bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Constraints and limitations on decoding positional information: the Bicoid case-study

Huy Tran[a,b], Aleksandra M. Walczak[b,*] and Nathalie Dostatni[a,*]

[a]Institut Curie, PSL Research University, CNRS, Sorbonne Université, Nuclear Dynamics, Paris, France

[b]Ecole Normale Supérieure, PSL Research University, CNRS, Sorbonne Université, Laboratoire de Physique, Paris, France

*Corresponding authors: email address: awalczak@lpt.ens.fr ; nathalie.dostatni@curie.fr

Abstract

The regulation of expression of the *hunchback* promoter by the maternal Bicoid gradient has been studied as a model system in development for many years. Yet, at the level of quantitative agreement between data and theoretical models, even the first step of this regulation, transcription, continues to be challenging. This situation is slowly progressing, thanks to quantitative live-imaging techniques coupled to advanced statistical data analysis and modelling. Here we outline the current state our knowledge of this apparently "simple" step, highlighting the newly appreciated role of bursty transcription dynamics and its regulation.

Key words: Bicoid-hunchback regulation, transcription regulation, early fly developing embryo, physical limits, modelling, live-experiments, FISH, noise

1. Introduction

Precision in the establishment of cell identity during development is critical for the emergence of properly proportioned individuals and their survival. The identity of each cell is determined by the expression of a few specific genes regulated in many cases by morphogen gradients. It is generally assumed that the morphogen is produced in a source cell and that it diffuses from the source along an axis or into a field of cells (2D or 3D) in such a way that each cell is confronted with a given concentration of the morphogen. This allows each cell to measure its position relative to the source cell and to accordingly turn on the expression of morphogen target genes responsible for the determination of its identity. Although morphogen gradients have been shown to be essential in many developmental systems for decades, several questions concerning their functioning at the mechanistic level remain largely debated. In most cases, we do not know how morphogen gradients are established, how morphogens are able to activate target gene expression in a dose-dependent manner. Also, it remains unclear how morphogens themselves contribute to the precision of axial patterning. Here, we discuss recent developments in describing how a noisy transcriptional process can be tuned very rapidly into a reproducible functional developmental pattern by focusing on one well-studied morphogen-target gene pair – the Bicoid (Bcd) morphogen controlling the *hunchback* (*hb*) gap gene in early fly development.

The *hb* gap gene is involved in patterning along the antero-posterior (AP) axis of young fruit fly embryos. *hb* expression is first detected at the onset of zygotic transcription at the eighth nuclear cycle (nc8), around one hour after fertilization. As early as nc11, *hb* transcription occurs in a step-like pattern with a sharp boundary separating highly expressing nuclei in the anterior from nearly expressionless nuclei in the posterior [1,2] (Figure 1A). This step-like pattern is also observed at the Hb protein level at nc14 [3] and the protein pattern exhibits very low variability in concentration between nuclei of the same position along the AP axis [4]. This robust Hb pattern, combined with other gap gene patterns, was proposed to contain enough positional information for nuclei to predict their position in the embryo with ~ 99% accuracy [5].

In young embryos, the major regulator of *hb* transcription is the homeodomain-containing Bcd [6]. Translated from maternally anchored mRNAs at the anterior pole, Bcd proteins form an AP exponential gradient and activate *hb* transcription in the anterior half of the embryo. Increasing or decreasing the amount of Bicoid in the embryo induces a posterior, respectively anterior, shift of the *hb* step-like pattern, arguing that the expression of *hb* is Bcd concentration dependent [6]. The discovery of Bcd's role in *hb* transcription was exciting as it provided the first example illustrating the elegant idea of the French Flag model for morphogenesis. However, 30 years after its discovery, the mechanisms responsible for the *hb* pattern sharpness and reproducibility are not yet fully

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

understood. The first quantitative studies of the Bcd gradient and Hb protein pattern reproducibility [3,7] raised doubts about the pattern emerging solely from diffusive biochemical interactions between transcription factors and the gene promoter region. In this simplest scheme of concentration sensing, the Berg-Purcell scheme [8], the time required for a Bcd binding site of the *hb* promoter to read the Bcd concentration with 10% accuracy was estimated to be of the order of 2 hours [4], much longer than the interphase duration of nuclear cycles. These early attempts to explain the Hb pattern robustness marked a shift in experimental studies towards more quantitative approaches. The advent of new measurement methods, based on live imaging especially of transcription dynamics [9], allowed for a deeper view of the processes of positional readout in young fruit fly embryos. Specifically it allowed one to focus on the outcome of the first steps of regulation, transcription at the *hb* locus, to determine how this stage of early development is controlled. Thanks to advances in imaging discussed below, the pattern of transcription from the *hb* locus is observed to be established sharply in about 3 min at nc11, despite being generated by noisy gene expression from a bursty promoter. In this review we discuss recent work and challenges for advancing our understanding of positional information propagation from the Bcd gradient to the *hb* expression pattern (Figure 1). We made the choice to only focus on the first step, the transcription process, which in itself contains several steps (Figure 1B-D), and highlight why this system remains challenging 30 years on. More details about the establishment of the Bcd gradient can be found in the chapter by Huang and Saunders in the same issue [10].

2. The Bcd gradient

2.1. Positional information

The Bcd concentration gradient arises from maternally anchored mRNA at the anterior pole of the syncytial embryo [11]. Large scale analysis of the Bcd gradient, through immunofluorescent staining of the endogenous protein or analysis of the fluorescent-fusion Bcd-eGFP protein [7] (Figure 2A), indicated that the Bcd gradient is exponential with a concentration decay length around 80 to 120µm (between one-fifth and one-fourth of embryo length) [7,12,13]. The concentration of this gradient is a source of positional information for each nuclei along the AP axis. All the experiments performed to measure the absolute concentration of the Bcd gradient along the AP axis used the Bcd-eGFP expressed from a transgene in the embryo which rescues the viability of the bcd^{E1} null allele. The total concentration of Bcd at the anterior pole, measured *via* fluorescent Bcd-eGFP ranges from 90 nM, measured by comparing the fluorescence inside nuclei with the fluorescence of an eGFP solution at a given concentration [14], to 140 nM [13], equivalent to respectively, 4.4 (estimated to be ~ 700 Bcd molecules per nuclei at nc14 [7]) to 7 molecules/μm^3 where the *hb* pattern boundary is established. Even though these measurements are consistent, they were obtained with the same fluorescent Bcd-eGFP and are likely to be underestimates of the real concentration because a proportion of the eGFP might not be fluorescent. Further analysis using a Bcd fusion carrying two fluorescent domains [15] might help resolve the issue of absolute Bcd concentration measurements in the embryo. Similarly, taking into account the maturation time of the eGFP points to a slight overestimation of the Bcd gradient decay length of about 15% [15,16]. After corrections, the length constant of the Bcd gradient is estimated as low as one sixth of embryo length (16.5 ± 0.7 % EL).

2.2. The motility of Bcd molecules

The *hb* locus extracts positional information from the local Bcd concentration via interactions with Bcd molecules. Given the short time window for positional readout in each interphase, the Bcd search time for the *hb* promoter τ_{bind} is critical in determining the limit of positional readout precision. Therefore, several studies analyzed Bcd motility, using FRAP [4] or FCS [13] on fluorescent Bcd-eGFP or single-molecule tracking (Figure 2A) [17,18]). In the initial FRAP experiments, Bcd motility in the cytoplasm turned out to be quite slow (~ 0.3 μm^2/s) [7]. FCS experiments performed both in the cytoplasm and the interphase nuclei revealed the existence of Bcd molecules with different motilities: best fitting of the data to the two-species diffusion model indicated that *i)* in the cytoplasm, 18% of the Bcd molecules are slow-moving while 82% of the Bcd molecules are fast-moving with an average diffusion coefficient of ~ 7.4 μm^2/s [13] and *ii)* in the nucleus, 43% of the Bcd molecuels are slow-moving (~ 0.22 μm^2/s) and 57% fast-moving (~ 7.7 μm^2/s) [19]. Fast moving Bcd molecules (~ 4 μm^2/s) were also observed using a photoactivable Dronpa-Bcd [18] and the existence of at least two populations of Bcd molecules was further confirmed by high resolution single molecule imaging suggesting that in nuclei, Bcd molecules spend the same amount of time on nuclear exploration (searching for a binding target) and on binding to chromatin with surprisingly high unbinding rates, distributed with long tails [17].

2.3. Bcd searching time τ_{bind} for *hb* promoter

In an initial attempt to estimate the Bcd search time for the *hb* promoter, Gregor and colleagues proposed that Bcd molecules can diffuse in 3D inside the nuclear space in search for the specific Bcd binding sites on the *hb*

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

promoter (Figure 2D). They proposed that the search time was inversely proportional to the diffusion coefficient of Bcd *D*, the local Bcd concentration *c* and the size of the Bcd binding site *a*.

$$\tau_{bind} = \tau_{3D} \sim 1/Dca \tag{1}$$

The binding site for transcription factors, including Bcd, are commonly 10bp long which corresponds to ~3nm. Given this value and assuming that the searching Bcd molecules are diffusing freely (i.e. are fast-moving molecules) and that Bcd binding is diffusion limited (each collision between a Bcd molecule and a binding site results in a successful binding event), the binding site search time τ_{3D} is estimated to be on the order of 10 s. However, this is an optimistic estimate given that *a* is likely 10 fold too large since displacement by a single bp may lead to an entirely different DNA sequence that is not recognizable by the protein [20]. In this later case, *a* would be ~ 0.3 nm and the search time for Bcd would be 100s.

The robustness of the Hb pattern, despite a long search time in 3D, leads to hypotheses that Bcd molecules may use a combination of 1D and 3D diffusion to search for their target sites [20,21]. This mode of searching was first observed in *Escherichia coli* [22–24] for the *lacI* repressor where a combination of a 1D and 3D search was found to reduce the search time by up to 100 times compared to a pure 3D search. In this scheme, proteins bind non-specifically to DNA and, while doing so, slide along the DNA segment in search for the specific target site (Figure 2E). Therefore, the effective size of the target site is increased by the TF's sliding footprint along the DNA a_{1D} (up to a hundred bp, compared to the size of a binding site a = 10 bp). Another hypothesis is that the target locus could be located in a micro-environments with enhanced TF concentration ($c_{\text{local}}>c$) (Figure 2F), thus speeding up the search, as proposed in the case of the Ultrabithorax protein [25]. The recent observation of Bcd concentration in dense hubs [17,26] opens up the possibility that micro-environments that enhance local Bcd concentration could contribute to reduce the Bcd search time for the *hb* promoter. However, one should note that these mechanisms can also introduce non-linearities into the position sensing process: Bcd hubs were found to persist even in the posterior region of low Bcd concentration, leading to a much flatter Bcd concentration profile in hubs than in the cytoplasm. In addition, Hammar *et al.* observed that bound lacI molecules may interfere with the 1D-sliding molecules when the distance between the target sites is shorter than the sliding footprint [24]. If Bcd employs this mode of searching, the very short distances between Bcd binding sites on the *hb* promoter (as short as 12 bp) may introduce negative feedback to Bcd binding, instead of the positive feedback normally linked to a sharp *hb* pattern [27].

2.4. Activation of transcription by Bcd

The Bcd protein is able, on its own, to activate transcription when bound to a promoter containing its DNA binding site [28]. However, how this is achieved remains largely unknown. Structure-function analyses of the Bcd transcription factor indicated that it contains many redundant functional domains [29]. Besides its homeodomain which allows binding to DNA [30,31], the Bcd protein contains several independent activation domains which can activate transcription on their own when multimerized and fused to a Gal4 DNA binding domain in vitro [32] or in the early embryo [33]. These include a Glutamine-rich domain, a ST-rich domain and a C-terminal acidic domain. The Bicoid protein also contains independent inhibitory domains which reduce its activation potential [33,34]. Finally, activation by Bcd is enhanced by other transcription factors binding to the promoter. These include the maternal contribution of the Hunchback protein itself [19,35] or Zelda [26,36]. Yet, the mechanisms underlying these essential synergistic effects are poorly understood.

3. *hb* transcription dynamics

3.1. Visualizing *hb* transcription dynamics

RNA-FISH on fixed embryos allowed for the monitoring of *hb* nascent transcript accumulation at their site of synthesis inside each nucleus of the embryo, making it an initial marker to study ongoing transcription and promoter dynamics at a given locus. This allowed subsequently for the detection of single mature mRNAs in the cytoplasm and in the nucleus [2,37,38]. The observed data suggested, despite low heterogeneity at the protein level [4], a very noisy transcription process occurring with periods of promoter activity and inactivity [39].

RNA FISH requires fixation of the sample and can only provide a snap shot view of the transcription process at a given time (the time of fixation) during nuclear interphase. Following the pioneering work of R. Singer [9], the MS2 fluorescent RNA-tagging system has been implemented to monitor transcription dynamics in living early *Drosophila* embryo development [40,41]. The system takes advantage of strong interactions between the MCP coating proteins and its RNA stem loops from the MS2 bacteriophage. As nascent RNA containing stem loops are being transcribed, they are bound by fusion proteins MCP-GFP, making the transcription loci visible as bright fluorescent spots under the confocal microscope [42].

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Recently, the MS2 system allowed for the direct visualization in real-time of position-dependent activation of the proximal *hb* P2 promoter (700 bp) [43]: at each nuclear interphase (from nc11 to nc13), *hb* expression first occurs in the anterior then proceeds to the boundary region. Of note, this difference in the *hb* activation time following mitosis is observed even in the anterior region where Bcd is presumably saturated. Analysis of individual MS2 time traces indicates that the transcription process is not only variable among nuclei at the boundary of the expression domain where Bcd is presumably limiting but also in the anterior region with very high Bcd concentration. This observation suggests that Bcd concentration is rate-limiting in both the anterior and boundary dynamics of *hb* expression. In addition, as variability is also observed in the anterior region with high Bcd concentration, Bcd is not the sole factor contributing to the noise in *hb* transcription.

3.2. Characterizing *hb* transcription dynamics

The fluorescent time traces acquired with the MS2 system provide an indirect observation of transcription dynamics. The signal is noisy, convoluting both experimental and intrinsic noise with the properties of the MS2 probe. To obtain a specific fluorescent signal sufficiently strong to overcome background fluorescence due to unbound MCP-GFP molecules, a long probe of 24 RNA stem loops was used [40,41]. As the signal is only detected while the probe is being transcribed (Figure 3A and B), this introduces a significant delay (buffering time) between each instant the promoter is ON and the corresponding fluorescent detection. This buffering time of ~1 min [41,44,45] and the short length of the traces (5-15 minutes) prevent traditional analysis based on OFF time distributions or autocorrelation functions to quantify the statistics of the activation and inactivation times.

Desponds *et al.* developed a tailored autocorrelation analysis of the fluorescent time traces to overcome these limitations [46]. Combining this analysis with models of transcription initiation (Figure 3C) and estimates of the precision of the transcriptional readout, provided evidence for bursty transcription initiation in nuclear cycles 12-13 [46]. Namely, they find the dynamics in agreement with a telegraph model, in which the promoter switches between the ON and OFF states. Only during the ON state can RNA polymerase arrive and initiate transcription successively (Figure 3C). The best-fit switching period (for a full ON-OFF cycle) is in the order of ~30 s, with the probability to be in the ON state of ~50% at the anterior and of ~10% at the boundary.

It should be noted that the autocorrelation function analysis alone is not able to distinguish reliably between different models for promoter activation and requires complementary information about the precision of the transcriptional readout to conclude that transcription is most likely bursty (Figure 3D). Recently, an inference method based on hidden Markov model and maximum likelihood has been developed [47] and tailored for the MS2 system in *Drosophila* [48]. This discrete-time model employs a hidden compound state, which records the previous promoter states during the elongation time (Figure 3B). This compound state is used to map RNA Polymerase (RNAP)'s position on the reporter gene segment and calculate the active loci intensity at a given time. The rates of switching between the promoter states in each time step are fitted based on maximum likelihood. While computationally expensive, the method allows for direct model selection and shows that transcription bursts are prevalent in stripe gene expression in later stages of fly development [48–50].

3.3. Transcription regulation of *hb* gene by Bcd proteins

Data obtained from the MS2 system provided insights not only at the molecular level about the kinetics of the promoter behavior but also at the cellular level when considering individual nuclei along the AP axis and individual loci in each of these nuclei. In particular, it was possible to analyze the transcription dynamics of each *hb-MS2* locus at the scale of the whole embryo. This analysis indicated that depending on its position along the AP axis, each locus was able to either turn ON when positioned in the anterior or remain silent when positioned in the posterior. Surprisingly, the steep border forms in under 3 min at each nuclear interphase 11 to 13 [43]. This indicates that the system is able to measure extremely rapidly very subtle differences of Bcd concentration and produce a complete sharp border. This rapid responsiveness is fascinating because it is almost ten times faster than predicted by previous theoretical models assuming that the Bcd gradient is the only driver for the *hb* transcription process.

The steep Bcd-dependent *hb* pattern, given the smooth Bcd gradient, demonstrates a strong nonlinear regulation of the *hb* gene by Bcd. The presence of multiple Bcd binding sites on the *hb* promoter [6] suggests that such strong nonlinearities can be achieved by high cooperativity of Bcd binding to the *hb* promoter site. Cooperative binding of Bcd to multimerized binding sites was observed *in vitro* [27,51] but remains too weak to account for the extremely steep Bcd-dependent *hb* pattern observed *in vivo*. Synthetic reporters with only Bcd binding sites are weakly expressed in very anterior domains which harbor, however, remarkably steep posterior boundaries [28,52]. This suggests that Bcd and Bcd binding sites are sufficient to generate a steep posterior border and models of regulation by Bcd binding/unbinding can help understand how this is achieved.

A general model of transcription regulation via binding/unbinding of TF to the binding sites on the target promoter (Figure 4A), demonstrates that the pattern's degree of steepness, conventionally characterized by a Hill

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

coefficient H of the fitted Hill function, is limited by the number of TF binding sites N [53]. If the model satisfies detailed balance (which corresponds to the model in Figure 4B where only the number of occupied Bcd binding sites matters, and not their identity), H is bound by N. Experimental H values obtained when observing the protein level [4] and several features of the *hb* transcription dynamics (e.g. total amount of RNA produced [41,43]) range from ~5 to ~7, roughly equal to the number of known Bcd binding sites on *hb* promoter. This leads to assumptions that the 6 binding sites, with an unstable first Bcd-bound state (large k_{-1} in Figure 4B) and a stable fully bound state (small k_{-N} in Figure 4B), are sufficient to explain the observed pattern steepness in static measurements.

However, Estrada *et al.* did not consider the search time issue [53] and their theory cannot explain how the *hb* pattern can be established in such a short time of 3 minutes following mitosis, as observed in live imaging data [43] (Figure 4C). Considering a model which accounts for the Bcd search time for the *hb* promoter, Tran *et al.* [54] found that, at the mid-boundary position, very high pattern steepness ($H \approx N$) requires a very slow promoter switching time (called τ_{active}) between Bcd-bound states allowing transcription and Bcd-free states prohibiting transcription. Therefore, according to the model, it should take a very long time for the pattern to be established and this is in contradiction with the experimental data from the MS2 system [43]. In addition, slow promoter dynamics results in high nuclei-to-nuclei variability in the amount of total RNA produced in each nuclear cycle. Thus, it would require even more averaging time to achieve the robust protein pattern (10% variability) observed in nuclear cycle 14 [4].

The failure of the simple model to explain both the observed high pattern steepness and fast formation time begs for the reconsideration of the model's assumptions. Most obvious candidates are either the underestimation of the number of Bicoid binding sites N or overestimation of the promoter search time τ_{bind}. Estrada *et al.* suggested that energy expenditure, which removes the detailed balance assumption from the binding and unbinding process, can expand the model's limit on pattern steepness H beyond the binding site number N, allowing both high steepness and fast formation time at the same time [53]. However, including non-equilibrium binding does not resolve the problem of obtaining a steep yet precise boundary in a short time. Alternatively, Desponds *et al.* [55] suggested that instead of probing the concentration for a fixed time and then making the decision about the positioning of the nucleus, constantly updating the odds of being in an anterior *vs* posterior position, always results in much faster decisions for a fixed accuracy. Assuming a promoter with 6 Bcd binding sites, they showed that the decision time can be reduced by an order of magnitude compared to the classical Berg-Purcell scheme, possibly below the 3 minute limit.

3.4. Dissecting noise in *hb* transcription

Noise in transcription dictates the variability of transcript and protein readouts after each interphase and might play a role in determining nuclei identity in downstream processes [56]. However, beyond its characterization from observed data, we still lack the mechanistic understanding of processes responsible for this noise.

As transcription bursts are prevalent across the embryo in the very short early nuclear cycles, *hb* transcription dynamics is well-fitted by a two-state model, in which the switching rates between the ON and OFF states are modulated by the nuclei's position or Bcd concentration [38,39,46,57]. It should be noted that these ON and OFF states do not correspond to Bcd-free and Bcd-bound states of the promoter as in [53,54]: in the Bcd-saturating anterior region, the *hb* promoter is constantly active (i.e. bound by Bcd molecules) but transcription still occurs in bursts with the switching time between ON and OFF states $\sim 50s$ [46]. This suggests that promoter bursting may be an inherent property of transcription in this phase of development [38,50]. The early transcription of *hb* is also regulated by other transcription factors such as maternal Hb [19,35,58] or Zelda [36,43,59,60]. Though these factors other than Bcd may not act as a source of positional information, their concentration may be rate-limiting and therefore responsible for bursts.

In the *hb* boundary region, where cell fate decision is critical, *hb* transcription readout is more variable than in the anterior region [43,46]. This was initially thought to be due to extrinsic noise from Bcd variability [7] being amplified in this region. However, given the very high steepness observed from the *hb* pattern [39,43], the switching time between Bcd-dependent active and inactive states of the promoter (τ_{active}) is expected to be at least one order of magnitude greater than the Bcd search time τ_{bind} [54]. If the search is done via 3D diffusion ($\tau_{bind} \sim 10s$), τ_{active} at *hb* boundary is at least of a similar time scale as the switching time between ON and OFF states at the anterior ($\sim 50s$). In the context of very rapid embryo development (interphase duration of 5 to 15 minutes in nc11 to nc13), Bcd-dependent promoter switching becomes a non-negligible source of intrinsic noise that contributes substantially to the higher readout variability observed.

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Perspectives

Despite several decades since the identification of the Bcd gradient, we still lack a quantitative description of the process allowing for the transcription of its main target gene *hb* in a step-like pattern. The short timescales of early development and its remarkable precision are questioning how fundamental limits coming from stochastic processes such as diffusion or bursty regulation influence the molecular encoding of regulation. To pursue these issues, recent experimental advances are allowing us to rigorously test theoretical ideas, and push call for the creation of new models. This simple example of developmental biology, is turning out to also be a very nice *in vivo* testbed for transcriptional regulation and advances in single molecule techniques to study protein motility [17,18] promise to bring a more definitive view on how TF can find their target promoter. Finally, while recent works have shown that positional information can be accurately decoded at the level of the gap genes, decoding as well as encoding mechanisms in the earlier cell cycles remain unknown. The current experimental and theoretical methods are ready to tackle these questions.

Acknowledgements

We thank M. Andrieu, M. Coppey, G. Fernandes, C. Fradin, C. Perez-Romero, A. Ramaekers for stimulating discussions. Work in the Walczak and Dostatni labs is supported by PSL IDEX REFLEX Grant for Mesoscopic Biology (AMW & ND), ANR-11-BSV2-0024 Axomorph (AMW & ND), ARC PJA20151203341 (ND) and ANR-11-LABX-0044 DEEP Labex (ND). HT was supported by the PSL IDEX REFLEX Grant, the Institut Curie and the DEEP Labex.

References

[1] A. Porcher and N. Dostatni, Curr. Biol. **20**, 249 (2010).

[2] S. C. Little, M. Tikhonov, and T. Gregor, Cell **154**, 789 (2013).

[3] B. Houchmandzadeh, E. Wieschaus, and S. Leibler, Nature **415**, 798 (2002).

[4] T. Gregor, D. W. Tank, E. Wieschaus, and W. Bialek, Cell **130**, 153 (2007).

[5] J. O. Dubuis, G. Tkacik, E. F. Wieschaus, T. Gregor, and W. Bialek, Proc. Natl. Acad. Sci. **110**, 16301 (2013).

[6] W. Driever and C. Nusslein-Volhard, Cell **54**, 95 (1988).

[7] T. Gregor, E. Wieschaus, A. P. McGregor, W. Bialek, and D. W. Tank, Cell **130**, 141 (2007).

[8] H. C. Berg and E. M. Purcell, Biophys. J. **20**, 193 (1977).

[9] E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long, Mol. Cell **2**, 437 (1998).

[10] A. Huang and T. E. Saunders, Curr. Top. Dev. Biol. (2019).

[11] O. Grimm, M. Coppey, and E. Wieschaus, Development **137**, 2253 (2010).

[12] T. Gregor, W. Bialek, R. R. de Ruyter van Steveninck, D. W. Tank, and E. Wieschaus, Proc. Natl. Acad. Sci. **102**, 18403 (2005).

[13] A. Abu-Arish, A. Porcher, A. Czerwonka, N. Dostatni, and C. Fradin, Biophys. J. **99**, 33 (2010).

[14] T. Gregor, H. G. Garcia, and S. C. Little, Trends Genet. **30**, 1 (2014).

[15] L. Durrieu, D. Kirrmaier, T. Schneidt, I. Kats, S. Raghavan, L. Hufnagel, T. E. Saunders, and M. Knop, Mol. Syst. Biol. **14**, e8355 (2018).

[16] F. Liu, A. H. Morrison, and T. Gregor, Proc. Natl. Acad. Sci. **110**, 6724 (2013).

[17] M. Mir, M. R. Stadler, M. M. Harrison, X. Darzacq, and M. B. Eisen, Elife **7**, (2018).

[18] J. A. Drocco, O. Grimm, D. W. Tank, and E. Wieschaus, Biophys. J. **101**, 1807 (2011).

[19] A. Porcher, A. Abu-Arish, S. Huart, B. Roelens, C. Fradin, and N. Dostatni, Development **137**, 2795 (2010).

[20] M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004).

[21] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, J. Phys. A Math. Theor. **42**,

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

434013 (2009).

[22] A. D. Riggs, S. Bourgeois, and M. Cohn, J. Mol. Biol. **53**, 401 (1970).

[23] J. Elf, G.-W. Li, and X. S. Xie, Science (80-.). **316**, 1191 (2007).

[24] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, Science (80-.). **336**, 1595 (2012).

[25] A. Tsai, A. K. Muthusamy, M. R. P. Alves, L. D. Lavis, R. H. Singer, D. L. Stern, and J. Crocker, Elife **6**, 1 (2017).

[26] M. Mir, A. Reimer, J. E. Haines, X. Y. Li, M. Stadler, H. Garcia, M. B. Eisen, and X. Darzacq, Genes Dev. **31**, 1784 (2017).

[27] X. Ma, D. Yuan, K. Diepold, T. Scarborough, and J. Ma, Development **122**, 1195 (1996).

[28] E. Ronchi, J. Treisman, N. Dostatni, G. Struhl, and C. Desplan, Cell **74**, 347 (1993).

[29] V. Schaeffer, F. Janody, C. Loss, C. Desplan, and E. A. Wimmer, Proc. Natl. Acad. Sci. **96**, 4461 (1999).

[30] J. Trelsman, P. Gönczy, M. Vashishtha, E. Harris, and C. Desplan, Cell **59**, 553 (1989).

[31] S. D. Hanes and R. Brent, Cell **57**, 1275 (1989).

[32] F. Sauer, S. K. Hansen, and R. Tjian, Science (80-.). **270**, 1825 (1995).

[33] F. Janody, R. Sturny, V. Schaeffer, Y. Azou, and N. Dostatni, Development **128**, 2281 (2001).

[34] C. Zhao, A. York, F. Yang, D. J. Forsthoefel, V. Dave, D. Fu, D. Zhang, M. S. Corado, S. Small, M. A. Seeger, and J. Ma, Development **129**, 1669 (2002).

[35] M. Simpson-Brose, J. Treisman, and C. Desplan, Cell **78**, 855 (1994).

[36] Z. Xu, H. Chen, J. Ling, D. Yu, P. Struffi, and S. Small, Genes Dev. **28**, 608 (2014).

[37] M. W. Perry, J. P. Bothma, R. D. Luu, and M. Levine, Curr. Biol. **22**, 2247 (2012).

[38] B. Zoller, S. C. Little, and T. Gregor, Cell **175**, 835 (2018).

[39] H. Xu, L. A. Sepúlveda, L. Figard, A. M. Sokac, and I. Golding, Nat. Methods **12**, 739 (2015).

[40] T. Lucas, T. Ferraro, B. Roelens, J. D. L. H. Chanes, A. M. Walczak, M. Coppey, and N. Dostatni, Curr. Biol. **23**, 2135 (2013).

[41] H. G. Garcia, M. Tikhonov, A. Lin, and T. Gregor, Curr. Biol. **23**, 2140 (2013).

[42] T. Ferraro, T. Lucas, M. Clémot, J. De Las Heras Chanes, J. Desponds, M. Coppey, A. M. Walczak, and N. Dostatni, Wiley Interdiscip. Rev. Dev. Biol. **5**, 296 (2016).

[43] T. Lucas, H. Tran, C. A. P. Romero, A. Guillou, C. Fradin, M. Coppey, A. M. Walczak, and N. Dostatni, PLoS Genet. **14**, e1007676 (2018).

[44] T. Fukaya, B. Lim, and M. Levine, Curr. Biol. **27**, 1387 (2017).

[45] A. Coulon, M. L. Ferguson, V. de Turris, M. Palangat, C. C. Chow, and D. R. Larson, Elife **3**, 1 (2014).

[46] J. Desponds, H. Tran, T. Ferraro, T. Lucas, C. Perez Romero, A. Guillou, C. Fradin, M. Coppey, N. Dostatni, and A. M. Walczak, PLOS Comput. Biol. **12**, e1005256 (2016).

[47] A. M. Corrigan, E. Tunnacliffe, D. Cannon, and R. C. Jonathan, Elife 1 (2016).

[48] N. C. Lammers, V. Galstyan, A. Reimer, S. A. Medin, C. H. Wiggins, and H. G. Garcia, BioRxiv (2019).

[49] A. Berrocal, N. C. Lammers, H. G. Garcia, and M. B. Eisen, BioRxiv 335901 (2018).

[50] J. P. Bothma, H. G. Garcia, E. Esposito, G. Schlissel, T. Gregor, and M. Levine, Proc. Natl. Acad. Sci. **111**, 10598 (2014).

[51] D. S. Burz, R. Rivera-Pomar, H. Jäckle, and S. D. Hanes, EMBO J. **17**, 5998 (1998).

[52] O. Crauk and N. Dostatni, Curr. Biol. **15**, 1888 (2005).

[53] J. Estrada, F. Wong, A. DePace, and J. Gunawardena, Cell **166**, 234 (2016).

[54] H. Tran, J. Desponds, C. A. P. Romero, M. Coppey, C. Fradin, N. Dostatni, and A. M. Walczak, PLoS

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Comput. Biol. **14**, e1006513 (2018).

[55] J. Desponds, M. Vergassola, and A. M. Walczak, BioRxiv (2019).

[56] D. M. Holloway, F. J. P. Lopes, L. da Fontoura Costa, B. A. N. Travençolo, N. Golyandina, K. Usevich, and A. V. Spirov, PLoS Comput. Biol. **7**, (2011).

[57] H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, Phys. Rev. Lett. **117**, 1 (2016).

[58] F. J. Lopes, A. V Spirov, and P. M. Bisch, Dev. Biol. **72**, 181 (2011).

[59] M. M. Harrison, X. Y. Li, T. Kaplan, M. R. Botchan, and M. B. Eisen, PLoS Genet. **7**, (2011).

[60] C. Y. Nien, H. L. Liang, S. Butcher, Y. Sun, S. Fu, T. Gocha, N. Kirov, J. R. Manak, and C. Rushlow, PLoS Genet. **7**, (2011).

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

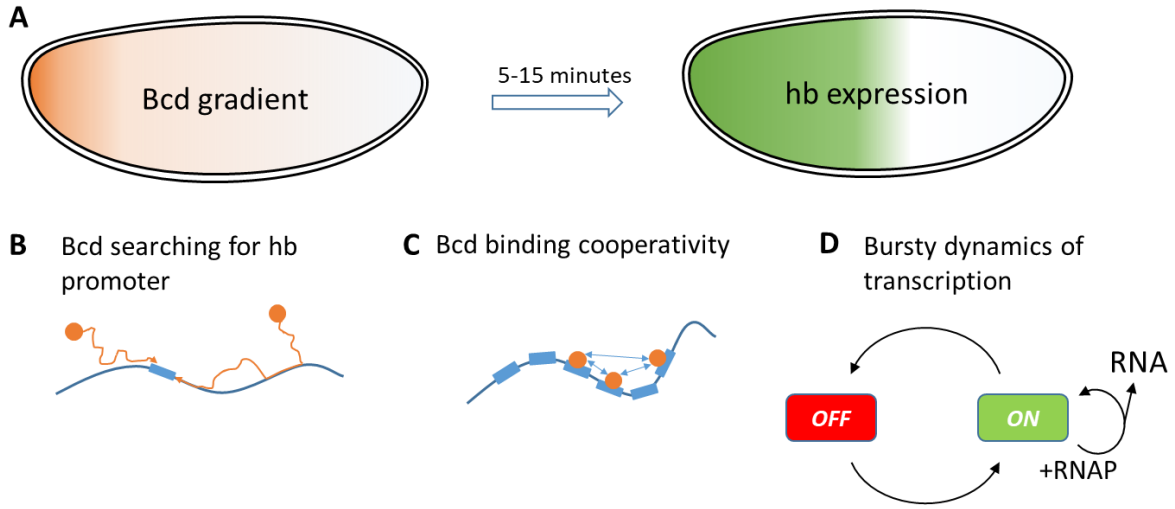

Figure 1. A) How is positional information contained in the shallow Bcd concentration gradient transformed into a robust step-like *hb* expression pattern in the 5 to 15 minutes of the nuclear cycle 11-13? **B-D)** The rate-limiting steps of *hb* expression in response to the Bcd gradient: **B:** The Bcd searches for the *hb* promoter, which can involve either 3D diffusion inside the nuclear space or 1D diffusion along the DNA. **C:** Bcd binds cooperatively to *hb* promoter, separating the embryo into a rich Bcd-bound anterior region and a no Bcd-bound posterior region. This process can involve exclusively Bcd molecules, other DNA bound transcription factors or factors facilitating chromatin accessibility. **D:** The production of *hb* RNA was shown to be bursty [46], with relatively infrequent switching of the *hb* promoter between the ON and OFF expression states.

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

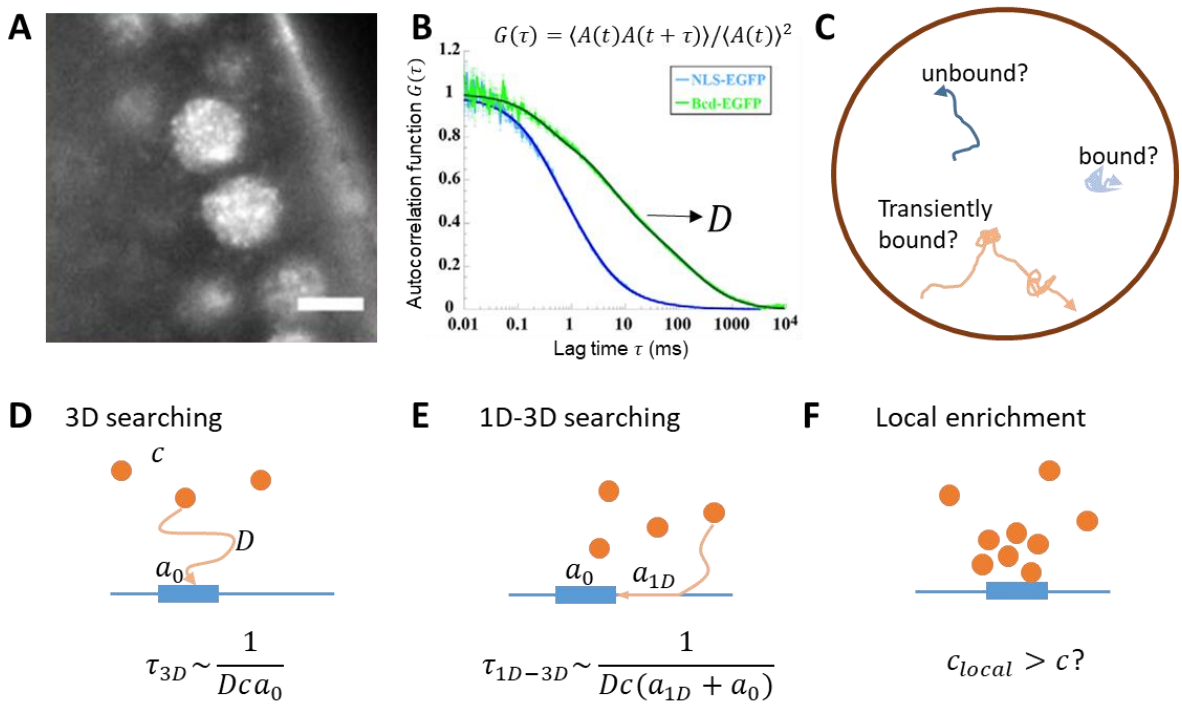

Figure 2. Studying Bcd motility in living fruit fly embryos. A) Fluorescent-tagged Bcd molecules were shown to be distributed in foci in the nuclear space of fixed embryos [39] and hubs of fluorescently-tagged Bcd [17,26]. Figure reused from [26]. **B-C**) Different motilities of Bcd molecules were inferred from Fluorescent Correlation Spectroscopy (FCS) analysis of the Bcd-eGFP signal in nuclei, with a Bcd population with a high (free) diffusion coefficient (**B**), Figure reused from [19] and **C**, by single-molecule tracking approaches revealing bound, unbound and transiently bound molecules [17]. **D-F)** various scenarios of Bcd searching for its binding sites in the *hb* promoter. **D**: Bcd molecules can diffuse in 3D nuclear space to search for Bcd binding sites on the target promoter. The time, $\boldsymbol{\tau_{3D}}$, for the binding site to be found is inversely proportional to the product of D (coefficient of free diffusion), c (concentration) and a_0 (size of the binding site); **E**: Bcd molecules can "hop" to non-specific locations on the DNA and slide along DNA segments in 1D to search for specific binding sites. The average sliding distance (or footprint) $\boldsymbol{a_{1D}}$ can be much bigger than the size of a Bcd binding site and therefore reduces the search time from the 3D case; **F**: Local enrichment of TF concentration ($c_{\text{local}}>c$) at the promoter region can also reduce the promoter searching time.

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

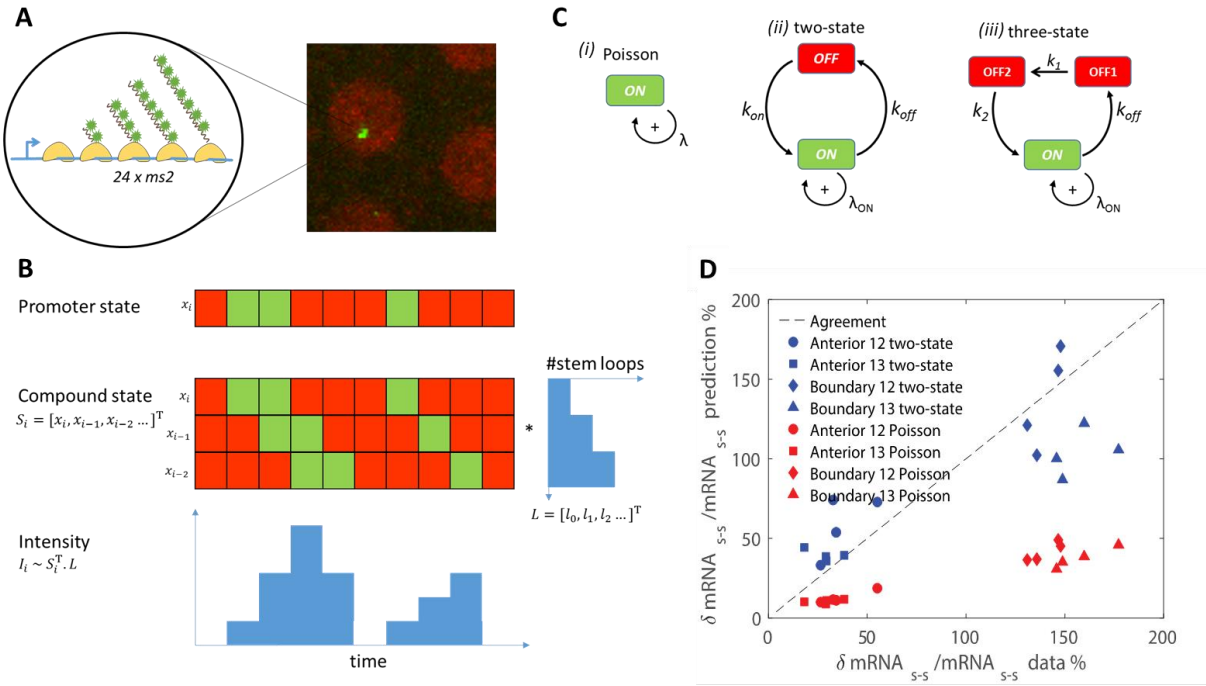

Figure 3. Inferring promoter dynamics from MS2 loci intensity: A) Visualization of active transcription loci: as RNAs containing MS2 stem loops are transcribed by RNA Polymerases (RNAP, in yellow), they are bound by fluorescent MCP-GFP molecules (in green). The succession of several RNAPs transcribing the gene allows for the accumulation of several fluorescently tagged MS2-containing RNA at the same location which become visible under the confocal microscope as green bright spots (right panel). **B**) Transformation from the promoter state dynamics to MS2 spot intensity dynamics: promoter state in discrete time $\boldsymbol{x_i}$ indicates whether the promoter is ON (green) or OFF (red). This state is encoded in the compound state vector $\boldsymbol{S_i} = [\boldsymbol{x_i}, \boldsymbol{x_{i-1}}, \boldsymbol{x_{i-2}} \dots]$, which also maps the position of RNAP on the MS2 cassette. RNAP arriving at time *i* will be transcribing a nascent RNA containing $\boldsymbol{L_j}$ MS2 stem loops at time $\boldsymbol{i+j}$. $\boldsymbol{L}$ depends on the length and on the arrangement of the MS2 stem loops on the reporter gene. The active loci intensity $\boldsymbol{I_i}$ at time $\boldsymbol{i}$ is given by the product of $\boldsymbol{S_i}$ and $\boldsymbol{L}$. **C**) Different models of promoter dynamics: *i)* Poisson model: random RNAP arrival and initiation of transcription; *ii)* two-state and *iii)* three-state models, where promoters switch successively between ON and OFF states. During the ON state, RNAPs arrive and initiate transcription in bursts, with maximum rates. **D**) Comparison of readout noise ($\boldsymbol{\delta mRNA/mRNA}$) at steady state generated from Poisson (red) and bursty two-state models (blue) and data (dashed). Figure reused from [46].

bioRxiv preprint doi: https://doi.org/10.1101/728840; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

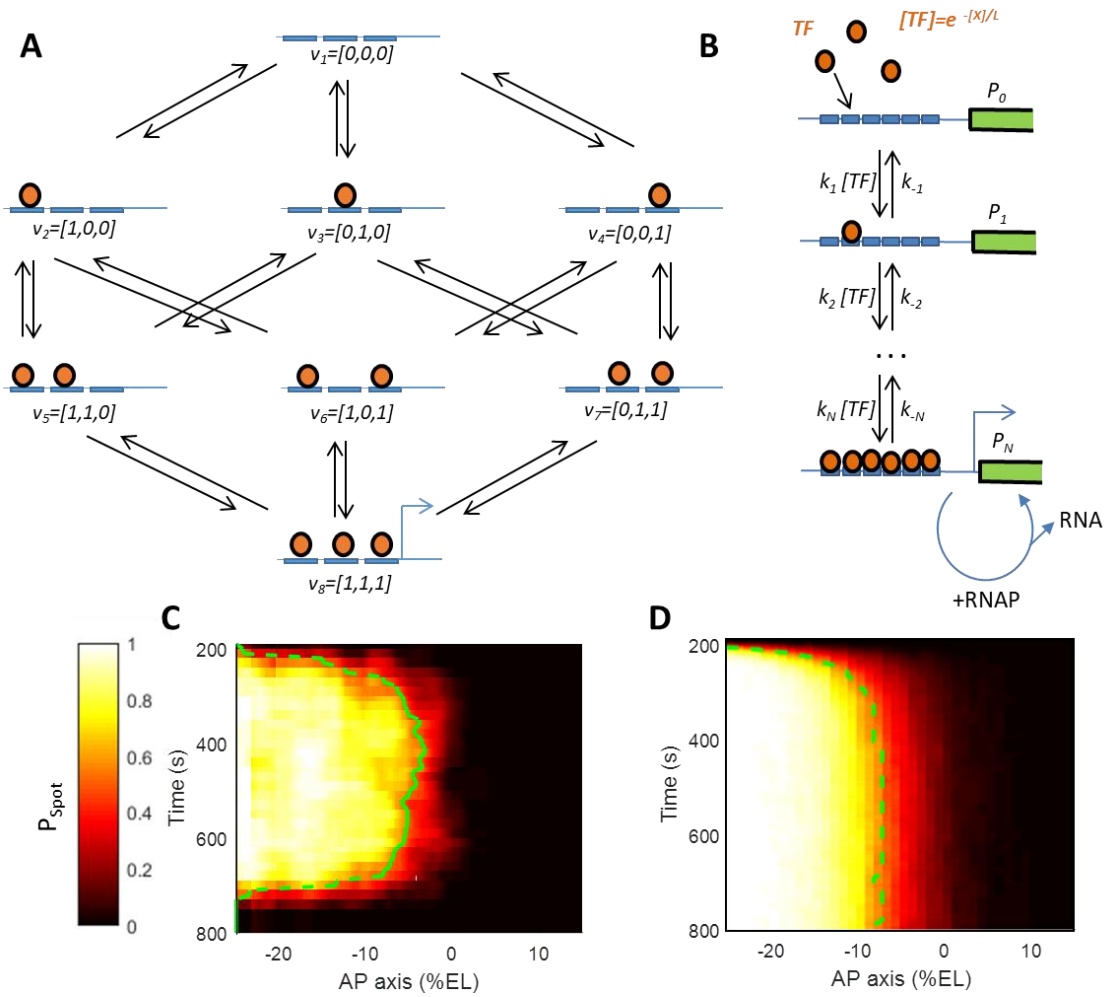

Figure 4: Modeling transcriptional regulation by the Bcd transcription factor through interactions with binding sites on *hb* promoter. A) A general model of transcription factors (TF) (orange) binding/unbinding to *N* binding sites of target promoter (*N*=3). Each node *v* corresponds to a unique ordered TF-bound promoter state. Figure adapted from [53]. **B)** A simplified promoter binding model assuming detailed balance to account for energy expenditure in the unbinding process [53], coupled with transcription initiation. When many Bcd binding sites on the *hb* promoter are occupied, RNAP can randomly bind to the promoter and initiate transcription. **C-D)** The probability of an active transcription locus (P_{Spot}, color bar) for the *hb* locus as a function of time in the nuclear cycle and position along the AP axis obtained from the MS2-MCP data (C, nc13) and from the model in B (D, with *N*=6, not accounting for mitosis at the end of interphase). Figure B-D are reused from [43].