Cell Systems

Article

Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient

Highlights

- A model for whether and when a gene will turn on and at what rate it will transcribe
- Minimal synthetic enhancers allow a systematic test of model predictions in an embryo
- DNA binding of the Dorsal transcriptional activator catalyzes transcriptional onset
- Once transcription ensues, Dorsal DNA occupancy can dictate the transcription rate

Authors

Simon Alamos, Armando Reimer, Clay Westrum, ..., Jiaxi Zhao, Emma Luu, Hernan G. Garcia

Correspondence

hggarcia@berkeley.edu

In brief

Developmental enhancers dictate whether a locus engages in transcription and, if so, at what rate it will produce mRNA. Using theory and live imaging in *Drosophila* embryos, Alamos and Reimer et al. propose a model for how a transcriptional activator binding to an enhancer may dictate these multiple regulatory dynamics.

Alamos et al., 2023, Cell Systems *14*, 1–17
March 15, 2023
https://doi.org/10.1016/j.cels.2022.12.008

CellPress

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

Cell Systems

CellPress

Article

Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient

Simon Alamos,[1,7] Armando Reimer,[2,7] Clay Westrum,[3] Meghan A. Turner,[1] Paul Talledo,[4] Jiaxi Zhao,[3] Emma Luu,[3] and Hernan G. Garcia[2,3,4,5,6,8,*]

[1]Department of Plant and Microbial Biology, University of California at Berkeley, Berkeley, CA, USA
[2]Biophysics Graduate Group, University of California at Berkeley, Berkeley, CA, USA
[3]Department of Physics, University of California at Berkeley, Berkeley, CA, USA
[4]Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA, USA
[5]Institute for Quantitative Biosciences-QB3, University of California at Berkeley, Berkeley, CA, USA
[6]Chan Zuckerberg Biohub, San Francisco, CA, USA
[7]These authors contributed equally
[8]Lead contact
*Correspondence: hggarcia@berkeley.edu
https://doi.org/10.1016/j.cels.2022.12.008

SUMMARY

How enhancers interpret morphogen gradients to generate gene expression patterns is a central question in developmental biology. Recent studies have proposed that enhancers can dictate whether, when, and at what rate promoters engage in transcription, but the complexity of endogenous enhancers calls for theoretical models with too many free parameters to quantitatively dissect these regulatory strategies. To overcome this limitation, we established a minimal promoter-proximal synthetic enhancer in embryos of *Drosophila melanogaster*. Here, a gradient of the Dorsal activator is read by a single Dorsal DNA binding site. Using live imaging to quantify transcriptional activity, we found that a single binding site can regulate whether promoters engage in transcription in a concentration-dependent manner. By modulating the binding-site affinity, we determined that a gene's decision to transcribe and its transcriptional onset time can be explained by a simple model where the promoter traverses multiple kinetic barriers before transcription can ensue.

INTRODUCTION

The adoption of distinct cellular identities in multicellular organisms relies on the formation of spatial gene expression domains driven, in large part, by transcriptional regulatory programs. The positional information giving rise to these mRNA patterns is typically provided by transcription factor gradients (Figure 1A) whose concentrations are interpreted by enhancer DNA sequences that, in turn, regulate transcription of developmental genes.[1,2] A long-standing goal in quantitative developmental biology is to precisely predict gene expression from knowledge of the DNA regulatory sequence and morphogen concentration.[3,4] Achieving this predictive understanding requires theoretical models that calculate how DNA sequence dictates the functional relation between input morphogen concentration and output transcriptional activity and calls for testing these predictions by measuring input-output functions.[3] Precise genetic manipulations[5,6] and powerful imaging technologies[7–9] have rendered the early embryo of the fruit fly *Drosophila melanogaster* (*Drosophila*) a prime model system for quantitatively dissecting these input-output functions in development.

In recent years, several studies have reported that *Drosophila* enhancers can control various, potentially independent aspects of transcriptional dynamics in early embryonic development.[8,10–17] First, for a given gene, a fraction of loci remain transcriptionally inactive throughout entire mitotic cycles in development, even when exposed to the same activator concentration as active loci (Figure 1B)—a behavior usually quantified through the fraction of active nuclei or loci. This stochastic decision for a locus to become active is a ubiquitous and potentially important regulatory feature for shaping gene-expression patterns in the embryo.[8,12,17,18] However, it remains unclear whether this feature constitutes a regulatory "knob" or whether inactive loci are artifacts of experimental detection thresholds. Second, the timing of transcription onset (and cessation, which is not addressed in the present investigation) can also be controlled by input transcription-factor dynamics[12,14,17–21] (Figure 1C). Finally, the rate of transcriptional initiation in active loci is under regulatory control (Figure 1D) and has been the focus of most studies to date.[8,11,12,15,16] Thus, multiple regulatory strategies together realize gene-expression patterns in space and time.

Intense theoretical scrutiny[14,18,19,22–27] has generated a compelling hypothesis that the regulation of transcriptional

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

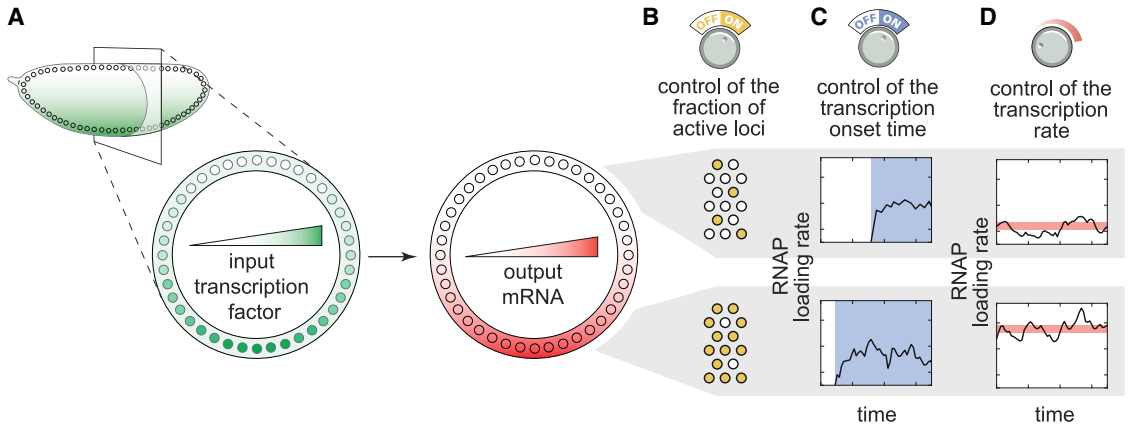

Figure 1. Transcriptional regulatory strategies of enhancers in response to transcription factor concentration gradients

(A) A *Drosophila* embryo with a transcription factor gradient along its dorsoventral axis. This input transcription factor dictates the emergence of output gene-expression patterns by controlling a combination of three enhancer regulatory "knobs": (B) the probability of loci becoming transcriptionally active, (C) the transcriptional onset time, and (D) the mean transcription rate of active loci.

dynamics can be separated into two stages. First, a promoter must pass through a series of kinetic barriers consisting of reactions catalyzed by transcription factors in order for loci to engage in transcription. Previous analyses of the mean and distribution of transcriptional onset times have suggested that the number of inactive promoter states can range from one to three.[14,17,18] These reactions could be associated with, for example, the stepwise unwrapping of DNA from nucleosomes[14,18,19] and the sequential recruitment of cofactors and elements of the general transcriptional machinery.[28] Second, after initial promoter activation, the rate of mRNA production is proportional to the probability of finding RNA polymerase II (RNAPII) bound to the promoter. Statistical mechanical (also called thermodynamic) models have been used to calculate this probability of finding RNAPII bound to the promoter and have shown to be successful in predicting mRNA production rates in bacteria.[29] However, whether these models can be applied to the more complex context of eukaryotic transcriptional regulation—let alone to the dynamical processes of cellular decision-making in development—is still an open question.[14,22,25,30–40]

One of the main challenges to systematically testing these models is the complexity of endogenous regulatory regions.[14,18,22–24,41] Because endogenous enhancers contain multiple binding sites for different transcription factors, accounting for these sites and their interactions leads to a combinatorial explosion of model parameters[3,42]; determining the values of these parameters from simple experiments constitutes a computational—and conceptual—challenge.[3,4,42] To render complex transcriptional regulatory systems tractable to theory, minimal synthetic enhancers have been engineered in bacteria,[29,43–45] eukaryotic cells,[46] and developing organisms.[23,24] In such experiments, a short, synthetic DNA sequence with only one to a few binding sites for a single transcription factor drives the expression of a reporter gene. As shown in detail in Box 1, measuring the concentration of the transcription-factor input and reporter mRNA output makes it possible to test models of transcriptional regulation and to infer molecular parameters that can be used to predict the behavior of more complex regulatory architectures.[45]

Here, we sought to use synthetic minimal promoter-proximal enhancers to challenge our integrated model of transcriptional control using the dorsoventral patterning system in *Drosophila* embryos, in which a concentration gradient of the Dorsal transcription factor specifies spatial domains of transcription, as a case study. To test the integrated model of transcriptional dynamics (Figures 3A and 3B), we performed simultaneous quantitative live-cell measurements of Dorsal concentration (input) and transcription (output) driven by minimal synthetic Dorsal-dependent promoter-proximal enhancers in single nuclei. By repurposing the *parS*-ParB DNA labeling technology[47,48] to quantify transcriptional activity independent of RNA detection, we determined that the inactive loci described by our model constitute a distinct transcriptional state under regulatory control and are not the result of detection artifacts. Further, our theoretical model predicted how, through the Dorsal-mediated catalysis of reactions prior to transcriptional onset, regulatory architecture dictates both the transcriptional onset time and the fraction of active loci. Finally, once promoters turn on, we found that our measurements are compatible with an equilibrium model. Thus, the present investigation provides quantitative evidence supporting a unified model of transcriptional regulation in eukaryotes that accounts for whether loci become transcriptionally active, when this activity ensues, and once transcription ensues, at what rate nascent RNA molecules are produced. More generally, our work demonstrates the feasibility of using minimal synthetic enhancers to engage in a dialog between theory and experiment in the context of transcriptional control in development.

RESULTS

An integrated model of transcriptional dynamics driven by a single activator binding site

To probe the transcriptional regulatory strategies of a minimal synthetic enhancer (Figure 1), we posit a theoretical model that predicts the fraction of loci that will become active, their transcriptional onset time, and RNAPII loading dynamics once

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

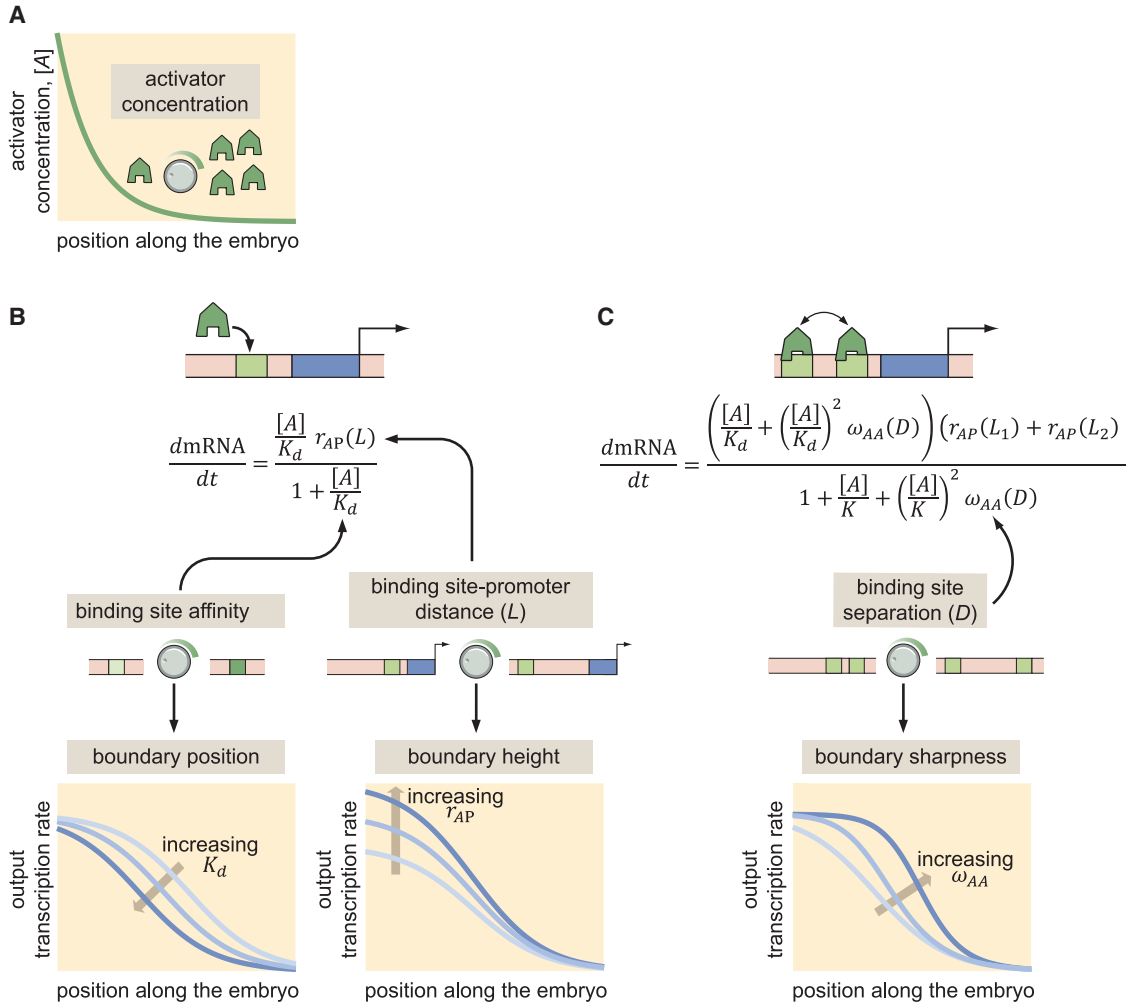

Figure 2. Iterative synthetic dissection of transcriptional control in development

(A) We consider an activator exponentially distributed along one of the axes of the embryo.

(B) A synthetic enhancer containing only one binding site can be described by a thermodynamic model with two parameters, the activator-DNA dissociation constant K_d and the transcription rate enhancement upon activator binding r_{AP}, which control the position and amplitude of the gene expression boundary driven by the enhancer, respectively.

(C) Adding one binding site to the synthetic enhancer introduces only one more free parameter ω_{AA} describing activator-activator interactions and dictating the sharpness of the developmental boundary (adapted from Garcia et al.[3]).

transcription ensues. Specifically, we consider a simplified case in which only one activator is present and can bind to one site only a few base pairs away from the promoter (Figure 3).

In order to explain the transcriptional onset dynamics of a locus and the probability of loci becoming active, we invoke recent experiments leading to a "kinetic barrier" model[14,18,19] proposing that, after exiting mitosis, all promoters are in an inactive state. In this state, labeled as "OFF_1" in Figure 3A, transcription is not possible. Promoters must then traverse a series of distinct inactive states (labeled "OFF_2" to "OFF_n" in Figure 3A) before reaching an active state in which transcription proceeds (labeled ON in Figure 3A).

The temporal evolution of the transcriptional dynamics as it traverses the states shown in Figure 3A can be simulated by computing the probability that the promoter occupies each state. Here, the transition rate between states, k, determines how the states probability spreads from the initial condition where the promoter is in state OFF_1 to the active state as time passes (see Methods S1.1 for details).

We propose that a transcriptional activator such as Dorsal can catalyze the transition between states in an affinity-dependent manner via binding to its cognate site in the enhancer. In this model, we assume that the transition rate k is much slower than transcription factor DNA binding, which has been shown to be on the order of a few seconds for transcription factors in the fly embryo,[9,56] and for the mammalian homolog of Dorsal.[57] As a result of this separation of timescales, we posit that Dorsal visits the enhancer multiple times before a transition between OFF states takes place such that the probability of finding Dorsal bound to the enhancer is proportional to its equilibrium occupancy. Under these conditions, the transition rate k is given by

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

Box 1. Bending nature to understand it

The inherent complexity of endogenous enhancers, with their plethora of binding sites for multiple transcription factors and protein-protein interactions, calls for complex theoretical descriptions with a multitude of free parameters. This explosion of free parameters makes it challenging to confront theoretical models against experiments.

An alternative to describing the complex reality of endogenous enhancers using complex theoretical models is to first reach a predictive understanding of simpler, synthetic regulatory architectures. As our predictive understanding increases, so too can the complexity of the regulatory regions assayed in an iterative cycle that, hopefully, will culminate with the understanding of endogenous regulatory regions.[3,42]

This idea of bending nature to make it simpler is illustrated in Figure 2. Consider, for example, an activator that is distributed in an exponential gradient along one of the axes of the fruit fly embryo (Figure 2A). Endogenous enhancers might contain multiple binding sites for this activator. However, a simpler synthetic enhancer bearing only one binding site for this activator could be created. As illustrated in Figure 2B, a theoretical description of the rate of mRNA production driven by this enhancer—based on thermodynamic models[49–55] for this particular illustrative example—would only have two free parameters: a dissociation constant for activator-DNA binding, K_d, and a parameter that captures the efficiency with which the activator increases transcription, r_{AP}. As shown in the figure, each of these free parameters dictate different aspects—boundary position and height—of the transcription profile. As a result, by measuring this profile and fitting to the model, a numerical estimate of each parameter can be obtained.

With a solid understanding of the single-activator enhancer system, the next iteration in this synthetic dissection calls for the addition of a second binding site for the same activator. As shown in Figure 2C, the prediction for the mRNA production rate looks more complicated than that for its single-binding site counterpart. However, a closer examination of the expression reveals that it only contains one free parameter: the activator-activator interaction term ω_{AA}, which dictates the sharpness of the boundary. As a result, by taking the parameters inferred from the previous iteration, the inference of ω_{AA} becomes much simpler.

Once a predictive understanding of this architecture with two binding sites is reached, the complexity can be further increased. Each iteration brings us closer to describing an endogenous enhancer. Of course, these regions do not exist in nature. However, we argue that there is little hope of predicting the input-output functions of endogenous enhancers if we cannot accomplish this feat in the much simpler context of the synthetic enhancers presented here.

$$k(t) = c \cdot \frac{\frac{[Dl](t)}{K_D}}{1 + \frac{[Dl](t)}{K_D}}, \qquad \text{(Equation 1)}$$

where c is a rate constant, [Dl](t) is the Dorsal concentration at time t, and K_D is the Dorsal-DNA dissociation constant.

Because Dorsal concentration varies in time, the model cannot be solved analytically. Thus, we numerically calculated the probability of the promoter being in each state as a function of time using a particular set of model parameters (see details in Methods S1.1). As seen in Figure 3C, since individual loci must traverse a sequence of intermediate states before reaching the ON state, this model introduces a delay in activation.

This kinetic barrier model accounts for loci that never transcribe during the nuclear cycle. Specifically, the model predicts that if nuclear cycles lasted indefinitely, all promoters would eventually reach the ON state as shown in Figure 3C. However, due to the rapid mitotic cycles that characterize early embryonic development in *Drosophila*, this duration is limited: transcription cannot initiate during mitosis and thus is only permissible during a time window within interphase (Figure 3C, vertical dashed line[8,14,58]). Consequently, if the time it takes a promoter to reach the ON state is longer than the duration of this window, then this promoter will not initiate transcription at all during the nuclear cycle (Figure 3C, horizontal dashed line).

The kinetic barrier model can be used to predict two of the three regulatory strategies, fraction of active loci and transcription onset times, that we aim to dissect quantitatively (Figure 1). First, the model predicts how the fraction of active loci is determined by Dorsal nuclear concentration and binding affinity (Figure 3D, left y axis). Second, this same model calculates the mean transcriptional onset time of those loci that turn on as a function of these same Dorsal parameters (Figure 3D, right y axis).

To model a locus once it is active, we follow Eck et al. [14] and propose a simple thermodynamic model[53,54] that assumes that the RNAPII loading rate, R, is proportional to the probability of finding RNAPII bound to the promoter p_{bound}, such that

$$R = R_{\max} \cdot p_{bound}, \qquad \text{(Equation 2)}$$

where $R_{\max}$ is a constant coefficient that dictates the maximum possible polymerase loading rate.

Thermodynamic models enable the calculation of p_{bound} by assigning a statistical weight to each possible state in which the regulatory system can be found. In the case of a minimal promoter-proximal enhancer with one activator binding site, the enhancer-promoter DNA can be empty, occupied by Dorsal, occupied by RNAPII, or simultaneously bound by Dorsal and RNAPII (Figure 3B). The statistical weight associated with each of these terms is shown in Figure 3B. Here, $[Dl]/K_D$ is the statistical weight associated with finding Dorsal (with concentration $[Dl]$ and binding dissociation constant K_D) bound to the promoter alone, whereas $[P]/K_P$ is the weight of finding RNAPII (with concentration $[P]$ and binding dissociation constant K_P) bound to the promoter alone. Note that the weight of having both Dorsal and RNAPII bound simultaneously includes an extra glue-like cooperativity coefficient, ω, that determines how strongly Dorsal recruits RNAPII to the promoter. The value of ω is constrained to be >1 so that higher Dorsal occupancy leads to higher RNAPII occupancy. This thermodynamic modeling approach also allows for more indirect forms of RNAPII recruitment by Dorsal such as binding mediated by cofactors. As shown in Figures S6 and S7, these more complex models make theoretical predictions that

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

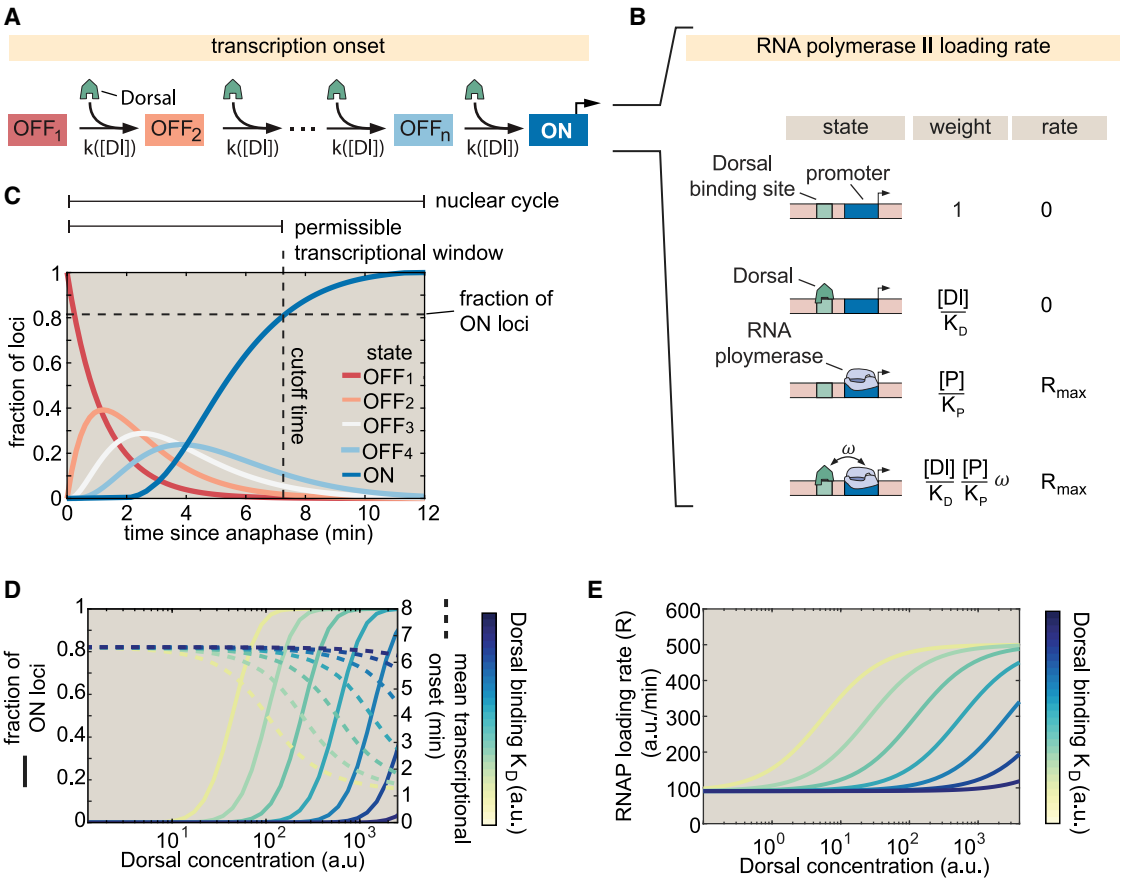

Figure 3. Integrated kinetic and thermodynamic model of simple activation by Dorsal

(A) The promoter undergoes kinetic transitions from transcriptionally inactive states (OFF_1 to OFF_n) to an active state (ON) with Dorsal accelerating the transition rate, k, by a factor proportional to the Dorsal occupancy at the promoter.

(B) Thermodynamic states and weights for the simple activator model. The probability of finding RNAPII bound to the promoter can be calculated from the statistical weights associated with all possible occupancy states of the proximal enhancer-promoter system.

(C) Visualization of a particular solution of the kinetic scheme from (A) showing the probability of finding a given locus in each of the states for an illustrative, representative set of parameters ($[Dl]$ = 1,000 a.u., K_D = 1,000 a.u., c = 10/min, n = 4 states, and 7 min nuclear cycle duration). The predicted fraction of active loci (dashed horizontal line) is calculated as the probability of being in the ON state by the end of the permissible time window (dashed vertical line) that is determined by mitotic repression.

(D) Predictions for the fraction of active loci (solid lines plotted against the left y axis) and mean transcriptional onset times (dashed lines plotted against the right y axis) as a function of Dorsal concentration for different, logarithmically spaced values of the Dorsal dissociation constant K_D in arbitrary units of Dorsal concentration. Note that under some parameter regimes, mean turn on times are similar across Dorsal concentrations.

(E) Rate of mRNA production across active loci as a function of Dorsal concentration for different values of K_D based on the model in (B) (R_{max} = 1,000 a.u., Dorsal K_D ranging from 10 to 10^5 a.u., ω = 10, $[P]/K_P$ = 0.1).

are essentially indistinguishable from those made by the simplest case considered in Figure 3B. As a result, throughout this work, we choose to entertain only the simplest model of direct Dorsal-RNAPII recruitment.

To calculate p_{bound}, we divide the sum of the weights featuring a bound RNAPII molecule by the sum of all possible weights. Substituting this calculation into Equation 2 yields

$$R = R_{max} \cdot p_{bound} = R_{max} \cdot \frac{\frac{[P]}{K_P} + \frac{[Dl]}{K_D}\frac{[P]}{K_P}\omega}{1 + \frac{[Dl]}{K_D} + \frac{[P]}{K_P} + \frac{[Dl]}{K_D}\frac{[P]}{K_P}\omega}, \qquad \text{(Equation 3)}$$

which is plotted in Figure 3E. As shown in the figure, increasing K_D shifts the concentration at which the RNAPII loading rate reaches half its maximum value toward higher Dorsal concentrations, but does not change the overall shape of the curve. We also note the presence of a non-zero baseline of RNAPII loading rate due to the Dorsal-independent $[P]/K_P$ term in the numerator of Equation 3. This baseline suggests that it could be possible for a promoter in the "ON" state to produce low, basal-level transcription in the absence of bound Dorsal.

Together, the kinetic barrier model outlined in Figure 3A and the thermodynamic model's Equation 3 define a comprehensive quantitative framework that predicts how the fraction of active loci, the transcriptional onset time, and the RNAPII loading rate as a function of Dorsal concentration vary as model parameters such as the Dorsal dissociation constant K_D are modulated (Figures 3D and 3E). These predictions constitute hypotheses

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

that we experimentally tested throughout the remainder of this work.

Establishing a minimal synthetic enhancer system to test theoretical predictions

To test our model's predictions, it is necessary to simultaneously measure transcription factor input and transcriptional output driven by a minimal regulatory system containing a single activator binding site. Thus, we sought to engineer and validate such a system in a developing embryo. To this end, we constructed single binding site promoter-proximal enhancers driven by the Dorsal activator, one of the best characterized transcription factors in *Drosophila* and a classic example of a morphogen.[59,60] Dorsal is provided maternally and forms a dorsoventral gradient of nuclear localization[61] (Figure 4A), acting as an activator by default[62,63] and as a repressor in the presence of nearby binding sites for corepressors.[64,65] Prior to activation of the zygotic genome (up to the 12th mitotic cycle), Dorsal is the only known transcription factor with a nuclear protein gradient across the dorsoventral axis.[66,67] Thus, the Dorsal nuclear concentration is the sole source of dorsoventral positional information for developmental enhancers at this stage in development. These features, combined, make Dorsal an ideal input transcription factor for activating a minimal synthetic reporter system.

In order to relate output transcriptional activity to the time-variant input Dorsal concentration throughout development, we measured the instantaneous Dorsal concentration in live embryos by creating a CRISPR knockin Dorsal-mVenus fusion allele based on a previous Dorsal fusion[60] that rescues embryonic development[68,69] (see Method details). Further, to increase the dynamic range of Dorsal concentration in our experiments, we combined this CRISPR allele with a Dorsal-mVenus transgene,[60] resulting in a line that will hereafter be referred to as 4× Dorsal flies. This fusion made it possible to quantify the concentration dynamics of the Dorsal protein input (Figures 4A and 4B) in individual nuclei (Video S1, left; materials and methods). Dorsal-mVenus nuclear fluorescence time traces quantified over nuclear cycle 12 confirmed the dynamic nature of Dorsal concentration and were quantitatively similar to previous measurements[60] (Figure 4B; details of Dorsal-mVenus quantification in Figures S8A and S8B). Nuclear cycle 12 nuclei in 4× Dorsal flies experience a Dorsal concentration gradient spanning several orders of magnitude, from less than 1 nM to more than 1 μM (Figure 4B; details of Dorsal-mVenus calibration in Figure S9).

To visualize the dynamics of Dorsal-dependent transcription, we built a reporter transgene containing a minimal synthetic promoter-proximal enhancer consisting of a single high affinity, consensus Dorsal binding site[70–72] (Figure 4C). Hereafter, we refer to this strong site enhancer as DBS_6.23 for Dorsal binding site, followed by its binding affinity score according to the Patser algorithm[73] (see Method details). To quantify the transcriptional activity of this enhancer, we used the MS2-MCP system to fluorescently label nascent RNA molecules in our reporter constructs, which appear as nuclear fluorescent puncta (hereafter "transcription spots") in laser-scanning confocal microscopy movies[8,10,74] (Video S1, right). We performed image analysis of the MS2 movies using a custom data analysis pipeline in Matlab and Fiji[12,75] (see Method details).

To validate this minimal synthetic system, we determined that DBS_6.23-MS2 drives quantifiable levels of transcription and that this transcriptional activity is mainly governed by Dorsal. We compared the transcriptional activity of DBS_6.23-MS2 in embryos laid by 4× Dorsal females with the activity in embryos laid by females homozygous for the *dl*^{1} *dorsal* null allele. Although transcription spots were clearly present in the 4× Dorsal background (Figure 4D, left), they were extremely rare in *dorsal* null embryos (Figure 4D, middle): not a single transcription spot was detected during nuclear cycle 12 in any of 4 replicates containing >60 nuclei in total. Dorsal is therefore necessary for transcriptional activity in our reporter constructs.

We next sought to determine whether the detected transcriptional activation is solely due to Dorsal interacting with the binding site we explicitly engineered into the construct or whether there are cryptic Dorsal binding sites contributing to gene expression. We generated a second reporter, DBS_4.29-MS2 in which the Dorsal binding site was strongly perturbed using known point mutations.[70] Transcription was rarely detectable in DBS_4.29-MS2 embryos (Figure 4D, right), with the average transcriptional activity (mean instantaneous fluorescence) per nucleus being less than 10% of the optimal DBS_6.23 enhancer at any Dorsal concentration (Figure S10). Thus, the Dorsal site placed within the synthetic enhancer is necessary for robust activation and is the main driver of its transcriptional activity.

Next, we asked whether the MS2 signal could be used as a reporter of Dorsal-dependent transcriptional activity that can be directly compared with our model predictions in terms of transcription onset time, transcription rate, and fraction of active nuclei. We collected DBS_6.23-MS2 time traces of MCP-mCherry fluorescence from transcription spots during nuclear cycle 12 along with the aforementioned three metrics of transcriptional activity (Figures 4E and 4F). First, the transcriptional onset time is defined as the time since the previous mitosis at which a transcription spot is first detected (Methods S1.4, Figure S4). Second, the maximum spot fluorescence corresponds to the 95th percentile of intensity over time, which is proportional to the transcription rate (Methods S1.2). Further, the integrated spot fluorescence corresponds to the time integral of the spot fluorescence and is directly proportional to the amount of mRNA produced by the locus[8] (see Method details). Finally, as previously observed in other genes in flies[8,12,17,18] and predicted by our model, not all nuclei exposed to similar nuclear Dorsal concentrations exhibited detectable transcription (Figure 4F). This failure of some nuclei to turn on and engage in transcription throughout the nuclear cycle is consistent with previous results from Dorsal-dependent synthetic enhancers that displayed a "salt and pepper" pattern even at peak Dorsal concentrations.[72] As a result, we quantified the fraction of active loci—regardless of their level of activity or temporal dynamics—by measuring the number of nuclei with observable transcription signal in at least one movie frame throughout nuclear cycle 12, divided by the total number of nuclei. Thus, we have established an experimental platform and quantitative metrics for Dorsal activity that enable us to engage in a dialog between experiment and theory.

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

CellPress

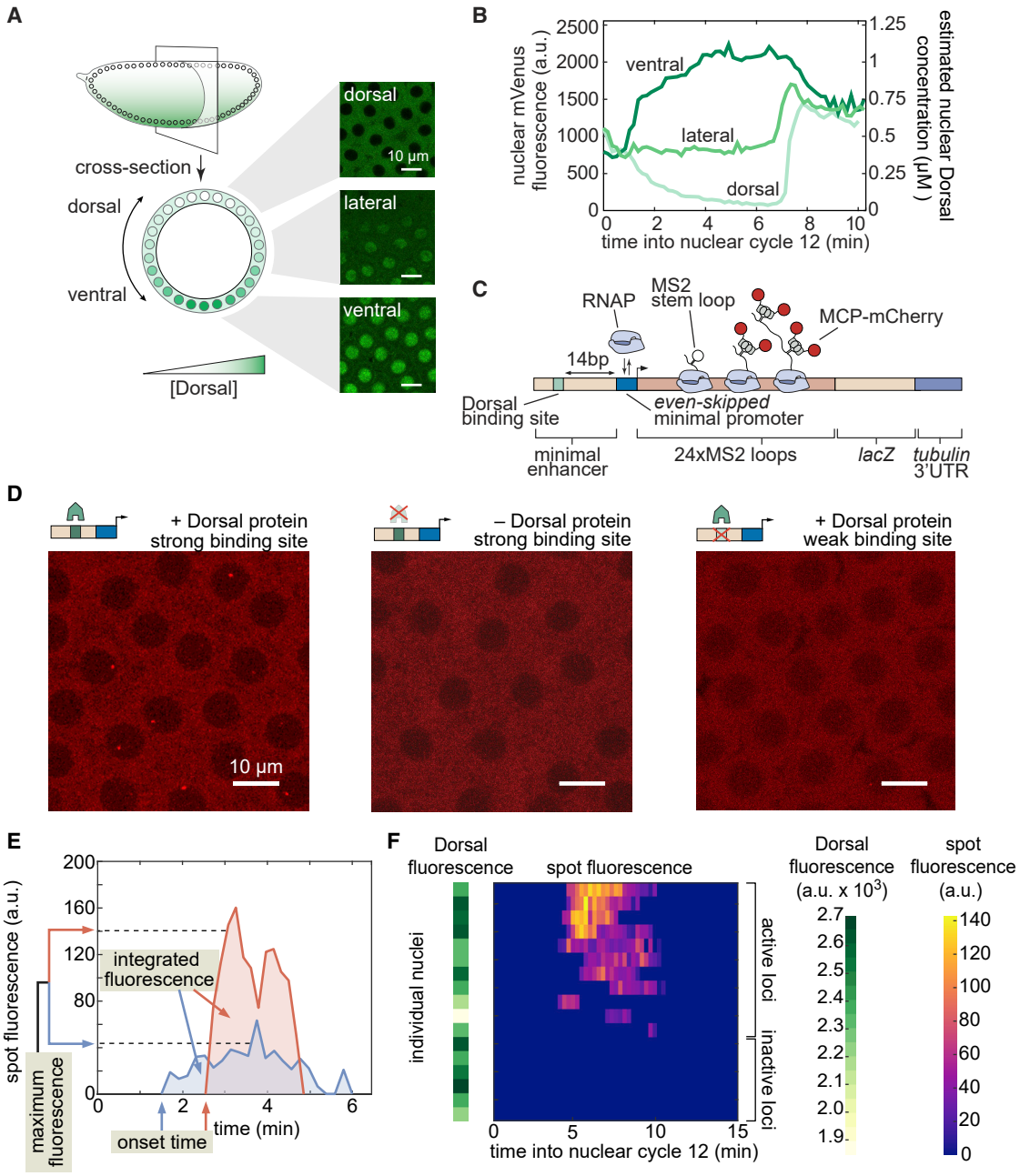

Figure 4. Simultaneously measuring transcription factor protein input and transcriptional output

(A) Schematic of the Dorsal protein gradient in early *Drosophila* embryos. Dorsal protein accumulates in ventral nuclei and is progressively excluded from more dorsal nuclei. Example snapshots show Dorsal-mVenus in various positions along the dorsoventral axis.

(B) Representative time traces of nuclear Dorsal-mVenus fluorescence in various positions along the dorsoventral axis in 2× Dorsal embryos. The right y axis shows the approximate nuclear Dorsal concentration according to the estimation described in Figure S9.

(C) Schematic of minimal synthetic promoter-proximal enhancer system containing a single binding site for Dorsal that drives transcription of a reporter tagged with MS2 loops, which are visualized through the binding of MCP-mCherry. The Dorsal binding site is placed 14 bp upstream of the *even-skipped* minimal promoter.

(D) Snapshots from embryos containing an optimal binding-site reporter in the presence (left) or absence (middle) of Dorsal or containing a strongly mutated Dorsal binding site (right).

(E) Examples of single-locus fluorescence time traces and quantitative metrics of transcriptional activity used throughout this work.

(F) Fluorescence of all transcription spots in individual nuclei in the field of view of one embryo as a function of time (heatmap) and their corresponding Dorsal-Venus fluorescence midway through the nuclear cycle (green bar on the left). If a transcription spot was detected within a nucleus at any point during the interphase of nuclear cycle 12, then the locus was considered active; otherwise, the locus was classified as inactive.

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

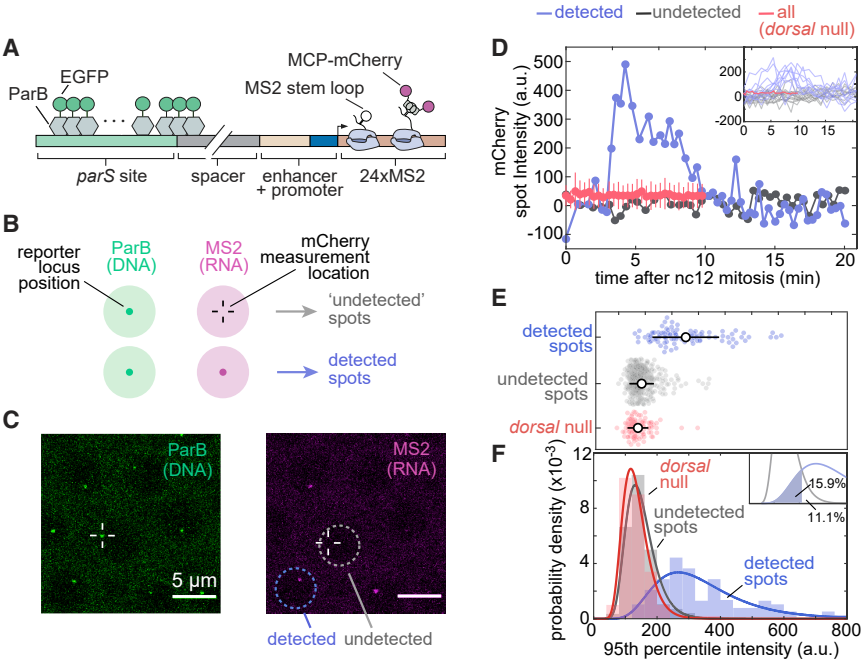

Figure 5. Transcriptionally independent ParB labeling confirms that transcriptionally inactive loci are functionally distinct from active loci

(A) Schematic of ParB-EGFP construct. ParB-EGFP molecules bind and polymerize out from *parS* sequences, which are placed ~400 bp upstream of the enhancer. The enhancer and promoter together drive transcription of MS2 loops that subsequently bind MCP-mCherry.

(B) Schematic of the experiment. Loci are located by detecting a signal in the ParB-EGFP channel; these locations were used to fit a 2D Gaussian to the same area in the MS2-mCherry channel to estimate fluorescence intensity, regardless of whether an MS2-mCherry signal was detected (Method details, ParB experiment fly crosses and microscopy).

(C) Example images of ParB-EGFP (left) and MCP-mCherry (right) channels. Detected and undetected loci are found based solely on the MCP-mCherry signal.

(D) Example time traces of MCP-mCherry fluorescence over time at the ParB-EGFP loci in nuclei with (blue) and without (gray) detected MS2-mCherry spots of the DBS_6.23 enhancer showing clear qualitative differences between the two populations. For comparison, the mean mCherry fluorescence at the ParB-EGFP loci in a representative Dorsal null embryo is also shown. Error bars correspond to the standard deviation across nuclei. Inset, all detected and undetected fluorescence traces obtained in the same embryo along with the mean fluorescence of all traces in a *dorsal* null embryo. Negative intensity values are due to spot intensities very close to the background fluorescence.

(E) Swarm plots of the maximum MCP-mCherry fluorescence over time at loci with detected (blue; n = 125 nuclei pooled from 20 embryos) and undetected MS2-mCherry transcription (gray; n = 425 nuclei pooled from 20 embryos) driven by the DBS_6.23 enhancer in wild-type Dorsal embryos. In red, all loci in a Dorsal null embryos are shown (n = 96 nuclei pooled from 6 embryos). Empty black circles correspond to the mean; bars are standard deviation. The maximum MCP-mCherry fluorescence over time is defined as the 95th percentile of values. Detected spots are significantly different from both null (ANOVA, $p < 0.01$) and undetected spots (ANOVA, $p < 0.01$).

(F) Histograms of the data shown in (E). Solid lines correspond to log-normal fits performed for ease of visualization. Inset, undetected and detected distribution fits and the area used to estimate the false-negative detection rate of 15.9% and the false-positive detection of 11.1%.

Transcriptionally active and inactive loci correspond to functionally distinct populations

To contrast to the predicted fraction of active loci with experimental observations, it is important to ensure that this fraction is the result of Dorsal action and not simply due to false negatives in our experimental setup. Transcriptionally silent loci that remain inactive throughout interphase, such as those revealed by our experiment (Figure 4F), have been observed using MS2 (and its sister mRNA labeling tool, PP7) in live-imaging experiments in flies,[8,12,15] plants,[76] and mammalian cells.[77] However, so far, it has not been possible to determine whether these inactive loci correspond to a separate transcriptional state from active loci or whether they are an artifact of the fluorescence detection thresholds associated with these microscopy techniques.

To answer this question, it is necessary to quantify MS2 fluorescence at loci undetected by our image analysis pipeline and determine whether they differ from loci not exposed to activators, which do not transcribe (Figure 4D, middle). However, to date, this approach has not been feasible because most MS2 measurements have relied on the presence of an MS2 signal itself to segment transcription spots and quantify their fluorescence. We hypothesized that if undetected loci correspond to a distinct and weaker, Dorsal-independent state, then detected and undetected spots in embryos carrying wild-type Dorsal would appear as two distinct populations. In this scenario, the mCherry fluorescence of inactive loci in wild-type Dorsal embryos would be similar to that observed in Dorsal null embryos and clearly distinct from the mCherry fluorescence of active loci in the presence of Dorsal.

To quantify MS2 fluorescence independently of whether an MS2 spot was detected, we implemented the *parS*-ParB DNA labeling system.[54,55] Here, fluorescently labeled ParB proteins bind the parS DNA sequence resulting in a fluorescence spot appearing at the locus independently of the transcriptional state of the locus (Figure 5A). We created flies with and without functional Dorsal expressing ParB2-EGFP (subsequently referred to as ParB-EGFP) and MCP-mCherry. We then crossed these flies to flies containing parS-DBS_6.23-MS2 to generate embryos that have our locus of interest labeled with ParB-EGFP colocalized with the transcriptional signal in the MCP-mCherry channel (Figures 5A and 5B; Video S2).

Guided by the spatial positions reported by ParB-EGFP, we measured the MCP-mCherry signal at all DBS_6.23 reporter loci in embryos carrying wild-type Dorsal (Figure 5C) or laid by mothers homozygous for the dl^1 null allele (Dorsal null embryos). We then classified loci from wild-type Dorsal embryos into two categories, detected and undetected, depending on whether they were identified as spots in the MCP-mCherry channel by

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

CellPress

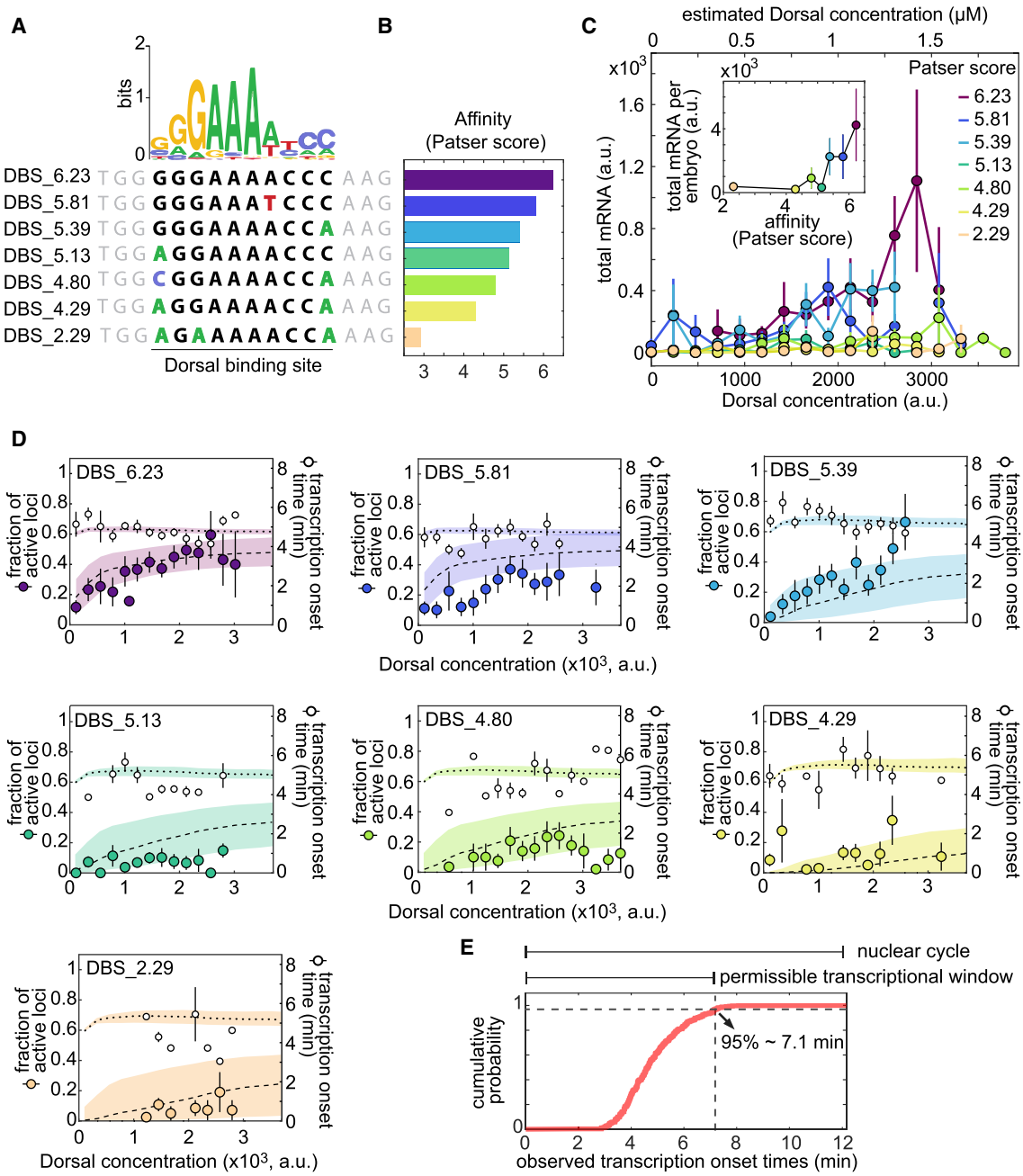

Figure 6. A multi-step kinetic barrier model predicts the Dorsal-dependent fraction of active loci with constant mean transcriptional onset times

(A) Top: Dorsal positional weight matrix logo from Ivan et al.[79] Bottom: sequence of the Dorsal binding sites engineered into our minimal synthetic enhancers. Bold letters, 10 bp Dorsal motif; black letters, consensus bases; colored letters, mutated bases; gray letters, sequence context.

(B) Relative affinities of Dorsal binding sites estimated from the Patser algorithm using the Dorsal position weight matrix.

(C) Overall transcriptional activity driven by the enhancers containing the binding sites in (A) measured as the total produced mRNA (fluorescence integrated over nuclear cycle 12) as a function of Dorsal concentration. Inset, mean total mRNA produced per embryo integrated across all Dorsal concentrations. Error bars, SEM over N > 3 embryos containing 3 or more nuclei belonging to that Dorsal fluorescence bin. The top x axis shows the estimated nuclear Dorsal concentration according to the calibration described in Figure S9.

(D) Data and model fits for the fraction of active loci (left y axis) and mean transcription onset time (right y axis) for each enhancer. Empty black circles, experimentally observed mean transcription onset time; filled circles, experimentally observed mean fraction of active loci. Fitted curves are represented as dashed lines (fraction of active loci) and dotted lines (mean onset times), corresponding to predictions using median parameter values from the joint posterior distribution. Shaded areas, 95% credible interval (see Table S1 for inferred parameter values). Error bars, SEM over N > 3 embryos containing 3 or

(legend continued on next page)

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

our analysis pipeline (Figures 5B and 5C; Method details, Image and time-series analysis). As shown in the examples presented in Figure 5D, there are clear qualitative differences between MCP-mCherry fluorescence time traces corresponding to detected and undetected transcriptional spots from wild-type embryos. Thus, our analysis made it possible to quantify MS2 fluorescence in three populations: detected loci and undetected loci in wild-type embryos, and all loci in Dorsal null embryos.

To compare these populations, we computed the 95th percentile value over each locus' MCP-mCherry fluorescence time trace (Figure 5E). The distribution of mCherry fluorescence from undetected spots in wild-type Dorsal embryos largely overlapped with that of all spots in Dorsal-null embryos (Figure 5F), consistent with these two populations corresponding to loci expressing Dorsal-independent levels of activity. Moreover, both distributions were clearly distinct from the distribution of detected spots in wild-type Dorsal embryos (Figures 5E and 5F). Thus, our results provide strong evidence that inactive loci are not artifacts of the detection limit of our imaging technique. Rather, loci can belong to one of two distinct populations: those that transcribe at a high, Dorsal-dependent level and those that are transcriptionally inactive (or active at a low, undetectable level that is comparable to that of embryos lacking Dorsal). We therefore conclude that the decision to transcribe made by each locus is an additional regulatory strategy controlled by Dorsal.

From the observations in Figures 5E and 5F, we estimated our error in classifying loci as inactive. This false-negative detection rate, corresponding to the area under the curve shaded in the inset of Figure 5F, is estimated as 15.9%. However, this false-negative rate is likely an underestimation. For example, this rate may depend on Dorsal concentration, which cannot be controlled for in this experiment. Additionally, the presence of ParB in the locus may itself affect transcriptional dynamics, impacting the false-negative rate. For these reasons, we do not attempt to correct our measurements of the fraction of active loci using this estimated false-negative rate.

Dorsal-dependent kinetic barriers explain transcription onset dynamics and modulation of the fraction of active loci

Having established that transcriptionally inactive promoters mostly constitute a separate population from transcriptionally active promoters (Figure 5), we sought to test whether our theoretical model (Figure 3A) can quantitatively recapitulate the fraction of active loci and their transcription onset times. Tuning transcription factor-DNA binding affinity has been a powerful tool to test models of transcriptional regulation in the past.[45,78] Inspired by these previous works, we probed our model by adjusting the Dorsal-DNA interaction energy in our minimal synthetic enhancer.

We constructed a series of enhancers containing a single binding site with varying affinities for Dorsal. Building on the optimal DBS_6.23 and the mutated DBS_4.29 sites (Figure 4D, left vs. right), we created five additional enhancers of varying intermediate strengths by introducing point mutations into the consensus Dorsal binding motif to obtain a range of predicted affinities (Figures 6A and 6B; Method details, Plasmids and reporter design). As described above, we refer to these enhancers as DBS, followed by their corresponding Patser score.

For the purpose of quantifying output transcriptional activity as a function of Dorsal concentration, we assigned a single Dorsal concentration value to each nucleus corresponding to the mVenus fluorescence in the center of that nucleus at a fiducial time point halfway through each nucleus' lifetime, approximately in the middle of nuclear cycle 12 when Dorsal levels are relatively stable (Figures S8A and S8B). We next grouped nuclei into 17 linearly spaced bins that span the dorsoventral axis based on their fiducial Dorsal fluorescence value (Figure S8B).

We assessed whether these point mutations were sufficient to generate a graded response to Dorsal and to determine the dynamic range of gene expression afforded by these enhancers. To make this possible, we integrated the total mRNA output over nuclear cycle 12 of each enhancer as a function of Dorsal concentration across all nuclei exposed to a given Dorsal concentration. The integrated mRNA output of the four weakest enhancers changed little across the dorsoventral axis (Figure 6C). However, an appreciable trend in integrated mRNA was observed for the three strongest affinities (Figure 6C). Further, plotting the total mRNA integrated across the entire dorsoventral axis of the embryo as a function of Patser score revealed that binding-site affinity (as reported by Patser score) is strongly correlated with transcriptional output in our single binding site enhancers (Figure 6C, inset). In the case of this measure, there was also a threshold affinity: enhancers containing binding sites with affinities below that of DBS_5.13 showed no substantial differences in transcriptional activity among them (Figure 6C, inset). We note that, although useful to drive qualitative insights about our synthetic constructs, the total mRNA is a quantity that is removed from the transcriptional dynamics that our models aim to predict. As a result, we do not attempt to draw quantitative insights from the analyses shown in Figure 6C.

We used these constructs to measure mean transcriptional onset time as a function of Dorsal concentration and binding affinity, one of the key magnitudes predicted by our model (Figure 3D). The measured mean onset time was relatively constant at approximately 5 min across all Dorsal concentrations and enhancer constructs (Figure 6D, white circles). This value is consistent with the measured onset times of other early embryonic genes such as the minimal *hunchback* promoter P2P.[8,10,14]

We also determined that the fraction of active loci is highly sensitive to Dorsal concentrations and Dorsal binding-site affinity (Figure 6D, filled circles). The strongest Dorsal binding sites showed a large modulation of the fraction of active loci across Dorsal concentrations, whereas the weakest drove a relatively constant and low fraction of active loci across all Dorsal concentrations (Figure 6D).

more nuclei belonging to that Dorsal fluorescence bin. The total numbers of embryos per enhancer from lowest to highest Patser score were 19, 27, 18, 26, 16, 35, and 46.

(E) Cumulative probability distribution of spot detection over all Dorsal fluorescence bins across all embryos and enhancers (N = 344 spots). Vertical dashed line, time at which 95% of spots have turned on (≈7.1 min) corresponding to the end of the permissible transcription time window.

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

Our kinetic barrier model assumes that loci that fail to become active during the permissible transcription time window will remain inactive during the rest of the nuclear cycle (Figure 3C). As a result, to determine whether the kinetic barrier model recapitulates the observations in Figure 6D, it was necessary to assign a value to this time window. We reasoned that the end of this time window determines the time point at which new transcription spots can no longer appear, possibly due to the onset of the next round of mitosis. To estimate the time point when nearly all spots have turned on, we calculated the 95th percentile of the observed spot onset times across all affinities to be approximately 7.1 min after the previous anaphase (Figure 6E).

Using the measured time window of permissible transcription, we performed a simultaneous fit to the fraction of active loci and mean transcription onset times across all enhancers based on the kinetic barrier model from our integrated model of transcriptional dynamics driven by a single activator binding site (Figure 6D; Method details). Consistent with our model, we forced all enhancers to share the same value for the rate constant, c, and only let the Dorsal dissociation constant, K_D, vary for each enhancer separately. By systematically exploring models with different numbers of OFF states n (Figures S11–S13), we determined that a biochemical cascade with at least 3–4 OFF states is capable of capturing the qualitative behavior of our observations: a Dorsal concentration- and binding affinity-dependent fraction of active loci (dashed lines in Figure 6D) and a mean transcription onset time that is mostly constant across Dorsal concentrations and affinities (dotted lines in Figure 6D). Alternative functional forms for k, such as modeling this transition rate as depending linearly on Dorsal concentration, instead of depending on Dorsal DNA occupancy, resulted in worse fits to the fraction of active loci at saturating concentrations of Dorsal (Methods S1.5, Figure S5). Thus, our observations can be explained by a model in which Dorsal, through DNA binding, accelerates the promoter's transition through a sequence of kinetic barriers to a state of active transcription. We note, however, that this model demanding the sequential transition across inactive states is not the only scenario capable of recapitulating our data. For example, a model in which multiple parallel switches need to be flipped on for transcription to ensue can also lead to a similar behavior as long as their switching rate is accelerated by Dorsal binding (Figures S14 and S15).

The experimentally measured RNAPII loading rates are compatible with a thermodynamic binding model

As a next step in our theoretical dissection, we tested the performance of our theoretical model in explaining the rate of transcription after loci become active. Typically, in MS2 experiments, the loading rate is measured from the initial slope of spot fluorescence traces.[8,14,80] However, due to the weak expression driven by our enhancers, it was not possible to perform this analysis with confidence (Figure S16). In lieu of directly measuring the transcription rate, we evaluated a related, more robust, and readily observable quantity: the maximum trace fluorescence (Figure 4E). A theoretical foundation for this approach can be found in Methods S1.2, where we show how we approximately relate the RNAPII loading rate predicted by the simple activator model (Equation 3) to the maximum fluorescence using a constant scaling factor, enabling direct comparison between theoretical predictions and experimental data. Examination of previously published live imaging data of transcription driven by the *hunchback* P2P reporter construct[14] confirms that the maximum fluorescence constitutes a good proxy for the RNAPII loading rate (Figure S2).

Our measurements revealed that the maximum spot fluorescence is relatively constant across Dorsal concentration for each of our seven minimal synthetic enhancers—particularly for the weakest of them, DBS_4.80, DBS_4.23, and DBS_2.92 (Figure 7). However, the sparse and noisy nature of our data makes it challenging to draw confident conclusions from the fits, even for the stronger binding sites (i.e., DBS_6.23, DBS_5.81, and DBS_5.39). In the case of the lower affinity binding sites, the constant maximum fluorescence suggests that the Dorsal concentration level in our embryos is far below the Dorsal dissociation constant K_D, even after increasing the Dorsal dosage by a factor of two with respect to wild type as in our 4× Dorsal line. The effect of very low Dorsal concentrations relative to their respective K_D values can be clearly seen in Equation 3 and in Figure 3, where, for $[Dl]/K_D \ll 1$, the RNAPII loading rate, R, adopts a basal level given by

$$R = R_{max}\frac{\frac{P}{K_P}}{1+\frac{P}{K_P}} \qquad \text{(Equation 4)}$$

that is independent of Dorsal concentration and binding affinity.

As shown on the right y-axes in Figure 7, this basal level corresponds to ≈20 RNAPII molecules actively transcribing the gene (≈15% of the maximum number of RNAPIIs that can fit on the gene, as described in Methods S1.3, Figure S3). We note that this estimate scales linearly with the magnitude of the RNAPII elongation rate (Equation S15), which can vary by a factor of two depending on the particular experiment.[8,80,81] For ease of visual comparison to the thermodynamic model predictions, we also plotted best-fit theoretical curves on top of the data using dashed curves (the insets in Figure 7 show the same plots but zoomed into the measured data and plotted on a linear scale). These fits further underscore that our data do not explore a wide dynamic range with the precision necessary to determine the magnitude of K_D for each construct and to thoroughly test the thermodynamic model.

DISCUSSION

A major obstacle to uncovering the mechanistic and quantitative underpinnings of enhancer action is the inherent complexity of endogenous regulatory sequences. Synthetic minimal enhancers are powerful alternatives to the complex experimental reality faced by modeling efforts in endogenous enhancers[3,42] (Box 1). Synthetic minimal enhancers contain binding sites for one or a handful of transcription factors, making them more amenable to theoretical dissection[23,24,82] and revealing the complex interplay among activators, repressors, and pioneer factors, as well as their contribution to mRNA transcript accumulation.[23,24,82] However, previous synthetic-based efforts to dissect enhancer function always involved fixed-embryo measurements, which cannot

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

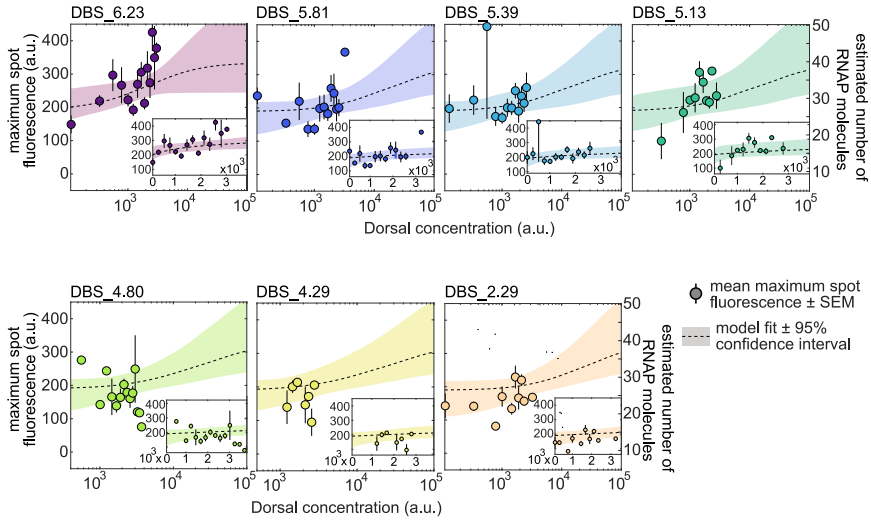

Figure 7. Testing RNAPII loading rate predictions of the thermodynamic model

Mean maximum spot fluorescence as a function of Dorsal concentration for minimal synthetic enhancers with different affinities for Dorsal. The right y axis denotes the calibrated number of actively transcribing RNAPII molecules. As shown in Equation S15, this calibration depends linearly on the elongation rate which can vary by a factor of two depending on the study.[8,80,81] For more details about this calibration, see Methods S1.3. Dashed curves correspond to a simultaneous Markov chain Monte Carlo curve fit to all data using Equation 3. Fits share all parameters except K_D. Shaded areas, 95% prediction intervals. Insets, same data and fits plotted on a linear scale with axis ranges zoomed in on the data. See Table S2 for inferred parameter values. Error bars, SEM across N > 3 embryos containing 3 or more nuclei in a given fluorescence bin. The total numbers of embryos per enhancer from lowest to highest Patser score were 19, 27, 18, 26, 16, 35, and 46.

reveal the three inherently dynamical features of transcription dictated by enhancer sequences (Figure 1).

Here, we augmented previous synthetic approaches by quantifying the real-time action of minimal enhancers with one binding site for the Dorsal activator in single cells of living, developing *Drosophila* embryos using the MS2 system. Contrary to theoretical speculations that single binding sites within eukaryotic genomes lack enough information to be recognized by transcription factors in the absence of other nearby binding sites,[83] we demonstrated that Dorsal can drive expression when bound to single binding sites (Figure 4D). Additionally, we demonstrated that the fraction of active loci is a feature under regulatory control in our synthetic system (Figures 4F and 5F), confirming the important role of this regulatory strategy in shaping the expression dynamics of endogenous enhancers.[8,12,17,18] Thus, although the signal driven by our minimal synthetic constructs is weak (Figure 7), it can be quantified and recapitulates biologically relevant dynamic features of transcription that are also at play in endogenous enhancers.

Although our minimal enhancer resembles endogenous promoter-proximal enhancers, it likely does not capture all the complexity of long-range distal enhancers where additional regulatory steps such as DNA looping are important. As a first step toward systematically studying the role of enhancer-promoter distance, we generated reporter constructs where we progressively moved the Dorsal binding site upstream up to 20 bp. These experiments showed that the binding site is fully functional 12 bp upstream of its original position (Figure S17). Thus, although our minimal enhancer system could also be used to study the effect of distal activators at short distances in future works, the fact that after moving the binding site 20 bp upstream, expression is almost completely lost suggests that our minimal regulatory sequence does not capture all the properties of endogenous distal enhancers.

It is important to note that the uncovering of a fraction of inactive loci in many reporter systems by us and others[8,12,17,18] did not necessarily imply that this modulation of transcriptional engagement constitutes a biological control variable. Indeed, because live cell imaging techniques typically lack single-molecule resolution, it was unclear whether undetected loci in our study—and all previous studies—corresponded to a distinct population or were a detection artifact. By simultaneously labeling the locus with the transcription-independent reporter ParB-EGFP and nascent mRNA with MCP-mCherry (Figure 5A), we demonstrated that a significant number of loci categorized as inactive do not constitute an experimental artifact and instead correspond to a distinct transcriptional state that is comparable with that measured in the absence of Dorsal protein (Figure 5). In the future, conducting all live transcription measurements with DNA loci labeled by ParB could make it possible to confidently quantify the activity of all loci, regardless of their activity.

Our minimal synthetic constructs and our validation of a distinct population of inactive loci enabled us to test an emerging theoretical model of enhancer action in development: a kinetic barrier model of transcriptional engagement[14,18,84] (Figure 3A). Our model deviated from previous theoretical efforts, which assumed that the transition rates between states preceding transcriptional engagement were either constant[18] or depended linearly on activator concentration.[14] Instead, to account for the effects of Dorsal binding affinity on transcriptional dynamics, we assumed that this rate was proportional to Dorsal occupancy at its target DNA site. Thus, although the mechanisms underlying several aspects of this model, such as the molecular identity of the various OFF states, remain unknown, this model can generate predictions for how the fraction of active loci and the transcriptional onset time are modulated by the Dorsal concentration and its binding affinity (Figures 3C–3E). Theoretical evidence such as the one presented here can guide the development of new experimental methods to directly test the hypotheses generated by these models.

We systematically challenged this model by generating a number of minimal synthetic enhancers spanning a large range of affinities for Dorsal (Figure 6A). Comparing the fraction of active loci and the transcription onset times of these enhancers revealed that the kinetic barrier model recapitulated our measurements (Figure 6D).

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

One interesting feature of our data is the fact that the mean transcriptional onset time is relatively constant as Dorsal concentration and binding affinity are varied. In past studies probing transcription dynamics in the *Drosophila* embryo,[14,18] the pioneer factor Zelda was found to be largely responsible for ensuring constant mean transcription onset times and for determining the fraction of active loci. We cannot rule out the potential existence of distal or low-affinity Zelda binding sites[85] in our constructs but believe that just like it has been recently demonstrated for the Bicoid activator,[86] Dorsal could also have a pioneering activity. Indeed, the Dorsal homolog NF-κB has been recently shown to displace nucleosomes.[27] To test the kinetic barrier model, it would be informative to directly perturb the temporal dynamics of nuclear Dorsal concentration to affect transcriptional engagement. For example, several optogenetics systems have been successfully deployed in the early fly embryo to inactivate transcription factors during discrete time widows.[87–90] In the future, a version of one of these systems may dissect how the temporal dynamics of Dorsal concentration affect transcriptional activation. To further probe the kinetic barrier model, it would be interesting to experimentally extend the nuclear cycle duration. A key prediction of our model is that given a longer permissible time window, seemingly silent nuclei would eventually engage in transcription. However, this is currently experimentally challenging and thus remains a thought experiment. Existing approaches to extend the duration of nuclear cycles such as lowering the temperature or genetic perturbations have pleiotropic effects on transcriptional dynamics.[91,92] Further, although it should be possible in principle to image our reporter in the much longer nuclear cycle 14, we observed extremely weak activity even for the strong DBS_6.23 at this stage. This contrasts with endogenous Dorsal targets that are driven by enhancers containing multiple Dorsal sites as well as binding sites for other transcription factors such as Twist.[93,94]

Although the kinetic barrier model predicted the fraction of active loci and onset times (Figure 6D) relatively well, we were unable to use our data to conclusively test the thermodynamic model's predictions of the rate of mRNA production (Figure 7). Such limitation stemmed in part from the fact that only a fraction of loci displays detectable transcription that can be used to quantify the mRNA production rate. In addition, for those loci with detectable transcription, the spot fluorescence signal is relatively low and highly variable. Finally, although the maximum spot fluorescence is a good proxy for the RNAPII loading rate, it is not a perfect one since $\approx 1/4$ of the variance in loading rates cannot be predicted from the maximum fluorescence alone (Figure S2). As a result, our statistics were limited such that it was not possible to perform an unequivocal test of the thermodynamic model.

The apparent lack of substantial Dorsal concentration dependence observed in our measurements of RNAPII loading rate could be explained in two possible ways. First, it is possible that there is a modulation of this rate in our measurements but that this modulation is obscured by our experimental noise. Second, the Dorsal concentrations accessed by our experiment could be below the K_D of our binding sites. In this scenario, a modulation in the mRNA production rate would become apparent only at Dorsal concentrations higher than those attainable by our experimental system. Although our embryos contained double the genetic dosage of Dorsal compared with wild type, perhaps 5–10 times the wild-type Dorsal concentration could be needed to exceed the K_D and modulate the rate of mRNA production. To express this high Dorsal concentration, which is certain to affect normal embryonic development, genetic approaches to increase Dorsal dosage in the embryos similar to those recently applied to flatten the Bicoid gradient might be necessary.[86]

It is important to note that despite not seeing a modulation in the rate of mRNA production, we do see a significant change in the fraction of active loci as Dorsal concentration is varied (Figure 6). This seeming contradiction could be explained through the presence of two effective dissociation constants in our model (Figure 3): one dissociation constant for the first part of the model governing the onset of transcription and a different dissociation constant for the second part of the model dictating the rate of RNAPII loading once transcription has ensued. Notably, previous works quantifying transcriptional dynamics of a minimal Bicoid-activated *hunchback* P2 enhancers also hinted at the existence of these two distinct dissociation constants.[8] These dissociation constants may arise from different binding kinetics depending on the chromatin state of the promoter, represented by the ON and OFF states in our model, which in turn could be modified by a pioneering factor like Zelda or by the Dorsal activator itself. Further, this model is consistent with our surprising observation of a basal level of transcription in the presence of even extremely weak binding sites (Figure 7) despite the lack of detected transcription in the absence of Dorsal protein (Figure 4D, middle). This observation could be explained if Dorsal acted both as a pioneer-like transcription factor triggering the onset of transcription, even at low concentrations, and as an activator of the transcription rate at high concentrations.

Going forward, synthetic minimal enhancers could constitute the foundation for exploring the behavior of more complex regulatory regions. Independently inferring biophysical parameters such as Dorsal-DNA binding and dissociation constants could help constrain models of Dorsal participating in the activation of promoters with additional activators and repressors.[23,24] For example, multiple Dorsal binding sites might allow for special binding configurations of both Dorsal and co-factors, enabling regulatory modes that are not possible with a single activator binding site. Although Dorsal is the sole known maternal nuclear-localized input specifying dorsoventral position in *Drosophila*, it rarely acts alone in endogenous enhancers.[95] For example, the interaction of Dorsal with Twist is a classic example of positive cooperativity in development.[72] Dorsal can also act as a repressor depending on the presence of nearby Capicua binding sites.[96] The minimal synthetic enhancers presented here could be used as scaffolds for more complex minimal enhancers incorporating a second binding site for Twist or Capicua. Finally, this minimal system could make it possible to further test the theoretical model beyond the minimal enhancer sequence by probing the effect of modifying the sequence of the minimal basal promoter.

In conclusion, we have developed a minimal synthetic enhancer system that has shed light on fundamental assumptions about transcription in development. By engaging in a dialog between theory and experiment, we have advanced

our understanding of how kinetic processes give rise to important features of transcriptional dynamics in the embryo and made progress toward predictive understanding of how regulatory DNA sequence dictates the functional relation between input transcription factor dynamics and output transcriptional activity in development.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Plasmids and reporter design
  - Transgenic flies
  - Microscopy
  - ParB experiment fly crosses and microscopy
  - Image and time-series analysis
  - Measuring Dorsal-mVenus concentration
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Curve fitting and parameter inference

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cels.2022.12.008.

ACKNOWLEDGMENTS

We thank Greg Reeves for providing the Dorsal-mVenus and *dl*[1] fly lines. We also thank Francois Payre and Philippe Valenti for sharing a ParB2-EGFP plasmid and a 2xIntB2 (aka *parS*) plasmid. We would like to thank Rob Phillips, Jane Kondev, and members of the Garcia lab for their helpful feedback on the manuscript.

H.G.G. was supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface, the Sloan Research Foundation, the Human Frontiers Science Program, the Searle Scholars Program, the Shurl and Kay Curci Foundation, the Hellman Foundation, the NIH Director's New Innovator Award (DP2 OD024541-01), and an NSF CAREER Award (1652236). H.G.G. is also a Chan Zuckerberg Biohub Investigator. A.R. was supported by NSF GRFP (DGE 1752814).

AUTHOR CONTRIBUTIONS

Conceptualization, H.G.G., A.R., and S.A.; methodology, H.G.G, A.R., and S.A.; software, H.G.G., A.R., S.A., and C.W.; investigation, A.R., S.A., C.W., P.T., J.Z., and E.L.; resources, A.R., S.A., and M.T.; writing, H.G.G., A.R., and S.A.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 7, 2021
Revised: September 3, 2022
Accepted: December 21, 2022
Published: January 24, 2023

REFERENCES

1. Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol. *25*, 1–47.
2. Briscoe, J., and Small, S. (2015). Morphogen rules: design principles of gradient-mediated embryo patterning. Development *142*, 3996–4009.
3. Garcia, H.G., Berrocal, A., Kim, Y.J., Martini, G., and Zhao, J. (2020). Lighting up the central dogma for predictive developmental biology. Curr. Top. Dev. Biol. *137*, 1–35.
4. Vincent, B.J., Estrada, J., and DePace, A.H. (2016). The appeasement of doug: a synthetic approach to enhancer biology. Integr. Biol. (Camb) *8*, 475–484.
5. Venken, K.J.T., and Bellen, H.J. (2005). Emerging technologies for gene manipulation in *Drosophila melanogaster*. Nat. Rev. Genet. *6*, 167–178.
6. Bier, E., Harrison, M.M., O'Connor-Giles, K.M., and Wildonger, J. (2018). Advances in engineering the fly genome with the CRISPR-Cas system. Genetics *208*, 1–18.
7. Gregor, T., Bialek, W., de Ruyter van Steveninck, R.R., Tank, D.W., and Wieschaus, E.F. (2005). Diffusion and scaling during early embryonic pattern formation. Proc. Natl. Acad. Sci. USA *102*, 18403–18407.
8. Garcia, H.G., Tikhonov, M., Lin, A., and Gregor, T. (2013). Quantitative imaging of transcription in living *Drosophila* embryos links polymerase activity to patterning. Curr. Biol. *23*, 2140–2145.
9. Mir, M., Reimer, A., Haines, J.E., Li, X.Y., Stadler, M., Garcia, H., Eisen, M.B., and Darzacq, X. (2017). Dense bicoid hubs accentuate binding along the morphogen gradient. Genes Dev. *31*, 1784–1794.
10. Lucas, T., Ferraro, T., Roelens, B., De Las Heras Chanes, J., Walczak, A.M., Coppey, M., and Dostatni, N. (2013). Live imaging of bicoid-dependent transcription in *Drosophila* embryos. Curr. Biol. *23*, 2135–2139.
11. Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer control of transcriptional bursting. Cell *166*, 358–368.
12. Lammers, N.C., Galstyan, V., Reimer, A., Medin, S.A., Wiggins, C.H., and Garcia, H.G. (2020). Multimodal transcriptional control of pattern formation in embryonic development. Proc. Natl. Acad. Sci. USA *117*, 836–847.
13. Fuqua, T., Jordan, J., Breugel, M.E., Halavatyi, A., Tischer, C., Polidoro, P., Abe, N., Tsai, A., Mann, R.S., Stern, D.L., et al. (2020). Dense encoding of developmental regulatory information may constrain evolvability. Preprint at bioRxiv. https://doi.org/10.1101/2020.04.17.046052.
14. Eck, E., Liu, J., Kazemzadeh-Atoufi, M., Ghoreishi, S., Blythe, S.A., and Garcia, H.G. (2020). Quantitative dissection of transcription in development yields evidence for transcription factor-driven chromatin accessibility. eLife *9*, e56429.
15. Berrocal, A., Lammers, N.C., Garcia, H.G., and Eisen, M.B. (2020). Kinetic sculpting of the seven stripes of the *Drosophila* even-skipped gene. eLife *9*, e61635.
16. Fukaya, T. (2021). Dynamic regulation of anterior-posterior patterning genes in living *Drosophila* embryos. Curr. Biol. *31* 2227.e6–2236.e6.
17. Harden, T.T., Vincent, B.J., and DePace, A.H. (2021). Defining kinetic roles of transcriptional activators in the early *Drosophila* embryo. Preprint at bioRxiv. https://doi.org/10.1101/2021.02.25.432925.
18. Dufourt, J., Trullo, A., Hunter, J., Fernandez, C., Lazaro, J., Dejean, M., Morales, L., Nait-Amer, S., Schulz, K.N., Harrison, M.M., et al. (2018). Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. Nat. Commun. *9*, 5194.
19. Desponds, J., Tran, H., Ferraro, T., Lucas, T., Perez Romero, C., Guillou, A., Fradin, C., Coppey, M., Dostatni, N., and Walczak, A.M. (2016). Precision of readout at the hunchback gene: analyzing short transcription time traces in living fly embryos. PLoS Comput. Biol. *12*, e1005256.
20. Tran, H., Desponds, J., Perez Romero, C.A., Coppey, M., Fradin, C., Dostatni, N., and Walczak, A.M. (2018). Precision in a rush: trade-offs between reproducibility and steepness of the hunchback expression pattern. PLoS Comput. Biol. *14*, e1006513.

21. Desponds, J., Vergassola, M., and Walczak, A.M. (2020). A mechanism for hunchback promoters to readout morphogenetic positional information in less than a minute. eLife *9*, e49758.

22. Park, J., Estrada, J., Johnson, G., Vincent, B.J., Ricci-Tam, C., Bragdon, M.D., Shulgina, Y., Cha, A., Wunderlich, Z., Gunawardena, J., et al. (2019). Dissecting the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity. eLife *8*, e41266.

23. Fakhouri, W.D., Ay, A., Sayal, R., Dresch, J., Dayringer, E., and Arnosti, D.N. (2010). Deciphering a transcriptional regulatory code: modeling short-range repression in the *Drosophila* embryo. Mol. Syst. Biol. *6*, 341.

24. Sayal, R., Dresch, J.M., Pushel, I., Taylor, B.R., and Arnosti, D.N. (2016). Quantitative perturbation-based analysis of gene expression predicts enhancer activity in early *Drosophila* embryo. eLife *5*, e08445.

25. Estrada, J., Ruiz-Herrero, T., Scholes, C., Wunderlich, Z., and DePace, A.H. (2016). Siteout: an online tool to design binding site-free dna sequences. PLoS One *11*, e0151740.

26. Scholes, C., DePace, A.H., and Sánchez, Á. (2017). Combinatorial gene regulation through kinetic control of the transcription cycle. Cell Syst. *4* 97.e9–108.e9.

27. Cheng, Q.J., Ohta, S., Sheu, K.M., Spreafico, R., Adelaja, A., Taylor, B., and Hoffmann, A. (2021).NF-κ B dynamics determine the stimulus specificity of epigenomic reprogramming in macrophages. Science 372, 1349–1353.

28. Zhou, J., Zwicker, J., Szymanski, P., Levine, M., and Tjian, R. (1998). TAFII mutations disrupt Dorsal activation in the *Drosophila* embryo. Proc. Natl. Acad. Sci. USA *95*, 13483–13488.

29. Razo-Mejia, M., Barnes, S.L., Belliveau, N.M., Chure, G., Einav, T., Lewis, M., and Phillips, R. (2018). Tuning transcriptional regulation through signaling: a predictive theory of allosteric induction. Cell Syst. *6* 456.e10–469.e10.

30. Polach, K.J., and Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. J. Mol. Biol. *254*, 130–149.

31. Schulze, S.R., and Wallrath, L.L. (2007). Gene regulation by chromatin structure: paradigms established in *Drosophila melanogaster*. Annu. Rev. Entomol. *52*, 171–192.

32. Lam, F.H., Steger, D.J., and O'Shea, E.K. (2008). Chromatin decouples promoter threshold from dynamic range. Nature *453*, 246–250.

33. Li, X.Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L., et al. (2008). Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. PLoS Biol. *6*, e27.

34. Kim, H.D., and O'Shea, E.K. (2008). A quantitative model of transcription factor-activated gene expression. Nat. Struct. Mol. Biol. *15*, 1192–1198.

35. Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. *20*, R754–R763.

36. Fussner, E., Ching, R.W., and Bazett-Jones, D.P. (2011). Living without 30 nm chromatin fibers. Trends Biochem. Sci. *36*, 1–6.

37. Bai, L., Ondracka, A., and Cross, F.R. (2011). Multiple sequence-specific factors generate the nucleosome-depleted region on cln2 promoter. Mol. Cell *42*, 465–476.

38. Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

39. Hansen, A.S., and O'Shea, E.K. (2015). cis determinants of promoter threshold and activation timescale. Cell Rep. *12*, 1226–1233.

40. Li, X.-Y., and Eisen, M.B. (2018). Zelda potentiates transcription factor binding to zygotic enhancers by increasing local chromatin accessibility during early *Drosophila melanogaster* embryogenesis. Preprint at bioRxiv. https://doi.org/10.1101/380857.

41. Foo, S.M., Sun, Y., Lim, B., Ziukaite, R., O'Brien, K., Nien, C.Y., Kirov, N., Shvartsman, S.Y., and Rushlow, C.A. (2014). Zelda potentiates morphogen activity by increasing chromatin accessibility. Curr. Biol. *24*, 1341–1346.

42. Garcia, H.G., Brewster, R.C., and Phillips, R. (2016). Using synthetic biology to make cells tomorrow's test tubes. Integr. Biol. (Camb) *8*, 431–450.

43. Garcia, H.G., and Phillips, R. (2011). Quantitative dissection of the simple repression input–output function. Proc. Natl. Acad. Sci. USA *108*, 12173–12178.

44. Brewster, R.C., Weinert, F.M., Garcia, H.G., Song, D., Rydenfelt, M., and Phillips, R. (2014). The transcription factor titration effect dictates level of gene expression. Cell *156*, 1312–1323.

45. Phillips, R., Belliveau, N.M., Chure, G., Garcia, H.G., Razo-Mejia, M., and Scholes, C. (2019). Figure 1 theory meets figure 2 experiments in the study of gene expression. Annu. Rev. Biophys. *48*, 121–163.

46. Popp, A.P., Hettich, J., Christof, J., and Gebhardt, M. (2020). Transcription factor residence time dominates over concentration in transcription activation. Preprint at bioRxiv. https://doi.org/10.1101/2020.11.26.400069.

47. Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. Proc. Natl. Acad. Sci. USA *79*, 1129–1133.

48. Buchler, N.E., Gerland, U., and Hwa, T. (2003). On schemes of combinatorial transcription logic. Proc. Natl. Acad. Sci. USA *100*, 5136–5141.

49. Vilar, J.M., and Leibler, S. (2003). Dna looping and physical constraints on transcription regulation. J. Mol. Biol. *331*, 981–989.

50. Bolouri, H., and Davidson, E.H. (2003). Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. Proc. Natl. Acad. Sci. USA *100*, 9371–9376.

51. Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005a). Transcriptional regulation by the numbers: models. Curr. Opin. Genet. Dev. *15*, 116–124.

52. Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005b). Transcriptional regulation by the numbers: applications. Curr. Opin. Genet. Dev. *15*, 125–135.

53. Sherman, M.S., and Cohen, B.A. (2012). Thermodynamic state ensemble models of cis-regulation. PLoS Comput. Biol. *8*, e1002407.

54. Germier, T., Kocanova, S., Walther, N., Bancaud, A., Shaban, H.A., Sellou, H., Politi, A.Z., Ellenberg, J., Gallardo, F., and Bystricky, K. (2017). Real-time imaging of a single gene reveals transcription-initiated local confinement. Biophys. J. *113*, 1383–1394.

55. Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T. (2018). Dynamic interplay between enhancer-promoter topology and gene activity. Nat. Genet. *50*, 1296–1303.

56. Mir, M., Stadler, M.R., Ortiz, S.A., Hannon, C.E., Harrison, M.M., Darzacq, X., and Eisen, M.B. (2018). Dynamic multifactor hubs interact transiently with sites of active transcription in *Drosophila* embryos. eLife *7*, e40497.

57. Callegari, A., Sieben, C., Benke, A., Suter, D.M., Fierz, B., Mazza, D., and Manley, S. (2018).Single-molecule dynamics and genome-wide transcriptomics reveal that NF-κB (p65)-DNA binding times can be decoupled from transcriptional activation. PLoS Genet. 15, e1007891.

58. Shermoen, A.W., and O'Farrell, P.H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell *67*, 303–310.

59. Roth, S., Stein, D., and Nüsslein-Volhard, C. (1989). A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. Cell *59*, 1189–1202.

60. Reeves, G.T., Trisnadi, N., Truong, T.V., Nahmad, M., Katz, S., and Stathopoulos, A. (2012). Dorsal-ventral gene expression in the *Drosophila* embryo reflects the dynamics and precision of the dorsal nuclear gradient. Dev. Cell *22*, 544–557.

61. Steward, R., Zusman, S.B., Huang, L.H., and Schedl, P. (1988). The dorsal protein is distributed in a gradient in early *Drosophila* embryos. Cell *55*, 487–495.

62. Thisse, C., Perrin-Schmitt, F., Stoetzel, C., and Thisse, B. (1991). Sequence-specific transactivation of the *Drosophila* twist gene by the dorsal gene product. Cell *65*, 1191–1201.

63. Jiang, J., Hoey, T., and Levine, M. (1991). Autoregulation of a segmentation gene in *Drosophila*: combinatorial interaction of the even-skipped homeo box protein with a distal enhancer element. Genes Dev. *5*, 265–277.

64. Kirov, N., Zhelnin, L., Shah, J., and Rushlow, C. (1993). Conversion of a silencer into an enhancer: evidence for a co-repressor in dorsal-mediated repression in *Drosophila*. EMBO J. *12*, 3193–3199.

65. Papagianni, A., Forés, M., Shao, W., He, S., Koenecke, N., Andreu, M.J., Samper, N., Paroush, Z., González-Crespo, S., Zeitlinger, J., and Jiménez, G. (2018). Capicua controls Toll/IL-1 signaling targets independently of RTK regulation. Proc. Natl. Acad. Sci. USA *115*, 1807–1812.

66. Sandler, J.E., and Stathopoulos, A. (2016). Quantitative single-embryo profile of *Drosophila* genome activation and the dorsal-ventral patterning network. Genetics *202*, 1575–1584.

67. Dufourt, J., Bellec, M., Trullo, A., Dejean, M., Rossi, S.D., and Lagha, M. (2020). Imaging translation dynamics in live embryos reveals spatial heterogeneities. Science *372*, 840–844.

68. Kremers, G.J., Goedhart, J., van Munster, E.B., and Gadella, T.W. (2006). Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and fret forster radius. Biochemistry *45*, 6570–6580.

69. Gratz, S.J., Rubinstein, C.D., Harrison, M.M., Wildonger, J., and O'Connor-Giles, K.M. (2015). Crispr-cas9 genome editing in *Drosophila*. Curr. Protoc. Mol. Biol. 31.2.1–31.2.20.

70. Ip, Y.T., Park, R.E., Kosman, D., Yazdanbakhsh, K., and Levine, M. (1992). dorsal-twist interactions establish snail expression in the presumptive mesoderm of the *Drosophila* embryo. Genes Dev. *6*, 1518–1530.

71. Jiang, J., and Levine, M. (1993). Binding affinities and cooperative interactions with bhlh activators delimit threshold responses to the dorsal gradient morphogen. Cell *72*, 741–752.

72. Szymanski, P., and Levine, M. (1995). Multiple modes of dorsal-bhlh transcriptional synergy in the *Drosophila* embryo. EMBO J. *14*, 2229–2238.

73. Stormo, G.D., and Hartzell, G.W. (1989). Identifying protein-binding sites from unaligned dna fragments. Proc. Natl. Acad. Sci. USA *86*, 1183–1187.

74. Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ash1 mrna particles in living yeast. Mol. Cell *2*, 437–445.

75. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

76. Alamos, S., Reimer, A., Niyogi, K.K., and Garcia, H.G. (2020). Quantitative imaging of RNA polymerase II activity in plants reveals the single-cell basis of tissue-wide transcriptional dynamics. Nat. Plants *7*, 1037–1049. https://doi.org/10.1038/s41477-021-00976-0.

77. Hafner, A., Reyes, J., Stewart-Ornstein, J., Tsabar, M., Jambhekar, A., and Lahav, G. (2020). Quantifying the central dogma in the p53 pathway in live single cells. Cell Syst. *10* 495.e4–505.e4.

78. Meijsing, S.H., Pufall, M.A., So, A.Y., Bates, D.L., Chen, L., and Yamamoto, K.R. (2009). Dna binding site sequence directs glucocorticoid receptor structure and activity. Science *324*, 407–410.

79. Ivan, A., Halfon, M.S., and Sinha, S. (2008). Computational discovery of cis-regulatory modules in *Drosophila* without prior knowledge of motifs. Genome Biol. *9*, R22.

80. Liu, J., Hansen, D., Eck, E., Kim, Y.J., Turner, M., Alamos, S., and Garcia, H.G. (2021). Real-time single-cell characterization of the eukaryotic transcription cycle reveals correlations between RNA initiation, elongation, and cleavage. PLoS Comp. Biol. *17*, e1008999.

81. Fukaya, T., Lim, B., and Levine, M. (2017). Rapid rates of Pol II elongation in the *Drosophila* embryo. Curr. Biol. *27*, 1387–1391.

82. Crocker, J., and Ilsley, G.R. (2017). Using synthetic biology to study gene regulatory evolution. Curr. Opin. Genet. Dev. *47*, 91–101.

83. Wunderlich, Z., and Mirny, L.A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. Trends Genet. *25*, 434–440.

84. Fritzsch, C., Baumgärtner, S., Kuban, M., Steinshorn, D., Reid, G., and Legewie, S. (2018). Estrogen-dependent control and cell-to-cell variability of transcriptional bursting. Mol. Syst. Biol. *14*, e7678.

85. Rushlow, C.A., and Shvartsman, S.Y. (2012). Temporal dynamics, spatial range, and transcriptional interpretation of the dorsal morphogen gradient. Curr. Opin. Genet. Dev. *22*, 542–546.

86. Hannon, C.E., Blythe, S.A., and Wieschaus, E.F. (2017). Concentration dependent chromatin states induced by the bicoid morphogen gradient. eLife *6*, e28275.

87. Huang, A., Amourda, C., Zhang, S., Tolwinski, N.S., and Saunders, T.E. (2017). Decoding temporal interpretation of the morphogen bicoid in the early *Drosophila* embryo. eLife *6*, e26258.

88. McDaniel, S.L., Gibson, T.J., Schulz, K.N., Fernandez Garcia, M., Nevil, M., Jain, S.U., Lewis, P.W., Zaret, K.S., and Harrison, M.M. (2019). Continued activity of the pioneer factor Zelda is required to drive zygotic genome activation. Mol. Cell *74*. 185.e4–195.e4.

89. Irizarry, J., McGehee, J., Kim, G., Stein, D., and Stathopoulos, A. (2020). Twist-dependent ratchet functioning downstream from dorsal revealed using a light-inducible degron. Genes Dev. *34*, 965–972.

90. Singh, A.P., Wu, P., Ryabichko, S., Raimundo, J., Swan, M., Wieschaus, E., Gregor, T., and Toettcher, J.E. (2022). Optogenetic control of the bicoid morphogen reveals fast and slow modes of gap gene regulation. Cell Rep. *38*, 110543.

91. Strong, I.J.T., Lei, X., Chen, F., Yuan, K., and O'Farrell, P.H. (2020). Interphase-arrested *Drosophila* embryos activate zygotic gene expression and initiate mid-blastula transition events at a low nuclear-cytoplasmic ratio. PLoS Biol. *18*, e3000891.

92. Kuntz, S.G., and Eisen, M.B. (2014). *Drosophila* embryogenesis scales uniformly across temperature in developmentally diverse species. PLoS Genet. *10*, e1004293.

93. Hong, J.W., Hendrix, D.A., and Levine, M.S. (2008). Shadow enhancers as a source of evolutionary novelty. Science *321*, 1314.

94. Reeves, G.T., and Stathopoulos, A. (2009). Graded dorsal and differential gene regulation in the *Drosophila* embryo. Cold Spring Harb. Perspect. Biol. *1*, a000836.

95. Hong, J.W., Hendrix, D.A., Papatsenko, D., and Levine, M.S. (2008). How the dorsal gradient works: insights from postgenome technologies. Proc. Natl. Acad. Sci. USA *105*, 20072–20076.

96. Shin, D.H., and Hong, J.W. (2014). Capicua is involved in Dorsal-mediated repression of Zerknüllt expression in *Drosophila* embryo. BMB Rep. *47*, 518–523.

97. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

98. Hertz, G.Z., and Stormo, G.D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics *15*, 563–577.

99. Haario, H., Laine, M., Mira, A., and Saksman, E. (2006). DRAM: efficient adaptive MCMC. Stat. Comput. *16*, 339–354.

100. Markstein, M., Markstein, P., Markstein, V., and Levine, M.S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. Proc. Natl. Acad. Sci. USA *99*, 763–768.

101. Tutucci, E., Vera, M., Biswas, J., Garcia, J., Parker, R., and Singer, R.H. (2018). An improved MS2 system for accurate reporting of the mRNA life cycle. Nat. Methods *15*, 81–89.

102. Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. *34*, W369–W373.

103. Bergman, C.M., Carlson, J.W., and Celniker, S.E. (2005). *Drosophila* DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, *Drosophila melanogaster*. Bioinformatics *21*, 1747–1749.

104. Bothma, J.P., Garcia, H.G., Esposito, E., Schlissel, G., Gregor, T., and Levine, M. (2014). Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living *Drosophila* embryos. Proc. Natl. Acad. Sci. USA *111*, 10598–10603.

105. Garcia, H.G., and Gregor, T. (2018). Live Imaging of mRNA Synthesis in Drosophila (Springer), pp. 349–357.

106. Witten, I.H., Frank, E., Hall, M.A., and Pal, C.J. (2016). Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Publishers).

107. Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K.W., Schindelin, J., Cardona, A., and Sebastian Seung, H. (2017). Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics *33*, 2424–2426.

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
yw; ParB2-EGFP; eNosx2-MCP-mCherry; +		N/A
yw; Dorsal-mVenus, pNos-MCP-mCherry; pNos-MCP-mCherry, His2Av-iRFP		N/A
yw; Dorsal-mVenus, pNos-MCP-mCherry; Dorsal-mVenus, pNos-MCP-mCherry, His2Av-iRFP		N/A
yw; *dl*1, *pNos-MCP-mCherry; pNos-MCP-mCherry, His2Av-iRFP*		N/A
yw; 1Dg(11); +		DBS_6.23
yw; 1DgS(2); +		DBS_5.81
yw; 1DgW(2); +		DBS_5.39
yw; 1DgAW(3); +		DBS_5.13
yw; 1DgSVW(2); +		DBS_4.8
yw; 1DgVW(1); +		DBS_4.73
yw; 1DgVVW(3); +		DBS_4.29
yw; 2xIntB2-1Dg(4); +		N/A
Oligonucleotides		
5' ggaacgaaggcagttagttgt		18.8
5'tagttccagtgaaatccaagcattttc		Ori-Seq-F1
5' ccattaaaacggaaccaagaggtgag		OutLHA
5' tctaacaatggctcgatttttgcca		OutDlRHA
Recombinant DNA		
pIB-1Dg-evePr-MS2v5-LacZ-Tub3UTR		DBS_6.23
pIB-1DgS-MS2v5-LacZ-Tub3UTR		DBS_5.81
pIB-1DgW-MS2v5-LacZ-Tub3UTR		DBS_5.39
pIB-1DgAW-MS2v5-LacZ-Tub3UTR		DBS_5.13
pIB-1DgSVW-MS2v5-LacZ-Tub3UTR		DBS_4.8
pIB-1DgVW-MS2v5-LacZ-Tub3UTR		DBS_4.73
pIB-1DgVVW-MS2v5-LacZ-Tub3UTR		DBS_4.29
pIB-4xIntB2-Neutral400-1Dg-MS2v5-LacZ-Tub3UTR		N/A
Dl-mVenus-dsRed		N/A
pU6-DlgRNA1		N/A
pBPhi-eNosx2-pTrans-NoNLS-MCP-mCherry-tub3'UTR		N/A
pCasper4-His2Av-iRFP		N/A
pCasper4-Pnos-NoNLS-MCP-mCherry-TUB3'UTR		N/A
pCasper-pNos-NoNLS-ParB2-GFP-TUB3'UTR		N/A
Software and algorithms		
ImageJ	Schneider et al.[97]	https://imagej.nih.gov/ij/
Matlab		N/A
Patser	Hertz and Stormo[98].	N/A
Image analysis pipeline	Garcia et al[8] Lammers et al.[12]	https://github.com/GarciaLab/mRNADynamics
Post image processing data analysis pipeline	This paper	https://github.com/GarciaLab/mRNADynamics_dorsal_synthetics/releases/tag/v1.0.0
MCMCSTAT pipeline	Haario et al.[99]	https://github.com/mjlaine/mcmcstat

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Hernan G. Garcia (hggarcia@berkeley.edu).

Materials availability

Plasmids and fly lines generated in this study are available upon request.

Data and code availability

- All data is available upon request.
- All code used to analyze confocal imaging files has been previously published[8,12] and can be found in the Github repository listed in the key resources table.
- All original code written for this paper for post-image processing analyses has been deposited at the Github repository listed in the key resources table and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Drosophila melanogaster (see key resources table)

METHOD DETAILS

Plasmids and reporter design

To design our minimal construct (Figure 4), we placed the 10 bp consensus Dorsal binding site[100] upstream of the *even-skipped* core promoter. This enhancer-promoter construct drives the expression of the MS2v5 sequence containing 24 nonrepetitive MS2 loops[101] followed by the *lacZ* coding sequence and the *tubulin* 3'UTR.[8]

In addition to the consensus Dorsal binding site (DBS_6.23), we created six enhancers of varying strength by introducing point mutations to the consensus Dorsal binding motif. Some of these binding sites were taken from known validated Dorsal motifs,[100] while others were generated based on mutations known to decrease Dorsal binding.[63,70] To guide the design of these binding sites, we used an already existing position weight matrix computed with the MEME algorithm[79,102] using motifs generated by DNAse I footprinting assays[103] and quantified the information content of each base pair using Patser.[98]

All plasmid sequences used in this study can be accessed from a public Benchling folder. Injections were carried out by Rainbow inc. or Bestgene inc.

Transgenic flies

Reporter plasmids were injected into BDSC fly line 27388 containing a landing site in position 38F1. Transgene orientation was confirmed by PCR using primers 18.8 (ggaacgaaggcagttagttgt) and Ori-Seq-F1 (tagttccagtgaaatccaagcattttc) binding outside of the 5' 38F1 *attP* site and the *even-skipped* promoter, respectively. All reporter lines were confirmed to be in the same orientation.

To generate the embryos used in the experiments shown in all figures except for Figure 5, we crossed 4x Dorsal or 2x Dorsal virgins to males carrying synthetic enhancers. The genotype of 4x Dorsal flies is *yw;Dl-mVenus (CRISPR), MCP-mCherry; Dorsal-mVenus, MCP-mCherry, His2Av-iRFP*. The genotype of 2x Dorsal flies is *yw;dl[1], MCP-mCherry; Dorsal-mVenus, MCP-mCherry, His2Av-iRFP*. Because there does not seem to be a difference in transcriptional activity between the CRISPR knock-in and the transgene Dorsal-mVenus alleles (Figure S18), we combined the 2x Dorsal and 4x Dorsal data for some enhancers.

MCP-mCherry and His-iRFP were described before by Liu et al.[80] The Dorsal-mVenus transgene was developed by Reeves et al.[60]

To generate the Dorsal-Venus knock-in allele we used the CRISPR/Cas9 protocol described by Gratz et al.[69] We generated a donor plasmid containing the mVenus sequence followed by a stop codon and a 3xP3-dsRed marker flanked by PiggyBac recombinase sites. This insert was flanked by two ≈1 kbp homology arms matching ≈2 kbp surrounding the Dorsal stop codon (plasmid Dl-mVenus-dsRed). The Cas9 expressing BDSC line 51324 was injected with the donor plasmid in combination with a plasmid carrying a sgRNA targeting the sequence GTTGTGAAAAAGGTATTACG located in the C-terminus of Dorsal (plasmid pU6-DlgRNA1). Survivors were crossed to *yw* and the progeny was screened for dsRed eye fluorescence. Several independent lines were established and tested for rescue. The insertion was confirmed by PCR using primers flanking the homology arms OutLHA (ccattaaaacggaaccaa gaggtgag) and OutDlRHA (tctaacaatggctcgatttttgcca). The dsRed eye marker cassette was flipped out of rescuing lines via crossing with a piggyBac recombinase line. The resulting Dorsal-mVenus locus was then resequenced using the same primers. We used the same procedure to generate Dl-mCherry knock-in fusion lines but failed to obtain fertile females.

The data shown in Figure 5 were obtained from embryos laid by *yw;ParB2-eGFP, eNosx2-MCP-mCherry;+* (wild-type Dorsal mothers) or *yw;ParB2-eGFP, eNosx2-MCP-mCherry, dl[1];+* (Dorsal null mothers).

Please cite this article in press as: Alamos et al., Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient, Cell Systems (2022), https://doi.org/10.1016/j.cels.2022.12.008

Microscopy

Fly cages were allowed to lay for 90 to 120 minutes prior to embryo collection. Embryos were then mounted on microscopy slides in Halocarbon 27 oil (Sigma-Aldrich, H8773) in between a coverslip and breathable membrane as described in Garcia et al.,[8] Bothma et al.,[104] and Garcia and Gregor.[105]

Confocal microscopy was performed on a Leica SP8 with HyD detectors and a White Light Laser. We used a 63x oil objective, and scanned bidirectionally with a scan rate of 420 Hz and a magnification of 3.4x zoom. We did not use line or frame accumulation. Time-lapse z-stacks were collected with ~10 s frame rate and 106 nm x-y pixel dimensions and 0.5 μm separation between z-slices (7 μM range, 16 slices). x-y resolution was 512x512 pixels. Pinhole was set to 1.0 Airy units at 600 nm. mVenus was excited by a 510 nm laser line calibrated to 5 μW using the 10x objective and detected in a 520-567 nm spectral window. mCherry was excited by a 585 nm laser line calibrated to 25 μW and detected in a 597-660 nm spectral window. To image His2av-iRFP, the 700 nm laser line was set to 10% and detected in a 700-799 nm spectral window. In all channels, detection was performed using the counting mode of the HyD detectors.

All movies were taken at ~50% along the anterior-posterior axis of the embryo.

ParB experiment fly crosses and microscopy

We created flies with and without functional Dorsal expressing ParB2-EGFP maternally driven by the *nanos* promoter and MCP-mCherry driven by two copies of a minimal *nanos* enhancer to label our locus DNA and nascent mRNA, respectively. In addition, we added a parS sequence followed by a 400 bp spacer (created with SiteOut[25]) to our DBS_6.23 enhancer. We then crossed male flies containing parS-DBS_6.23-MS2 to *yw; ParB2-EGFP; eNosx2-MCP-mCherry;* + females to create embryos that have our locus of interest labeled with EGFP colocalized with transcriptional loci in the MCP-mCherry channel (Figures 5A and 5B).

After mounting embryos using the protocol described above in microscopy, we used the sequential scanning mode on the Leica SP8 confocal microscope to eliminate bleedthrough from eGFP into the mCherry channel, and imaged at approximately 20 s per stack, half the rate used in other imaging experiments in this study.

Image and time-series analysis

Image analysis was performed in Matlab using the custom pipeline described in Garcia et al.[8] and Lammers et al.[12] This pipeline is publicly available and can be found in the Github repository listed in the key resources table. Image segmentation was also aided by the Trainable Weka Segmentation plugin in FIJI.[106,107] Further analysis of time-series and other data were likewise performed in Matlab. Movies for publication were made in FIJI.[75,97]

Measuring Dorsal-mVenus concentration

Dorsal-mVenus concentration was calculated as in (Figure S8). As shown in the figure, we measured the average mVenus fluorescence intensity in a circle of 2 μm radius at the center of the nucleus in every z-slice of each nucleus. This results in a z-profile of fluorescence values covering the nucleus itself and the cytoplasm below and above it. The reported concentration corresponds to the value at the middle z-plane of each nucleus. To find this plane, we fit a parabola to the fluorescence z-profile. We use as the nuclear concentration the fluorescence value at the plane corresponding to the fitted parabola's vertex (Figure S8B). We then plotted this value over time and selected a single time point for each trace corresponding to the middle of each nucleus's observed trajectory (Figure S8B). To determine the background fluorescence in the mVenus channel we imaged flies with the same genotype as 4x Dorsal except for the Dorsal-Venus fusions. We calculated the average nuclear fluorescence in the mVenus channel across nuclear cycle 12 and subtracted this value from our Dorsal-Venus measurements.

QUANTIFICATION AND STATISTICAL ANALYSIS

Curve fitting and parameter inference

Curve fitting and parameter inference were performed in Matlab using the MCMCSTAT Matlab package using the DRAM Markov Chain Monte Carlo algorithm.[99] For simplicity, uniform priors were assumed throughout.