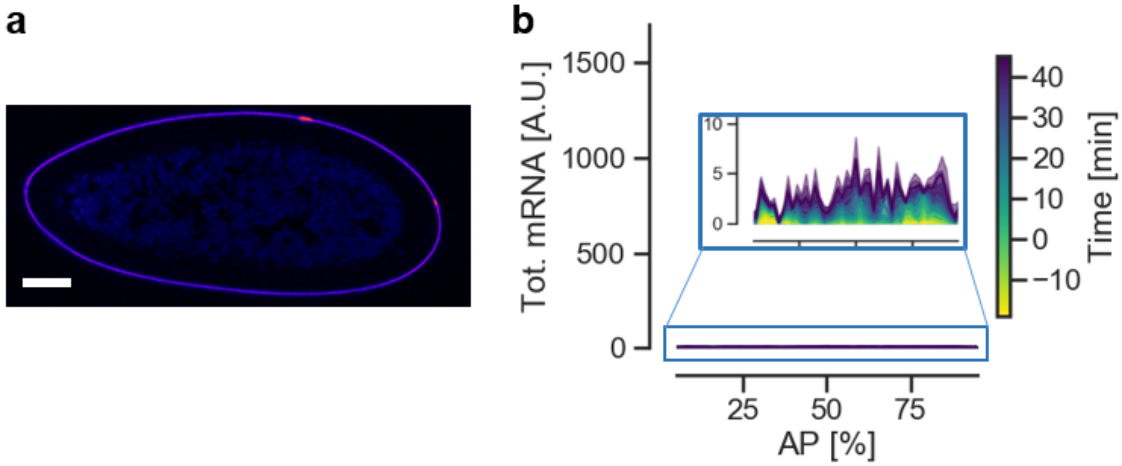

Figure 2.11: Imaging a wild type embryo as a negative control for the image and data analysis pipeline. a) A confocal fluorescence section of a wild type embryo at the end of stage 5. Whereas a weak autofluorescence signal can be observed in the vitelline membrane and in the yolk in the middle of the embryo, the background signal is extremely low in the relevant region where the nuclei are. b) Reconstruction of the total mRNA levels along the AP axis from wild-type embryos is shown. Confocal images from a wild type embryo has been analyzed using the same analysis pipeline applied to embryos expressing the mNeon reporter, including image segmentation, background correction and mRNA reconstruction. The final result does not show any evident artifact or systematic error introduced by the data analysis pipeline. The noise level of our method results in 5 A.U. of total mRNA, which is 10 times lower than the expression level of the weakest enhancer we measured.

2.4.9 Analysis of mNeonGreen reporter maturation time and mRNA degradation rate

To measure the maturation rate of mNeonGreen and the degradation rate of its mRNA, we analyzed the reporter signal from hb_ant-mNeonRep embyros that have been injected with either the translation inhibitor cycloheximide or the transcription inhibitor alpha-amanitin (Fig. 2.12). To interpret these datasets and extract the kinetics parameters, we used the model described by Equation 2.6.

Injection of cycloheximide blocks protein translation and, therefore, the timecourse of protein fluorescence can be described by setting $k_p = 0$. Under this assumption, the

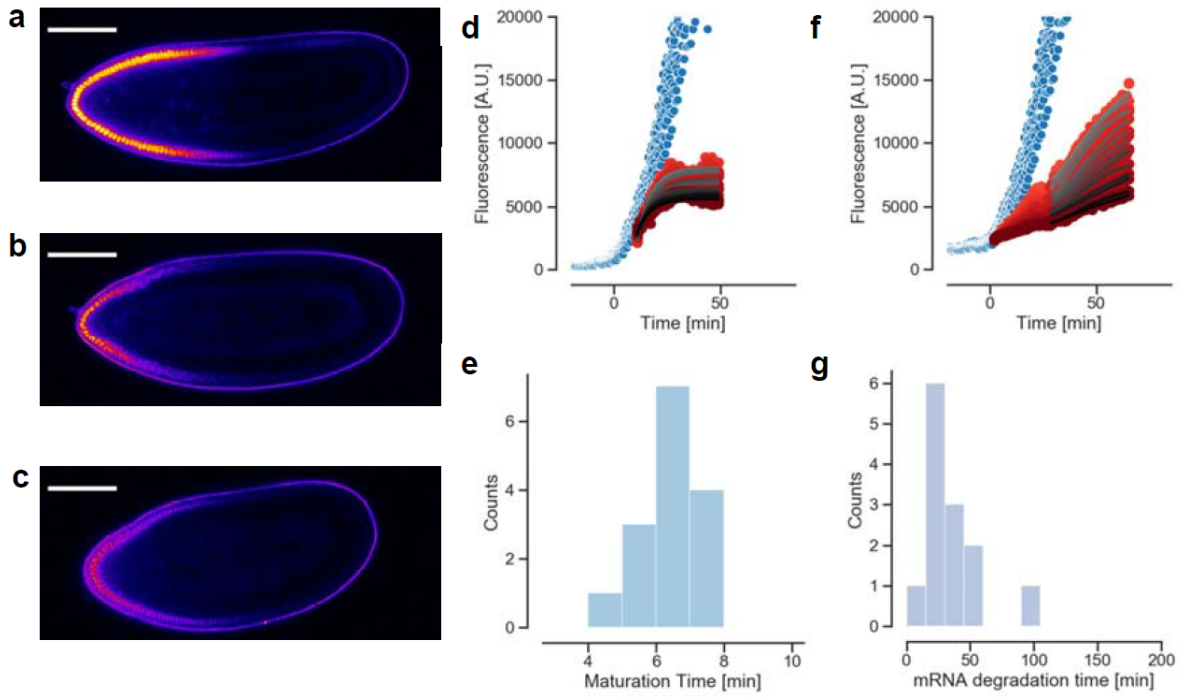

Figure 2.12: mNeonGreen reporter calibration. Calibration experiments for the mNeon expression reporter. a-b-c Embryos have been injected with either a) water b) the translation inhibitor cycloheximide or c) the transcription inhibitor alpha-amanitin. Fluorescence from the mNeonGreen reporter has been monitored using confocal microscopy. (Scale bars 100μm). d) Time course of the signal at different positions in a cycloheximide injected embryo (in red) and a water injected embryo (in blue). Time series data from different positions in a cycloheximide injected embryo are modelled with exponential curves characterized by a maturation time t. (See method for details about the fitting) e) Estimates of the maturation time. The average maturation time over different positions is 6.58 minutes. f) Time course of the signal at different positions in an alpha-amanitin injected embryo (in red) and a water injected embryo (in blue) The time courses of fluorescence at different positions in a alpha-amanitin injected embryo are modelled to infer the degradation rate of the reporter mRNA (See methods for details about the fitting). g) Estimates of the reporter mRNA half-life. The average reporter mRNA half-life over different positions is 35 minutes.

concentration of fluorescent protein is expected to follow:

$$F(t) = Ae^{-k_{dp}(t-t_0)} - B(e^{-k_{mat}(t-t_0)-1})e^{-k_{dp}(t-t_0)} \tag{2.12}$$

Fitting this equation to the fluorescence time course of cycloheximide-injected embryos, we inferred the values of mNeonGreen maturation rate in the embryo (Fig. 2.12 d-e) which results $k_{mat} = 0.15\text{min}^{-1}$.

Injection of alpha-amanitin blocks mRNA transcription and, therefore, the timecourse of protein fluorescence can be described by setting $\text{mRNA}_p = 0$ after the time of injection in Equation 2.6. Under this assumption, the concentration of fluorescent protein is expected to follow:

$$F(t) = \frac{A(e^{k_{mat}t}((k_{dm} - k_{mat})e^{k_{dm}t} + k_{mat}) - k_{dm}e^{k_{dm}t})}{k_{dm}(k_{dm} - k_{mat})e^{(k_{dm}+k_{mat})t}} + B(1 - e^{-k_{mat}t}) \tag{2.13}$$

By fitting this equation to the time course of fluorescence, we inferred the value of the reporter mRNA degradation rate in the embryo (Fig. 2.12 f-g) which results $k_{dm} = 0.028\text{min}^{-1}$.

We assume the rates to be independent from the position in the embryo and constant over the observed time scale.

2.4.10 Analysis of MS2 data

MS2 fluorescent spots were detected in 3D using the *Definiens XD 2.0* software package. We first identified the boundary of the embryo, similarly to what was described above. We then segmented in 3D the transcription foci exhibiting accumulated GFP fluorescence signal with the procedure given below. As there is considerable variation in background intensity, a precise identification of GFP-labeled foci with weak signal intensity is difficult. Therefore, application of a global threshold does not produce good segmentation. Instead, we developed a strategy based on background reduction. As a first step, we applied a 3DGaussian filter with a kernel size of 3x3x3 pixels (GaussianFilter1), then applied a second 3D-Gaussian filter, again with a kernel size of 5x5x3 pixels, (GaussianFilter2). We then subtracted GaussianFilter2 from Gaussianfilter1, which resulted in a background subtracted image. As a last step, we applied a global threshold and carried out segmentation using a so-called Multi-Threshold Segmentation algorithm implemented in the *Definiens* software platform. Briefly, this algorithm splits the image domain and classifies the resulting image objects based on a defined pixel value threshold. To reject particles resulting from segmentation errors, we filtered the resulting image objects by setting a threshold based on particle volume. Hence, with this procedure we avoided artifacts due to detector shot noise, fluctuations of GFP background, or embryo auto-fluorescence.

Following a previous study (Garcia et al., 2013), we calculated the MS2 signal as the integral of the fluorescence of each particle minus the local GFP background that was estimated from a spherical shell surrounding each detected spot. To compensate

for differences in imaging depth, this signal is further normalized by the average local background.

To calculate the expression profiles, the data were binned based on their AP position, by using 2% AP bins. We defined the total mRNA production rate in each bin as the sum of MS2 signal in the bin. This allowed for integrating both the information on the fraction of active nuclei in the area and on the intensity of transcription in each cell. Since the imaged portion of the embryo at different AP positions is not uniform (Fig. 2.4 a-b), we normalized the total mRNA production rate in each bin over the width of the embryo at the respective AP position. At last, we defined the total mRNA production as the integral over time of the mRNA production rate inside each bin.

Chapter 3

The effect of Hunchback binding sites in segmentation enhancers

3.1 Introduction

In the introductory chapter of this thesis, we reviewed many complex biochemical mechanisms relevant for the interactions between TFs and enhancers and for the effect of TFs on transcription. In particular, we have seen that the activity of some TFs turns out to be context dependent. Not surprisingly, this observation also holds true for some TFs involved in *D. melanogaster* segmentation. One particularly relevant example is given by Hunchback (Hb), a key regulator of many gap and segment polarity genes of the segmentation network, which is expressed, both zygotically and maternally, mostly in the anterior half of the embryo. Hb has been reported to act either as an activator or a repressor on different enhancers (Zuo et al., 1991)(Staller et al., 2015). Early reports postulated that Hb could become an activator instead of a repressor when bound to the enhancer in close proximity to Bicoid (Simpson-Brose et al., 1994). Other studies have instead suggested that different binding modalities of Hb (e.g., as a dimer vs. a monomer) could be a key determinant of its effect on gene expression (Papatsenko and Levine, 2008). Even when Hb is clearly acting as a repressor, for example in setting the anterior boundary of activity of various enhancers controlling the expression of gap or segment-polarity genes (Kruppel (Kr), Knirps (Kni) and Giant (Gt)), its repressive activity can show two different context dependent behaviors. While, for some of these enhancers, Hb is sufficient for repression and works in a simple concentration dependent manner, for other enhancers Hb only creates a permissive environment for the action of additional repressive factors (Yu and Small, 2008). A potential explanation is that Hb could be a short-range repressor, for which the arrangement and spacing of binding sites is critical (Payankaulam and Arnosti,

2008). Moreover, the ability of Hb to influence the activity of other TFs could be linked to its ability to recruit a chromatin remodeler such as dMi-2, which is one component of a chromatin remodeling and de-acetylation complex (Kehle et al., 1998).

Despite this relatively large set of observations, we haven't yet reached a satisfactory understanding of the role of Hb binding sites in segmentation enhancers and, in particular, how different features of the enhancer sequence can coordinate the different behaviors of Hb.

Synthetic biology could offer the potential to clarify the role of Hb binding sites in enhancers by measuring the activity of synthetic sequences carrying different arrangements of binding sites for Hb combined with binding sites for other activator TFs. The use of synthetic enhancers has proved useful in characterizing various properties of enhancer's architecture. For example, it made it possible to elucidate the role of cooperativity among closely spaced binding sites for the maternal activator Bicoid (Bcd) (Burz et al., 1998)(Hanes et al., 1994). However, failures in the attempt to reconstitute entire enhancers by combining transcription factor binding sites (Vincent et al., 2016) and the difficulties in interpreting the results of some synthetic enhancers experiments (Barr et al., 2017) suggested that the simple binding preferences of activating/repressing factors are not the sole determinant of enhancer activity. Multiple experimental hints (Lebrecht et al., 2005)(Yu and Small, 2008)(Thomas et al., 2011)(Bozek et al., 2019) clearly indicated that a more complex level of regulation involving the arrangement and spacing of binding sites as well as additional features of the DNA context sequence (the bases in between and surrounding the TF binding sites) also play a key role. Multiple biochemical mechanisms have been suggested to be involved in this process. Binding site occupancy can be promoted by cooperativity (Lebrecht et al., 2005), or hindered by competition with other factors (Small et al., 1996) and by the presence of nucleosomes. Enhancers have been found to overlap with nucleosomes depleted regions (Grossman et al., 2017)(Thomas et al., 2011)(Bozek et al., 2019), thus pointing to an important role of DNA accessibility as a prerequisite for the binding of input TFs to their cognate sites. Nucleosome binding is also influenced by the DNA sequence. For example, Poly-dA-dT sequences are considered to be the strongest nucleosome disfavoring motifs (Kaplan et al., 2009)(Segal and Widom, 2009) and they have been found to influence the activity of promoters, by reducing nucleosome occupancy in their vicinity (Raveh-Sadka et al., 2012).

Since not only the strength of the binding sites but also their arrangement and orientation can influence enhancer activity, any attempt at reconstituting a complex and long native enhancer requires to work with a large number of degrees of freedom in its sequence. Thus, studying synthetic variants of such complex enhancers might not be the easiest strategy to understand the rules underlying the organization of enhancer sequences.

In contrast, using fully synthetic enhancers with a controlled, small number of binding sites for only a few distinct TFs proved to be a more successful approach to investigate the organizational rules of enhancers sequences. For example, a fully synthetic enhancer system has been used to study how the combination of activators and repressors generate precise expression patterns (Crocker et al., 2017) and short synthetic enhancers made it possible to characterize various transcriptional repressors of the segmentation network as short range transcriptional repressors (Fakhouri et al., 2010)(Li and Arnosti, 2011).

To fully exploit the potential of synthetic enhancer constructs, it is necessary to precisely track their activity in space and time, with a system that is sensitive enough to measure both weak and strong enhancers and to reliably detect subtle quantitative differences. As we have seen in the previous chapters, various techniques to measure transcription have been developed over the past decades, each with different advantages as well as different limitations. Among these methods, only the MS2-MCP systems allows for tracking enhancer activity in real time *in-vivo*. However, as we have shown in Chapter 2, the sensitivity of this method is not high enough to observe the activity of weak synthetic enhancers. In this respect, the new reporter system we presented in Chapter 2 offers the potential to navigate an uncharted territory. In fact, this reporter system proved to be advantageous not only in terms of sensitivity but also with respect to throughput compared to previous methods. Therefore, it is ideally suited to quantitatively study the activity of multiple, short synthetic constructs *in-vivo*. The possibility to quantitatively study these simpler and shorter enhancers is a considerable advantage, since it simplifies the interpretation of the results compared with the study of natural sequences.

In this chapter, we elucidate the role of Hb binding sites by measuring the spatiotemporal dynamics of 20 synthetic enhancer sequences combining binding sites for Hb with those for different activators. First, following a similar experimental strategy as introduced by Fakhouri et al. (Fakhouri et al., 2010), we combined Hb binding sites with sites for two orthogonal activating transcription factors of the DV patterning system Twist (Twi) and Dorsal (Dl). Importantly, this design allows for simultaneously monitoring the enhancer activity both in the presence or absence of the Hb protein in a single embryo since Twi and Dl are present on the ventral side of the embryo from the anterior to the posterior tip while Hb is only present in the anterior half. This setting allows us to study the impact of binding site spacing and orientation on the enhancer's activity. In a second set of experiments, we exchanged the Twi and Dl binding sites with binding sites for a completely different TF, the anterior activator Bcd, and found similar effects of the Hb binding sites on expression. Our results reveal a dual role of Hb binding sites in shaping segmentation enhancers activity: on the one hand, when it is bound to its target sites Hb acts as a typical short range repressor; on the other hand, the sequence containing

multiple Hb binding sites, independently from Hb binding, enhances expression of both Twi-Dl and Bcd driven enhancers, which is a novel effect that we report. Since Hb binding sites coincide with Poly-dA stretches, an appealing explanation for the activating effect of this sequence is that it promotes a permissive environment for the enhancer activity by disfavoring nucleosome occupancy. Overall, the effect of Hb binding sites depends on a balance between the activating effect of the Hb binding site sequence and repression of Hb binding, with a net effect which we found to be always positive. Moreover, the distance dependencies of Hb repression and Hb binding sites activation are different, thus creating a strong non-linear behavior of the enhancer's activity as a function of the enhancer's architecture.

3.2 Results

3.2.1 Spatiotemporal characterization of enhancer activity.

We started by measuring the activity of a well-established 57bp-long synthetic construct derived from the *snail* proximal enhancer containing two binding sites for each of the D-V activators Twi and Dl (2Twi-2Dl; Fig. 3.1 A) (Szymanski and Levine, 1995). The activity of this sequence has been characterized using *in-situ* hybridization staining and exhibits homogenous activity in the embryo ventral side (Szymanski and Levine, 1995)(Fakhouri et al., 2010).

In order to detect subtle quantitative effects in the activity of different constructs we monitored their activity using the new mNeon reporter we described in Chapter 2. Briefly, the mNeon reporter system is based on the expression of an optimized reporter fluorescent protein. The fluorescence intensity is measured by means of live confocal fluorescence microscopy and a data analysis pipeline infers the information about mRNA levels by analyzing the time course of protein fluorescence with a model of ordinary differential equations as described in detail in the methods section of the previous chapter (see Section 2.4).

Since the surrounding sequence could play an important role for small enhancers, we first included in the enhancer 300bp of the flanking background sequence from the original reporter plasmid used in Fakhouri et al. (Fakhouri et al., 2010). However, when we analyze this flanking sequence using positional weight matrices representing the binding preferences of TFs of the segmentation network (Jung et al., 2018), we found multiple and strong predicted binding sites, in particular for Hb. All the observed binding sites are located just upstream of the 2Twi-2Dl enhancer (Fig. 3.1 A and arrow in Fig. D.6). Following precisely the same steps described in detail in Chapter 2, we monitored protein expression levels by confocal fluorescence microscopy (Fig. 3.1 B). We quantified the fluorescence signal in the cortical region of the embryo and reconstructed the information about mRNA levels. Similarly to what has been reported by other studies (Fakhouri et al., 2010), we measure a relatively weak and homogenous activity on the ventral side of the embryo for the 2Twi-2Dl enhancer (Fig. 3.1 B,C and D). However, our data show a somewhat stronger activity towards the embryo posterior, which may not have been evident in older *in-situ* stainings due to saturation.

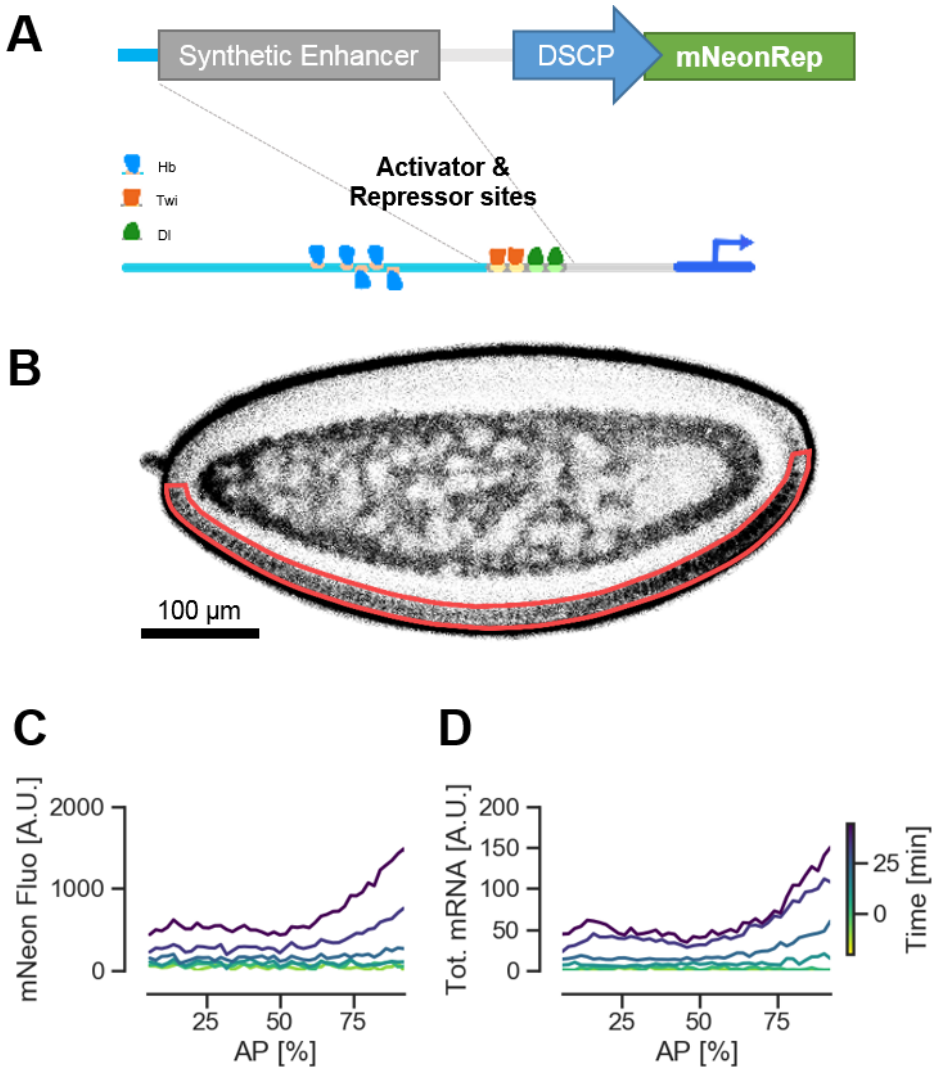

Figure 3.1: Quantifying synthetic enhancer activity with the mNeonRep reporter. A) Synthetic regulatory sequences, containing binding sites for selected transcription factors, are cloned upstream of a Drosophila Synthetic Core Promoter (DSCP) controlling the expression of the mNeon reporter. B) A representative confocal slice of a living embryo, laid by homozygous transgenic parents carrying the reporter construct (Original-2Twi-2Dl-mNeonRep), showing mNeon fluorescence in grayscale. The fluorescence of the mNeon reporter can be detected on the ventral side (lower half in the image) in the nuclei in the cortical region of the embryo. C) Quantification of the mNeon fluorescence pattern on the ventral side of the embryo as a function of the position along the AP axis. mNeon fluorescence has been quantified from three single confocal slices and averaged in 2% bins along the AP axis. D) Patterns of cumulative mRNA production reconstructed from the time course of mNeon fluorescence. In panel C and D each line corresponds to a different color-coded time of embryo development.

3.2.2 The effect of Hb binding sites in Twist-Dorsal driven enhancers.

The effect of Hb binding sites on enhancer activity has been mainly investigated in the context of complex natural enhancers like the eve enhancers (Struffi et al., 2011)(Vincent et al., 2018). In order to analyze their effect systematically, we started by designing synthetic enhancer constructs combining Hb binding sites with the 2Twi-2Dl enhancer. Moreover, in order to work in a clean setting free of putative binding sites for additional TFs, we introduced in all our constructs, immediately upstream of the enhancer, a 340bp neutral flanking sequence. This spacer sequence does not contain any predicted binding sites for all relevant TFs (as checked using PySite, a custom-written python script based on first order positional weight matrices (PWMs) (Jung et al., 2018)) (Fig. D.7) and has been demonstrated to not drive any expression *in-vivo* (Sayal et al., 2011). In addition to the 2Twi-2Dl enhancer having the neutral background sequence (2Twi-2Dl; Fig. 3.2 B), we generated 4 additional enhancers by introducing a sequence containing 3 functional binding sites for Hb (Simpson-Brose et al., 1994). In particular, we inserted the sequence containing the 3 Hb binding sites just upstream of the 2Twi-2Dl enhancer (3Hb-2Twi-2Dl; Fig. 3.2 C), or at increasing distances: 70bp (3Hb-70-2Twi-2Dl; Fig. 3.2 D), 150bp (3Hb-150-2Twi-2Dl; Fig. 3.2 E) and 250bp (3Hb-250-2Twi-2Dl ; Fig 3.2 F) away from the activator sites.

The results of this set of experiments were surprising: when exchanging the flanking sequence used in older studies with the neutral background, expression was substantially reduced to a barely detectable level (Fig. 3.2 A and B). Even more surprisingly, introducing 3 Hb sites in this new setting substantially increased enhancer activity in the entire embryo (Fig. 3.2 C and Fig. 3.3). Interestingly, the strong activating effect of Hb binding is not localized to the anterior half of the embryo where Hb is present and, therefore, cannot be attributed to Hb binding. Moreover, the expression pattern of 3Hb-2Twi-2Dl is not homogenous along the AP axis. In particular, the region where Hb is expressed (represented with a gray shading in Fig. 3.3) corresponds to a region of relatively weak expression, around 3 times weaker compared to the activity of the same enhancer in the posterior region. However, even in the anterior region, the overall balance of inserting 3Hb binding sites in the enhancer still results in an increased activity compared to the 2Twi-2Dl enhancer. What we observe in this region of the embryo could be attributed to an overlap of an activating effect from the presence of Hb binding sites and the repressive effect from Hb binding to them.

The expression pattern significantly varies when the distance between the 3 Hb and the activator sites is increased, with important differences between the anterior and posterior

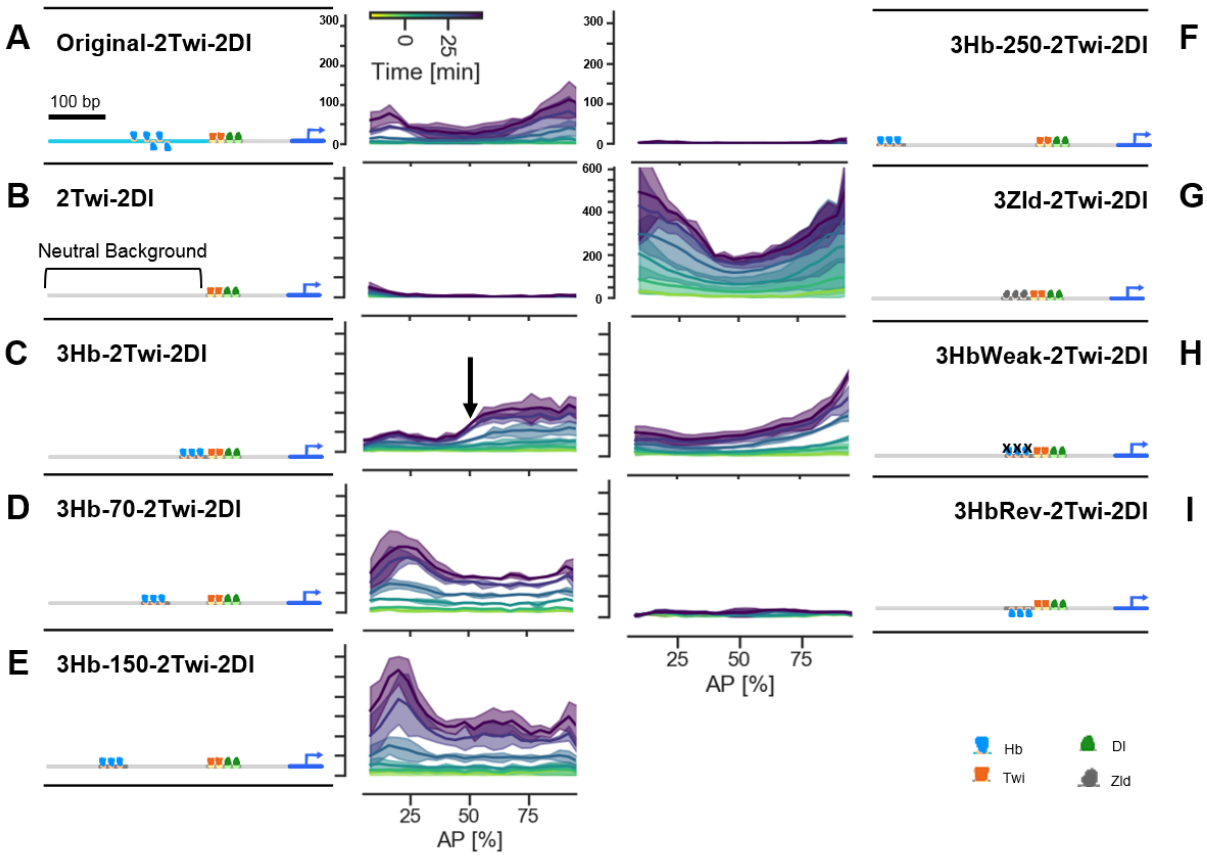

Figure 3.2: Sketches of the binding site structure of synthetic enhancers (A-I) driven by Twist and Dorsal and an overview of their expression patterns. Solid lines represent the average cumulative mRNA production from 2 or 3 embryos. The data are grouped into bins corresponding to 4% of the embryo length. The shaded areas represent $\pm 1\sigma$ confidence intervals. Expression profiles at different time points are represented in different colors. All data represent the reporter expression in the ventral side of the embryo.

halves of the embryo. In the posterior region, the activating effect of the Hb binding sites remains constant by increasing the distance from a few base pairs to 70 or 150bp (Fig. 3.2 D and E), while, at the distance of 250bp any activating effect is lost and the expression returns to the baseline level (Fig. 3.2 F). The anterior part of the embryo shows a very different behavior. The relative repression due to Hb binding is lost when the 3 Hb and Twi-Dl sites are separated by 70bp, and the enhancer activity becomes even stronger than

in the posterior. Anterior activity increases even further when the separation is increased to 150bp. Finally, also in this part of the embryo, when the distance is further increased to 250bp, expression returns to baseline levels. These observations in the embryo anterior, together with the overlap of the Hb expression domain with the region of relatively weak activity for the 3Hb-2Twi-2Dl enhancer, suggest that Hb acts as a typical short-range repressor while the sequence of Hb bindings sites foster enhancer activity.

To further investigate whether the relatively weaker activity of the 3Hb-2Twi-2Dl enhancer in the embryo anterior is indeed due to Hb binding, we introduced single point mutations in each of the 3Hb binding sites. In particular, we mutated the core of each Hb binding site switching A->T, thus leading to a 10 folds decrease in the predicted binding site strength, without affecting the enhancer GC content. The activity of this mutated construct (3HbWeak-2Twi-2Dl enhancer) does not show any clear drop with respect to the Hb expression domain (represented with a gray shading in Fig. 3.3). This finding further supports the idea that Hb binds to the 3Hb-2Twi-2Dl enhancer and represses its activity in the anterior half of the embryo.

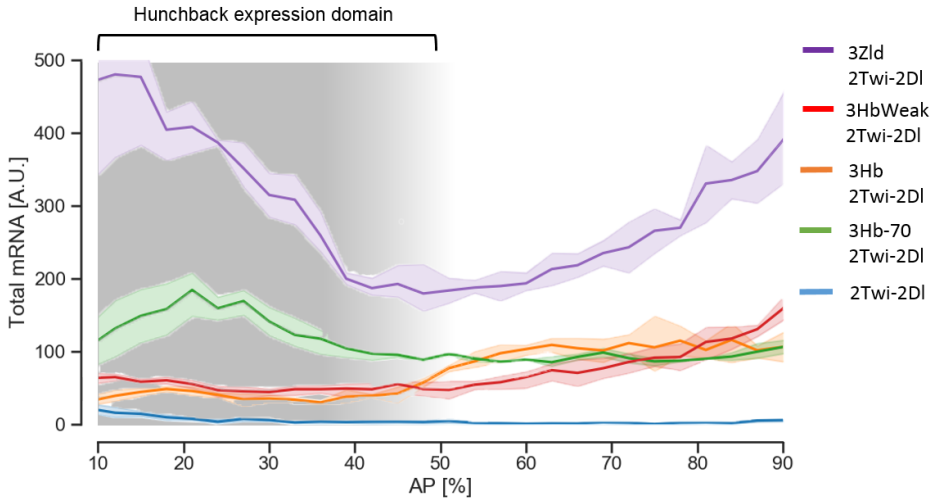

Figure 3.3: Comparison of the total cumulative mRNA production driven by a subset of enhancers. The gray shading represents the area of the embryo in which Hb is expressed.The shaded areas represent $\pm 1\sigma$ confidence intervals.

The interesting observation that a sequence containing 3Hb binding sites increases enhancer activity in the absence of Hb could indicate that it carries DNA features promoting a permissive state of the enhancer. This could be achieved, for example, by influencing the enhancer accessibility. In this respect, an additional interesting observation is that Hb binding preferences coincide with a stretch of As(Jung et al., 2018) and can therefore

be considered a poly-dA-dT sequence. Poly-dA-dT sequences are considered to be the strongest nucleosome disfavoring motifs (Kaplan et al., 2009)(Segal and Widom, 2009) and have been found to influence the activity of promoters, by reducing nucleosome occupancy in their vicinity (Raveh-Sadka et al., 2012). They also have been found to affect the enhancers activity by influencing the binding of TFs to nearby sites through their effect on DNA-shape (Levo et al., 2015). To look for additional evidence in this direction, we decided to test if the enhancer sequence with only Twi and Dl binding sites is in a relatively 'closed' state or if it is already highly accessible. To this end, we inserted upstream the 2Twi-2Dl enhancer 3 binding sites for the TF Zelda, which is known to increase enhancer accessibility (Liang et al., 2008)(Sun et al., 2015)t. Among the synthetic enhancers studied in this work,his construct drove the highest activity along the whole AP (3Zld-2Twi-2Dl Fig. 3.2 G) suggesting that indeed the 2Twi-2Dl enhancer weak activity is affected by a low accessibility.

Moreover, to investigate if the 3Hb sequence has the potential to significantly influence nucleosome occupancy, we looked at the predicted nucleosome occupancy based on *in-vitro* nucleosome sequence preferences as predicted by a model developed by Kaplan et al. (Kaplan et al., 2009). This model predicts a substantially different nucleosome occupancy landscape among the synthetic enhancers we studied (Fig. D.6, D.7, D.8). In particular, the 2Twi-2Dl enhancer containing the original flanking sequence, which was used in previous works (Li and Xie, 2011), has a low nucleosome occupancy (Fig. D.6 B). In contrast, the 2Twi-2Dl enhancer with a clean background sequence, which has a very low activity, has a high predicted nucleosome occupancy (Fig. D.7 B). Insertion of the 3Hb sequence into this enhancer, substantially reduces nucleosome occupancy (Fig. D.8 B), as expected given that Poly (dA) tracts have the lowest average nucleosome occupancy (Kaplan et al., 2009). To further support these observations, we measured the *in–vitro* nucleosome binding energy of three 150bp long DNA sequences, encompassing either the 2Twi-2Dl, 3Hb-2Twi-2Dl or 3HbWeak-2Twi-2Dl enhancers and part of their flanking sequence. To measure DNA-histone binding free energies we used a recently developed fluorescence anisotropy assay (Schnepf et al., 2020). Briefly, a robotic system was used to obtain competitive nucleosome formation of histones with a fluorescently labeled reference DNA sequence or a non-fluorescent competitor DNA sequence in a microwell plate. A modified epifluorescence microscope was used to measure fluorescence anisotropy in each well and derive the fraction of bound vs unbound DNA. Full titration curves were obtained by varying the concentration of the competitor sequence in different wells. Even though the differences of binding energies among these enhancers are not statistically significant, we could still observe a correlation between nucleosome binding energy and enhancer activity (see Fig. 3.4).

Finally, we also looked into the effect of the orientation of the Hb binding sites. It is generally accepted that enhancer activity is not sensitive to orientation. Contrary to what is expected, we observed a major impact of the orientation of Hb binding sites: reversing the orientation of the sequence containing 3Hb sites substantially reduced expression in the entire embryo (3HbRev-2Twi-2Dl; Fig 3.2 I).

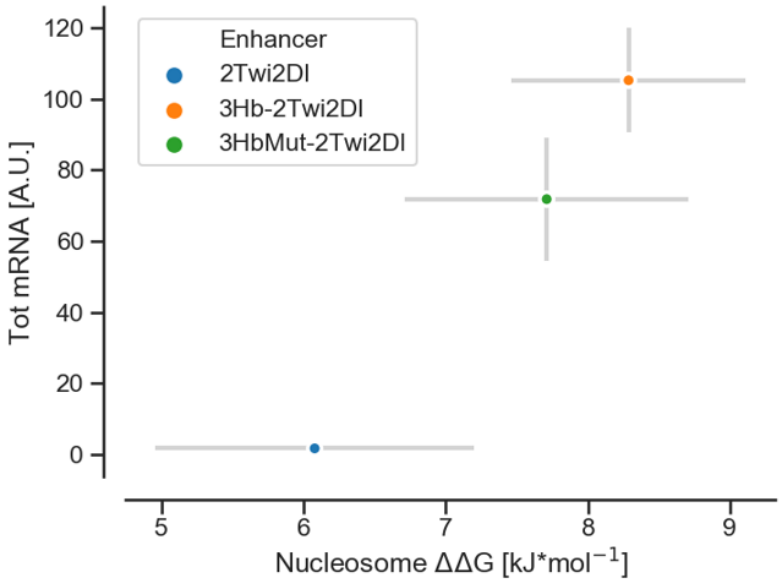

Figure 3.4: Correlation between cumulative mRNA production driven by the 2Twi-2Dl, 3Hb-2Twi-2Dl and 3HbWeak-2Tw-2Dl enhancers in the embyro posterior and the *in-vitro* nucleosome binding energy of their enhancer sequence measured with a fluorescence anisotropy based assay. The error bars in the binding energy measurements represent a standard error of the mean over, on average, three replicates.

3.2.3 The effect of Hb binding sites in Bicoid driven enhancers.

To test if the observed effect of Hb binding sites is consistent in different enhancers containing binding sites for other activators besides Twi and Dl, we designed a second group of enhancers. We chose to combine the same set of Hb binding sites with sites for the Bcd activator. This design is more relevant for the understanding of native enhancers, since Hb and Bcd often regulate the same enhancers. Moreover, this is the same setting in which Hb bifunctionality has been originally reported, leading to the postulation of a synergistic effect between these two factors (Barr et al., 2017) (Simpson-Brose et al., 1994). We expect the activity of these enhancers to be driven by Bcd binding. Unfortunately, since both Hb and Bcd are localized in the embryo anterior, in the same set of experiments it will not be possible to observe the effect of unoccupied Hb binding sites on Bcd dependent enhancer activation without perturbing the enhancer sequence to weaken the binding sites.

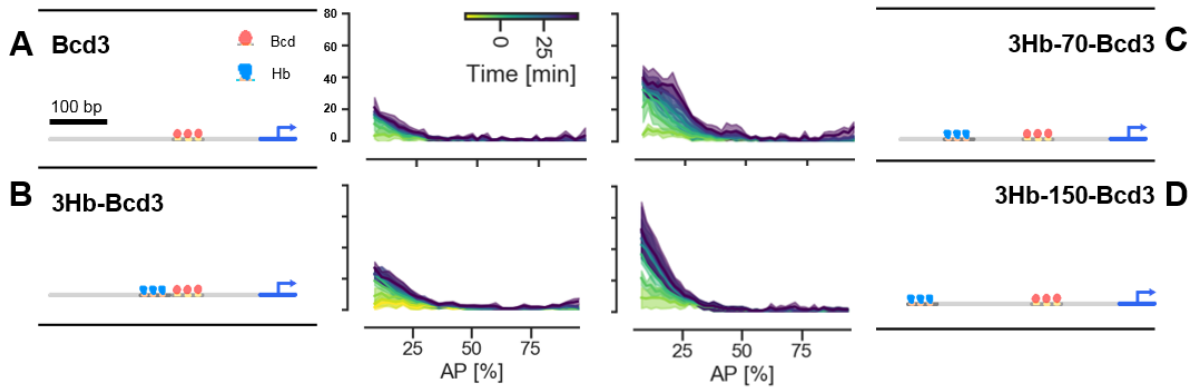

Figure 3.5: Hb binding sites similarly affect Bcd and Twi-Dl activity. Spatial-temporal dynamics of the activity of A) Bcd3 B) 3Hb-Bcd3 C) 3Hb-70-Bcd3 and D) 3Hb-150-Bcd3 synthetic enhancers. Solid lines represent the average cumulative mRNA production from 2 or 3 embryos grouped into bins corresponding to 4% of the embryo length. The shaded areas represent $\pm 1\sigma$ confidence intervals. All data represent the reporter expression in the ventral side of the embryo.

Similarly to the synthetic constructs driven by the Twi and Dl activators, we inserted the sequence containing 3 Hb binding sites at increasing distances from a sequence carrying 3 binding sites for Bcd. An enhancer with only 3 Bcd binding sites (Bcd3) drives

expression in the anterior most region of the embryo (Bcd3; Fig. 3.5 A), as previously reported by other studies (Ronchi et al., 1993). Introducing 3 Hb binding sites just upstream of the 3 Bcd sites slightly but significantly ($p = 0.0048$, two-tailed Welch's t-test) increased enhancer activity (3Hb-Bcd3; Fig. 3.5 B and Fig. 3.7 A). This is in agreement with our previous observation that, at short range, the balance between Hb binding sites activating effect and Hb binding repressing effect is shifted towards activation.

One study already looked at the activity of the Bcd3 and the 3Hb-Bcd3 enhancers (Simpson-Brose et al., 1994). This study reported an expansion towards the embryo posterior of the expression domain of the Bcd3 enhancer upon inserting 3 binding sites for Hb. Importantly, an expansion of the expression domain with a constant peak activity could support the idea that the enhancer reacts to a threshold in the concentration of its activator Bcd. Moreover, a change in this threshold following the inclusion of 3Hb sites in the enhancer could support the idea of a direct cooperative interaction of Hb and Bcd. In order to compare our results with this previous report, we checked if the insertion of the 3Hb sequence caused an expansion of the expression domain driven by the enhancer. In our data, including 3Hb sites increases expression homogeneously and does not substantially alter the shape of the expression domain and therefore its boundary (Fig. 3.6 A,B). This difference might be a result of the qualitative nature of older *in-situ* staining techniques which offered a high sensitivity but had a non-linear response that could easily cause saturation.

Similarly to what we observed for the Twi-Dl driven enhancers, activity increased substantially when we introduced a 70bp (3Hb-70-Bcd3; Fig. 3C) or 150bp (3Hb-150-Bcd3; Fig. 3D) long neutral spacer sequence between the Hb and Bcd sites. To characterize this distance dependence in both groups of enhancers, we looked at the fold change of the total amount of mRNA produced by each enhancer as a function of the distance between the Hb and Twi-Dl or Bcd binding sites. For the Bcd driven enhancers, we looked at a window spanning from 10% to 30% AP and we calculated the fold change in bins corresponding of 2% of the embryo length (Fig. 3.7 A). Similarly, for the Twi-Dl driven enhancers, we looked at two regions, one from 10% to 30% AP and the other from 70% to 90% AP corresponding to the presence or absence of the Hb protein, respectively (Fig. 3.7 B). Interestingly, the fold change in the enhancer's activity induced by the presence of Hb binding sites qualitatively shows a similar trend for both Bcd or Twi-Dl driven enhancers. In both cases, we observed an increase when introducing a spacer sequence of 70bp or 150bp, although the quantitative overall impact is different.

To investigate if inserting the 3Hb sequence can only affect the enhancer or can also directly influence the activity of the promoter, we looked again at the activity of the Bcd3 and 3Hb-150-Bcd3 enhancers but we created two new constructs in which we varied the

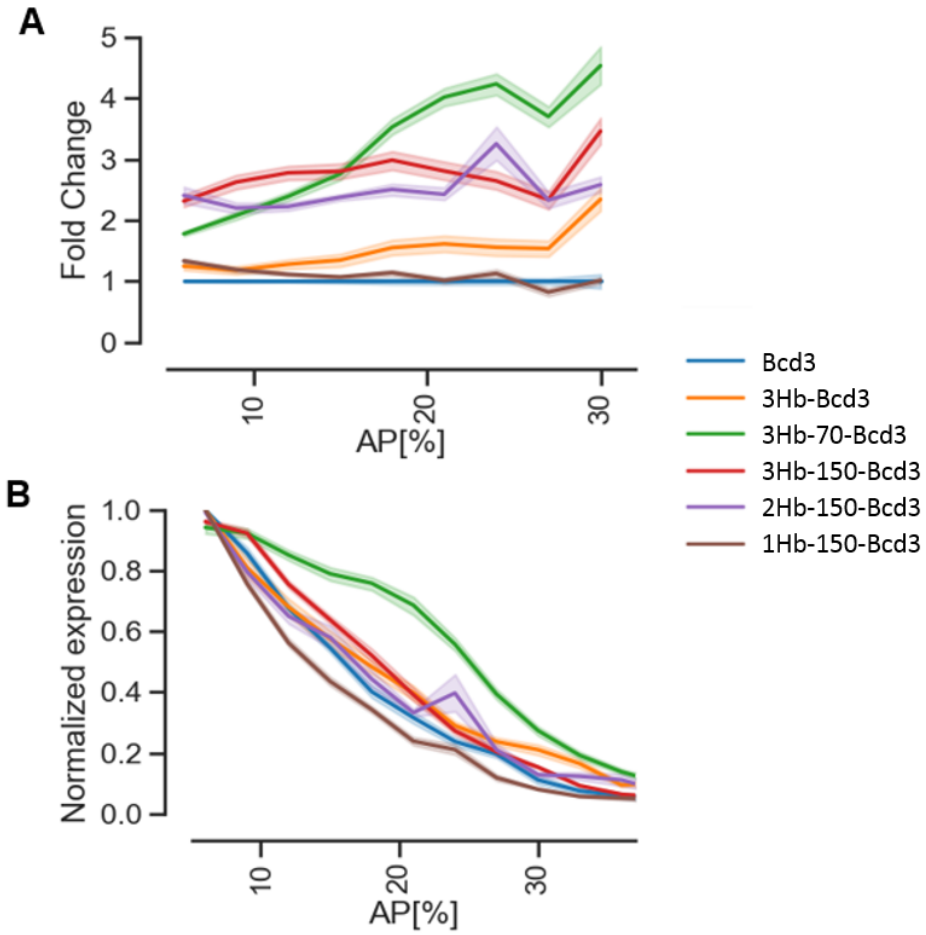

Figure 3.6: Effect of the 3Hb sequence on the shape of the expression profile of Bcd driven enhancers. A) Fold change of the cumulative mRNA production driven by the Bcd driven synthetic enhancers compared to the Bcd3 enhancer. The fold change has been computed at each position and all time points and averaged. The homogeneity of the fold change throughout the AP axis, with the exception of the 3Hb-70-Bcd3 enhancer, proves that the 3Hb sequence rescales the expression pattern without altering its shape. B) Normalized expression patterns of cumulative mRNA production driven by the Bcd driven synthetic enhancers. The expression patterns for each enhancer have been normalized to their spatial maximum at each time point and then averaged.

enhancer-promoter distance. As we have seen in Chapter 2, enhancer-promoter distance can significantly influence the expression level driven by an enhancer. However, if the 3Hb sequence does not directly influence the promoter, we expect that it will influence enhancer activity by the same relative amount for different enhancer-promoter distances.

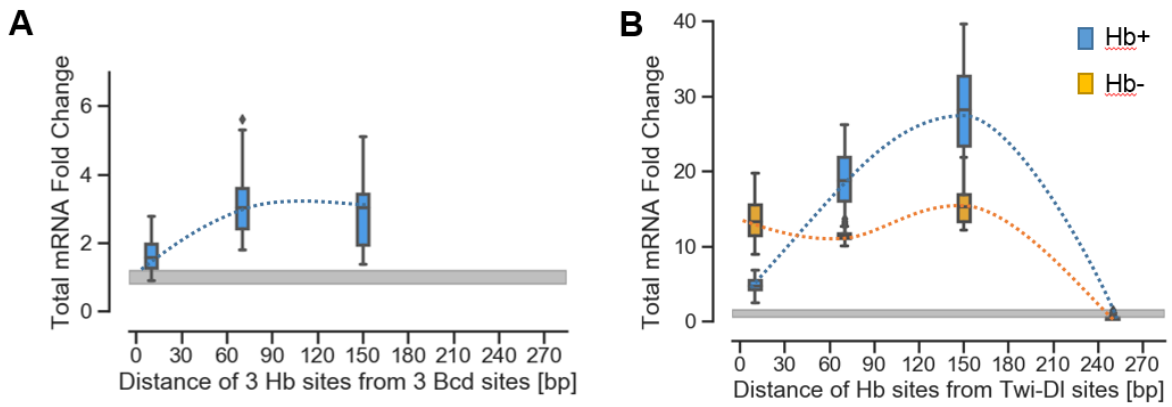

Figure 3.7: Effect of the 3Hb sequence at increasing distance from activator sites. A) Fold change of the total cumulative mRNA production in the embryo anterior for the enhancers with 3 Hb binding sites at various distances from the 3 Bcd binding sites, compared to the Bcd3 enhancer. The fold change has been computed independently in bins corresponding to 2% AP and the data have been then pooled together in the region 10-30% AP. The gray area illustrates the variability in the total cumulative mRNA production level for the Bcd3 enhancer and represents an interval of $\pm 1\sigma$. B) The fold change of the total cumulative mRNA production in the embryo anterior (Hb+) and posterior (Hb-) for the enhancers with 3 Hb binding sites at various distances from the 2Twi and 2Dl binding sites, compared to the 2Twi-2Dl enhancer. The fold change has been computed independently in bins corresponding to 2% AP and data have been then pooled together in the regions of 10-30% AP and 70-90% AP. The gray area illustrates the variability in the total cumulative mRNA production level for the 2Twi-2Dl enhancer and represents an interval of $\pm 1\sigma$.

This is indeed what we observed: although the absolute amount of mRNA production is increased when the enhancer is closer to the promoter (compare Fig.3.8 A-B and Fig. 3.5 A-D), the fold change in activity due to the insertion of the 3Hb sequence at 150bp from the Bcd binding sites is the same in both cases (Fig. 3.8 C).

Using the same setting, we also explored the effect of changing the number, strength and orientation of Hb binding sites on activity. Removing one Hb binding site from the 3Hb-150-Bcd3 enhancer only slightly reduced the enhancer's activity, which remained sustained (Fig. 3.9 A,B and G). However, removing a second Hb binding site reduced the activity to a level compatible with the baseline level of the Bcd3 enhancer (Fig. 3.9 C

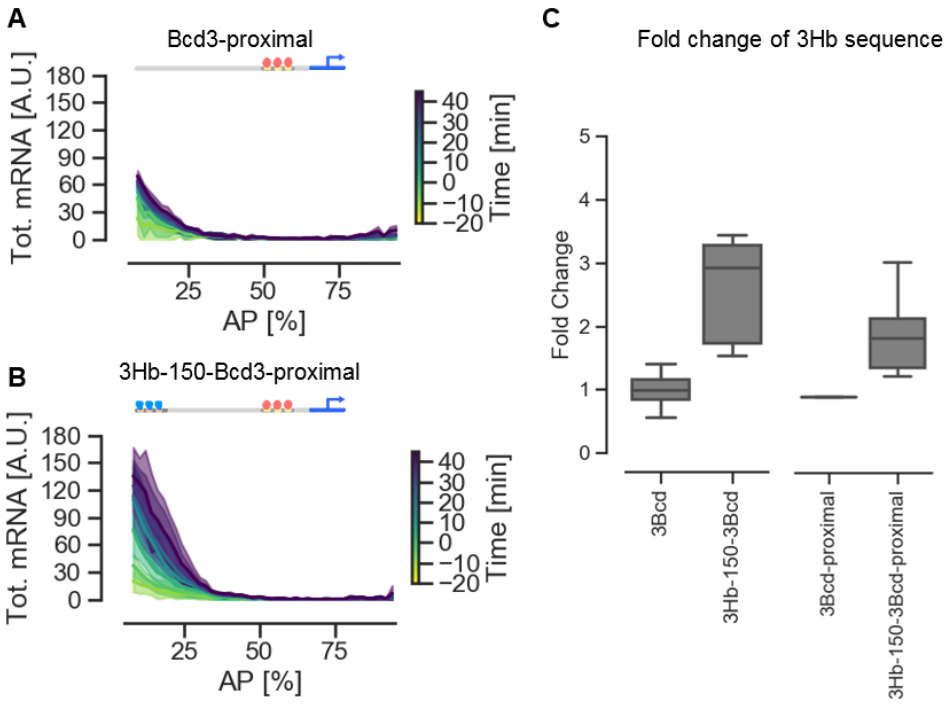

Figure 3.8: Effect of the 3Hb sequence at different distances from the promoter. A) The cumulative mRNA production driven by the Bcd3 enhancer positioned just upstream of the DSCP promoter (Bcd3-proximal). B) The cumulative mRNA production driven by the 3Hb-150-Bcd3 with the whole enhancer positioned just upstream of the DSCP promoter (3Hb-150-Bcd3-proximal). C) The 3Hb sequence at 150bp from 3Bcd binding sites induces the same increase in expression level when the enhancer is positioned just upstream of the promoter or further away.

and G). Single point mutations in all three Hb binding sites further increased the activity, revealing that, even at 150bp away, Hb binding could still have some residual repressing effect (Fig. 3.9 D). Reversing the orientation of 3Hb binding sites proved once again to have a severe impact on expression, although less pronounced than in the case of the Twi-Dl activators. At 150bp from the activator sites, the reversed 3Hb sequence was still able to increase expression compared to the baseline level, but the activity was only half of that obtained in the forward orientation (Fig. 3.9 E-G). Similarly, when we inserted the inverted Hb sites just upstream of the Bcd binding sites, expression was also reduced and turned out to be comparable to that of the Bcd3 sequence alone (Fig. 3.9 F,G).

Experiments with synthetic or reconstituted enhancers have often proven difficult to interpret. It is therefore remarkable that, in this study, we find consistent effects when

including the 3Hb sequence into synthetic enhancers based on different activators. In particular, in the anterior part of the embryo, the inclusion in the enhancer of multiple Hb binding sites have, in general the same activating effect on the Twi-Dl and Bcd activators. Our results suggest that this activating input stems from two opposing mechanisms: short-range repression due to Hb binding to its cognate sites and the activating effect of the Hb binding sites sequence, which is possibly due to an increase in enhancer accessibility.

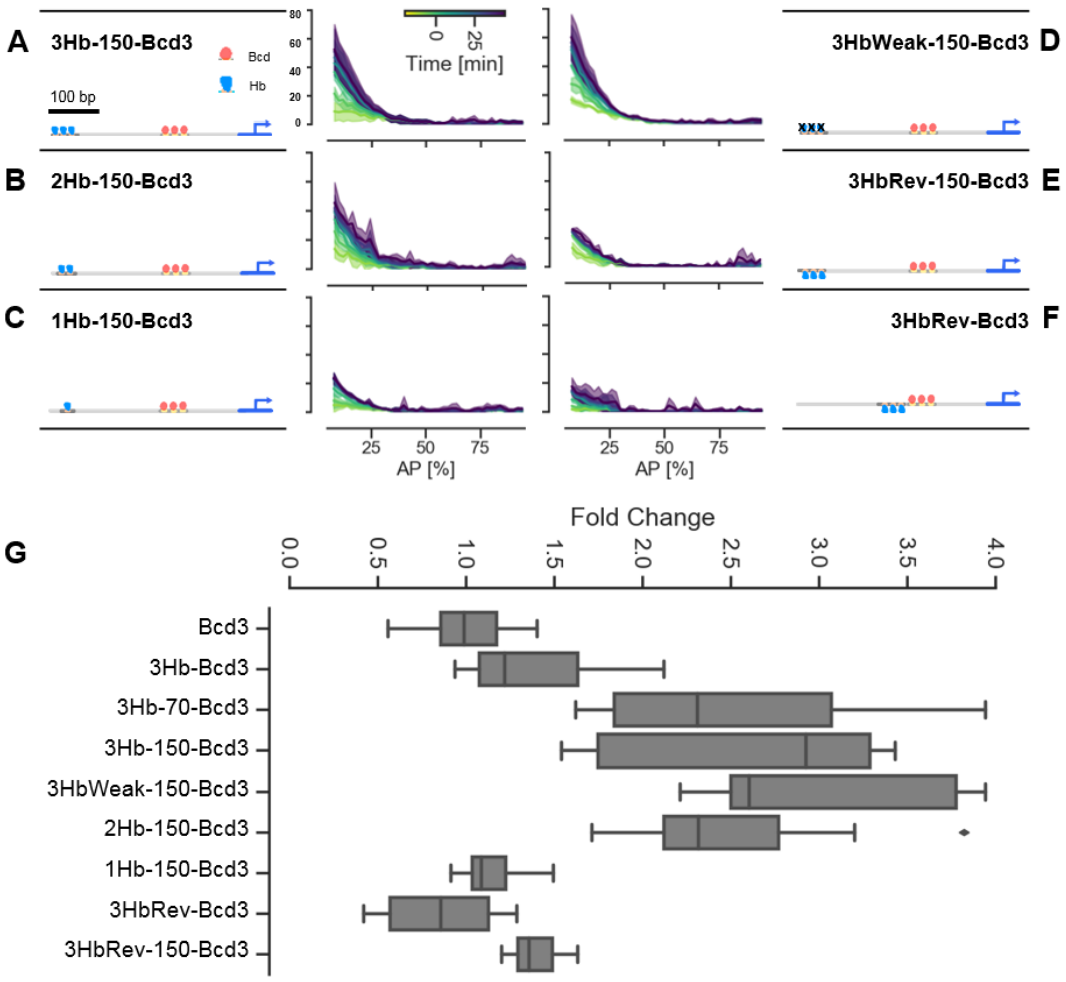

Figure 3.9: Effect of stoichiometry and orientation of Hb binding sites on Bicoid dependent activity. Spatiotemporal dynamics of the activity of A-C) synthetic enhancers containing 3, 2 or 1 binding sites for Hb at 150bp from the 3 Bcd sites. D) Activity of a synthetic enhancer containing 3 weaker binding sites for Hb obtained with point mutations of the consensus binding motif. E-F) Synthetic enhancers containing 3 binding sites for Hb in the reverse orientation either right upstream of, or at 150bp from, the 3 Bcd sites. Solid lines represent the average cumulative mRNA production from 2 or 3 embryos grouped in bins corresponding to 4% of the embryo length. The shaded areas represent $\pm 1\sigma$ confidence intervals. All data represent the reporter expression in the ventral side of the embryo. G) The fold change of the total cumulative mRNA production in the embryo anterior for all Bcd-driven synthetic enhancers. The fold change has been computed independently in bins corresponding to 2%AP and the data have been then pooled together in the region 10-30%AP.

3.2.4 Synthetic enhancers dynamics.

Finally, we took advantage of the time resolution of the mNeon reporter system to investigate the temporal dynamics of the synthetic enhancers. To better highlight the time at which each enhancer is active, we now look at the instantaneous rate of mRNA production as opposed to the cumulative mRNA production. Our data clearly show that Twi and Dl driven enhancers have a similar dynamics in the embryo anterior (Fig. 3.10 A) and posterior (Fig. 3.10 B). Moreover, they are active later compared with the Bcd driven synthetic enhancers (Fig. 3.10 C). The observed dynamics is consistent with the dynamics of the concentration of input TFs previously reported in literature. While Twi and Dl concentrations rise through early embryo development and reach a peak at the end of n.c. 14 (Liberman et al., 2009)(Sandler and Stathopoulos, 2016), the concentration of Bcd instead declines (Gregor et al., 2007b), and its activity is further decreased by Bcd sumoylation (J. and Ma, 2012). Importantly, we couldn‘t observe any consistent effect on the temporal dynamics caused by the insertion of Hb binding sites in any of the constructs under study. The only construct that shows a marked difference in its dynamics is the 3Zld-2Twi2Dl enhancer, which is active much earlier than all other Twi and Dl driven enhancers (Fig. 3.10 A and B). This observation also agrees with a study in which Zld binding sites were added to the snail enhancer, substantially accelerating enhancer activation (Yamada et al., 2019).

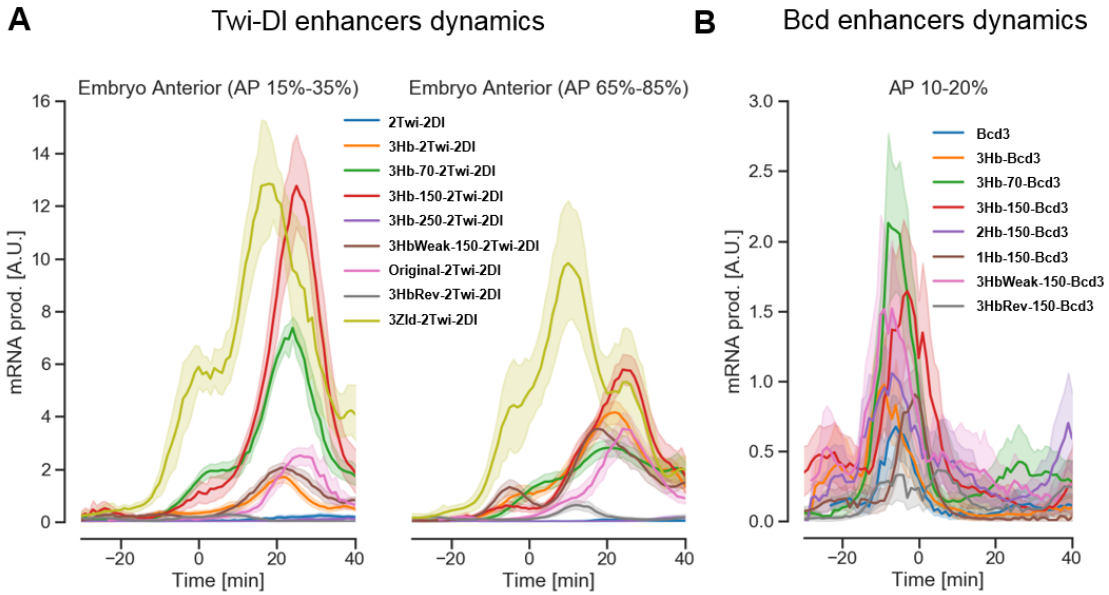

Figure 3.10: Temporal dynamics of synthetic enhancers expression. The average rate of instantaneous mRNA production reconstructed from the time course of protein fluorescence for all the enhancers in this study. The shaded areas represent $\pm 1\sigma$ confidence intervals. A) The temporal dynamics of enhancer activity for the Twi-Dl driven enhancers in the embryo anterior (left) and posterior (right), respectively. B) The temporal dynamics of Bcd driven enhancers in the embryo anterior.

3.3 Discussion

Interpreting the regulatory code of transcriptional enhancers requires an understanding of how multiple transcription factors and the chromatin mutually interact to modulate gene expression. This process is ultimately orchestrated by the DNA sequence of the enhancers. Specifically, it is driven by the architecture of the binding sites they contain and by the context sequence, which can influence their activity. In the second part of this thesis, we focused on Hb, a context dependent transcription factor active in *D. melanogaster* embryo segmentation. We quantitatively measured the transcriptional activity of 20 synthetic enhancers combining Hb binding sites with sites for either the dorso-ventral activators Twi and Dl, or the anterior activator Bcd. We found that (i) a sequence containing 3 Hb binding sites is able, independently from the presence of the Hb protein, to create a permissive environment for enhancer activity, strengthening the activity of other activators; (ii) when Hb binding sites are at close distance from the activator sites, Hb binding causes a relative repression of enhancer' activity. However, the overall balance of inserting the 3Hb sequence still leans towards activation; (iii) the 3Hb sequence acts consistently in both Twi and Dl, and Bcd driven enhancers and (iv) it influences enhancer activity levels without altering their temporal dynamics.

The expression patterns of synthetic and reconstituted enhancers have often been difficult to interpret (Vincent et al., 2016)(Barr et al., 2017), although with some notable exceptions (Fakhouri et al., 2010). Here, we show that focusing on shorter and simpler synthetic enhancers and using a quantitative and sensitive reporter to measure enhancer activity can help to extract valuable information from experiments based on synthetic enhancers. In fact, most of the differences we observed, in particular in the case of Bcd driven enhancers, are subtle quantitative effects on the expression of rather weak enhancers. These effects may not have been resolved without the use of a reporter for enhancer activity that is both quantitative and sensitive, which is therefore pivotal for this kind of study. Our findings also highlight once again that every study on synthetic enhancers should take great care in the selection of the flanking and the background sequences. We feel that the flanking sequences in particular might have been overlooked in older studies. While their impact is probably limited when studying long native enhancers, the flanking sequences can strongly influence enhancer activity of shorter enhancers. The importance of the flanking sequences for short enhancers is well illustrated by the comparison of the expression of the 2Twi-2Dl and 2Twi-2Dl-Original enhancers (see Fig. 3.2 A and B). The 2Twi-2Dl-Original enhancer included 300bp of the sequence that is flanking the 2Twi-2Dl enhancer in the reporter construct used in previous studies. This sequence contains multiple binding sites for Hb and increases enhancer activity. Importantly, this

sequence is part of a plasmid backbone that has been extensively used to generate reporter constructs to study enhancer expression (Kvon et al., 2014).

Several studies identified Hb as a bifunctional TF: generally acting as a transcriptional repressor but able to switch to an activator in specific contexts. These studies, however, did not reach an agreement regarding the mechanism driving this behavior. Some postulated that Hb bifunctionality might be driven by the formation of Hb dimers (Papatsenko and Levine, 2008), similarly to what is known for other TFs. Others suggested that Hb bifunctionality would require specific protein-protein interactions with additional factors, for example when Hb binds to the enhancer in close proximity to a binding site for Bcd (Simpson-Brose et al., 1994). Our data investigated this landscape more closely, providing us with a more complete picture of the role of Hb binding sites in segmentation enhancers. In particular, we report a novel effect of Hb binding sites which are, independently from Hb binding, able to influence enhancer activity. Specifically, we observed that a sequence containing 3Hb binding sites is able to substantially increase expression of enhancers in the embryo posterior, where Hb is only transiently expressed in a small domain and only late in blastoderm development. Moreover, the 3Hb sequence is not able to drive expression by itself (Simpson-Brose et al., 1994), but it rather boosts the activity of other transcriptional activators. This is further confirmed by the observation that the activity of the synthetic enhancers studied here is always limited to the expression domains of the activating factors driving their expression.

An appealing explanation for the direct effect of the 3Hb sequence is that it could directly influence nucleosome occupancy, thus promoting the activity of transcriptional activators through an increase of DNA-accessibility. We can summarize the evidence in support of this model in three main points: (i) the 3Hb sequence has the potential to significantly influence nucleosome occupancy. In fact, the predicted nucleosome occupancy of our synthetic enhancers is substantially reduced when the 3Hb sequence is included (Fig. D.6, D.7, D.8). (ii) The in-vitro nucleosome binding energy of three 150bp long DNA sequences, encompassing either the 2Twi-2Dl, 3Hb-2Twi-2Dl or 3HbWeak-2Twi-2Dl enhancers and part of their flanking sequence, correlates with enhancer activity, although the observed differences in binding energies observed among these sequences are not statistically significant. (iii) The distance dependence of the activating effect of the 3Hb sequence is compatible with what has been reported for the influence of Poly(dA:dT) tracts in promoters. The 3Hb sequence is able to influence expression even when positioned at 70 or 150bp away from the activator sites; it becomes ineffective only when the spacing is increased to 250bp. The observed distance dependence corresponds with that observed for the influence of Poly(dA:dT) tracts in the activity of yeast promoters (Raveh-Sadka et al., 2012), an effect that has been shown to be mediated by changes in

nucleosome occupancy and also to be sensitive to sequence orientation (Lubliner et al., 2015). Therefore, this behavior further supports the idea of Hb binding sites influencing enhancer activity by altering the enhancer accessibility. The effect of Poly(dA:dT) tracts has been already studied in the context of enhancers, in particular regarding their effect as flanking sequences for TF binding sites (Levo et al., 2015). Poly(dA:dT) can, in fact, influence TF binding by their influence on DNA-shape. However, this effect is only observed if the Poly(dA:dT) tract is immediately adjacent to the binding sites, which is not the case in our synthetic constructs.

Our analysis of synthetic enhancers also confirmed the repressive activity of Hb, which we characterize as a short range repressor (Fig. 3.2 K). However, even if Hb binding can cause a substantial reduction in the enhancer activity, the net effect of including Hb binding sites is anyway shifted towards an increase in enhancer activity. The observation that Hb acts as short-range repressor is not surprising. In fact, short range repressors cause local histone deacetylation (Li and Xie, 2011), which reduce DNA-accessibility in a region corresponding to roughly the size of one nucleosome (Fakhouri et al., 2010). Moreover, it has been known for a while that Hb can interact with the chromatin remodeler and deacetylation complex NURD (Kehle et al., 1998), thus already suggesting that Hb could act as a short-range repressor (Payankaulam and Arnosti, 2008).

Overall, the simplest model to explain our observations assumes that Hb binding sites are able to increase enhancer activity by directly disfavoring nucleosome occupancy and thus increasing the enhancer accessibility. When these binding sites are occupied by Hb, expression is relatively reduced because of Hb activity as a short range repressor. However, the net balance between these effects still leans towards activation. Remarkably, we found the effect of the 3Hb sequence to be consistent when combined with two different groups of transcriptional activators. The fact that the effect of Hb seems to be independent from which other activator targets the same enhancer, although limited to the cases of Bcd and Twi and Dl, is an important observation: it suggests that these processes are not driven by protein-protein interactions among TFs, a common mechanism of context dependent activity that was already proposed to explain Hb activity (Staller et al., 2015) (Simpson-Brose et al., 1994)(Papatsenko and Levine, 2008). Our observation of the action of Hb in Bcd driven enhancers allows us to think of the Hb-Bcd interaction as a simpler and more general effect due to Hb binding sites, instead of a direct interaction between the two proteins.

An alternative model to explain our results would require the presence of unintended binding sites for additional TFs in the 3Hb sequence. If present, these binding sites would have to be recognized by an ubiquitous activator like Zld or D-STAT. However no strong binding sites for these or other factors involved in the A-P or D-V axis segmentation

could be found in the 3Hb sequence (Fig. D.8). Therefore, one would have to postulate the existence of an additional ubiquitous factor recognizing a motif in the 3Hb sequence and synergistically increasing the effect of other activators, while not being able to drive any expression by itself. In addition, the fact that including the 3Hb sequence does not influence enhancer dynamics but only expression levels, would also imply that this factor is constantly expressed at the same concentration or in excess. Moreover, if sites for a general activator are present, one would expect them to act independently of the binding sites orientation, which is not the case in our dataset. The surprisingly strong effect of the orientation of the 3Hb sequence is an aspect that requires further investigation and cannot be well captured in our model. Although the effect of Poly(dA:dT) sequences in promoters has been found to be sensitive to sequence orientation, this effect was only weak (Lubliner et al., 2015).

In summary, we believe that an interpretation of these results based on a direct effect of Hb binding sites on nucleosome occupancy offers a simpler explanation for our observations. However, a direct proof of this mechanism would require further investigations based on different methods, in order to directly measure the impact of the 3Hb sequence as well as other sequences on both nucleosome occupancy and expression *in vivo*. Moreover, from a technical point of view, the quantitative analysis of the activity of 20 synthetic enhancers further demonstrates the potential of the mNeon reporter in studying enhancers activity. The advantages of this method in both sensitivity and throughout allowed us to precisely quantify the activity of short synthetic enhancers. As the work presented in this chapter has shown, this turns out to be a major advantage, since measuring the activity of carefully designed sequences provided critical information for understanding specific mechanisms of transcriptional regulation.

Conclusions

This thesis aimed at investigating transcriptional regulation in the context of the development of *D. melanogaster* embryos. We developed a new method to quantitatively track enhancer activity in living *D. melanogaster* embryos based on a new fluorescent protein reporter and the analysis of the timecourse of its fluorescence. We employed this new method to study various aspects of transcriptional regulation in the context of *D. melanogaster* development by measuring the activity of synthetic enhancers, with a special focus on the bi-functional role of binding sites for the transcription factor Hunchback.

Currently, there are only a few approaches available for measuring the activity of enhancers in both fixed and living *D. melanogaster* embryos, with distinct advantages and limitations: *in-situ* hybridization stainings offers the highest sensitivity, but they are laborious and rely on the staining of fixed embryos, and therefore cannot measure transcriptional dynamics. In contrast, the MS2-MCP system is an *in-vivo* mRNA labeling technique that captures directly the temporal dynamics of enhancers' activity in living cells. However, it relies on the detection of nascent transcripts as fluorescent spots over a fluorescent background of unbound MCP-GFP molecules, which dramatically impairs signal-to-noise ratios and, as a consequence, limits the sensitivity of detection that can be reached with this system. Moreover, the high resolution required to detect the fluorescent spots requires the use of high numerical aperture objectives, thus limiting the field of view to a small portion of the embryo and practically reducing experimental throughput. In conclusion, the demand is high for a quantitative and sensitive method that offers higher throughput and, at the same time, high enough resolution to track enhancer activity in space and time.

In the second chapter of this thesis, we introduced a new method that overcomes these various limitations. We developed an optimized version of the bright and fast-maturing fluorescent protein mNeonGreen as a real-time, quantitative reporter of gene expression. We derived enhancer activities and mRNA concentrations from the dynamics of reporter fluorescence with high spatial and temporal resolution. By comparing our results with data obtained with the MS2-MCP system, we demonstrated the higher detection sensitiv-

ity of our reporter. Additionally, we demonstrated that our new reporter system can be used to study how the architectural features of an enhancer affect its activity, thanks to the reporter's ability to detect subtle quantitative differences in the activity of synthetic enhancer sequences. This proved that our new method is perfectly suited to measure the activity of a larger number of synthetic enhancers compared to previous techniques, a technical advantage that was instrumental in developing the work presented in the third chapter of this thesis.

In Chapter 3, we applied our new reporter to measure the activity of 20 synthetic enhancers designed to gain new insights into the effect of binding sites for Hb on enhancer activity. In particular, we combined a sequence of 3Hb binding sites with binding sites for two groups of activators, varying binding site spacing, strength and orientation. We observed a novel effect of the 3Hb sequence, which is able, independently from the presence of the Hb protein, to create a permissive environment for enhancer activity and thus strengthen the activity of other activators; Moreover, we observed that, when Hb binding sites are at close distance from the activator sites, Hb binding causes relative repression of the enhancer's activity. However, the overall balance of inserting the 3Hb sequence still leans towards activation. These results allowed us to interpret the bifunctionality of the Hb binding sites in segmentation enhancers as the result of the opposing effects of the short-range repression due to Hb binding to its cognate sites and the activating effect of the Hb binding site sequence. Since Hb binding sites coincide with Poly-dA sequences, the simplest explanation for the observed activating effect could be a direct influence of Poly-dA sequences on nucleosome occupancy, increasing the enhancer accessibility.

In a broader context, this work is a small step forward in the larger effort to quantitatively understand how the activity of enhancers and some of the features of their sequence are linked together. From a methodological and technical point of view, we believe that the mNeonGreen reporter system presented in this thesis is a powerful new tool that could be leveraged to obtain larger quantitative datasets on the enhancer activity of synthetic sequences. Our results on synthetic enhancers highlighted that a careful design of enhancer sequences from scratch can provide critical information for understanding specific mechanisms of transcriptional regulation. Possibly, this approach could be more efficient than alternative approaches based on the mutational analysis of complex natural enhancers or on the synthetic evolution of enhancer sequences. Our results also highlighted once more that special care needs to be taken in selecting background and flanking sequences of synthetic enhancer to avoid confounding effects due to the presence of accidental binding sites.

Finally, we believe that our data will also be an important resource for modeling studies. In fact, modeling enhancer activity has been complicated by the fact that most of the

available enhancer activity data in *D. melanogaster* embryos only offer relative expression patterns. These data characterize the regions of the embryo in which an enhancer is mostly active, but not how strong this activity is compared with other enhancers. Only a few studies have quantitatively compared the activity of a limited number of multiple enhancers. Moreover, combining data from these studies is very challenging, if possible at all, because reporter constructs are not standardized and, for example, use different promoters. Our dataset is among the largest quantitative and spatiotemporally resolved dataset of enhancer activity in *D. melanogaster* embryos, and is also the first dataset comparing the activities of natural and synthetic enhancers, highlighting that short synthetic sequences with few activator binding sites drive much lower expression levels compared to their natural counterpart.

In future studies, an interesting challenge would be to include our data in the training set of existing sequence-to-expression models and verify that the information we can learn from synthetic constructs is relevant for predicting the activity of natural sequences. Moreover, the simplest explanation for our data on the activating effect of Hb binding sites, as a direct effect of Poly-dA sequences on accessibility, would imply that nucleosome sequence preferences are relevant for the prediction of enhancers activity. Future studies should investigate this aspect in greater detail, both to directly demonstrate that the nucleosome sequence preferences have an impact for enhancer accessibility *in-vivo* and to evaluate the impact of including this additional information in sequence-to-expression models.

Appendices

Appendix A

Simulations of the data analysis pipeline

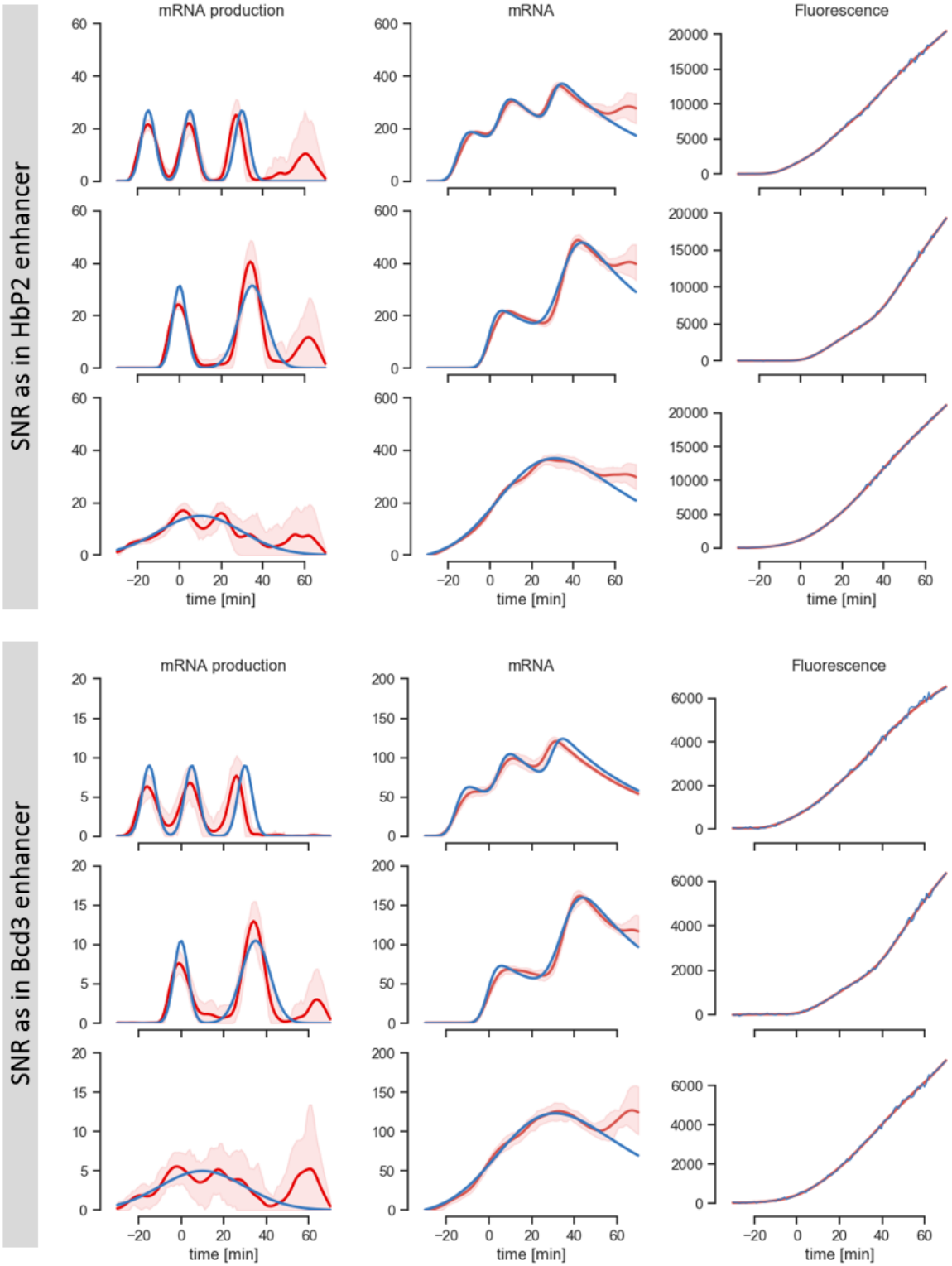

Figure A.1: (Caption on the following page.)

Figure A.1: Robustness tests of the mRNA reconstruction analysis on simulated data. The leftmost column represents instantaneous mRNA production, the middle column represents mRNA concentration and the right column represents protein fluorescence. In each plot the blue lines represent simulated data and the red lines are the result of the reconstruction algorithm. To simulate the data we consider three different scenarios for the dynamics of mRNA production with three short bursts (top row), two bursts (second row) and one single large burst (third row). Then, we simulate the total mRNA concentration and the protein fluorescence accordingly. We add noise to the protein fluorescence with the same amplitude we observed in our actual experimental data for the hb_ant enhancer (first block) and Bcd3 enhancer (second block), using the relation obtained from the data presented in Fig. 2.9. The mRNA reconstruction algorithm is then used to reconstruct the total amount of mRNA and the rate of mRNA production rates by fitting the noisy fluorescence data. The result of the reconstruction and corresponding 90% confidence intervals are plotted in red in each panel. The reconstruction algorithm is able to accurately capture the dynamics of mRNA production both in the case of high and low signal to noise levels. Generally, the reconstruction of instantaneous mRNA production is noisier than the total mRNA level, since rapid fluctuations in the rate of mRNA production are smoothed at the protein level. Note that the reconstruction is not reliable towards the end of the time series. This is due to the fact that the reconstruction of mRNA production at a given time is dependent only on the protein fluorescence at later times. At the end of the time series, there are simply fewer data points, leading to uncertainty in the reconstruction.

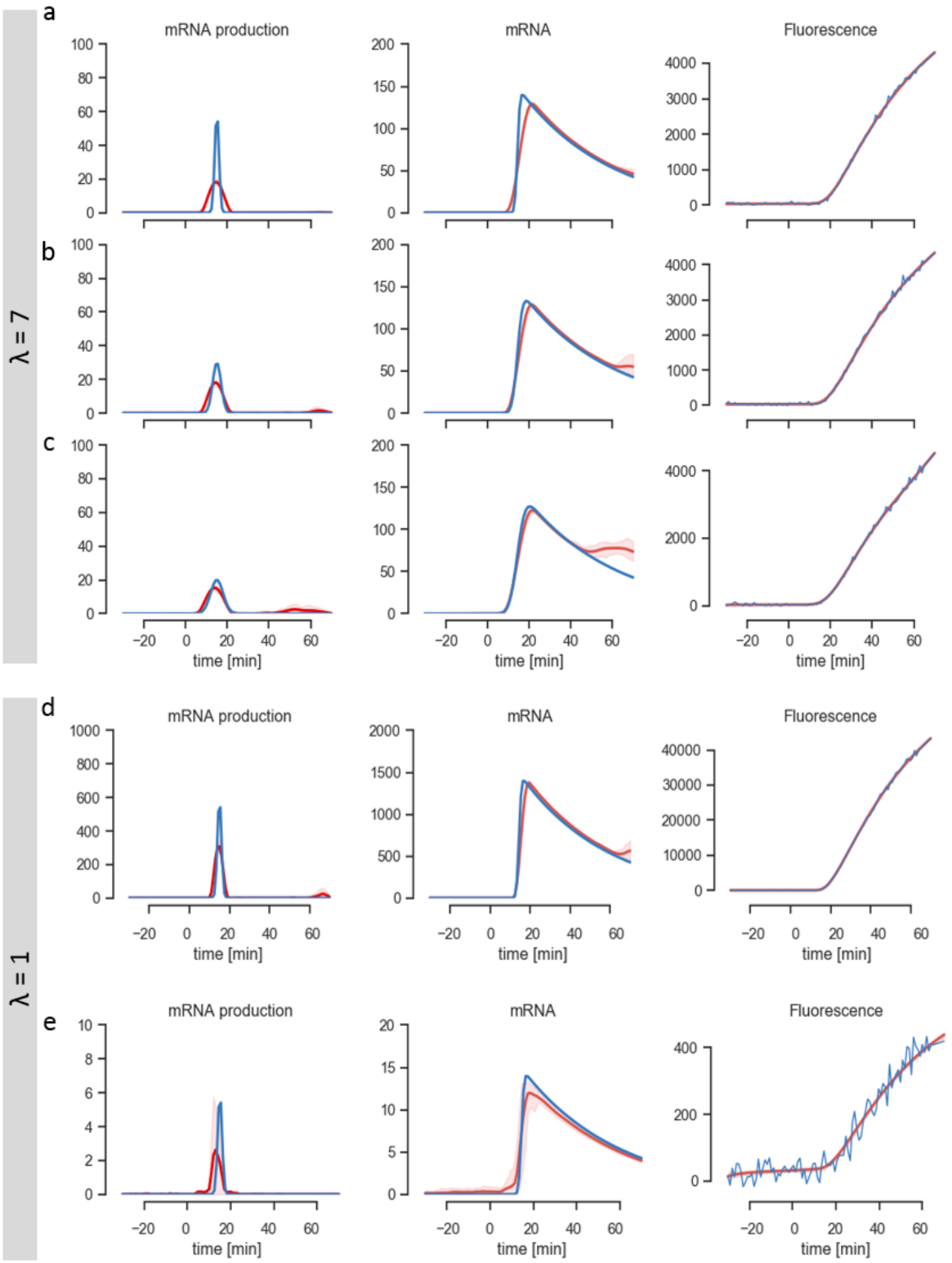

Figure A.2: (Caption on the following page.)

Figure A.2: Effect of the parameter λ on temporal resolution.
The leftmost column represents instantaneous mRNA production, the middle column represents mRNA concentration and the right column represents protein fluorescence. In each plot the blue lines represent simulated data and the red lines are the result of the reconstruction algorithm. To simulate the data we considered bursts of mRNA production with different durations producing the same total amount of mRNA. The duration of the bursts (FWHM) is set to 5 minutes in panels a,d and e, 10 minutes in panel b and 15 min in panel c. Then, we simulate the total mRNA concentration and the protein fluorescence accordingly. We add noise to the protein fluorescence with the same amplitude we observed in our actual experimental data for the Bcd3 enhancer (panels a, b and c), whereas 10x weaker noise is added in panel d and 10x stronger noise in panel e. The mRNA reconstruction algorithm is used to reconstruct the total amount of mRNA and the rate of mRNA production rates by fitting the noisy fluorescence data. The regularization parameter λ is set to 7 in panels a-c and to 1 in panels d and e. The result of the reconstruction and corresponding confidence intervals are plotted in red in each panel. The high signal-to-noise ratio considered in panel d allows a robust reconstruction even at a setting of $\lambda = 1$, which substantially improves the temporal resolution. In contrast, at lower signal-to-noise ratios, such as those shown in panel e, it is not possible to obtain a robust reconstruction with $\lambda = 1$; in this case the bootstrapping error of the instantaneous mRNA production rate becomes as large as the signal.

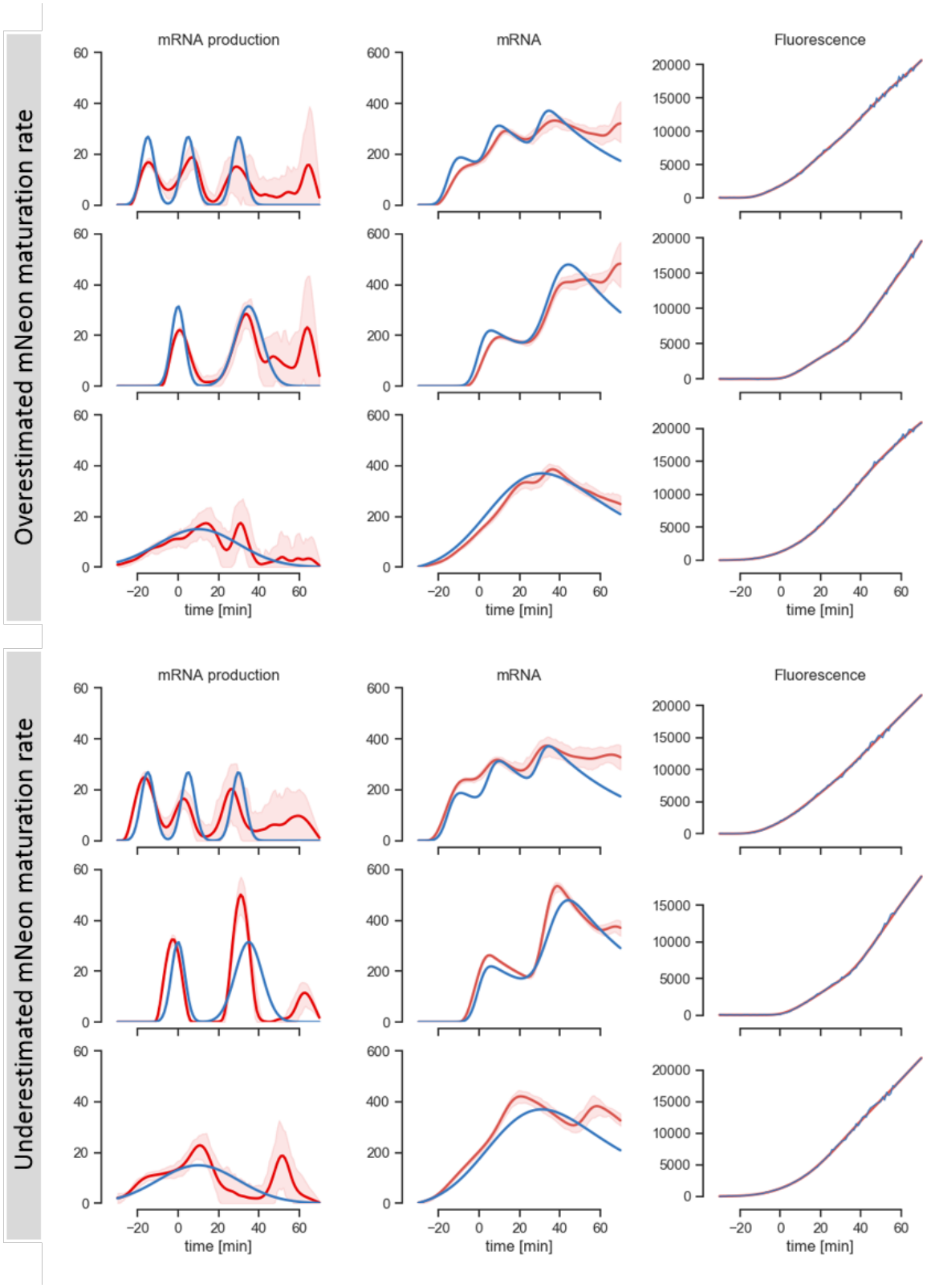

Figure A.3: (Continued on the following page.)

Figure A.3: Sensitivity tests of mRNA reconstruction analysis to variations in the protein maturation rate.
The leftmost column represents instantaneous mRNA production, the middle column represents mRNA concentration and the right column represents protein fluorescence. In each plot the blue lines represent simulated data and the red lines are the result of the reconstruction algorithm. To simulate the data we consider three different scenarios for the dynamics of mRNA production with three short bursts (top row), two bursts (second row) and one single large burst (third row). Then, we simulate the total mRNA concentration and the protein fluorescence accordingly. We add noise to the protein fluorescence with the same amplitude we observed in our actual experimental data for the hb_ant enhancer (first block) and Bcd3 enhancer (second block), using the relation obtained from the data presented in Fig. 2.9. The mRNA reconstruction algorithm is then used to reconstruct the total amount of mRNA and the rate of mRNA production rates by fitting the noisy fluorescence data. The result of the reconstruction and corresponding 90% confidence intervals are plotted in red in each panel. The result of the reconstruction and corresponding confidence intervals are plotted in red in each panel. The rate of maturation of the mNeon reporter in the reconstruction algorithm has been set to twice its real value in the upper panel, and to half in the lower panel, λ and N are set to 7 and 40.

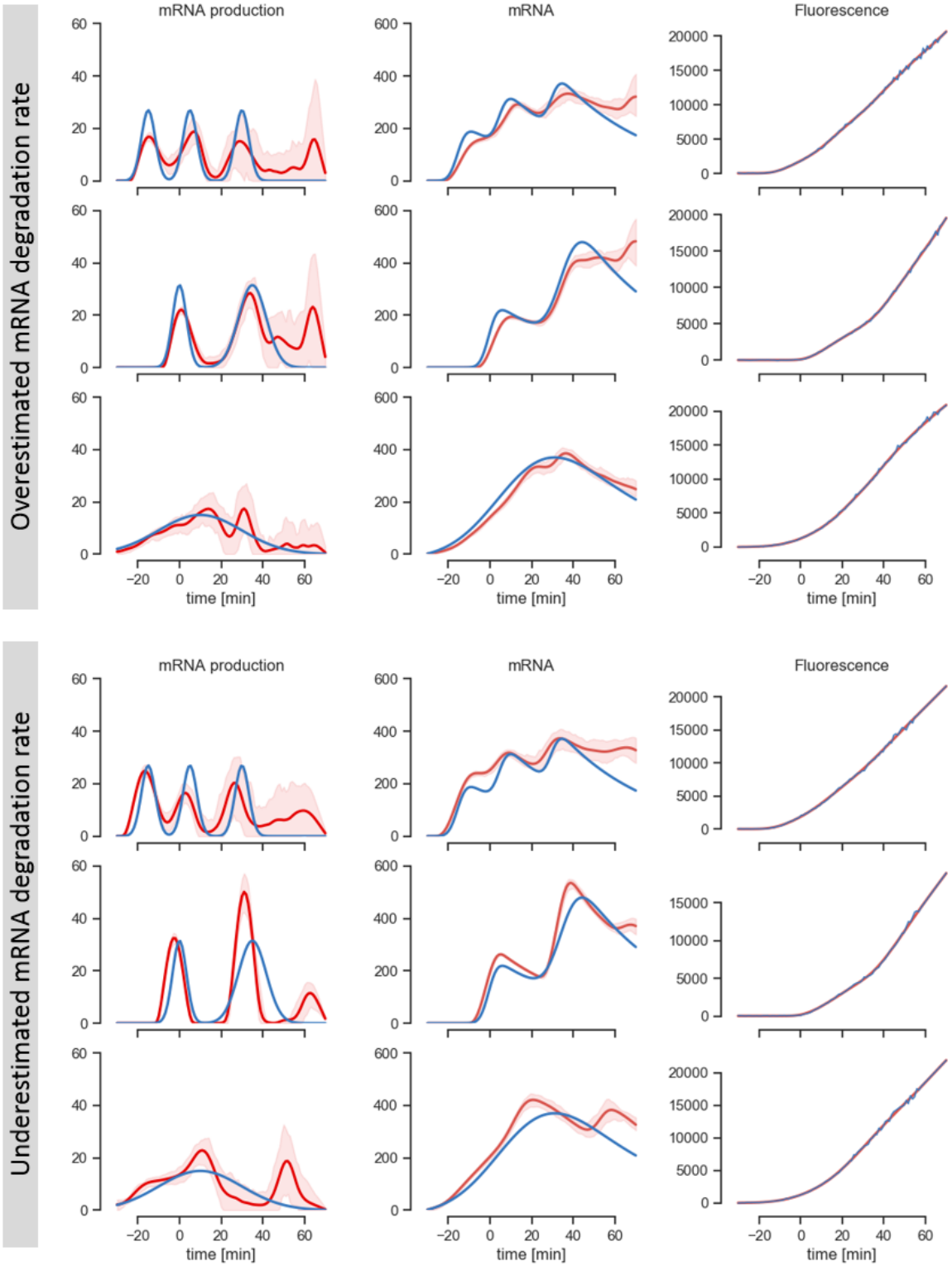

Figure A.4: (Continued on the following page.)

Figure A.4: Sensitivity tests of mRNA reconstruction analysis to variations in the mRNA degradation rate.
The leftmost column represents instantaneous mRNA production, the middle column represents mRNA concentration and the right column represents protein fluorescence. In each plot the blue lines represent simulated data and the red lines are the result of the reconstruction algorithm. To simulate the data we consider three different scenarios for the dynamics of mRNA production with three short bursts (top row), two bursts (second row) and one single large burst (third row). Then, we simulate the total mRNA concentration and the protein fluorescence accordingly. We add noise to the protein fluorescence with the same amplitude we observed in our actual experimental data for the hb_ant enhancer (first block) and Bcd3 enhancer (second block), using the relation obtained from the data presented in Fig. 2.9. The mRNA reconstruction algorithm is then used to reconstruct the total amount of mRNA and the rate of mRNA production rates by fitting the noisy fluorescence data. The result of the reconstruction and corresponding 90% confidence intervals are plotted in red in each panel. The rate of degradation of the mNeon mRNA in the reconstruction algorithm has been set to twice its real value in the upper panel, and to half in the lower panel, λ and N are set to 7 and 40.

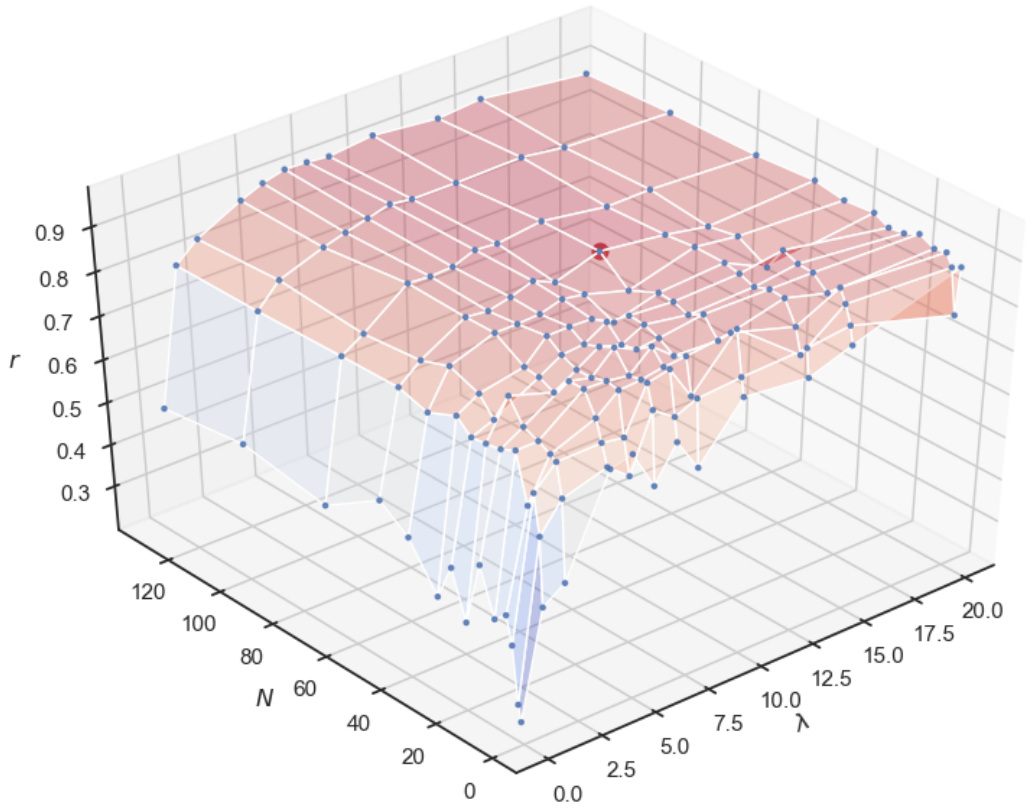

Figure A.5: The average correlation (r) between the simulated mRNA production time-course and the result of the reconstruction algorithm for three different dynamics of mRNA production, similar to those represented in Fig. A.1, as a function of the regularization parameter λ and the number of bootstraps. We find that the average correlation reaches a maximum (highlighted with a red dot) for $\lambda = 7$ and that no further significant improvement is achieved for $N > 40$.

Appendix B

Reproducibility of the mNeonGreen reporter

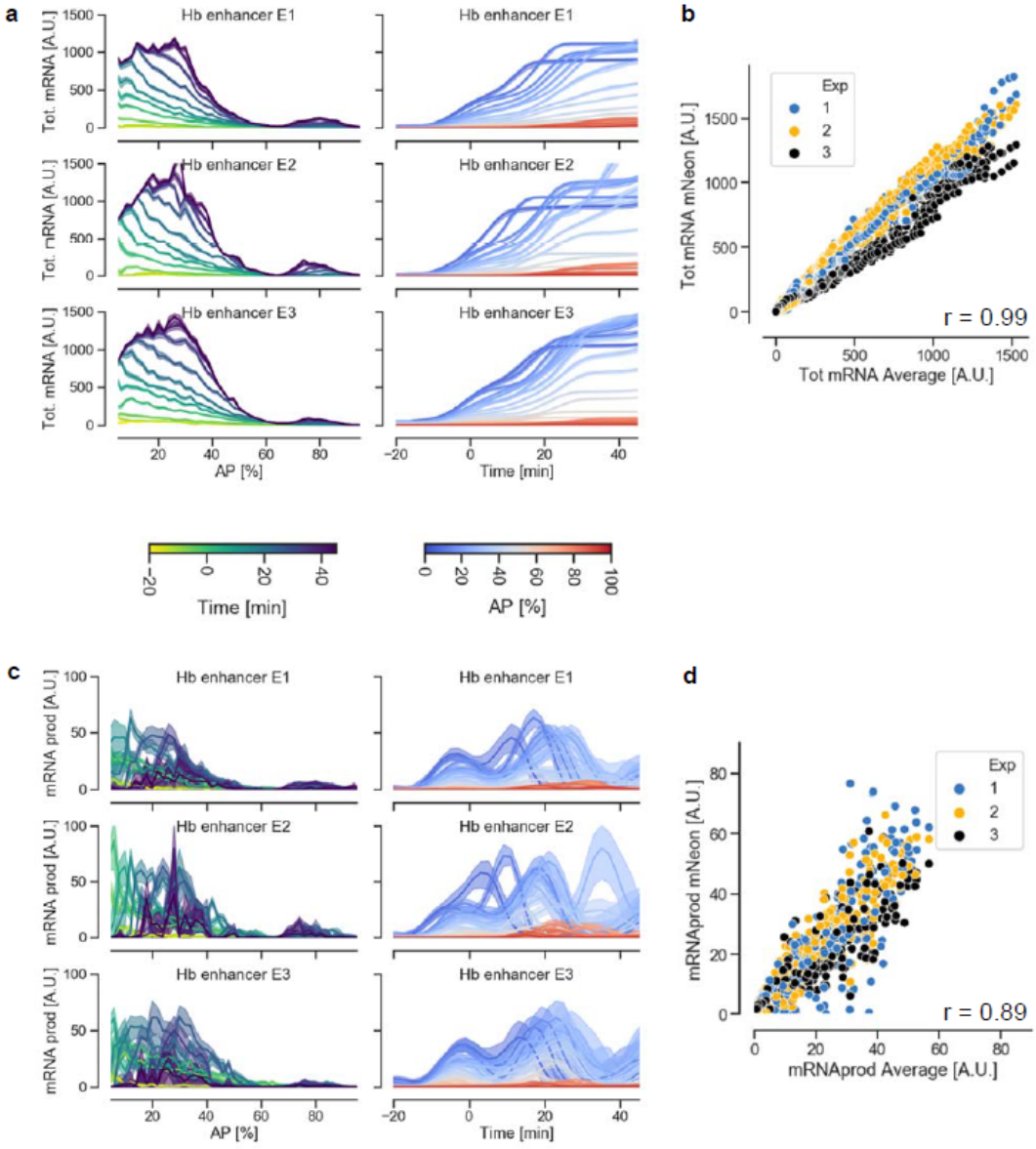

Figure B.1: Reproducibility of the mNeon reporter measurements for the Hunchback Anterior enhancer. a) Cumulative mRNA production patterns at different time points of embryo development (left) and time course of cumulative mRNA production in 2% bins along the AP axis of the embryo (right) for three different embryos (top, middle and bottom) carrying the hb_ant-DSCP-mNeonRep construct. b) Comparison of the three replicates for the measurement of the cumulative mRNA production at all times and positions. r is the Pearson correlation coefficient. c) Pattern of the instantaneous mRNA production rate at different time points of embryo development (left) and time course of mRNA production rate in 2% bins along the AP axis of the embryo (right) for three different embryos (top, middle and bottom). d) Comparison of the three replicates for the measurement of the mRNA production rate at all times and positions. r is the Pearson correlation coefficient.

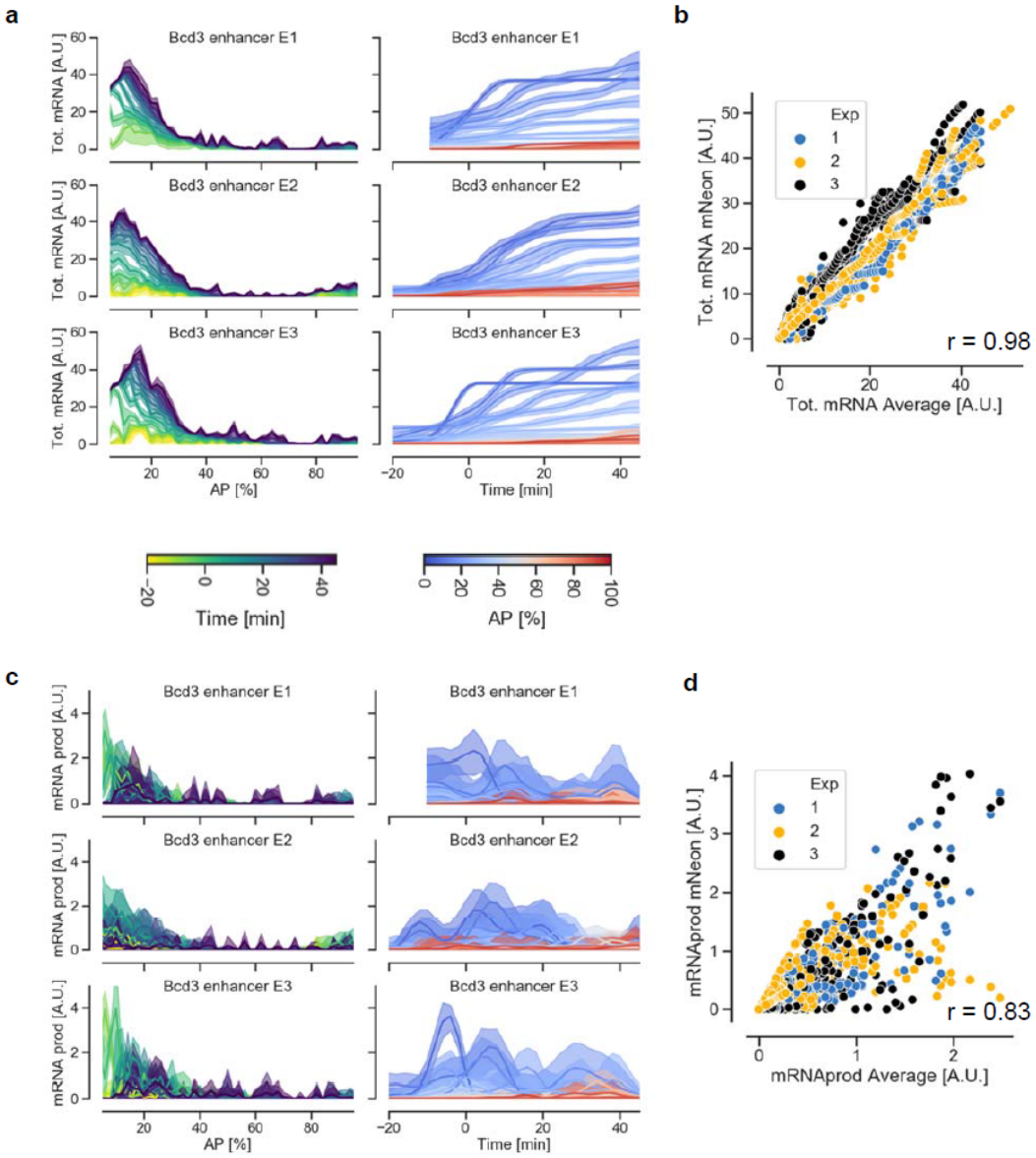

Figure B.2: Reproducibility of the mNeon reporter measurements for the Bcd3 enhancer. a) Cumulative mRNA production patterns at different time points of embryo development (left) and time course of cumulative mRNA production in 2% bins along the AP axis of the embryo (right) for three different embryos carrying the Bcd3-DSCP-mNeonRep construct (top, middle and bottom). b) Comparison of the three replicates for the measurement of the cumulative mRNA production at all time points and positions.r is the Pearson correlation coefficient. c) Pattern of the instantaneous mRNA production rate at different time points of embryo development (left) and time course of mRNA production rate in 2% bins along the AP axis of the embryo (right) for three different embryos (top, middle and bottom). d) Comparison of the three replicates for the measurement of the mRNA production rate, at all times and positions. r is the Pearson correlation coefficient.

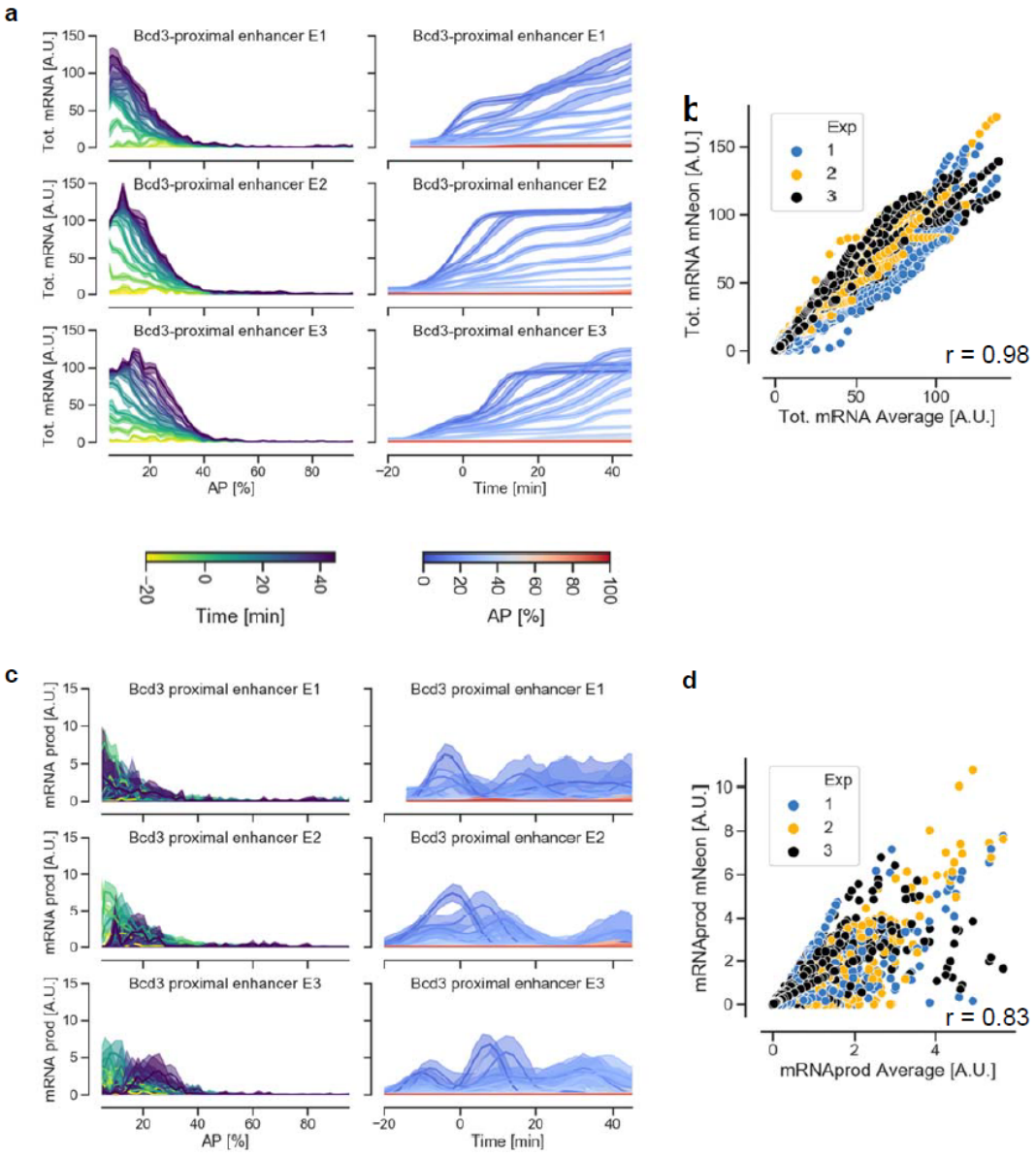

Figure B.3: Reproducibility of the mNeon reporter measurements for the Bcd3-proximal enhancer. a) Cumulative mRNA production patterns at different time points of embryo development (left) and time course of cumulative mRNA production in 2% bins along the AP axis of the embryo (right) for three different embryos carrying the Bcd3-proximal-DSCP-mNeonRep construct (top, middle and bottom). b) Comparison of the three replicates for the measurement of the cumulative mRNA production at all time points and positions.r is the Pearson correlation coefficient. c) Pattern of the instantaneous mRNA production rate at different time points of embryo development (left) and time course of mRNA production rate in 2% bins along the AP axis of the embryo (right) for three different embryos (top, middle and bottom). d) Comparison of the three replicates for the measurement of the mRNA production rate, at all times and positions. r is the Pearson correlation coefficient.

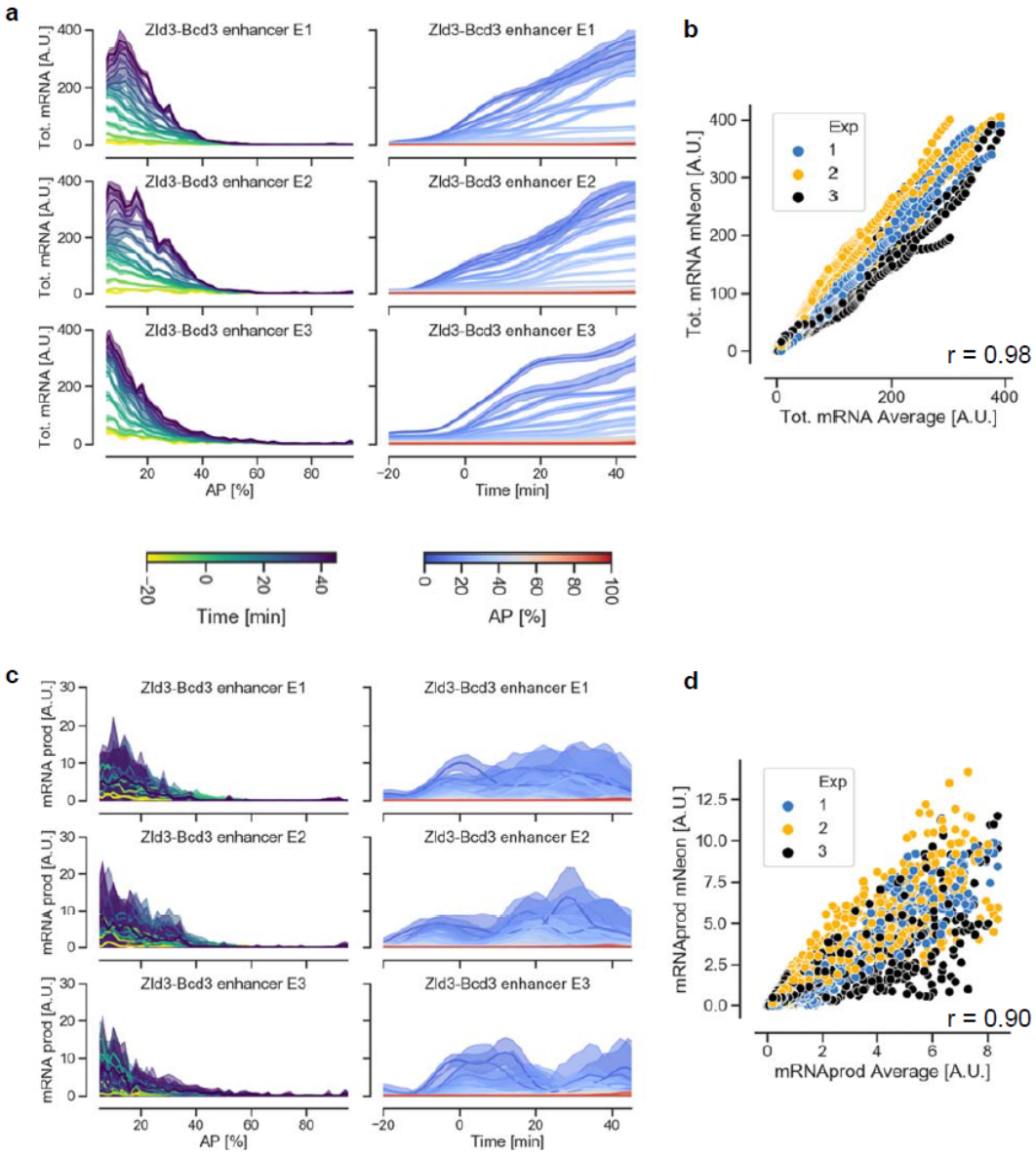

Figure B.4: Reproducibility of the mNeon reporter measurements for the Zld3-Bcd3 enhancer. a) Cumulative mRNA production patterns at different time points of embryo development (left) and time course of cumulative mRNA production in 2% bins along the AP axis of the embryo (right) for three different embryos carrying the Zld3-Bcd3-DSCP-mNeonRep construct (top, middle and bottom). b) Comparison of the three replicates for the measurement of the cumulative mRNA production at all time points and positions.r is the Pearson correlation coefficient. c) Pattern of the instantaneous mRNA production rate at different time points of embryo development (left) and time course of mRNA production rate in 2% bins along the AP axis of the embryo (right) for three different embryos (top, middle and bottom). d) Comparison of the three replicates for the measurement of the mRNA production rate, at all times and positions. r is the Pearson correlation coefficient.

Appendix C

Protein and DNA sequences

C.1 mNeonGreen Reporter

The mNeonGreen sequence has been fused C and N terminally to three different nuclear localization signals (NLS):

Bipartite class, N-term NLS(Magico and Bell, 2011)	RKQVSSHIQVLARRKLR
Sv40-NLS, C-term – class1(Kosugi et al., 2009)	PKKKRKV
Class 3 NLS, C-term(Kosugi et al., 2009)	AAAKRSWSMAF

In order to determine the arrangement of NLSs providing the strongest nuclear localization, various arrangements of these motifs fused to mNeonGreen have been characterized through preliminary experiments in *D. melanogaster* S2 cells. The strength of the nuclear localization has been characterized using the ratio of nuclear vs cytoplasmic mNeon fluorescence.

					Nuc/Cyt Min	Nuc/Cyt Max
A	Bip	mNeonGreen	SV40	-	16	20
B	Bip	mNeonGreen	-	-	1	2
C	Bip	mNeonGreen	SV40	Class3	17.5	19.5
D	-	mNeonGreen	SV40	-	2	3
E	-	mNeonGreen	SV40	Class3	2	6

All NLSs have been fused together and to mNeonGreen (in green) through short GS linker sequences. The protein sequence of the construct achieving the strongest localization in S2 cells (Construct C) has been codon-optimized for expression in *D. melanogaster*

using the Eurofins genomics GENEius software package - Munich, Germany , and obtained by gene synthesis.

The amminoacid sequence of the reporter reads:

>mNeonRep

MRKQVSSHIQVLARRKLRGSVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQG
TGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAMVD
GSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVKGTGFPADGPVMTNSL
TAADWCRSKKTYPNDKTIISTFKWSYTTGNGKRYRSTARTTYTFAKPMAANYL
KNQPMYVFRKTELKHSKTELNFKEWQKAFTDVMGMDELYKGSPKKKRKVGSA
AAKRSWSMAF

Which corresponds to the following DNA sequence:

>mNeonRep

ATGCGGAAGCAAGTCTCGAGCCACATCCAGGTGTTGGCCCGACGCAAACTG
CGTGGCTCCATGGTGTCTAAGGGCGAAGAGGACAACATGGCGAGTCTACCA
GCCACGCATGAGCTTCACATCTTTGGCTCCATCAATGGAGTGGATTTCGATA
TGGTGGGTCAAGGAACTGGCAATCCGAATGACGGCTATGAGGAACTGAACC
TGAAGTCGACCAAGGGCGATTTGCAGTTTTCTCCCTGGATTCTGGTACCGCA
TATTGGCTATGGCTTTCACCAGTACTTGCCATATCCTGATGGCATGTCACCT
TTCCAAGCGGCAATGGTGGACGGTAGCGGCTATCAAGTCCATCGAACTATG
CAGTTCGAGGATGGAGCCAGTCTGACCGTGAACTACCGCTATACCTATGAA
GGCTCCCACATTAAGGGTGAAGCTCAGGTGAAAGGAACAGGCTTTCCAGCT
GATGGTCCCGTTATGACGAACTCCTTGACTGCAGCCGACTGGTGCAGATCCA
AGAAAACCTACCCCAATGACAAGACCATCATCTCGACGTTCAAATGGTCGTA
TACGACAGGAAATGGTAAGCGCTACCGTTCAACAGCGAGGACAACGTACAC
CTTCGCCAAACCGATGGCTGCCAACTACCTGAAGAACCAGCCCATGTACGTC
TTTCGCAAGACCGAGCTCAAACACAGCAAGACTGAGCTGAACTTCAAAGAG
TGGCAGAAGGCCTTCACCGATGTTATGGGCATGGATGAGCTCTACAAGGGA
TCGCCGAAGAAGAAGCGGAAAGTCGGATCGGGCTCGGCTGCGGCCAAGCGC
TCCTGGAGCATGGCCTTC

C.2 Spacers

Enhancer-promoter linker sequence, present in the all the constructs studied in this thesis with the exception of those named ‘proximal’:

>Linker

AGGTTCCAGTGTGGTGGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCT
CGAGTCTAGAGGGCCCTTCGAA

Spacer sequence, inserted upstream the enhancer sequence to create a neutral background around enhancers:

>Neutral_Spacer
GAATTCGGCCGGCCCACCAAACCGCTTTAGTCCCGCCAGCCAGCCACCAGTG
GTGCAGCCCACCTCCTCGGCCCGCCAGTCGCCCATCGATGTCTGCCTGGAGG
AGGATGTTCACTCCGTGCACAGCCATCAGTCGTCCGCAAGCCTCCTGCATCC
CATTGCCATCCGAGCCACGCCAACCACTCCGACTAGCAGCAGCCCGCTGAGT
TTTGCGGCCAAGATGCAGAGCTTGTCGCCCGTTTCGGTTTGCTCCATTGGCG
GCGAAACCACCAGCGTTGTACCAGTGCATCCTCCCACCGTTTCCGCTCAAGA
AGGACCCATGGATCTGAGCATGAAGACCTCGCGGAGCTCCGT

Flanking sequence from the reporter plasmid used in (Fakhouri et al., 2010), inserted upstream of the 2Twi-2Dl enhancer in the Original-2Twi-2Dl construct:

>Original_Flanking
AAGCTTGCACAGCACTTTGTGTTTAATTGATGGCGTAAACCGCTTGGAGCTT
CGTCACGAAACCGCTGACAAAATGCAACTGAAGGCGGACATTGACGCTACG
TAACGCTACAAACGGTGGCGAAAGAGATAGCGGACGCAGCGGCGAAAGAGA
CGGCGATATTTCTGTGGACAGAGAAGGAGGCAAACAGCGCTGACTTTGAGT
GGAATGTCATTTTGAGTGAGAGGTAATCGAAAGAACCTGGTACATCAAATA
CCCTTGGATCGAAGTAAATTTAAAACTGATCAGATAAGTTCAATGATATCCA
GTGCAGTAAAAAATAAAAAAAAAATATGTTTTTTTAAATCTACATTCTCCAA
AAAAGGGTTTTATTAACTTACATACATACTAGAATTCGGTACCGAG

C.3 Enhancers

>Hb_ant
AACAATTGCAACAGGCATTAGTTTATATATCGCTCAGGTAGACGGATGCAC
GCGTCAAGGGATTAGATGGGCAGAGGTGACGGGAAGTCAGGTACAGGTCGC
GGATCGGTCGGGAATCGAGGATCACGGATCGCGGATTGAGGATTGCGCTCT
TGATCCATTCTGGATTAGAGCAGAAACAAAAAATTATGCGCACTTGGATTTG
GATGATCCGGGAGCTTAGCGGATGGCCAGCTTAGCAGCGAGCTGCGAATTT
TCCACCGGTTTTCTATGGGGATTACGTTGGTCAGGAGTCGACAGCAGGAGT
AGGCAGCTAGCGTGGGCAGTTTCGTAGTTAATAATAAAAAGTAAAAAGGAT
TGCGGGACTTAACTAAATTAACGGATCAGAACTGCTTACACCTGCGGGAAA
ACTCTAAGGACCAACTAAACTATATGCATAATATGTGCAGTATAATTATTAC

ACACCCATTTGAAAAACATTTTCCTGACAACAATTTTCCGCCAGACATTTCA
CTTTGATTTGCGTAGTTTTTCTAATAATTCTCGCATTAAAATTGCTTGTTGC
CTATATTTTTTCCATTTCCAATTTCACACTGAAAAATTGTGCAGTTGCTGCA
TTTTTGGCTAATTGTTTGTGCTTTCAAGTAAATATTATTAAAAACGCAAAAC
GGGAAAAAGGGGCATTTACGGAATATTATTATGGGAGGATGGTGCTGTGCT
A

>Bcd3

AGGTTCTAATCCCGGTCTAATCCCTCGAGTCTAATCCCATGAGTCGAC

>Zld3Bcd

CGGAAGTTCAGGTATTGCTATTCAGGTAGAGGCCGTACGTGCAGGTAACCT
TGCGTAGGTTCTAATCCCGGTCTAATCCCTCGAGTCTAATCCCATGAGTCGA
CG

>2Twi-2Dl

CTCGCATATGTTGAGCATATGTTTTGGGGGATTTTCCCAAATCGAGGGAAA
ACCCAA

>3Hb-2Twi-2Dl

AGGTACTCAAAAAACTATGGACTCAAAAAACTAATCGATACTCAAAAAACTA
TGAGTCGAACTCGCATATGTTGAGCATATGTTTTGGGGGATTTTCCCAAATC
GAGGGAAAACCCAAACTAGT

>3Hb-70-2Twi-2Dl

AGGTACTCAAAAAACTATGGACTCAAAAAACTAATCGATACTCAAAAAACTA
TGAGTCGAAACCGCTTTAGTCCCGCCAGCCAGCCACCAGTGGTGCAGCCCAC
CTCCTCGGCCCGCCAGTCGCCCATCGCTCGCATATGTTGAGCATATGTTTTG
GGGGATTTTCCCAAATCGAGGGAAAACCCAA

>3Hb-150-2Twi-2Dl

AGGTACTCAAAAAACTATGGACTCAAAAAACTAATCGATACTCAAAAAACTA
TGAGTCGAAACCGCTTTAGTCCCGCCAGCCAGCCACCAGTGGTGCAGCCCAC
CTCCTCGGCCCGCCAGTCGCCCATCGATGTCTGCCTGGAGGAGGATGTTCAC
TCCGTGCACAGCCATCAGTCGTCCGCAAGCCTCCTGCATCCCATTGCCATCC
GACTCGCATATGTTGAGCATATGTTTTGGGGGATTTTCCCAAATCGAGGGA
AAACCCAA

>3Hb-250-2Twi-2Dl

AGGTACTCAAAAAACTATGGACTCAAAAAACTAATCGATACTCAAAAAACTA
TGAGTCGAAACCGCTTTAGTCCCGCCAGCCAGCCACCAGTGGTGCAGCCCAC
CTCCTCGGCCCGCCAGTCGCCCATCGATGTCTGCCTGGAGGAGGATGTTCAC
TCCGTGCACAGCCATCAGTCGTCCGCAAGCCTCCTGCATCCCATTGCCATCC
GAGCCACGCCAACCACTCCGACTAGCAGCAGCCCGCTGAGTTTTGCGGCCAA

GATGCAGAGCTTGTCGCCCGTTTCGGTTTGCTCCATTGGCGGCGAAACCACT
CGCATATGTTGAGCATATGTTTTGGGGGATTTTCCCAAATCGAGGGAAAAC
CCAA

>3HbWeak-2Twi-2Dl

AGGTACTCAATAAACTATGGACTCAATAAACTAATCGATACTCAATAAACTA
TGAGTCGAACTCGCATATGTTGAGCATATGTTTTGGGGGATTTTCCCAAATC
GAGGGAAAACCCAA

>3HbRev-2Twi-2Dl

TTCGACTCATAGTTTTTTGAGTATCGATTAGTTTTTTGAGTCCATAGTTTTT
TGAGTACCTCTCGCATATGTTGAGCATATGTTTTGGGGGATTTTCCCAAATC
GAGGGAAAACCCAA

>3Zld-2Twi-2Dl

CGGAAGTTCAGGTATTGCTATTCAGGTAGAGGCCGTACGTGCAGGTAACCT
TGCGTCTCGCATATGTTGAGCATATGTTTTGGGGGATTTTCCCAAATCGAG
GGAAAACCCAA

>Hb3-Bcd3

AGCTTAGGTACTCAAAAAACTATGGACTCAAAAAACTAATCGATACTCAAAA
AACTATGAGTCGAAGGTTCTAATCCCGGTCTAATCCCTCGAGTCTAATCCCA
TGAGTCGACG

>3Hb-70-3Bcd

AGGTACTCAAAAAACTATGGACTCAAAAAACTAATCGATACTCAAAAAACTA
TGAGTCGAAACCGCTTTAGTCCCGCCAGCCAGCCACCAGTGGTGCAGCCCAC
CTCCTCGGCCCGCCAGTCGCCCATCGAGGTTCTAATCCCGGTCTAATCCCTC
GAGTCTAATCCCATGAGTCGACG

>3Hb-150-3Bcd

AGGTACTCAAAAAACTATGGACTCAAAAAACTAATCGATACTCAAAAAACTA
TGAGTCGAAACCGCTTTAGTCCCGCCAGCCAGCCACCAGTGGTGCAGCCCAC
CTCCTCGGCCCGCCAGTCGCCCATCGATGTCTGCCTGGAGGAGGATGTTCAC
TCCGTGCACAGCCATCAGTCGTCCGCAAGCCTCCTGCATCCCATTGCCATCC
GAAGGTTCTAATCCCGGTCTAATCCCTCGAGTCTAATCCCATGAGTCGACG

>2Hb-150-3Bcd

ACTCAAAAAACTAATCGATACTCAAAAAACTATGAGTCGAAACCGCTTTAGT
CCCGCCAGCCAGCCACCAGTGGTGCAGCCCACCTCCTCGGCCCGCCAGTCGC
CCATCGATGTCTGCCTGGAGGAGGATGTTCACTCCGTGCACAGCCATCAGTC
GTCCGCAAGCCTCCTGCATCCCATTGCCATCCGAAGGTTCTAATCCCGGTCT
AATCCCTCGAGTCTAATCCCATGAGTCGACG

>1Hb-150-3Bcd

TACTCAAAAAACTATGAGTCGAAACCGCTTTAGTCCCGCCAGCCAGCCACCA
GTGGTGCAGCCCACCTCCTCGGCCCGCCAGTCGCCCATCGATGTCTGCCTGG
AGGAGGATGTTCACTCCGTGCACAGCCATCAGTCGTCCGCAAGCCTCCTGCA
TCCCATTGCCATCCGAAGGTTCTAATCCCGGTCTAATCCCTCGAGTCTAATC
CCATGAGTCGACG

>3HbW-150-3Bcd

AGGTACTCACAAAACTATGGACTCACAAAACTAATCGATACTCACAAAACTA
TGAGTCGAAACCGCTTTAGTCCCGCCAGCCAGCCACCAGTGGTGCAGCCCAC
CTCCTCGGCCCGCCAGTCGCCCATCGATGTCTGCCTGGAGGAGGATGTTCAC
TCCGTGCACAGCCATCAGTCGTCCGCAAGCCTCCTGCATCCCATTGCCATCC
GAAGGTTCTAATCCCGGTCTAATCCCTCGAGTCTAATCCCATGAGTCGACG

>3HbRev-150-3Bcd

TCGACTCATAGTTTTTTGAGTATCGATTAGTTTTTTGAGTCCATAGTTTTTT
GAGTACCTAACCGCTTTAGTCCCGCCAGCCAGCCACCAGTGGTGCAGCCCAC
CTCCTCGGCCCGCCAGTCGCCCATCGATGTCTGCCTGGAGGAGGATGTTCAC
TCCGTGCACAGCCATCAGTCGTCCGCAAGCCTCCTGCATCCCATTGCCATCC
GAAGGTTCTAATCCCGGTCTAATCCCTCGAGTCTAATCCCATGAGTCGACG

>3HbRev-3Bcd

TCGACTCATAGTTTTTTGAGTATCGATTAGTTTTTTGAGTCCATAGTTTTTT
GAGTACCTAGGTTCTAATCCCGGTCTAATCCCTCGAGTCTAATCCCATGAGT
CGACG

Appendix D

Predicted binding sites in enhancer sequences

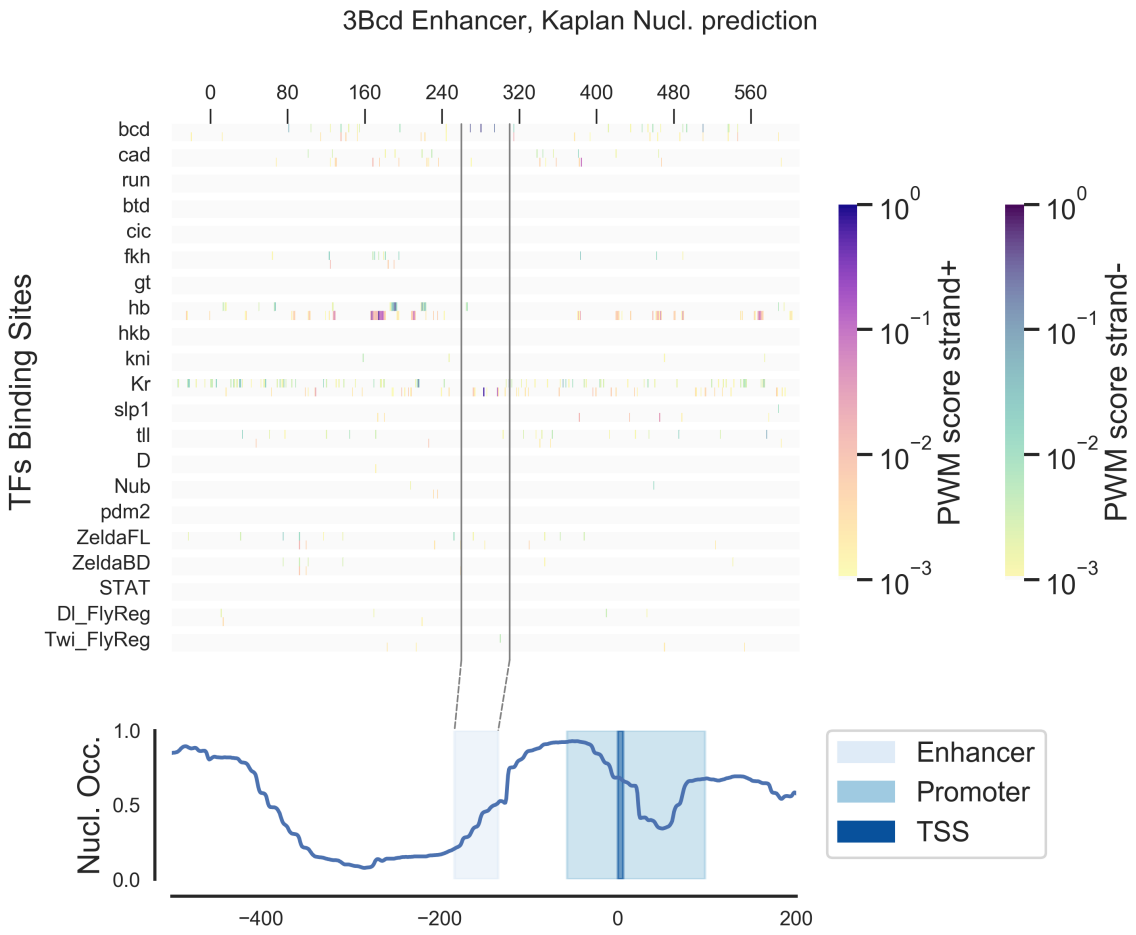

Figure D.1: Bcd3 enhancer in the original pBDP backbone. The predicted binding sites for TFs of the segmentation network and the nucleosome occupancy for the Bcd3 enhancer. The heatmap in the first panel represents bindings sites strength at each position along the sequence. Two different color schemes are used to represent binding sites in the forward or reverse strands and thus the binding sites orientation. The lower panel shows the predicted nucleosome occupancy over a larger region that covers the enhancer as well as the surrounding sequence, including the promoter. The positions of the enhancer, promoter and TSS are highlighted with different shades of blue. Multiple strong binding sites for segmentation TFs are present in the plasmid backbone, just upstream of the Bcd3 enhancer.

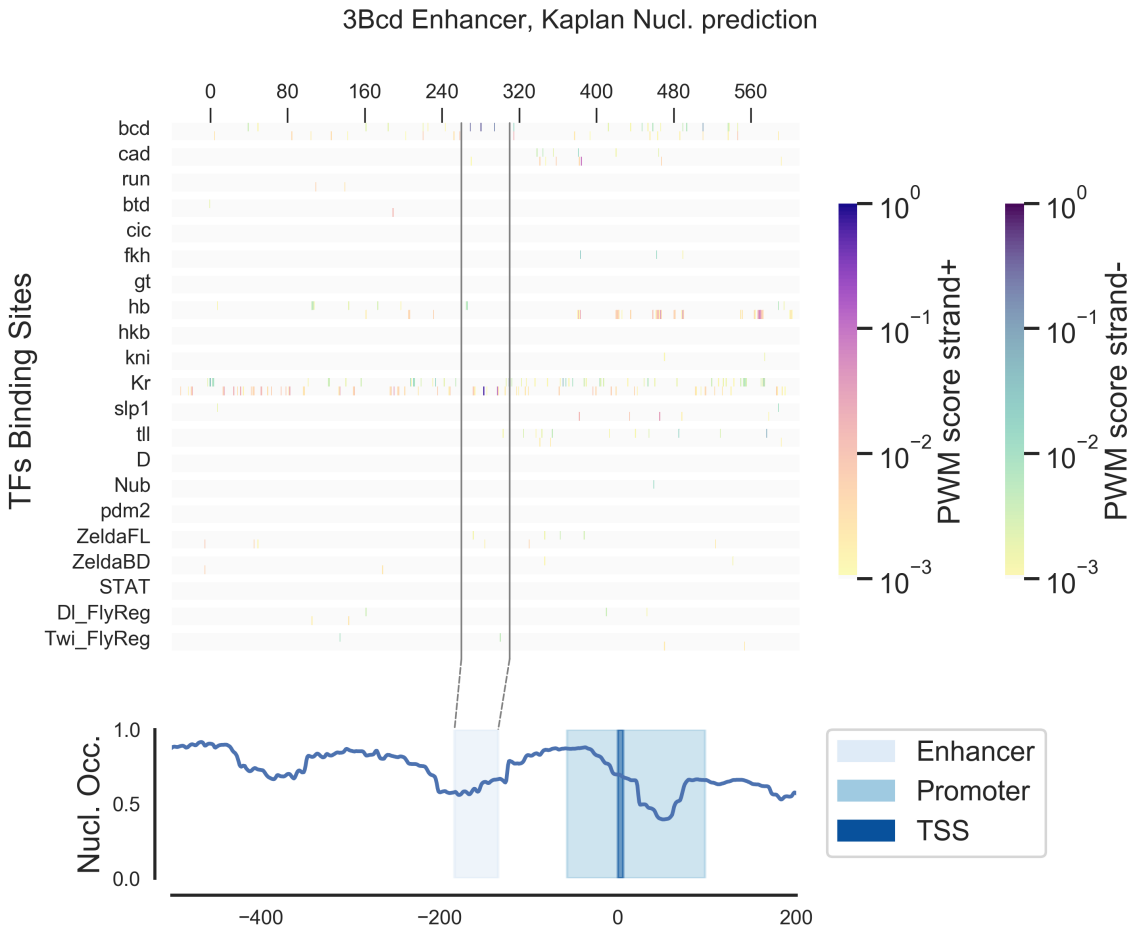

Figure D.2: Bcd3 enhancer in the clean-pBDP backbone. The predicted binding sites for TFs of the segmentation network and the nucleosome occupancy. The heatmap in the first panel represents bindings sites strength at each position along the sequence. Two different color schemes are used to represent binding sites in the forward or reverse strands and thus the binding sites orientation. The lower panel shows the predicted nucleosome occupancy over a larger region that covers the enhancer as well as the surrounding sequence, including the promoter. The positions of the enhancer, promoter and TSS are highlighted with different shades of blue.

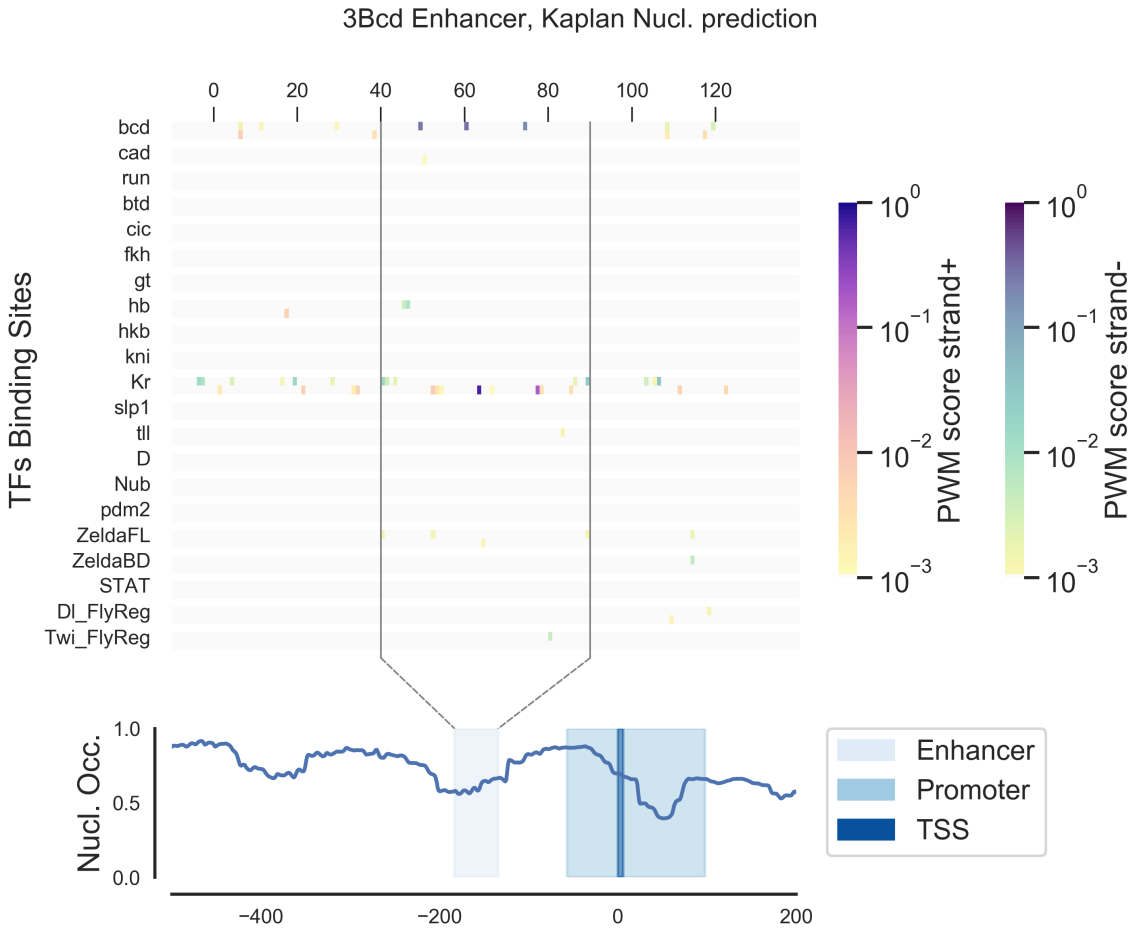

Figure D.3: Bcd3 enhancer in the clean-pBDP backbone (zoom-in). The predicted binding sites for TFs of the segmentation network and the nucleosome occupancy. The heatmap in the first panel represents bindings sites strength at each position along the sequence. Two different color schemes are used to represent binding sites in the forward or reverse strands and thus the binding sites orientation. The lower panel shows the predicted nucleosome occupancy over a larger region that covers the enhancer as well as the surrounding sequence, including the promoter. The positions of the enhancer, promoter and TSS are highlighted with different shades of blue.

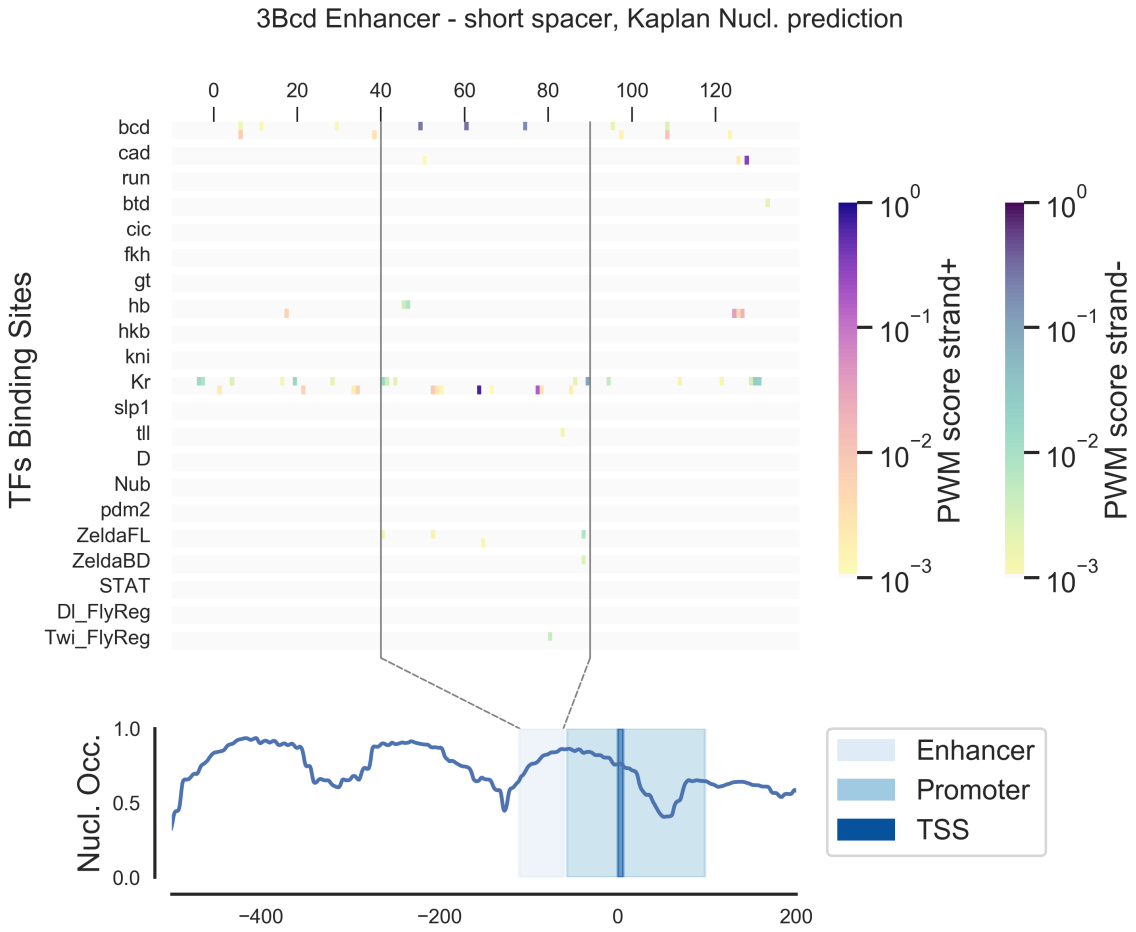

Figure D.4: Bcd3-proximal enhancer in the clean-pBDP backbone. The predicted binding sites for TFs of the segmentation network and the nucleosome occupancy. The heatmap in the first panel represents bindings sites strength at each position along the sequence. Two different color schemes are used to represent binding sites in the forward or reverse strands and thus the binding sites orientation. The lower panel shows the predicted nucleosome occupancy over a larger region that covers the enhancer as well as the surrounding sequence, including the promoter. The positions of the enhancer, promoter and TSS are highlighted with different shades of blue.

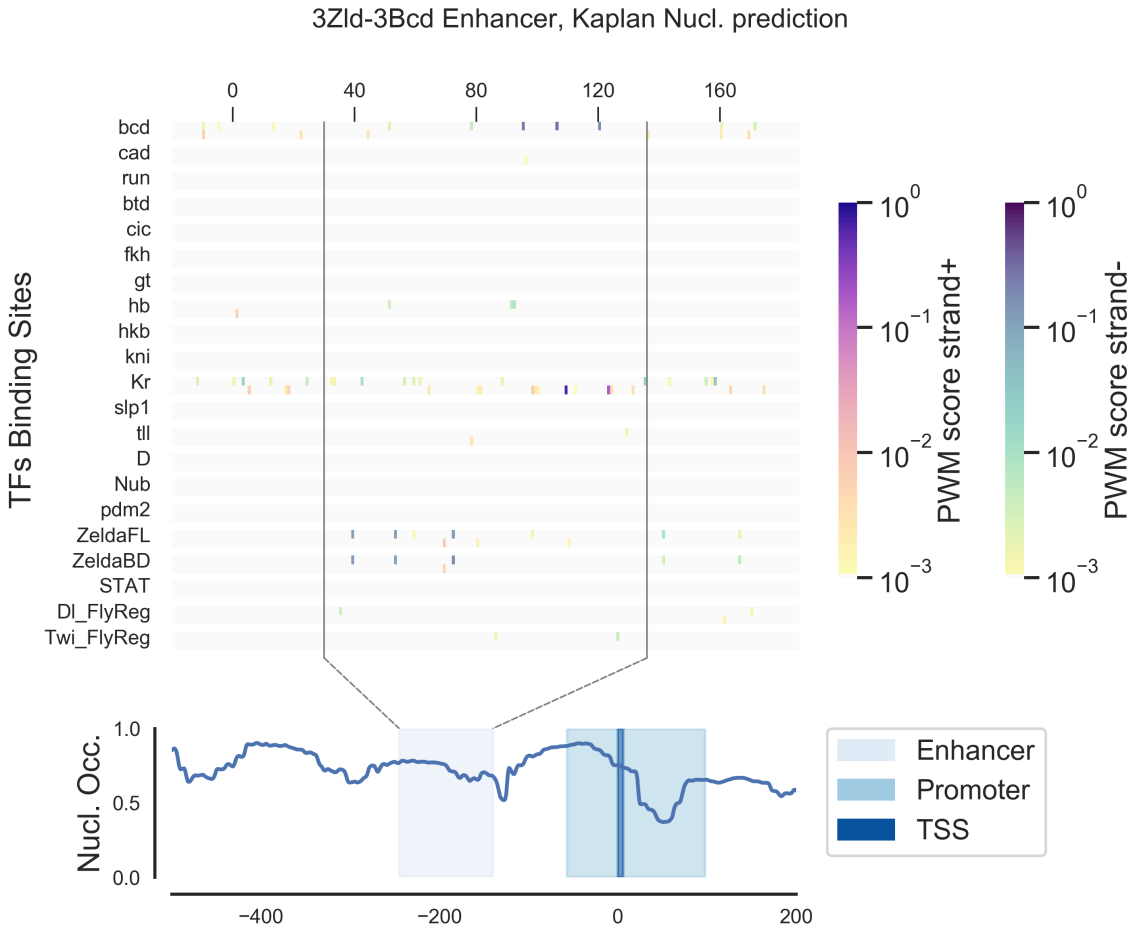

Figure D.5: Zld3Bcd3 enhancer in the clean-pBDP backbone. The predicted binding sites for TFs of the segmentation network and the nucleosome occupancy. The heatmap in the first panel represents bindings sites strength at each position along the sequence. Two different color schemes are used to represent binding sites in the forward or reverse strands and thus the binding sites orientation. The lower panel shows the predicted nucleosome occupancy over a larger region that covers the enhancer as well as the surrounding sequence, including the promoter. The positions of the enhancer, promoter and TSS are highlighted with different shades of blue.

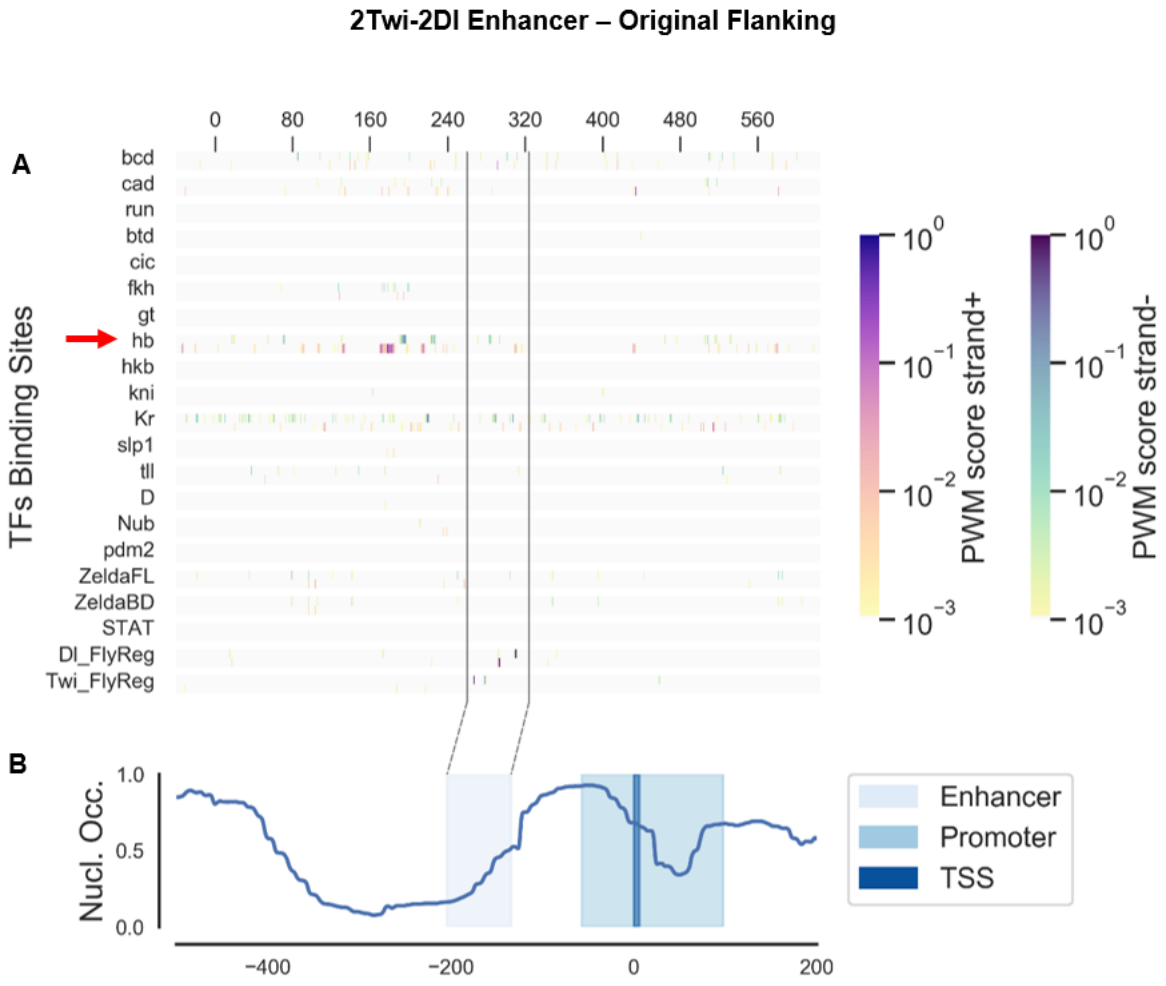

Figure D.6: 2Twi-2Dl enhancer in the original backbone. The predicted binding sites for TFs of the segmentation network and the nucleosome occupancy. The heatmap in the first panel represents bindings sites strength at each position along the sequence. Two different color schemes are used to represent binding sites in the forward or reverse strands and thus the binding sites orientation. The lower panel shows the predicted nucleosome occupancy over a larger region that covers the enhancer as well as the surrounding sequence, including the promoter. The positions of the enhancer, promoter and TSS are highlighted with different shades of blue. The arrow highlight the presence of multiple strong binding sites for the TF Hb in the plasmid backbone.

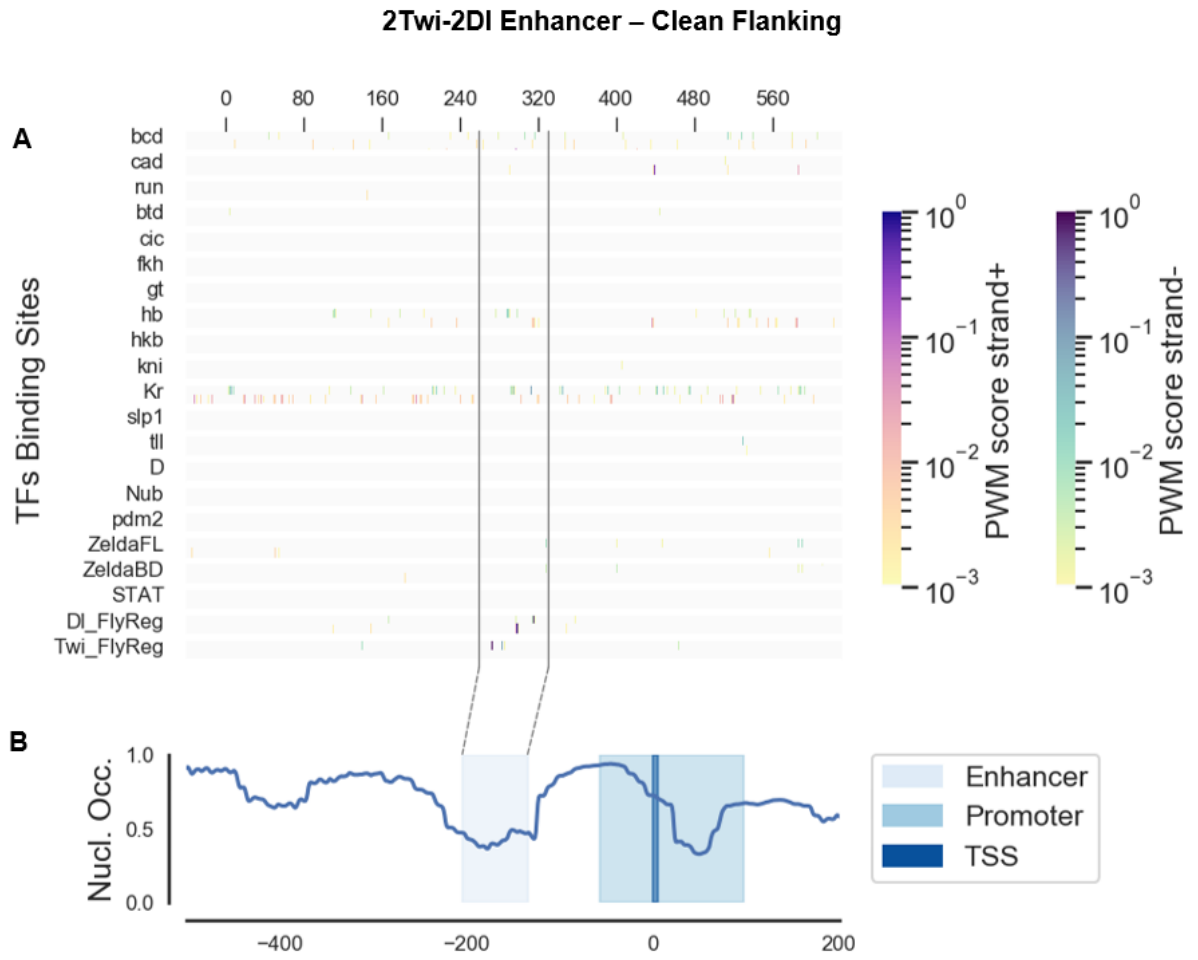

Figure D.7: 2Twi-2Dl enhancer in the clean-pBDP backbone. The predicted binding sites for TFs of the segmentation network and the nucleosome occupancy. The heatmap in the first panel represents bindings sites strength at each position along the sequence. Two different color schemes are used to represent binding sites in the forward or reverse strands and thus the binding sites orientation. The lower panel shows the predicted nucleosome occupancy over a larger region that covers the enhancer as well as the surrounding sequence, including the promoter. The positions of the enhancer, promoter and TSS are highlighted with different shades of blue.

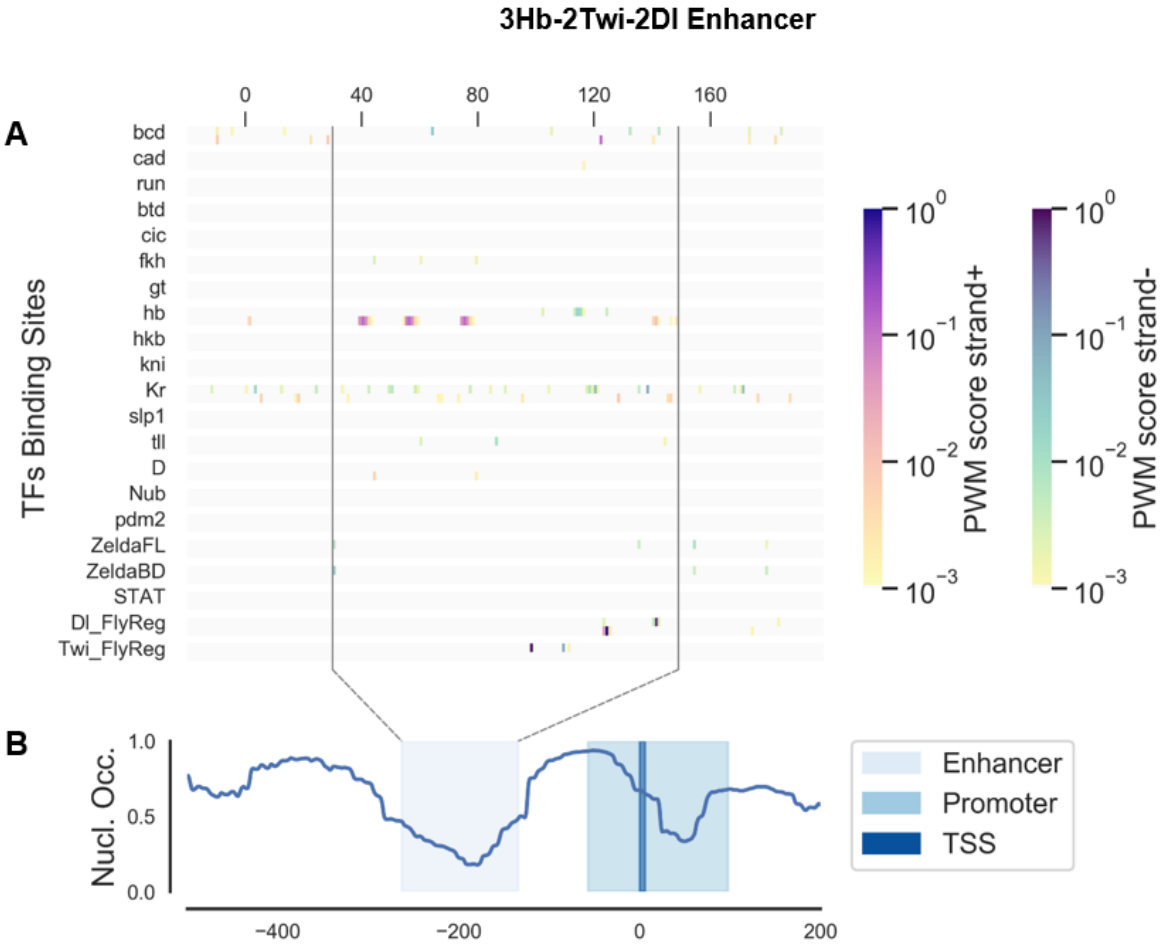

Figure D.8: 3Hb-2Twi-2Dl enhancer in the clean-pBDP backbone. The predicted binding sites for TFs of the segmentation network and the nucleosome occupancy. The heatmap in the first panel represents bindings sites strength at each position along the sequence. Two different color schemes are used to represent binding sites in the forward or reverse strands and thus the binding sites orientation. The lower panel shows the predicted nucleosome occupancy over a larger region that covers the enhancer as well as the surrounding sequence, including the promoter. The positions of the enhancer, promoter and TSS are highlighted with different shades of blue.

Bibliography

R. Amit, H. G. Garcia, R. Phillips, and S. E. Fraser. Building enhancers from the ground up: a synthetic biology approach. Cell, 146(1):105–118, 2011. ISSN 1097-4172 (Electronic) 0092-8674 (Linking). doi: 10.1016/j.cell.2011.06.024. URL https://www.ncbi.nlm.nih.gov/pubmed/21729783.

D. N. Arnosti and M. M. Kulkarni. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards?, 2005. ISSN 07302312.

J. Banerji, S. Rusconi, and W. Schaffner. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell, 27(2 Pt 1):299–308, 1981. ISSN 0092-8674 (Print) 0092-8674 (Linking). doi: 10.1016/0092-8674(81)90413-x. URL https://www.ncbi.nlm.nih.gov/pubmed/6277502.

K. A. Barr, C. Martinez, J. R. Moran, A. R. Kim, A. F. Ramos, and J. Reinitz. Synthetic enhancer design by in silico compensatory evolution reveals flexibility and constraint in cis-regulation. BMC Systems Biology, 2017. ISSN 17520509. doi: 10.1186/s12918-017-0485-2.

B. P. Berman, Y. Nibu, B. D. Pfeiffer, P. Tomancak, S. E. Celniker, M. Levine, G. M. Rubin, and M. B. Eisen. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. Proceedings of the National Academy of Sciences of the United States of America, 2002. ISSN 00278424. doi: 10.1073/pnas.231608898.

B. P. Berman, B. D. Pfeiffer, T. R. Laverty, S. L. Salzberg, G. M. Rubin, M. B. Eisen, and S. E. Celniker. Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in Drosophila melanogaster and Drosophila pseudoobscura. Genome biology, 2004. ISSN 14656914. doi: 10.1186/gb-2004-5-9-r61.

E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long. Localization of ASH1 mRNA particles in living yeast. Molecular Cell, 1998. ISSN 10972765. doi: 10.1016/S1097-2765(00)80143-4.

T. G. Bialek William, Gregor Thomas. Action at a distance in transcriptional regulation. arxiv, 2019.

J. P. Bothma, H. G. Garcia, S. Ng, M. W. Perry, T. Gregor, and M. Levine. Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. eLife, 2015. ISSN 2050-084X. doi: 10.7554/elife.07956.

J. P. Bothma, M. R. Norstad, S. Alamos, and H. G. Garcia. LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos. Cell, 173(7):1810–1822 e16, 2018. ISSN 1097-4172 (Electronic) 0092-8674 (Linking). doi: 10.1016/j.cell.2018.03.069. URL https://www.ncbi.nlm.nih.gov/pubmed/29754814.

M. Bozek, R. Cortini, A. E. Storti, U. Unnerstall, U. Gaul, and N. Gompel. ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the Drosophila blastoderm. Genome Res, 29(5):771–783, 2019. ISSN 1549-5469 (Electronic) 1088-9051 (Linking). doi: 10.1101/gr.242362.118. URL https://www.ncbi.nlm.nih.gov/pubmed/30962180.

D. S. Burz, R. Rivera-Pomar, H. Jäckle, and S. D. Hanes. Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the Drosophila embryo. EMBO Journal, 1998. ISSN 02614189. doi: 10.1093/emboj/17.20.5998.

H. N. Cai, D. N. Arnosti, and M. Levine. Long-range repression in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 1996. ISSN 00278424. doi: 10.1073/pnas.93.18.9309.

H. Chen, Z. Xu, C. Mei, D. Yu, and S. Small. A system of repressor gradients spatially organizes the boundaries of bicoid-dependent target genes. Cell, 2012. ISSN 00928674. doi: 10.1016/j.cell.2012.03.018.

J. Crocker, A. Tsai, and D. L. Stern. A Fully Synthetic Transcriptional Platform for a Multicellular Eukaryote. Cell Rep, 18(1):287–296, 2017. ISSN 2211-1247 (Electronic). doi: 10.1016/j.celrep.2016.12.025. URL https://www.ncbi.nlm.nih.gov/pubmed/28052257.

A. B. Cubitt, R. Heim, S. R. Adams, A. E. Boyd, L. A. Gross, and R. Y. Tsien. Understanding, improving and using green fluorescent proteins. Trends in Biochemical Sciences, 1995. ISSN 09680004. doi: 10.1016/S0968-0004(00)89099-4.

S. De Renzis, O. Elemento, S. Tavazoie, and E. F. Wieschaus. Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. PLoS Biol, 5(5):e117, 2007. ISSN 1545-7885 (Electronic) 1544-9173 (Linking). doi: 10.1371/journal.pbio.0050117. URL https://www.ncbi.nlm.nih.gov/pubmed/17456005.

J. Dufourt, A. Trullo, J. Hunter, C. Fernandez, J. Lazaro, M. Dejean, L. Morales, S. Nait-Amer, K. N. Schulz, M. M. Harrison, C. Favard, O. Radulescu, and M. Lagha. Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. Nat Commun, 9(1):5194, 2018. ISSN 2041-1723 (Electronic) 2041-1723 (Linking). doi: 10.1038/s41467-018-07613-z. URL https://www.ncbi.nlm.nih.gov/pubmed/30518940.

B. A. Edgar and P. H. O'Farrell. Genetic control of cell division patterns in the Drosophila embryo. Cell, 57(1):177–187, 1989. ISSN 0092-8674 (Print) 0092-8674 (Linking). doi: 10.1016/0092-8674(89)90183-9. URL https://www.ncbi.nlm.nih.gov/pubmed/2702688.

B. A. Edgar, M. P. Weir, G. Schubiger, and T. Kornberg. Repression and turnover pattern fushi tarazu RNA in the early Drosophila embryo. Cell, 47(5):747–754, 1986a. ISSN 0092-8674 (Print) 0092-8674 (Linking). doi: 10.1016/0092-8674(86)90517-9. URL https://www.ncbi.nlm.nih.gov/pubmed/3096577.

B. A. Edgar, M. P. Weir, G. Schubiger, and T. Kornberg. Repression and turnover pattern fushi tarazu RNA in the early Drosophila embryo. Cell, 1986b. ISSN 00928674. doi: 10.1016/0092-8674(86)90517-9.

E. El-Sherif and M. Levine. Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the Drosophila Embryo. Current Biology, 2016. ISSN 09609822. doi: 10.1016/j.cub.2016.02.054.

W. D. Fakhouri, A. Ay, R. Sayal, J. Dresch, E. Dayringer, and D. N. Arnosti. Deciphering a transcriptional regulatory code: modeling short-range repression in the Drosophila embryo. Mol Syst Biol, 6:341, 2010. ISSN 1744-4292 (Electronic) 1744-4292 (Linking). doi: 10.1038/msb.2009.97. URL https://www.ncbi.nlm.nih.gov/pubmed/20087339.

J. V. Falvo, D. Thanos, and T. Maniatis. Reversal of intrinsic DNA bends in the IFNβ gene enhancer by transcription factors and the architectural protein HMG I(Y). Cell, 1995. ISSN 00928674. doi: 10.1016/0092-8674(95)90137-X.

C. C. Fowlkes, C. L. Luengo Hendriks, S. V. Keränen, G. H. Weber, O. Rübel, M. Y. Huang, S. Chatoor, A. H. DePace, L. Simirenko, C. Henriquez, A. Beaton, R. Weiszmann, S. Celniker, B. Hamann, D. W. Knowles, M. D. Biggin, M. B. Eisen, and J. Malik. A Quantitative Spatiotemporal Atlas of Gene Expression in the Drosophila Blastoderm. Cell, 2008. ISSN 00928674. doi: 10.1016/j.cell.2008.01.053.

M. Fujioka, Y. Emi-Sarker, G. L. Yusibova, T. Goto, and J. B. Jaynes. Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. Development, 1999. ISSN 09501991.

T. Fukaya, B. Lim, and M. Levine. Enhancer Control of Transcriptional Bursting. Cell, 166(2):358–368, 2016. ISSN 1097-4172 (Electronic) 0092-8674 (Linking). doi: 10.1016/j.cell.2016.05.025. URL `https://www.ncbi.nlm.nih.gov/pubmed/27293191`.

H. G. Garcia, M. Tikhonov, A. Lin, and T. Gregor. Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr Biol, 23 (21):2140–2145, 2013. ISSN 1879-0445 (Electronic) 0960-9822 (Linking). doi: 10.1016/j.cub.2013.08.054. URL `https://www.ncbi.nlm.nih.gov/pubmed/24139738`.

T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek. Probing the Limits to Positional Information. Cell, 2007a. ISSN 00928674. doi: 10.1016/j.cell.2007.05.025.

T. Gregor, E. F. Wieschaus, A. P. McGregor, W. Bialek, and D. W. Tank. Stability and Nuclear Dynamics of the Bicoid Morphogen Gradient. Cell, 2007b. ISSN 00928674. doi: 10.1016/j.cell.2007.05.026.

T. Gregor, H. G. Garcia, and S. C. Little. The embryo as a laboratory: quantifying transcription in Drosophila. Trends Genet, 30(8):364–375, 2014. ISSN 0168-9525 (Print) 0168-9525 (Linking). doi: 10.1016/j.tig.2014.06.002. URL `https://www.ncbi.nlm.nih.gov/pubmed/25005921`.

S. R. Grossman, X. Zhang, L. Wang, J. Engreitz, A. Melnikov, P. Rogov, R. Tewhey, A. Isakova, B. Deplancke, B. E. Bernstein, T. S. Mikkelsen, and E. S. Lander. Systematic dissection of genomic features determining transcription factor binding and enhancer function. Proceedings of the National Academy of Sciences of the United States of America, 2017. ISSN 10916490. doi: 10.1073/pnas.1621150114.

S. D. Hanes, G. Riddihough, D. Ish-Horowicz, and R. Brent. Specific DNA recognition and intersite spacing are critical for action of the bicoid morphogen. Molecular and Cellular Biology, 1994. ISSN 0270-7306. doi: 10.1128/mcb.14.5.3364.

M. Hoch, C. Schröder, E. Seifert, and H. Jäckle. cis-acting control elements for Krüppel expression in the Drosophila embryo. The EMBO Journal, 1990. ISSN 0261-4189. doi: 10.1002/j.1460-2075.1990.tb07440.x.

M. Hoch, E. Seifert, and H. Jäckle. Gene expression mediated by cis-acting sequences of the Krüppel gene in response to the Drosophila morphogens bicoid and hunchback. The EMBO Journal, 1991. ISSN 0261-4189. doi: 10.1002/j.1460-2075.1991.tb07763.x.

L. Hostettler, L. Grundy, S. Kaser-Pebernard, C. Wicky, W. R. Schafer, and D. A. Glauser. The Bright Fluorescent Protein mNeonGreen Facilitates Protein Expression Analysis In Vivo. G3 (Bethesda), 7(2):607–615, 2017. ISSN 2160-1836 (Electronic) 2160-1836 (Linking). doi: 10.1534/g3.116.038133. URL `https://www.ncbi.nlm.nih.gov/pubmed/28108553`.

S. C. Hughes and H. M. Krause. Double labeling with fluorescence in situ hybridization in Drosophila whole-mount embryos. BioTechniques, 1998. ISSN 07366205. doi: 10.2144/98244bm01.

P. S. I. Gray and M. Levine. Short-range repression permits multiple enhancers to function autonomously within a complex promoter. Genes and Development, 8(15):1829–1838, 1994.

L. J. J. and Ma. Drosophila Bicoid is a substrate of sumoylation and its activator function is subject to inhibition by this post-translational modification. FEBS Lett., 586(12):1719–1723, 2012. doi: 10.1016/j.febslet.2012.04.059.

J. Jaeger. The gap gene network. Cell Mol Life Sci, 68(2):243–274, 2011. ISSN 1420-9071 (Electronic) 1420-682X (Linking). doi: 10.1007/s00018-010-0536-y. URL `https://www.ncbi.nlm.nih.gov/pubmed/20927566`.

C. Jung, P. Bandilla, M. Von Reutern, M. Schnepf, S. Rieder, U. Unnerstall, and U. Gaul. True equilibrium measurement of transcription factor-DNA binding affinities using automated polarization microscopy. Nature Communications, 2018. ISSN 20411723. doi: 10.1038/s41467-018-03977-4.

G. Junion, M. Spivakov, C. Girardot, M. Braun, E. H. Gustafson, E. Birney, and E. E. Furlong. A transcription factor collective defines cardiac cell fate and reflects lineage history. Cell, 2012. ISSN 00928674. doi: 10.1016/j.cell.2012.01.030.

N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom, and E. Segal. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature, 2009. ISSN 00280836. doi: 10.1038/nature07667.

N. Karaiskos, P. Wahle, J. Alles, A. Boltengagen, S. Ayoub, C. Kipar, C. Kocks, N. Rajewsky, and R. P. Zinzen. The Drosophila embryo at single-cell transcriptome resolution. Science, 358(6360):194–199, 2017. ISSN 1095-9203 (Electronic) 0036-8075 (Linking). doi: 10.1126/science.aan3235. URL https://www.ncbi.nlm.nih.gov/pubmed/28860209.

J. Kehle, D. Beuchle, S. Treuheit, B. Christen, J. A. Kennison, and M. Bienz. dMi-2, a hunchback-interacting protein that functions in Polycomb repression. Science, 1998. ISSN 00368075. doi: 10.1126/science.282.5395.1897.

S. Kosugi, M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita, and H. Yanagawa. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J Biol Chem, 284(1):478–485, 2009. ISSN 0021-9258 (Print) 0021-9258 (Linking). doi: 10.1074/jbc.M807017200. URL https://www.ncbi.nlm.nih.gov/pubmed/19001369.

J. P. Kreiss and S. N. Lahiri. Bootstrap Methods for Time Series. In Handbook of Statistics. 2012. doi: 10.1016/B978-0-444-53858-1.00001-6.

E. Z. Kvon, T. Kazmar, G. Stampfel, J. O. Yáñez-Cuna, M. Pagani, K. Schernhuber, B. J. Dickson, and A. Stark. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature, 2014. ISSN 14764687. doi: 10.1038/nature13395.

T. J. Lambert. FPbase: a community-editable fluorescent protein database. Nat Methods, 16(4):277–278, 2019. ISSN 1548-7105 (Electronic) 1548-7091 (Linking). doi: 10.1038/s41592-019-0352-8. URL https://www.ncbi.nlm.nih.gov/pubmed/30886412.

C. L. Lawson and R. J. Hanson. Solving Least Squares Problems. 1995. doi: 10.1137/1.9781611971217.

D. Lebrecht, M. Foehr, E. Smith, F. J. Lopes, C. E. Vanario-Alonso, J. Reinitz, D. S. Burz, and S. D. Hanes. Bicoid cooperative DNA binding is critical for embryonic

patterning in Drosophila. Proc Natl Acad Sci U S A, 102(37):13176–13181, 2005. ISSN 0027-8424 (Print) 0027-8424 (Linking). doi: 10.1073/pnas.0506462102. URL https://www.ncbi.nlm.nih.gov/pubmed/16150708.

M. Levine. Transcriptional enhancers in animal development and evolution. Curr Biol, 20(17):R754–63, 2010. ISSN 1879-0445 (Electronic) 0960-9822 (Linking). doi: 10.1016/j.cub.2010.06.070. URL https://www.ncbi.nlm.nih.gov/pubmed/20833320.

M. Levo, E. Zalckvar, E. Sharon, A. C. D. Machado, Y. Kalma, M. Lotam-Pompan, A. Weinberger, Z. Yakhini, R. Rohs, and E. Segal. Unraveling determinants of transcription factor binding outside the core binding site. Genome Research, 2015. ISSN 15495469. doi: 10.1101/gr.185033.114.

G. W. Li and X. S. Xie. Central dogma at the single-molecule level in living cells. Nature, 475(7356):308–315, 2011. ISSN 0028-0836. doi: 10.1038/nature10315. URL {%}3CGoto.

L. M. Li and D. N. Arnosti. Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes. Current Biology, 2011. ISSN 09609822. doi: 10.1016/j.cub.2011.01.054.

X. Y. Li, M. M. Harrison, J. E. Villalta, T. Kaplan, and M. B. Eisen. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. Elife, 3, 2014. ISSN 2050-084X (Electronic) 2050-084X (Linking). doi: 10.7554/eLife.03737. URL https://www.ncbi.nlm.nih.gov/pubmed/25313869.

H. L. Liang, C. Y. Nien, H. Y. Liu, M. M. Metzstein, N. Kirov, and C. Rushlow. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature, 456(7220):400–403, 2008. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/nature07388. URL https://www.ncbi.nlm.nih.gov/pubmed/18931655.

L. M. Liberman, G. T. Reeves, and A. Stathopoulos. Quantitative imaging of the Dorsal nuclear gradient reveals limitations to threshold-dependent patterning in Drosophila. Proceedings of the National Academy of Sciences of the United States of America, 2009. ISSN 00278424. doi: 10.1073/pnas.0906227106.

S. C. Little and T. Gregor. Single mRNA Molecule Detection in Drosophila. Methods Mol Biol, 1649:127–142, 2018. ISSN 1940-6029 (Electronic) 1064-3745 (Linking). doi: 10.1007/978-1-4939-7213-5_8. URL https://www.ncbi.nlm.nih.gov/pubmed/29130194.

J. Liu and J. Ma. Modulation of temporal dynamics of gene transcription by activator potency in the Drosophila embryo. Development, 142(21):3781–3790, 2015. ISSN 1477-9129 (Electronic) 0950-1991 (Linking). doi: 10.1242/dev.126946. URL https://www.ncbi.nlm.nih.gov/pubmed/26395487.

S. Lubliner, I. Regev, M. Lotan-Pompan, S. Edelheit, A. Weinberger, and E. Segal. Core promoter sequence in yeast is a major determinant of expression level. Genome Research, 2015. ISSN 15495469. doi: 10.1101/gr.188193.114.

T. Lucas, T. Ferraro, B. Roelens, J. De Las Heras Chanes, A. M. Walczak, M. Coppey, and N. Dostatni. Live imaging of bicoid-dependent transcription in Drosophila embryos. Curr Biol, 23(21):2135–2139, 2013. ISSN 1879-0445 (Electronic) 0960-9822 (Linking). doi: 10.1016/j.cub.2013.08.053. URL https://www.ncbi.nlm.nih.gov/pubmed/24139736.

M. Z. Ludwig, Manu, R. Kittler, K. P. White, and M. Kreitman. Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. PLoS Genet, 7(11):e1002364, 2011a. ISSN 1553-7404 (Electronic) 1553-7390 (Linking). doi: 10.1371/journal.pgen.1002364. URL https://www.ncbi.nlm.nih.gov/pubmed/22102826.

M. Z. Ludwig, Manu, R. Kittler, K. P. White, and M. Kreitman. Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. PLoS Genetics, 2011b. ISSN 15537390. doi: 10.1371/journal.pgen.1002364.

A. C. Magico and J. B. Bell. Identification of a classical bipartite nuclear localization signal in the Drosophila TEA/ATTS protein scalloped. PLoS One, 6(6):e21431, 2011. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi: 10.1371/journal.pone.0021431. URL https://www.ncbi.nlm.nih.gov/pubmed/21731746.

L. H. Margaritis, F. C. Kafatos, and W. H. Petri. The eggshell of Drosophila melanogaster. I. Fine structure of the layers and regions of the wild-type eggshell. J Cell Sci, 43:1–35, 1980. ISSN 0021-9533 (Print) 0021-9533 (Linking). URL https://www.ncbi.nlm.nih.gov/pubmed/6774986.

C. A. Martinez and D. N. Arnosti. Spreading of a Corepressor Linked to Action of Long-Range Repressor Hairy. Molecular and Cellular Biology, 2008. ISSN 0270-7306. doi: 10.1128/mcb.01203-07.

G. A. Maston, S. K. Evans, and M. R. Green. Transcriptional Regulatory Elements in the Human Genome. Annual Review of Genomics and Human Genetics, 2006. ISSN 1527-8204. doi: 10.1146/annurev.genom.7.080505.115623.

M. Mavrakis, R. Rikhy, M. Lilly, and J. Lippincott-Schwartz. Fluorescence imaging techniques for studying Drosophila embryo development. Current Protocols in Cell Biology, 2008. ISSN 19342616. doi: 10.1002/0471143030.cb0418s39.

M. L. McCleland and P. H. O'Farrell. RNAi of mitotic cyclins in Drosophila uncouples the nuclear and centrosome cycle. Curr Biol, 18(4):245–254, 2008. ISSN 0960-9822 (Print) 0960-9822 (Linking). doi: 10.1016/j.cub.2008.01.041. URL `https://www.ncbi.nlm.nih.gov/pubmed/18291653`.

S. L. McDaniel, T. J. Gibson, K. N. Schulz, M. Fernandez Garcia, M. Nevil, S. U. Jain, P. W. Lewis, K. S. Zaret, and M. M. Harrison. Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation. Mol Cell, 74(1):185–195 e4, 2019. ISSN 1097-4164 (Electronic) 1097-2765 (Linking). doi: 10.1016/j.molcel.2019.01.014. URL `https://www.ncbi.nlm.nih.gov/pubmed/30797686`.

M. Merika and D. Thanos. Enhanceosomes, 2001. ISSN 0959437X.

M. Merika, A. J. Williams, G. Chen, T. Collins, and D. Thanos. Recruitment of CBP/p300 by the IFNβ enhanceosome is required for synergistic activation of transcription. Molecular Cell, 1998. ISSN 10972765. doi: 10.1016/S1097-2765(00)80028-3.

M. Mir, A. Reimer, J. E. Haines, X. Y. Li, M. Stadler, H. Garcia, M. B. Eisen, and X. Darzacq. Dense Bicoid hubs accentuate binding along the morphogen gradient. Genes Dev, 31(17):1784–1794, 2017. ISSN 1549-5477 (Electronic) 0890-9369 (Linking). doi: 10.1101/gad.305078.117. URL `https://www.ncbi.nlm.nih.gov/pubmed/28982761`.

A. Nasiadka, B. H. Dietrich, and H. M. Krause. Anterior-posterior patterning in the Drosophila embryo. Advances in Developmental Biology and Biochemistry, 12:155–204, 2002.

C. Nusslein-Volhard. Determination of the embryonic axes of Drosophila. Dev Suppl, 1: 1–10, 1991. URL `https://www.ncbi.nlm.nih.gov/pubmed/1742496`.

C. Nusslein-Volhard, H. G. Frohnhofer, and R. Lehmann. Determination of anteroposterior polarity in Drosophila. Science, 238(4834):1675–1681, 1987. ISSN 0036-8075

(Print) 0036-8075 (Linking). doi: 10.1126/science.3686007. URL https://www.ncbi.nlm.nih.gov/pubmed/3686007.

D. Pan and A. Courey. The same dorsal binding site mediates both activation and repression in a context-dependent manner. The EMBO Journal, 1992. ISSN 0261-4189. doi: 10.1002/j.1460-2075.1992.tb05235.x.

D. Papatsenko and M. S. Levine. Dual regulation by the Hunchback gradient in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 2008. ISSN 00278424. doi: 10.1073/pnas.0711941105.

S. Payankaulam and D. N. Arnosti. Gene Regulation: Boundaries within Limits, 2008. ISSN 09609822.

L. A. Pennacchio, W. Bickmore, A. Dean, M. A. Nobrega, and G. Bejerano. Enhancers: five essential questions. Nat Rev Genet, 14(4):288–295, 2013. ISSN 1471-0064 (Electronic) 1471-0056 (Linking). doi: 10.1038/nrg3458. URL https://www.ncbi.nlm.nih.gov/pubmed/23503198.

M. W. Perry, A. N. Boettiger, J. P. Bothma, and M. Levine. Shadow enhancers foster robustness of drosophila gastrulation. Current Biology, 2010. ISSN 09609822. doi: 10.1016/j.cub.2010.07.043.

M. W. Perry, A. N. Boettiger, and M. Levine. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 2011. ISSN 00278424. doi: 10.1073/pnas.1109873108.

I. S. Peter and E. H. Davidson. Evolution of gene regulatory networks controlling body plan development. Cell, 144(6):970–985, 2011. ISSN 1097-4172 (Electronic) 0092-8674 (Linking). doi: 10.1016/j.cell.2011.02.017. URL https://www.ncbi.nlm.nih.gov/pubmed/21414487.

M. D. Petkova, S. C. Little, F. Liu, and T. Gregor. Maternal origins of developmental reproducibility. Current Biology, 2014. ISSN 09609822. doi: 10.1016/j.cub.2014.04.028.

M. D. Petkova, G. Tkacik, W. Bialek, E. F. Wieschaus, and T. Gregor. Optimal Decoding of Cellular Identities in a Genetic Network. Cell, 176(4):844–855 e15, 2019. ISSN 1097-4172 (Electronic) 0092-8674 (Linking). doi: 10.1016/j.cell.2019.01.007. URL https://www.ncbi.nlm.nih.gov/pubmed/30712870.

B. D. Pfeiffer, T. T. Ngo, K. L. Hibbard, C. Murphy, A. Jenett, J. W. Truman, and G. M. Rubin. Refinement of tools for targeted gene expression in Drosophila. Genetics, 186 (2):735–755, 2010. ISSN 1943-2631 (Electronic) 0016-6731 (Linking). doi: 10.1534/genetics.110.119917. URL https://www.ncbi.nlm.nih.gov/pubmed/20697123.

B. D. Pfeiffer, J. W. Truman, and G. M. Rubin. Using translational enhancers to increase transgene expression in Drosophila. Proc Natl Acad Sci U S A, 109(17):6626–6631, 2012. ISSN 1091-6490 (Electronic) 0027-8424 (Linking). doi: 10.1073/pnas.1204520109. URL https://www.ncbi.nlm.nih.gov/pubmed/22493255.

C. D. C. R. D. R. Yan S. Small and J. E. Darnell. Identification of a Stat gene that functions in Drosophila development. Cell, 84(3):421–430, 1996.

T. Raveh-Sadka, M. Levo, U. Shabi, B. Shany, L. Keren, M. Lotan-Pompan, D. Zeevi, E. Sharon, A. Weinberger, and E. Segal. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nature Genetics, 2012. ISSN 10614036. doi: 10.1038/ng.2305.

J. C. Reese. Basal transcription factors, 2003. ISSN 0959437X.

E. Ronchi, J. Treisman, N. Dostatni, G. Struhl, and C. Desplan. Down-regulation of the Drosophila morphogen bicoid by the torso receptor-mediated signal transduction cascade. Cell, 74(2):347–355, 1993. ISSN 0092-8674 (Print) 0092-8674 (Linking). doi: 10.1016/0092-8674(93)90425-p. URL https://www.ncbi.nlm.nih.gov/pubmed/8343961.

D. N. A. S. Small and M. Levine. Spacing ensures autonomous expression of different strip enhancers in the even-skipped promoter. Development, 119(3):762–772, 1993.

J. E. Sandler and A. Stathopoulos. Quantitative single-embryo profile of drosophila genome activation and the dorsal–ventral patterning network. Genetics, 2016. ISSN 19432631. doi: 10.1534/genetics.116.186783.

T. Sandmann, L. J. Jensen, J. S. Jakobsen, M. M. Karzynski, M. P. Eichenlaub, P. Bork, and E. E. Furlong. A Temporal Map of Transcription Factor Activity: Mef2 Directly Regulates Target Genes at All Stages of Muscle Development. Developmental Cell, 2006. ISSN 15345807. doi: 10.1016/j.devcel.2006.04.009.

R. Sayal, S. M. Ryu, and D. N. Arnosti. Optimization of reporter gene architecture for quantitative measurements of gene expression in the drosophila embryo. Fly, 2011. ISSN 19336942. doi: 10.4161/fly.5.1.14459.

M. Schnepf, C. Ludwig, P. Bandilla, S. Ceolin, U. Unnerstall, C. Jung, and U. Gaul. Sensitive Automated Measurement of Histone-DNA Affinities in Nucleosomes. iScience, 2020. ISSN 25890042. doi: 10.1016/j.isci.2020.100824.

C. Scholes, K. M. Biette, T. T. Harden, and A. H. DePace. Signal Integration by Shadow Enhancers and Enhancer Duplications Varies across the Drosophila Embryo. Cell Reports, 2019. ISSN 22111247. doi: 10.1016/j.celrep.2019.01.115.

M. D. Schroeder, M. Pearce, J. Fak, H. Q. Fan, U. Unnerstall, E. Emberly, N. Rajewsky, E. D. Siggia, and U. Gaul. Transcriptional control in the segmentation gene network of Drosophila. PLoS Biology, 2004. ISSN 15449173. doi: 10.1371/journal.pbio.0020271.

M. D. Schroeder, C. Greer, and U. Gaul. How to make stripes: Deciphering the transition from nonperiodic to periodic patterns in Drosophila segmentation. Development, 2011. ISSN 09501991. doi: 10.1242/dev.062141.

E. Segal and J. Widom. Poly(dA:dT) tracts: major determinants of nucleosome organization, 2009. ISSN 0959440X.

E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall, and U. Gaul. Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature, 451(7178):535–540, 2008. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/nature06496. URL https://www.ncbi.nlm.nih.gov/pubmed/18172436.

N. C. Shaner, G. G. Lambert, A. Chammas, Y. Ni, P. J. Cranfill, M. A. Baird, B. R. Sell, J. R. Allen, R. N. Day, M. Israelsson, M. W. Davidson, and J. Wang. A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat Meth, 10(5):407–409, 2013. ISSN 1548-7091. doi: 10.1038/nmeth.2413http://www.nature.com/nmeth/journal/v10/n5/abs/nmeth.2413.html#supplementary-information. URL http://dx.doi.org/10.1038/nmeth.2413.

K. Shrinivas, B. R. Sabari, E. L. Coffey, I. A. Klein, A. Boija, A. V. Zamudio, J. Schuijers, N. M. Hannett, P. A. Sharp, R. A. Young, and A. K. Chakraborty. Enhancer Features that Drive Formation of Transcriptional Condensates. Molecular Cell, 2019. ISSN 10974164. doi: 10.1016/j.molcel.2019.07.009.

M. Simpson-Brose, J. Treisman, and C. Desplan. Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. Cell, 1994. ISSN 00928674. doi: 10.1016/S0092-8674(94)90622-X.

S. Small, A. Blair, and M. Levine. Regulation of even-skipped stripe 2 in the Drosophila embryo. The EMBO Journal, 1992. ISSN 0261-4189. doi: 10.1002/j.1460-2075.1992.tb05498.x.

S. Small, A. Blair, and M. Levine. Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. Dev Biol, 175(2):314–324, 1996. ISSN 0012-1606 (Print) 0012-1606 (Linking). doi: 10.1006/dbio.1996.0117. URL `https://www.ncbi.nlm.nih.gov/pubmed/8626035`.

F. Spitz and E. E. Furlong. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet, 13(9):613–626, 2012. ISSN 1471-0064 (Electronic) 1471-0056 (Linking). doi: 10.1038/nrg3207. URL `https://www.ncbi.nlm.nih.gov/pubmed/22868264`.

M. V. Staller, B. J. Vincent, M. D. Bragdon, T. Lydiard-Martin, Z. Wunderlich, J. Estrada, and A. H. DePace. Shadow enhancers enable hunchback bifunctionality in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America, 2015. ISSN 10916490. doi: 10.1073/pnas.1413877112.

P. Struffi, M. Corado, L. Kaplan, D. Yu, C. Rushlow, and S. Small. Combinatorial activation and concentration-dependent repression of the drosophila even skipped stripe 3+7 enhancer. Development, 2011. ISSN 09501991. doi: 10.1242/dev.065987.

Y. Sun, C. Y. Nien, K. Chen, H. Y. Liu, J. Johnston, J. Zeitlinger, and C. Rushlow. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. Genome Res, 25(11):1703–1714, 2015. ISSN 1549-5469 (Electronic) 1088-9051 (Linking). doi: 10.1101/gr.192542.115. URL `https://www.ncbi.nlm.nih.gov/pubmed/26335633`.

I. Sur and J. Taipale. The role of enhancers in cancer. Nat Rev Cancer, 16(8):483–493, 2016. ISSN 1474-1768 (Electronic) 1474-175X (Linking). doi: 10.1038/nrc.2016.62. URL `https://www.ncbi.nlm.nih.gov/pubmed/27364481`.

O. Symmons, L. Pan, S. Remeseiro, T. Aktas, F. Klein, W. Huber, and F. Spitz. The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. Dev Cell, 39(5):529–543, 2016. ISSN 1878-1551 (Electronic) 1534-5807 (Linking). doi: 10.1016/j.devcel.2016.10.015. URL `https://www.ncbi.nlm.nih.gov/pubmed/27867070`.

P. Szymanski and M. Levine. Multiple modes of dorsal-bHLH transcriptional synergy in the Drosophila embryo. The EMBO Journal, 1995. ISSN 0261-4189. doi: 10.1002/j.1460-2075.1995.tb07217.x.

K. B. E. Z. W. A. H. D. Tara Lydiard-Martin Meghan Bragdon. Locus architecture affects mRNA expression levels in Drosophila embryos. bioRxiv, 2014. doi: https://doi.org/10.1101/005173.

S. Thomas, X. Y. Li, P. J. Sabo, R. Sandstrom, R. E. Thurman, T. K. Canfield, E. Giste, W. Fisher, A. Hammonds, S. E. Celniker, M. D. Biggin, and J. A. Stamatoyannopoulos. Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. Genome Biology, 2011. ISSN 14747596. doi: 10.1186/gb-2011-12-5-r43.

D. W. Thompson. On Growth and form. Cambridge University Press, 1917.

P. Tomancak, A. Beaton, R. Weiszmann, E. Kwan, S. Q. Shu, S. E. Lewis, S. Richards, M. Ashburner, V. Hartenstein, S. E. Celniker, and G. M. Rubin. Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome biology, 2002. ISSN 14656914. doi: 10.1186/gb-2002-3-12-research0088.

A. Turing. The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 1952. ISSN 2054-0280. doi: 10.1098/rstb.1952.0012.

B. J. Vincent, J. Estrada, and A. H. DePace. The appeasement of Doug: a synthetic approach to enhancer biology. Integr Biol (Camb), 8(4):475–84, 2016.

B. J. Vincent, M. V. Staller, F. Lopez-Rivera, M. D. Bragdon, E. C. Pym, K. M. Biette, Z. Wunderlich, T. T. Harden, J. Estrada, and A. H. DePace. Hunchback is counter-repressed to regulate even-skipped stripe 2 expression in Drosophila embryos. PLoS Genetics, 2018. ISSN 15537404. doi: 10.1371/journal.pgen.1007644.

S. A. Vokes, H. Ji, W. H. Wong, and A. P. McMahon. A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. Genes and Development, 2008. ISSN 08909369. doi: 10.1101/gad.1693008.

T. C. Voss, R. L. Schiltz, M. H. Sung, P. M. Yen, J. A. Stamatoyannopoulos, S. C. Biddie, T. A. Johnson, T. B. Miranda, S. John, and G. L. Hager. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. Cell, 2011. ISSN 00928674. doi: 10.1016/j.cell.2011.07.006.

R. Warn, B. Bullard, and S. Maleki. Myosin as a constituent of the Drosophila egg cortex [18], 1979. ISSN 00280836.

T. T. Weil, R. M. Parton, and I. Davis. Making the message clear: Visualizing mRNA localization, 2010. ISSN 09628924.

S. Yamada, P. H. Whitney, S. K. Huang, E. C. Eck, H. G. Garcia, and C. A. Rushlow. The Drosophila Pioneer Factor Zelda Modulates the Nuclear Microenvironment of a Dorsal Target Enhancer to Potentiate Transcriptional Output. Curr Biol, 29(8):1387–1393 e5, 2019. ISSN 1879-0445 (Electronic) 0960-9822 (Linking). doi: 10.1016/j.cub.2019.03.019. URL `https://www.ncbi.nlm.nih.gov/pubmed/30982648`.

D. Yu and S. Small. Precise Registration of Gene Expression Boundaries by a Repressive Morphogen in Drosophila. Current Biology, 2008. ISSN 09609822. doi: 10.1016/j.cub.2008.05.050.

V. Zulkower, M. Page, D. Ropers, J. Geiselmann, and H. de Jong. Robust reconstruction of gene expression profiles from reporter gene data using linear inversion. Bioinformatics, 31(12):i71–9, 2015. ISSN 1367-4811 (Electronic) 1367-4803 (Linking). doi: 10.1093/bioinformatics/btv246. URL `https://www.ncbi.nlm.nih.gov/pubmed/26072511`.

P. Zuo, D. Stanojevic, J. Colgan, K. Han, M. Levine, and J. L. Manley. Activation and repression of transcription by the gap proteins hunchback and Kruppel in cultured Drosophila cells. Genes and Development, 1991. ISSN 08909369. doi: 10.1101/gad.5.2.254.

Acknowledgements

Dear Ulrike and Ulrich, when our paths first crossed I was captured by the enthusiasm of the community you created here in Munich: the lab and the graduate school that you founded and shaped were a beautiful group of people brought together by their common passions. The school and the lab were thriving and full of life. As sometimes happens during our lives, events take unpredictable turns, but the memories of that enthusiasm will accompany me forever. Thank you so much for giving me the opportunity to work in this fantastic group and for introducing me to the most interesting questions of a research field that was totally new to me when I started this journey.

Dear Ulrich, I have learned from you a lot in these five years and not only professionally or about science, but well beyond. I admired the way you kept together the lab, without ever sacrificing the scientific content of our work and defending the spirit of each project despite all the constraints our lab had to deal with. I will be forever grateful for all the support you gave me and the time you spent to discuss my work, always with a joyful smile, during these years.

Dear Christophe, thanks for your friendship, your expert help, your hard work, constant support and pragmatic advice. Your contribution has been fundamental to turn this project into a success. Working together almost on a daily basis to overcome all the difficulties we encountered during this journey has been a great fun and joy.

A heartfelt thanks goes to all the former members of the Gaul lab that I had the great luck to meet during these years. In particular, thanks to Monika, for her incredible organization, great help and hard work; to Peter, who is one of the most knowledgeable persons I ever met and has the incredible ability to fix literally anything; to Andrea and Marta, for teaching me everything I got to know about molecular biology; to Max, that could always help me moving forward in the right direction by asking the right question during every meeting; and to Sara, Alessio, Zhan, Miro, Sabine, Roberto, Marc, Anja, Claudia, Shao-Yen and Liliana, for creating such a great fun, friendly and supportive environment.

I am very grateful to the members of my thesis advisory committee, Prof. Don Lamb

and Prof. Nicolas Gompel, for their valuable feedback. I am extremely grateful to the Graduate School of Quantitative Biosciences Munich for creating a very supportive and stimulating environment. Thanks to Markus, Filiz, Mara and Julia for having managed to create a wonderful interdisciplinary group of young scientists by teaching a bunch of physicists and biologists how to talk to each other.

If I never lost my good spirit throughout these years all the merit goes to my good friends and my extended family in Pordenone and Oristano. You make up everything that is really important in my life and you all supported me much more than you know with uncountable dinners, dream holidays, exciting concerts, amazing day trips, frequent phone calls and simple relaxing moments.

Finally, thank you once more, Fede, for being by my side during all these years through good and tough times, for putting up with the ups and downs of my mood, for your help and encouragement and, most importantly, for your friendship and love.