Numerical simulations and modelling in biological species development

UNIVERSIDAD DE GRANADA

TESIS DOCTORAL

Manuel Cambón Gandarias

Departamento de Matemática Aplicada
Facultad de Ciencias
Programa Doctorado FisyMat
Universidad de Granada

10 de Octubre de 2020

Editor: Universidad de Granada. Tesis Doctorales
Autor: Manuel Cambón Gandarias
ISBN: 978-84-1306-707-0
URI: http://hdl.handle.net/10481/65312

Documento maquetado con TEXIS v.1.0+.

Este documento está preparado para ser imprimido a doble cara.

Numerical simulations and modelling in biological species development

A thesis presented for the degree of Doctor of Philosophy
Tesis presentada para obtener el título de Doctor

FisyMat, Universidad de Granada

Departamento de Matemática Aplicada
Facultad de Ciencias
Programa Doctorado FisyMat
Universidad de Granada

10 de Octubre de 2020

Declaración de derechos

Este trabajo ha sido parcialmente financiado por las becas de investigación MINECO-Feder (Gobierno de España) números FPI2015/074837 y RTI2018-098850-B-100 y la Consejería de Economía, Innovación, Ciencia y Empleo, Proyecto PY18-RT-2422 de la Junta de Andalucía (Gobierno de Andalucía), y el proyecto de la Universidad de Granada-Feder A-FQM-311-UGR18.

El doctorando, Manuel Cambón Gandarias, y el director de tesis, Óscar Sánchez Romero, garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección del director de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 10 de Octubre de 2020.

Director de la tesis | Doctorando

Fdo: Óscar Sánchez Romero | Fdo: Manuel Cambón Gandarias

A mi familia

Agradecimientos

Antes de nada, me gustaría agradecer a Óscar Sánchez, Juan Soler y el resto del Departamento de Matemática Aplicada de la Universidad de Granada, por darme la oportunidad de trabajar con ellos.

Por supuesto, no puedo olvidarme de mis compañeros del Despacho 60, aquellos personajes que tanta vidilla me han dado, y con los que tantísimos buenos ratos he podido disfrutar. Sí, os estoy hablando a vosotros, cotillas. Seguramente esta es la primera página de la tesis que habeis abierto para ver los secretillos que cuento de vosotros, ¿no? Vale, vale, como querais. Ahí van:

Quiero darle las gracias a Mauricio, o como todos le llamamos: el P. Grinch, o el Hawking del Swing. Cada uno de estos motes tienen su historia, pero ya sería contar demasiado. Quiero agradecerte de verdad esas canciones tan amenas y-no-tan-chirriantes que me hacían la tarde frente al ordenador. También esos momentos de trolleo, bailes en el pata-palo, tapeos y discusiones frente a la pizarra, a los cuales el Toledano le gustaba meterse para... bueno, quién sabe lo que hace el Toledano.

Y ya que estamos, sí, también quiero darle las gracias a Víctor (el susodicho Toledano). De verdad, no sé qué habría hecho en el Doctorado sin el Toledano. Has sido, sin exagerar, la voz que ha guiado la mayoría de mis pasos. Pasos que solían acabar unas bombitas del Bambi, unas empanadas del Cambalache, y rematando con unas Alhambras y billarcicos en el Liberia (o villancicos, como te gusta decir). Gracias, de verdad.

Pasando a la siguiente de la lista, y probablemente la persona que más ansiosa está de que hablen de ella, tenemos a Claudia. Ya puedo oír su vocecilla diciendo 'qué tontillo eres'. A ti también hay bastantes cosas que agradecerte, pero... nah, ya las sabes todas. Lo único que nunca te agradeceré es que me hayas sustituído por ese chico. ¿Cómo le llamabas? Ah, sí, 'New Manu'.

Y es que, el pobre no tiene la culpa de tener el mejor nombre del mundo. Manu, también conocido como el Tower por su habilidad en el Liberia en el ajedrez (¿qué esperábais? ¡Ahí se hace más que tomar Alhambras y jugar al billar!). A ti tengo que agradecerte precisamente el que te decidieras a hacer la tesis con nosotros. Tu llegada al Despacho (una pena que fuese en mi último año) nos ha alegrado a todos. Tengo muy buenos recuerdos contigo de noches de flamenco en el Albaicín, guitarrita española en tu terracita frente a la Alhambra, y kebaps del Mar Chica. Ya que estamos, aprovecho para darle las gracias también a ese hombre. Prepara posiblemente los mejores kebaps de toda calle Elvira. Si alguna vez lees estas líneas, gracias de corazón por quitarme la resaca en más de una ocasión. Uno se cree que puede seguirle el ritmo al Toledano, y claro, pasa lo que pasa.

Pero recordando días duros de resaca, creo que quien se lleva la palma es cierta personilla que nos abandonó a todos demasiado pronto. No, no está muerto, sólo leyó la tesis en mi primer año de Doctorado. Hablo de ti, Luis. Porque si no hay nada que más recuerde en esta vida es ese cóctel explosivo que te marcaste aquel día en la playa. Y no, no recuerdo lo que llevaba, pero sí

recuerdo que tuvimos que esperar mucho, pero que mucho, para volver a casa. En realidad te lo agradezco, porque ese día nos lo pasamos genial todos los del Despacho, y estrechamos grandes lazos que aún duran a día de hoy. De hecho, tuvimos una nueva incorporación al despacho ese día. Hablo de nuestro queridísimo amigo, βfumeiro. Hiciste un gran trabajo cuidando de él, todos esos meses en tu terraza a pleno sol. Nunca te lo agradeceré lo suficiente.

También ese día salió bastante perjudicadillo David. Sí David, a ti te dejo para el final. Tú me pusiste el primero en tu lista de agradecimientos del Despacho, así que te la devuelvo pero a la inversa. Dicen que se suele dejar lo mejor para el final, ¿no? A ver, por una parte sí es así, pero tengo que decirle a todos los demás aludidos que realmente os he ordenado por nivel de toca-ejem. ¡Alguien tenía que hacerlo! A ti David, primero te respondo de tesis a tesis: sí, me gustaban mucho nuestras conversaciones de citonemas a altas horas de la noche. Y sí, lo hacía realmente para llevarte a cenar. Pero gracias a esas cenillas, y esas conversaciones, creo que hemos llegado a entablar una gran amistad. Es posible que ahora nuestros caminos diverjan, y cada uno estará con sus respectivos post-docs en la otra punta del mundo. Pero estoy seguro de que incluso así, en algún punto del espacio y del tiempo, nuestras trayectorias tenderán a converger. Y algo me dice que ese punto será un punto estable.

Me gustaría hacer mención especial también a todas esas personas que se pasaban por el Despacho de vez en cuando. Gracias a Salva, por esas clases magistrales de ocarina con las manos en Liberia (Gabi me debería pagar comisión por esto). También a Ricarda, por ayudarme para encontrar una mesa que nos viniera bien a los dos. Por supuesto, gracias también a Juan Calvo y Juan Campos, por venir a buscarme para ir a desayunar, almorzar en los Comedores, y alguna que otra cervecilla tras una dura jornada de trabajo. Y a Cristina, gracias por esa fauna de origamis gigantes que aún pueblan mi mesa.

Ahora me gustaría hacer un pequeño enciso, y salirme de Andalucía un momento para dirigir mis agradecimientos a gente que están lejos de aquí. Pero para aquellos que aún estén esperando a ser nombrados, sed pacientes. Volveré.

Por un lado, me gustaría darle las gracias al Centro de Biología Molecular Severo Ocha de la Universidad Autónoma de Madrid. En particular, a los integrantes del Laboratorio de Isabel Guerrero que tan bien me han acogido todas las veces que he estado por allí. Muchas gracias a la misma Isabel, por sus claras explicaciones e interesantes ideas. Y por supuesto, especial mención a Adrián, mi compañero de... bueno, de bastantes cosas, la verdad. Gracias por tu gran esfuerzo y dedicación con los experimentos. Sin ellos gran parte del trabajo discutido aquí no se habría podido realizar. Recuerdo esas locas noches en el CBM, ciegos de tanto mirar por el confocal a ver si podíamos medir vesículas y citonemas. También, gracias por llevarme para arriba y abajo en las escaleras del CBM, desde la planta más baja hasta la más alta. Es un buen deporte, mi corazón te lo agradecerá. Y ahora fuera bromas, muchas gracias por acogerme tan bien allí, por recogerme y llevarme al trabajo, de cena, y todo lo demás. Creo que que mis estancias por allí no podrían haber sido mejores. Ah, y por supuesto, dale las gracias a tu madre por parirte el día de MI cumpleaños.

Aunque no sé si debería alegrarme mucho eso...

ずっとスペイン語で書いて恐縮です。お待たせしました。今のチャプターで、この論文に関係する重要な方々ににお礼を申し上げます。なので、もちろん、深谷研究室の皆さんへ感謝を伝えさせていただきたいと思います。まずは深谷さん、今までありがとうございます。初対面の私に、チャンスを与えてくださって、そして深谷さんの研究室で働かせてくださり、心から感謝しています。あの数ヶ月の間にたくさんのことを学びました。将来また一緒に働けることを願っております。そして、深谷研究室のメンバーの皆さんにもお礼を伝えたいです。余越さん、困った時に助けてくださり、そしていつも応援してくださって、本当にありがとうございました。川崎君、ラーメンを食べてビールを飲みながら、スペイン語と日本語（スペ本語!）で色んな話してくれてありがとう。また「Hasta mañana」を聞かせてね！濱本君、ゲームや車の話が楽しかったよ。でも今度日本に行けたら話じゃなく、生で車レーシングを見ましょう！そして、皆のお母さん、すなわち滝下さん、いつも困った時に助けてくれて本当にありがとうございました。皆さん、ありがとう。これからもよろしくお願いします。 Quisiera también hacer mención al Laboratorio del doctor Tetsuya Tabata. 田羽多先生へ。4年前このPhDを始めた時に、形態形成について何も知りませんでした。しかし、2004年に先生が出した「Development」の記事の読んで、すごく勉強になりました。そのお陰で今この論文をかけるようになったと思います。本当に感謝しております。IQBで田羽多先生と直接話せてとても嬉しかったです。 También quisiera darle las gracias a los integrantes del Laboratorio de Teruhiro Okuyama. 奥山先生、すっごく美味しいお酒や、ビールや、ワインや、あと記憶があんまり残っていない位楽しい夜に先生と一緒に過ごせて、たのしかったです。本当にありがとうございました。今度、日本に戻ったらいいスペインのワインをご馳走させてください！もちろん、たおさんもね！そして、美禰子さんにも感謝しています。まさか憧れの歌手に会えるとは思っていなかった。今度私が日本へ来たときには、一緒にカラオケへ行きましょうね！

A mis amigos de la Facultad de Física de la Universidad de Sevilla: Dock, Jumax, Migue TAN, Migue, Alberto y Alessandra. Con vosotros, estudiar Física fue toda una aventura.

También quiero darle especialmente las gracias a Maria José Ruiz Montero, que en paz descanse, sin ella esta tesis no habría sido posible.

A todos mis compañeros de intercambio que conocí durante la carrera en Sevilla: 皆がいなかったら、私は日本語を話せるようにならなかった。でもまあ、「compañero de intercambio」ってのはアレだけど。だって、私たちは言語交換の相棒だけじゃないでしょ？みおちゃん、なっちゃん、純基、ソフィア、かな、みやこちゃん、わかなちゃん:日本の世界を初めて見せてくれてありがとう。皆のお陰で日本語を使い始め、「一歩」できた。まさふみ、しおり、しほこ、ぱいせん、もとむ、なつみちゃん、あやちゃん、修平君、佳菜葉: 皆のお陰で日本語をもっと使えて、自由に「歩く」と「走る」ことができた。そして、あずさ、君のお陰で「飛ぶ」ことができた。本当に、心から、皆ありがとう。

Tampoco puede faltar mencionar a mis profesoras de Japonés, con quien todo empezó: おきた先生、えりこ先生、ありがとう！

Gracias también a todos aquellos que he conocido durante mi vida en Granada.

Gracias a Carolina (かえでちゃんしーた). この二年間日本語と他にも教えてくれて、そしてたくさん助けてありがとう (これからもよろしくおお願いします！). Por supuesto, gracias a ti también Susana (すずちゃん), 三人で過ごした夏は楽しかったね。また三人でどこかで会おう！También quiero dar una especial mención al mejor bar japonés de toda Granada: Izakaya Kobachi. Para aquellos que os guste la comida japonesa, teneis que ir. Gracias りょうさん (¡seññoooorrrrr!)。あのtapasや、生ビールや、夜中でいろんな話できてすっごく楽しかった。グラナダではもう働かないだろうけど、絶対にまた遊びに行きますね. 後、みちこさんにもお礼を言わなくちゃ。みちこさんのお陰で日本酒の真髄を学んで、味わえるようになった。人生が変わったぐらい！だから、ありがとう。また一緒にいい純米酒や大吟醸酒を飲もう！Y por supuesto, ゆいちゃん、あかねちゃん、ゆきちゃん、りさちゃんもありがと！

A todos los demás que por falta de espacio o memoria no he podido citar. Gracias de verdad.

Siendo consciente de que he dejado lo más importante para el final, quisiera darle las gracias a mi familia. Sin ellos no habría llegado a ninguna parte en la vida. Muchas gracias a mis hermanos, únicos e irreemplazables. Gracias a Javier, por aguantar todas y cada una de mis rarezas durante tantos años. Sé que ha sido duro, pero poco a poco iré mejorando. Muchas gracias a Jorge, por ser el hermano mayor que cualquiera habría deseado. He aprendido muchísimo de ti. Aprovecho también para darle las gracias a Bárbara, por hacer feliz a mi hermano y a todos nosotros. Y por supuesto, gracias a Choko por ser la bombita más bonita del universo. Aunque ándate con ojo, ahora ha llegado alguien que te puede hacer sombra. Con el poco tiempo que lleva, ya se merece estar entre estas páginas: ¡gracias Yitán! Gracias a mi Abuela, tíos y tías, por estar ahí desde que tengo uso de razón. Y por último, gracias a mis padres, los verdaderos merecedores de esta tesis. Se requiere mucho esfuerzo y determinación para poder guiar y educar correctamente a un hijo, y mucho más para hacerlo tan bien como vosotros habéis hecho conmigo. De no ser por vosotros, directamente ninguna de estas páginas habrían sido posibles. Es por ello que quiero reservar estas últimas líneas a deciros lo que, por vergüenza o sobreentendimiento, pocas veces me habréis oído decir. Vuestra sección ha sido sin duda la más complicada de escribir. Y es que hay tantas cosas que agradeceros que tan solo enumerarlas me llevaría cien veces más el número de páginas totales que abarca toda la tesis. Por eso, tras darle vueltas y vueltas a lo que quiero deciros, creo que lo resumiré solo en tres palabras. Son tres palabras que deseo con todas mis fuerzas se queden grabadas aquí y en todos los tiempos, en todos los posibles universos paralelos, y en todos los momentos que hemos vivido y viviremos juntos. Papá y Lola (lo siento, pero no vamos a cambiar las costumbres ahora): gracias. Os quiero.

Y con esto, damos por finalizada esta sección tan personal. En los siguientes capítulos entraremos en materia, no sin antes lanzar un último agradecimiento:

Gracias a ti, mi lector. No sé si sabes dónde te estás metiendo, pero de todos modos gracias por dedicarme tu preciado tiempo.

Dissertation Summary

This thesis focuses on the development of models applied to different areas of morphogenesis. These areas refer to different stages of the development of the living being. For this reason, the models presented must be able of working at different scales, both spatial and temporal. This gives rise to a modeling work where, depending on the problem to be dealt with, particular mathematical tools have to be used for each one. Going into more particular details, we can summarize the biological scales covered by this thesis into two: the tissular scale and the molecular scale. On the one hand, the tissular scale collects global biological events, which occur in a wide area of the developing tissue. At this scale, the importance of the problem to be modeled resides not in a sole element, but in a set of elements. Generally it is usually related with the mathematical concept of macroscopic scale, where it tends to pose continuous models defined by partial derivative equations. A clear example of these models is the diffusion of proteins (signals) in a tissue. In this type of problem, the temporal evolution of protein concentrations is modeled using equations, usually parabolic, such as the diffusion equation. However, in this thesis we study the signaling problem from a perspective more focused on the biological elements itself. In this way, although it is still a problem on a tissular scale, the way to approach it and model it mathematically will depend basically on the biological machinery behind it: in this case, the cytonemes. Cytonemes are biological components of which, at the time of deposit of this document, little is still known both at the biological and mathematical levels. Applying functional minimization techniques, and in constant agreement between both biological and numerical experimentation, the present thesis proposes novel modeling techniques to better analyze these elements. On the other hand there exists the molecular scale. This focuses on individual units, with their own functioning, such as the nucleus of cells that make up a tissue. These, depending on the external information they receive, behave in one way or another. In turn, the behavior of a cell is encoded in the genes in its DNA. Therefore, a proper understanding of how gene transcription (copying) works is a fundamental basis to better understand the ins and outs of morphogenesis. At this scale, the present thesis tries to address the problem of gene transcription through thermostatistic modeling.

List of articles

During the development of the thesis the following articles have been published or submitted:

- *Analysis of the transcriptional logic governing differential spatial expression in Hh target genes*, in collaboration with Óscar Sánchez, published on PLOS ONE (2019),
- *Thermodynamic modelling of transcriptional control: a sensitivity analysis*, in collaboration with Óscar Sánchez, submitted to Applied Mathematics and Computation (2020),
- *Regulation of transcriptional bursting by core promoter elements*, in collaboration with Takashi Fukaya and Moe Yokoshi, submitted to Cell Reports (2020),

and working on:

- *Modeling of morphogen transportation along moving cytonemes in Droshophila melanogaster*, in collaboration with Adrián Aguirre-Tamaral, Isabel Guerrero and Juan Soler,
- *Burst regulation by transcription factors*, in collaboration with Takashi Fukaya.

Thesis contents and methodology

The document presents a total of 4 chapters, structured in 2 parts. Chapter 1 summarizes the biological concepts that will be treated throughout the thesis. Two parts that collect the main content of the document. Each part contains chapters structured as follows: an introduction to the problem to be modeled, deduction and properties of the proposed mathematical model, and an experimental application to a biological system (Drosophila fly). In this way, the first part is focused on the tissular scale, where Chapter 2 (Cytonemes) is collected. The second part deals with content related to the molecular scale, in Chapter 3 (Transcriptional Dynamics). Finally, Chapter 4 includes some discussions of possible extensions applicable to the models presented, as well as future work to be carried out after the thesis submission. Additionally, an Appendix section has been included, where the reader can find detailed proofs and calculations that have not been introduced in the main text.

Resumen en castellano

La presente tesis se centra en el desarrollo de modelos aplicados a distintos ámbitos de la morfogénesis. Dichos ámbitos se dan en diferentes etapas del desarrollo del ser vivo. Por ello, los modelos presentados han de ser capaces de trabajar en distintas escalas tanto espaciales, como temporales. Esto da lugar un trabajo de modelado donde, dependiendo del problema a tratar, se han de utilizar herramientas matemáticas particulares para cada uno. Entrando en detalle, podemos resumir en dos las escalas biológicas que tratan la presente tesis: la escala tisular, y la escala molecular. Por un lado, la escala tisular recoge eventos biológicos globales, que se dan en una amplia zona del tejido en desarrollo. En esta escala la importancia del problema a modelar reside no en un elemento en particular del problema, sino en un conjunto de elementos. Generalmente suele atribuirse con el concepto matemático de escala macroscópica, donde se tiende a plantear modelos continuos definidos por ecuaciones en derivadas parciales. Un ejemplo claro de estos modelos es la difusión de proteínas (señales) en un tejido. En este tipo de problemas se suele modelar la evolución temporal de las concentraciones de proteínas mediante ecuaciones usualmente tipo parabólicas, como la ecuación del difusión. Sin embargo, en la presente tesis se ha optado por estudiar el problema de señalización desde una perspectiva más centrada en el funcionamiento biológico en sí. De esta manera, si bien sigue siendo un problema a escala tisular, la forma de abordarlo y modelarlo matemáticamente dependerá mucho de la maquinaria biológica que hay detrás: en este caso, los citonemas. Los citonemas son componentes biológicas de las que, a día de depósito de este documento, aún se conoce poco tanto a nivel biológico como matemático. Aplicando técnicas de minimización de funcionales, y en constante convenio entre la experimentación tanto biológica como numérica, la presente tesis propone novedosas técnicas de modelado para analizar mejor estos elementos. Por otro lado está la escala molecular. Ésta se centra en unidades individuales, con un funcionamiento propio, tales como son los núcleos de las células que componen un tejido. Éstas, dependiendo de la información externa que reciben, se comportan de una manera u otra. A su vez, el comportamiento de una célula está codificado en los genes que conforman su ADN. Por ello, comprender de forma adecuada cómo funciona la transcripción (copiado) de genes es una base fundamental

para poder entender mejor los entresijos de la morfogénesis. En esta escala, la presente tesis trata da abordar el problema de la transcripción génica mediante modelados termoestadísticos.

Lista de artículos de la tesis

Durante el desarollo de la tesis se han publicado o sometido los siguientes artículos:

- *Analysis of the transcriptional logic governing differential spatial expression in Hh target genes*, en colaboración con Óscar Sánchez, publicado en PLOS ONE (2019),
- *Thermodynamic modelling of transcriptional control: a sensitivity analysis*, en colaboración con Óscar Sánchez, sometido a Applied Mathematics and Computation (2020),
- *Regulation of transcriptional bursting by core promoter elements*, en colaboración con Takashi Fukaya y Moe Yokoshi, sometido a Cell Reports (2020),

y actualmente se está trabajando en:

- *Modeling of morphogen transportation along moving cytonemes in Droshophila melanogaster*, en colaboración con Adrián Aguirre-Tamaral, Isabel Guerrero y Juan Soler,
- *Burst regulation by transcription factors*, en colaboración con Takashi Fukaya.

Organización y metodología del contenido de la tesis

El documento se presenta un total de 4 capítulos, estructurados en 2 partes. El Capítulo 1 resume los conceptos biológicos que se irán tratando a lo largo de la tesis. A continuación se presentan las dos partes que recogen el contenido principal del documento. Cada parte contiene capítulos estructurados de la siguiente manera: una introducción del problema a modelizar, deducción y propiedades del modelo matemático propuesto, aplicación experimental a un sistema biológico (mosca Drosophila). De esta manera, la primera se centra en la escala tisular, donde se recoge el Capítulo 2 (Citonemas). La segunda parte trata el contenido relacionado con la escala molecular, recogiendo el capítulo 3 (Dinámica transcripcional). Por último, el Capítulo 4 recoge algunas discusiones de posibles extensiones aplicables a los modelos presentados, así como trabajo futuro a realizar tras el depósito de la tesis. Adicionalmente se ha incluído una sección de Apéndices, donde el lector

puede encontrar demostraciones y deducciones detalladas de cálculos que no se han introducido en el texto principal.

Contents

Chapter 1

Biological introduction

Morphogenesis [104] (from the Greek *morphê* shape and *genesis* creation), as its name suggests, is the biological process that causes a cell, tissue or organism to develop its different parts and tissues. It is one of three fundamental aspects of developmental biology along with the control of tissue growth and patterning of cellular differentiation. It can also take place in a mature organism, such as in the normal maintenance of tissue homeostasis by stem cells, or in regeneration of tissues after damage [76]. Highly abnormal and pathological tissue Morphogenesis is involved in degenerative diseases, and cancer [4, 81].

The Morphogenesis has been a long-standing question from 19-th century beginnings, from both experimental and theoretical point of view. Some of the earliest ideas and mathematical descriptions on how physical processes and constraints affect biological growth were written by D'Arcy Wentworth Thompson in his 1917 book '*On Growth and Form*' [104]. In his work, he explained animal body shapes as being created by varying rates of growth in different directions, for instance to create the spiral shell of a snail. Alan Turing in his work '*The Chemical Basis of Morphogenesis*' (1952) [108] predicted a mathematical mechanism that describes Morphogenesis. In his work he proposes a model of diffusion of two different chemical signals, one activating and one deactivating biological processes, that predicts theoretically the formation of developmental patterns (the so called Turing Patterns). Some decades later, these patterns have been observed experimentally, and fruitfully, in many relevant works [68, 56].

To examine Morphogenesis in depth, we need a sufficiently flexible (and visible) biological system that can accommodate a range of experiments. As the reader can understand, analyzing Morphogenesis involves studying living beings at various stages of their development. That is, we need to have access to internal biological variables (levels of protein expression, cell tissue, etc.),

with a resolution and in reasonable laboratory time scales. The Drosophila fly has been one of the experimental setting par excellence.

1.1 Drosophila

Drosophila [109] is a genus of small flies, belonging to the Drosophilidae family, whose members are often called 'fruit flies'. They are also called vinegar flies since many species tend to remain near ripe or rotten fruit. The Drosophila family is made up of about 1500 varieties, among which we have a particular species: *Drosophila melanogaster*. With a small number of chromosomes (4 pairs) and a short life cycle (around 2-3 weeks), D. melanogaster is one of the most widely used species in the field of genetics. And it is that, although it seems implausible, this little fly has a genome with 13,600 genes that is certainly human-like. It is known that no more and no less than 75% of human disease genes have a version in the D. melanogaster genome, as well as more than 50% protein sequences that have analogs in mammals [109]. All this, added to its high reproduction rate, make *Drosophila* [1] one of the best working tools for genetic experimentation and the study of Morphogenesis. Throughout this thesis, we will work extensively with experimental data from *Drosophila*. Therefore, it is necessary to make a brief introduction to certain basic concepts that define it.

At one hand, *Drosophila*'s body is defined in terms of four axes (Fig. 1.1):

- *Dorso-Ventral*: Axis that goes from the belly (Ventral) to the back (Dorsal) of the fly.
- *Anterior-Posterior*: Axis that goes from the front (Anterior) to the back (Posterior) of the fly.
- *Right-Left*: Axis that goes from the right view to the left of the fly.
- *Proximo-Distal*: Axis that goes from the closest part (Proximo) to the outermost part (Distal) to the fly, in each of its extremities.

On the other hand, the fly exhibits a metamorphic development cycle, similar to that of butterflies and other insects (Fig. 1.2). That is, from the laying and fertilization of the eggs, the fly goes through the stages [31]:

- *Embryonic stage*: Characterized by 17 cellular stages, and maternal-provided morphogens such as Dorsal and Bicoid (see definition of morphogen in next section).

[1] From now on we will omit the surname *Melanogaster*

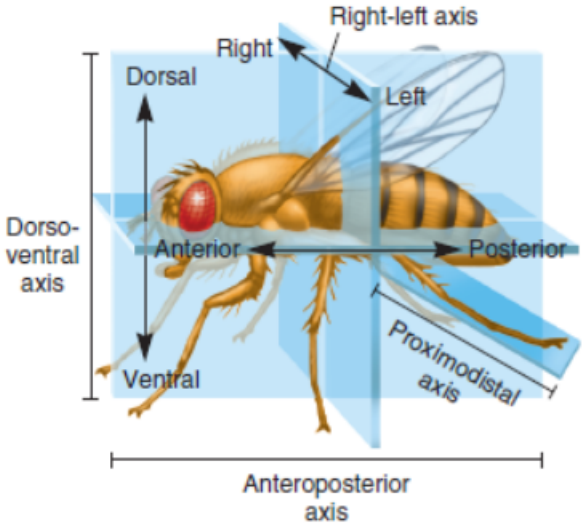

Figure 1.1: *Drosophila*'s axes. Image obtained from [92]

- *First, Second and Third instar larva stage*: Formation of primal structures, such as the wing and eye imaginal discs. Each of these structures are also equipped by morphogens, such as Hedgehog.

- *Pupa stage*: Last stage where the primal structures transition to the adult form.

- *Adult stage*: Final stage. 8-10 hours after eclosion, females become sexually mature and the cycle starts again.

All these stages are governed by dynamical cellular structures, and signaling proteins that orchestrate the process. In the next section we will go deeper, and define what is the main functionality of these proteins, and their importance in morphogenetic processes.

1.2 Morphogen signaling

If there were an essential word in Morphogenesis, it would be 'signaling'. However, it must be clarified that this word in biology differs slightly from the common term which we are familiar with. In biology, the signals are usually attributed essentially to how certain proteins are distributed and configure the system. Proteins are chains of aminoacids that occur in nature in different shapes and structures. Depending on their chemical composition, they are able to anchor and contact each other. These interactions are called

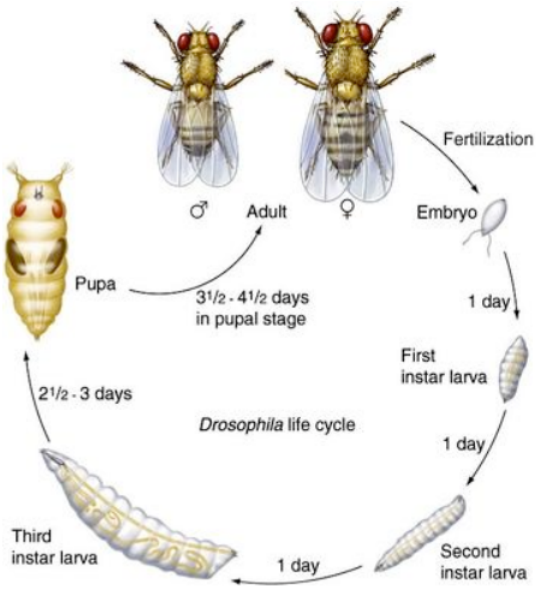

Figure 1.2: *Drosophila* life cycle. The time from egg to adult is temperature dependent. The above cycle is for a range of 21-23°C . The higher temperature, the faster generation time, whereas a low (to 18°C) temperature causes a longer generation time. Image obtained from [24]

protein interactions, and they are the basic information transmission medium used by nature in developmental biology [22].

In Morphogenesis, there is a type of protein that is called morphogen. This has a fundamental role to control spatially how different cells of the system can differentiate into others [102]. Cell differentiation is the ability of a cell to perform a specific function in the development of the living being. In the early stages of development, embrionic cells are stem cells that are undifferentiated and require some kind of external signal to guide them. This signal, normally provided externally during the fertilization of the organism, is the morphogen. First, the morphogen invades a region of cells distributed in the system (hereinafter, *extracellular matrix*, ECM). The morphogen invasion results in a protein concentration profile called a *gradient* (do not confuse with the mathematical concept). The spatial extension and concentration of the gradient along the extracellular matrix play a fundamental role in the correct differentiation of cells. And it is that, depending on the amount of morphogen, and its exposure time, the cells acquire different genetic fates. This will lead to the segmentation and compartmentalization of cells of different types, which will be important for the successive stages of the development of the organism.

As the morphogen concentration gradient forms, cells can internalize the protein within their nucleus. Internalization usually takes place through what is known as transmembrane receptors [102]. These are proteins distributed through the entire thickness of the plasma membrane of the cell, and they act as an access bridge from the outside to the inside of the cell. The structure of these proteins is very specific: they are only able to contact with those proteins that 'fit' with them. That is why cells are equipped with different transmembrane receptors, each of them with the function of capturing specific proteins and morphogens from the environment. It is interesting to note that, until relatively recently, the cell membrane was viewed as a passive structure. That is, the morphogen was the one that had to reach the cell to be able to produce the 'coupling' with the transmembrane receptor, to later be internalized by the cellular machinery. However, recent advances in confocal microscopy have made possible to obtain much more precise images of how this process occurs. And it is that, contrary to how it was initially believed, cells are capable of modifying their cell membrane to capture morphogen [50]. These extensions of the membrane are called cytonemes, and they have structures that are very reminiscent of neurons. They are microtubules with an actin backbone, which direct the membrane towards the morphogen to promote protein-protein contact and promote signaling.

These elements are present in the Morphogenesis of every living being. In Drosophila [109], for example, we have the Dorsal morphogen (Dl) provided by the mother during the fertilization process. As its name indicates, it is the morphogen responsible for the differentiation of cells in the Dorso-Ventral axis during the embryonic phase and is absorbed by Toll receptors. Other morphogens play their role in other stages of *Drosophila*'s development. For example, in the larval stage, there are circular structures that are called Imaginary Disks. These represent the primordia of cuticular structures of the adult that will form in the pupa during metamorphosis. They include wings, legs, antennae, eyes, head, thorax, and genitalia. The cells allocated in the imaginal discs are called imaginal cells, and although they are not stem cells, they are not differentiated either. That is why the imaginal discs are equipped with morphogens that allow the differentiation of cells for their function in the structure in question. In this thesis, we will work with Hedgehog (Hh), one of the morphogens present in the imaginal discs of wings, eyes, and thorax, together with its receptor Patched (Ptc). In addition to its essential role in developmental biology in *Drosophila*, the study of Hh is important since it has a direct counterpart for vertebrate animals: Sonic Hedgehog[2] (Shh). It is present in the formation of structures such as the hypothalamus in mice or the neural tube in chicken and humans. How-

[2]Indeed, in honor of the SEGA video game character ®.

ever, the role of Shh is not only limited to embryonic development. Shh is related to tumor formation in a significant number of human cancers. These include oral squamous cell carcinoma (OSCC) [103], present in more than 90% of head and throat tumors, and non-small cell lung carcinoma (NSCLC) [48], present in lung cancer. And it is that, in cancer, aberrant expression of Shh has been observed correlated with the extension and spread of the tumor. For this reason, it is believed that Shh must have some function related to tumor growth, and it lays the groundwork for possible alternative anti-cancer therapies to conventional ones [4].

1.3 Genetic transcription

Once the receptors capture the morphogen, it forms a morphogen-receptor complex that is internalized in the nucleus of the cell. Inside the nucleus is the DNA, which encodes all the possible functions that the cell can perform. In this way, the morphogen-receptor complex acts as an activating signal for different genes (called target genes of the signal), which will give rise to the expression of other proteins necessary for the successive stages of the process. The decoding of the information of these target genes is a process that receives the name of gene transcription. To fully understand how this process works, we must first introduce how a gene is structured within the DNA chain. For this, it is necessary to remember that DNA is a chain of 4 nucleotides characterized by its nitrogenous bases: adenine (A), thymine (T), cytosine (C), and guanine (G). These are paired following very specific complementarity rules, where adenine binds with thymine (A-T), and guanine with cytosine (G-C), forming what is known as base pairs. In this way, a double chain of nucleotides (a helix of base pairs) is built, which gives shape to DNA (see Fig. 1.3). A gene is a segment of DNA. It is characterized by a sequence of base pairs that encodes the synthesis of a product (typically a protein). Here we have to make an important remark. Regarding the gene concept, we are not referring to a specific sequence of pairs. The same gene can be defined by different base pairs sequences, called alleles[3]. This is important when we define the structure of DNA. Genes are distributed throughout DNA in separate blocks called chromosomes (see Fig. 1.3). Most living beings have an even number of chromosomes, which is known as a Ploidy number. For example, the *Drosophila* genome is made up of 4 pairs of chromosomes (i.e., Ploidy number of 2), which hold a total of approximately 13,600 genes [109], and humans 23 pairs of chromosomes with approximately 27,000 genes [27]. The reason for this parity is that every chromosome has a copy (homologous chromosome) that contains versions of

[3]Please recall the classic example of Mendel and the peas. There the gene for the color of pea seeds had two alleles (green and yellow).

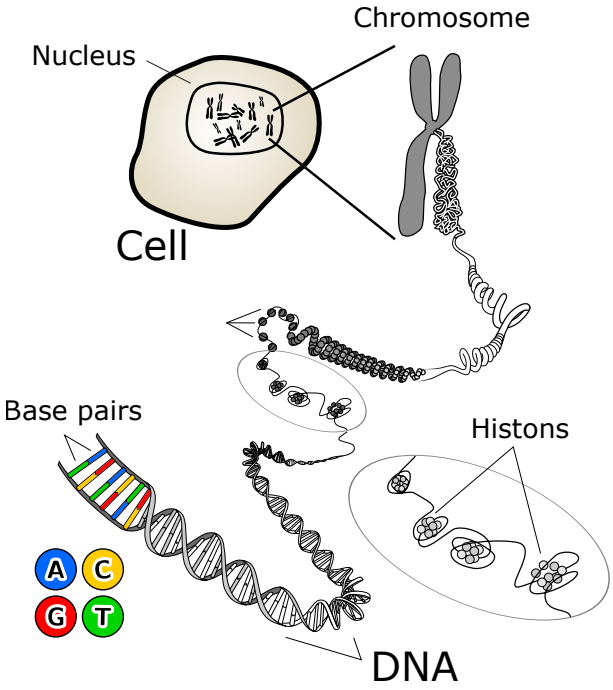

Figure 1.3: DNA structural scheme in the cell nucleus. Image adapted from [117]

the same genes, so that DNA is made up of two copies (alleles) of the same gene. These alleles come from sexual reproduction, where the female delivers one of her pairs of chromosomes (and therefore, one of her alleles for each gene) and the male the other.

Gene transcription is the process by which the base pairs that makeup genes are copied, and typically ends up producing proteins. DNA is often compact within the chromosome, coiled up in structures called histones (see Fig. 1.3). That is why most genes are usually inaccessible and hidden, and they need the DNA to be opened and copied. This process is controlled by a protein called RNA polymerase. When DNA is accessible, PolII (a kind of polymerase) binds to a specific region of the gene known as the promoter, and from there begins to unwind the DNA. The unrolled DNA is copied base pair by base pair, to form a copy of the gene called messenger RNA (mRNA). This will travel outside the nucleus, entering the cell's cytoplasm, while the RNA polymerase disengages from the DNA. The cytoplasm is the place where the messenger RNA is synthesized into proteins, via what is called the ribosomes, in other post-transcriptional events. The binding of the RNA polymerase protein is, therefore, essential to have transcription. The amount of messenger RNA sent to the cytoplasm depends on the num-

ber of times the RNAp has copied the DNA. That is, the concentration of protein synthesized at the end of the process depends strongly on the number of anchors, per unit of time, that the PolII makes in the gene's promoter. As we have discussed before, certain protein concentration levels play important roles in developmental biology. It is for this reason that there is another type of protein responsible for regulating the anchoring of RNA polymerase. These proteins are called Transcription Factors (TFs) [120], and they are also capable of anchoring to specific DNA sites called enhancers [100]. Depending on their nature, transcription factors are usually characterized by two families: those that promote PolII binding (activators), and those that prevent it (repressors). The specific mechanisms that govern these functions are not fully understood. It is known that the transcription levels of the protein in question are altered depending on the concentration of activators or repressors. However, it has not been possible to observe directly how these proteins interact with RNA polymerase, at the molecular level. There is a large family of promoters and enhancers that are separated by a large number of base pairs. This implies that TF-RNAp interactions can occur over long distances, making the question much more intriguing.

Drosophila has different transcription factors depending on the target gene of the morphogen. For example, in the embryonic stage, the Dorsal morphogen acts simultaneously as an activator of more than 50 genes involved in the formation of the dorso-ventral axis. On the other hand, the stage of the larva *Drosophila* is controlled by the transcription factor Cubitus Interruptus (Ci), which appears both in an activating and repressive form in gene expression such as *patched* (*ptc*) and *decapentaplegic* (*dpp*). As with Hedgehog, these transcription factors have their counterparts on vertebrates and are called Gli proteins. These are also involved in tumor development, along with the regulatory proteins P53 and Nanog [113].

1.4 A multi-scale problem

In previous sections, we have discussed in a basic way the elements that make up the problem that we deal with in this thesis. Before going into more details, it is convenient to stop to reflect on the scales where these elements unfold. We have introduced protein as the fundamental element of information transmission and interaction in developmental biology. An important part of characterizing any protein molecule is to determine its size and shape [29]. The interior of protein subunits (aminoacids) and domains consists of closely packed atoms. As a consequence, all proteins have approximately the same density, about 1.37g/cm^3. This gives a partial specific volume of $\nu_2 = 0.73\text{cm}^3/\text{g}$, which can be used to calculate the volume V (in nm^3) occupied by a protein of mass M (in Dalton) as:

$$V = \frac{\nu_2 \times 10^{21}\mathrm{nm}^3/\mathrm{cm}^3}{6.023 \times 10^{23}\mathrm{Da/g}} \times M = 1.212 \times 10^{-3}\mathrm{nm}^3/\mathrm{Da} \times M\,. \tag{1.1}$$

However, what we really want is a physically intuitive parameter for the size of the protein. If we assume the protein is packed in a sphere, we can calculate its radius R_{min}:

$$R_{min} = \sqrt[3]{\frac{3V}{4\pi}} = 0.066\sqrt[3]{M}\,, \tag{1.2}$$

for the protein mass M in Dalton and the radius R_{min} in nm. For example, in *Drosophila*'s embryo, we have introduced the morphogen Dorsal and its receptor Toll. Dorsal has a mass of 111.551 kDa [110], which is a radius of about 3.17 nm, while Toll has a mass of 124.656 kDa [110] (around 3.30 nm). The same calculation can be applied to *Drosophila*'s wing imaginal disk. Hedgehog morphogen has a mass of 52.150 kDa [110], and Patched receptor of 142.831 kDa [110]. This gives rise to similar sizes, with 2.46 nm for Hh and 3.44 nm for Ptc. It is interesting to see that these scales hold even when we scale up the system (with a similar radius of 2.40 nm for Shh, of 47.773 kDa [110], in the case of vertebrates). These numbers mean nothing if we don't put them in context. To understand a little better what implications these dimensions have on the scale of the problem, we can do a simple modeling exercise in the case of the Dorsal morphogen in the *Drosophila* embryo. This morphogen is known to spread across the outer surface of the embryo, occupying approximately 40% of the entire dorso-ventral axis. The surface of the embryo can be approximated by an ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{1.3}$$

where, taking the Fig. 1.1 as a reference, we can consider x as the anterior-posterior axis, y the right-left axis, and z the dorso-ventral axis. As in normal conditions, the anterior-posterior axis is the longest, in this way we will naturally have $a > b$ and $a > c$. In biology, the coordinates are also usually considered so that the point $(0, 0, 0)$ coincides with the most ventral and centered position. Thus, taking

$$\begin{cases} \tilde{x} = x \\ \tilde{y} = y \\ \tilde{z} = z + c \end{cases} \tag{1.4}$$

we get the equation of the embryo in the biological coordinates

$$\frac{\tilde{x}^2}{a^2} + \frac{\tilde{y}^2}{b^2} + \frac{(\tilde{z} - c)^2}{c^2} = 1\,. \tag{1.5}$$

Knowing that the Dorsal gradient covers 40% of the entire dorso-ventral axis (that is, $\tilde{z} \in [0, 0.8c]$) we could calculate the region invaded by the morphogen by integrating (1.5). To do the calculation rigorously, we could make a change of variables to ellipsoidal coordinates, and then integrate on the surface of the ellipsoid in the range we want. However, in this section, we are more interested in doing a quick calculation, which gives us an approximation of the magnitudes we are working with. The experimental technique for these cases is based on reconstructing three-dimensional surfaces, from 2D projections. Following the experimental line, for each dorsal-ventral plane $\tilde{z} \in (0, 2c)$ we can define an ellipse

$$\vec{e}(\theta; \tilde{z}) = (A(\tilde{z}) \cos(\theta), B(\tilde{z}) \sin(\theta)) \tag{1.6}$$

with

$$\begin{cases} A(\tilde{z}) = a\sqrt{1 - \frac{(\tilde{z}-c)^2}{c^2}}\,, \\ B(\tilde{z}) = b\sqrt{1 - \frac{(\tilde{z}-c)^2}{c^2}}\,, \end{cases} \tag{1.7}$$

and θ the parameterization angle. Note that we have to impose the value on the poles to fulfill (1.5), this is

$$\vec{e}(\theta; 0) = \vec{e}(\theta; 2c) = (0, 0)\,, \tag{1.8}$$

so we can reconstruct the embryo's surface by cutting around $\tilde{z} \in [0, 2c]$. The length of the ellipses for each plane is given by the integral

$$L_e(\tilde{z}) = \int_0^{2\pi} |\vec{e}\,'(\theta; \tilde{z})| d\theta = 4 \int_{\pi/2}^{\pi} \sqrt{A^2(\tilde{z}) \sin^2(\theta) + B^2(\tilde{z}) \cos^2(\theta)}\, d\theta\,. \tag{1.9}$$

Please note that the the eccentricity of all the ellipses that make up the embryo is the same, that is,

$$e = \sqrt{1 - \frac{B^2(\tilde{z})}{A^2(\tilde{z})}} = \sqrt{1 - \frac{b^2}{a^2}}\,. \tag{1.10}$$

Making use of this property and introducing the change of variable $t = \sin(\theta + \pi/2)$, we can rewrite (1.9) in the form of an elliptic integral

$$L_e(\tilde{z}) = 4A(\tilde{z}) \int_0^1 \frac{\sqrt{1 - e^2 t^2}}{\sqrt{1 - t^2}}\, dt \tag{1.11}$$

which solution can be expressed in terms of the hypergeometric Gaussian function

$$L_e(\tilde{z}) = 2\pi A(\tilde{z})\ {}_2F_1\left(\frac{1}{2}, -\frac{1}{2}; 1; e^2\right)\,. \tag{1.12}$$

Integrating once more along $\tilde{z} \in [0, 0.8c]$ we obtain the surface that invades the Dorsal gradient

$$Dl_S = \int_0^{0.8c} L_e(\tilde{z}) d\tilde{z} \tag{1.13}$$
$$= \frac{\pi ac}{2} \, {}_2F_1\left(\frac{1}{2}, -\frac{1}{2}; 1; e^2\right) \left(2\arcsin(-0.2) + \sin\left(2\arcsin(-0.2)\right) + \pi\right) .$$

A *Drosophila* embryo has an anterior-posterior axis of approximately $a = 500\mu$m, and is relatively symmetric around it with $b = c = 90\mu$m. This would give us an area of approximately $Dl_S = 10^5 \mu\text{m}^2$. If we were to study the embryo in the whole 40% of its surface, this magnitude fixes the spatial scale of the problem. We previously have calculated the approximate radius of a sphere packing Dorsal ($R_{min} = 3.17\,\text{nm}$), which gives a projected area of $31.57\,\text{nm}^2$. Supposing that Dorsal occupies homogeneously the embryo surface, we see that the surface of the 40% of the embryo represents an average of 3×10^9 packed Dorsal proteins[4].

This quick calculation allows us to establish the first scale of the problem. If we want to study the morphogen gradient in its entirety, then it makes sense to treat the gradient as a whole instead of studying its components separately. On this new scale, proteins clump together in small regions of space, but collectively their high numbers invade a considerable region of tissue. This is what we will call throughout the thesis as the *tissular* scale of Morphogenesis. In it, we will treat morphogen not as a given number of proteins, but as a concentration of these. This is what is generally known as a macroscopic limit of the problem. In biology, this type of approach is not unusual, where concentration (molar) units are extensively used. If we are interested in studying the evolution of the morphogenetic signal throughout the tissue (in this case, the embryo), we can model the Dorsal morphogen concentration as a continuous substance. This is dispersed over the surface of the embryo so that at each point of this surface we can measure the concentration that we have in Dorsal. In our model, this would mean that $[Dl] \in \mathbb{R}_0^+$ is a function of the position $[Dl] = [Dl](\vec{r})$, for all $\vec{r} = (\tilde{x}, \tilde{y}, \tilde{z})$ defined by (1.5). This macroscopic treatment of the problem has its mathematical advantages: instead of having an equation for each protein in the system, we reduce all the information into a single equation that measures the evolution of the concentration. There are quite examples of mathematical models applied to morphogenesis using parabolic equations such as Fick (Heat) equation [68], and Flux-Limited equations [111]. However, we should keep in mind that the elements that interact with morphogen are the cells

[4] This is, of course, an idealization. Tissues are formed by other elements, such cells and other proteins. However, it is interesting to note that even under this naive approximation, this number is of the order of magnitude that have been tested experimentally [57]

that populate the ECM. Taking into account that the size of a *Drosophila* cell is around $10\mu m^2$ [35], we would be talking about an extracellular matrix, for the embryo, of approximately 1000 cells. If we study the problem at the scale of a cell, we have lowered the mean number of proteins in Dorsal by three orders of magnitude. This does not imply that the measurement based on concentrations stops working, but experimentally important differences concerning the tissue scale are beginning to be observed. These differences reside, fundamentally, in the fact that at the cellular level one can appreciate what is known as protein vesicles. A vesicle is a structure consisting of liquid or cytoplasm, enclosed by a lipid bilayer. Vesicles are produced naturally during the processes of secretion (exocytosis), uptake (endocytosis) and transport of materials within the cellular plasma membrane. These complexes usually are equipped by different concentrations of proteins, which are transferred from cell to cell. Cytonemes, previously introduced as extensions of the cell membrane, is the medium that cells use to exchange proteins. In this way, the vesicles can travel inside the cytonemes and be captured by the receptors that extend along the membrane. For this reason, although we are not yet treating morphogen as individual proteins, the macroscopic approach cannot be used so lightly, since the concentration is no longer continuous. On this scale, morphogen is treated as 'protein packages', which we will denote as the *cellular* scale (we will go deeper in Chapter 2).

Once the vesicles have been internalized by the cell, we enter the last scale of the problem: the *molecular* scale. Here the processes become discrete, as we begin to deal with interactions that take place in specific regions of DNA where proteins anchor (promoters and enhancers). These regions are certainly small. Recall that DNA is made up of nucleotide base pairs, each of them on Armstrong order. This means that a molecular scale is considerably below the tissular scale. However, as we will see in Chapter 3, we can relate these two using appropriate modeling hypotheses.

Part I

Tissular scale

Chapter 2

Cytonemes

In previous chapters, we have highlighted the importance of morphogens during the development of living organisms. These are distributed throughout the developing tissue, forming gradients whose concentration levels (among other elements) decide the correct differentiation of cells. As we introduced in Chapter 1, when we try to scale the problem down into the cellular level, macroscopic descriptions begin to lose some validity. This is, instead of having a continuous gradient in a region of space, at these scales, independent agglomerations of morphogen (vesicles) can be appreciated. These vesicles are packets of protein concentration, and are trapped by cells. Once the morphogen vesicle is internalized, it regulates concentrations of other proteins (transcription factors) and continues with other developmental steps. In Chapter 1 we already introduced the cytonemes. These are extensions of the cell membrane that regulate the exchange process of these vesicles. They are dynamic structures in the shape of nanotubules, equipped with an actin skeleton capable of elongating, retracting, and adequately orienting the membrane to ensure the capture of proteins. A direct correlation between the dynamics of these structures and that of the morphogen gradient has recently been experimentally verified [12, 121, 15, 16] (see Fig. 2.1). This has caused the scientific community to consider cytonemes as tools that cells use to interchange information between them. However, how cytonemes orient themselves in the extracellular matrix is a problem about which little is known. In this thesis we have tried to give answers to this question, making theoretical-experimental approaches that shed some light on this very interesting problem.

2.1 Cytoneme mediated cellular communication

To carry out the study of cytonemes, we focus our analysis on Hedgehog signaling, in the imaginal disc of *Drosophila* wing. As we already introduced in Chapter 1, imaginal discs are structures that contain the primitive geometry,

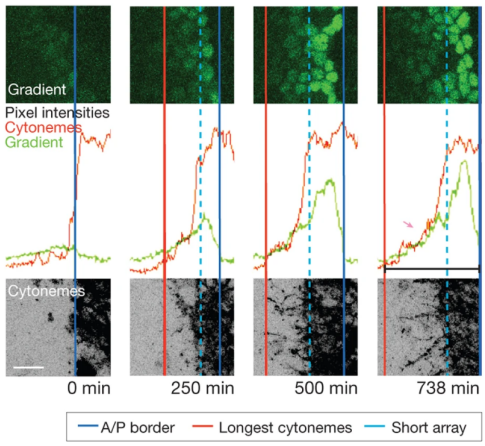

Figure 2.1: Four frames of time showing cytonemes dynamics. Top panels: Hh activity gradient shown by expression of the Ptc. Bottom panels: cytonemes. Middle panels: profile plots of vertically averaged pixel intensities for the panels above (green curve) and below (red curve). The blue line marks the A/P border. Image obtained from [12].

and protein information, necessary for the establishment of different parts of *Drosophila*. In particular, the imaginal disc during the formation of the fly wing is composed of two cellular regions: the anterior and posterior compartments. Cells located in the posterior region express the Hedgehog (Hh) morphogen, which invades the anterior region of the disc from what is called the A/P border (developed around the 60% of the imaginal disc). There, cells equipped with Patched (Ptc) receptors capture the Hh vesicles. At the same time, the gen of Ptc, named *ptc*, is among the target genes of Hh. This implies that, while Hh is captured by Ptc proteins, its concentration in the membrane increases due to its genetic expression, stabilizing the morphogen gradient between the A/P border and 20% of the anterior region. If we observe this process at a cellular scale, we can see cytonemes that extend the membranes from both Hh-producing and receiving cells [12, 16]. These cytonemes grow and invade both Anterior and Posterior regions, sending and collecting morphogen vesicles.

The question then resides in how cells establish this cytoneme-mediated communication. In this sense, there are previous studies that have identified proteins involved in the cytonemes dynamics [50]. On one hand, we have the transmembrane protein Ihog. This acts as a Hedgehog co-receptor, working in coordination with Patched, and it is also known to increase the half-life of cytonemes in overexpression situations [49]. On the other hand, we have the Heparan sulfate proteoglycans proteins (HSPGs). Among these, Dally and Dally-Like (Dlp) stand out, which seems to be involved in the

stabilization of cytonemes [50]. This motivates us to use these proteins as possible fundamental ingredients to model the cytonemes orientation. We will use these proteins as a starting point for modeling. Let us define the quantity

$$\begin{aligned}\rho(\vec{r},t) &= \beta_{Ihog}\Big([Ihog](\vec{r}) + [Ihog]_{cyt}(\vec{r},t)\delta_{cyt}(\vec{r},t)\Big) \\ &+ \beta_{Dlp}\Big([Dlp](\vec{r}) + [Dlp]_{cyt}(\vec{r},t)\delta_{cyt}(\vec{r},t)\Big) \\ &+ \beta_{Dally}\Big([Dally](\vec{r}) + [Dally]_{cyt}(\vec{r},t)\delta_{cyt}(\vec{r},t)\Big),\end{aligned} \tag{2.1}$$

where

- $[X](\vec{r})$ denotes concentration of protein X at disc position $\vec{r} \in \mathbb{R}^2$,
- $[X]_{cyt}(\vec{r},t)\delta_{cyt}(\vec{r},t)$ is the concentration of protein X in a cytoneme positioned at disc position $\vec{r} \in \mathbb{R}^2$, at time t, and
- β_X is a proportionality constant for the protein X,

for $X = \{Ihog, Dlp, Dally\}$. In fact, in collaboration with Isabel Guerrero's laboratory (Severo Ochoa Molecular Biology Center, Madrid), we have been able to verify that the distribution of these proteins in the imaginal disc is correlated with the activation of Hh target genes. In particular, Ihog concentration shows a significant drop in those regions where Ptc is activated by Hh uptake, while Dally and Dlp concentrations grow (see Fig. 2.2).

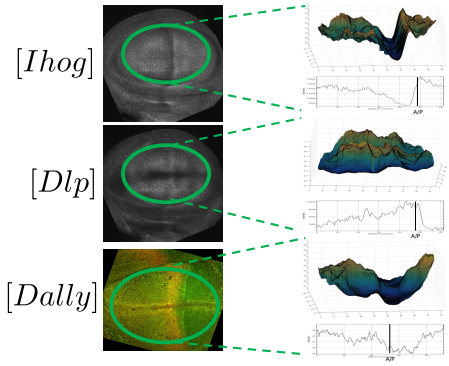

Figure 2.2: Wild-Type (WT) concentrations of Ihog ($[Ihog]$), Dally-Like ($[Dlp]$) and Dally ($[Dally]$). Left: experimental profile measured in different imaginal discs. Right: Disc-averaged mean concentrations in 2D and 1D, projected in the A/P axis. Image obtained by Isabel Guerrero's laboratory.

2.1.1 Cytoneme orientation field and potential

The next in our cytoneme model development is to check to what extent $\rho(\vec{r}, t)$ is able to control the cytonemes dynamics. For this, altered versions of the extracellular matrix are usually generated, what are called clones. Clones are cells in the tissue whose functionality have been genetically modified. In our case, Isabel Guerrero's laboratory generated different clones with malfunctions in the production of Ihog, Dlp and Dally (Dlp-Dally from now on), expressing higher levels of these proteins. These higher levels, 'overexpression', are equivalent to twice the protein expression with respect the Wild-Type concentration of the other cells in the tissue (see scheme in Fig. 2.3). During the collaboration, Isabel Guerrero's laboratory has been able to show that cytonemes behave differently depending on the concentration levels of the proteins. More specifically, their experiments seem to show that

- Cytonemes from cellular regions of overexpression of Ihog and WT-low levels of Dlp-Dally, orient themselves towards regions of overexpression of Dlp-Dally and WT-low levels of Ihog.
- Cytonemes from cellular WT-low Dlp-Dally overexpressing Ihog regions deviate from other WT-low Dlp-Dally overexpressing Ihog regions.
- Cytonemes from cellular Dlp-Dally overexpression clones and WT-low levels of Ihog deviate from other Dlp-Dally overexpression regions and WT-low levels of Ihog.

See Table 2.1 and Fig. 2.3 for an scheme of these behaviors. The rest of combinations either lacked of significance, or didn't represent any important effect in the cytonemes orientation. On the other hand, these effects were observed only if the interacting cytonemes where separated a maximum distance (about $5\,\mu$m). This seems to indicate that there is a short-range and dual interaction between the proteins. To model this interaction, we must first specify a relationship between the constants β_X in (2.1). On the one hand, as shown by the WT distribution in the disc, Ihog seems to play an opposite role to that of Dally and Dlp in the activation of the target genes. Furthermore, it appears from previous clone experiments that opposite protein levels give rise to attraction or repulsion effects between cytonemes. This motivates us to impose on our model a condition on the sign of its corresponding constants:

$$\begin{cases} \text{sign}\{\beta_{Ihog}\} = -\text{sign}\{\beta_{Dlp}\}\,, \\ \text{sign}\{\beta_{Dlp}\} = \text{sign}\{\beta_{Dally}\}\,. \end{cases} \tag{2.2}$$

In our model we have fixed $\beta_{Ihog} > 0$ and $\beta_{Dlp}\,, \beta_{Dally} < 0$. On the other hand, experiments seem to indicate that interactions occur over short distances.

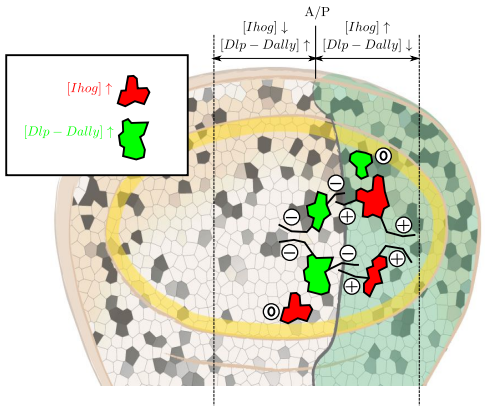

Figure 2.3: Scheme of an experiment with clones in a *Drosophila* imaginal disc. Cells marked in red are clones of Ihog overexpression ($[Ihog] \uparrow$). Cells marked in green are clones of Dlp-Dally overexpression ($[Dlp - Dally] \uparrow$). The Wild-Type concentrations in anterior region close to the A/P border are high levels of Ihog ($[Ihog] \uparrow$) and low levels of Dlp-Dally ($[Dlp - Dally] \downarrow$). The Wild-Type concentrations in posterior region close to the A/P border are low levels of Ihog ($[Ihog] \downarrow$) and high levels of Dlp-Dally ($[Dlp - Dally] \uparrow$). The behavior of citonemes emanating from the clones depends on the balance between the WT and overexpressed concentrations. I.e.: Ihog clones in the anterior region will have a balance between the overexpressed levels of Ihog and the WT high levels of Dlp-Dally (charge ⓪); Dlp-Dally clones in the posterior region will have a balance between the overexpressed levels of Dlp-Dally and the WT high levels of Ihog (charge ⓪); Ihog clones in the posterior region will have a positive behaviour, due to the overexpressed levels of Ihog and the WT low levels of Dlp-Dally (charge ⊕); Dlp-Dally clones in the anterior region will have a negative behaviour, due to the overexpressed levels of Dlp-Dally and the WT low levels of Ihog (charge ⊖). Cytonemes of the same charge repel each other, and cytonemes of different charges attract each other. 'Nothing' stands for no significant effect in the cytonemes orientation.

	$[Ihog]\uparrow$ $[Dlp-Dally]\uparrow$	$[Ihog]\uparrow$ $[Dlp-Dally]\downarrow$	$[Ihog]\downarrow$ $[Dlp-Dally]\uparrow$	$[Ihog]\downarrow$ $[Dlp-Dally]\downarrow$
$[Ihog]\uparrow$ $[Dlp-Dally]\uparrow$	Nothing ⓪ VS ⓪	Nothing ⓪ VS ⊕	Nothing ⓪ VS ⊖	Nothing ⓪ VS ⓪
$[Ihog]\uparrow$ $[Dlp-Dally]\downarrow$	Nothing ⊕ VS ⓪	Repulsion ⊕ VS ⊕	Atraction ⊕ VS ⊖	Nothing ⊕ VS ⓪
$[Ihog]\downarrow$ $[Dlp-Dally]\uparrow$	Nothing ⊖ VS ⓪	Atraction ⊖ VS ⊕	Repulsion ⊖ VS ⊖	Nothing ⊖ VS ⓪
$[Ihog]\downarrow$ $[Dlp-Dally]\downarrow$	Nothing ⓪ VS ⓪	Nothing ⓪ VS ⊕	Nothing ⓪ VS ⊖	Nothing ⓪ VS ⓪

Table 2.1: Observations from the clones experiments done in Guerrero's laboratory. Different levels of Ihog, Dlp and Dally lead to different behaviors in the cytoneme dynamics. **Legend:** $X\uparrow$, $X\downarrow$ follow the notation introduced in Fig. 2.8. Nothing, Atraction and Repulsion are the visible effects between the clones cytonemes. Following our sign convention in (2.2), ⓪, ⊕, ⊖ stand for null-charge, positive-charge and negative-charge respectively. x VS y stands for the interaction between regions with charges x and y, for $x, y = \{$ ⓪, ⊕, ⊖ $\}$, in the interaction range of around $5\,\mu$m.

To model this effect, we can define an interaction nucleus with compact support $W(r)$ fixed at the maximum distance that a cytoneme can interact, where r is the distance between the cytoneme and $\rho(\vec{r},t)$. In this way we define the magnitude $\phi(\vec{r},t)$ as

$$\phi(\vec{r},t) = (W * \rho)(\vec{r},t) = \int_{\vec{s}\in D} W(|\,\vec{r}-\vec{s}\,|)\rho(\vec{s},t)d\vec{s}\,, \tag{2.3}$$

being D the domain of the imaginal disc. That is, $\phi(\vec{r},t)$ represents an interaction potential, which measures all the interactions sensed by a cytoneme located at position $\vec{r}$ of the extracellular matrix, at time t. Since the interaction kernel $W(r)$ has compact support (remember that this distance is about $5\,\mu$m, but for simplicity we will leave the notation free), we are truncating the sensing of ρ to a finite scope restricted precisely at said maximum interaction distance, at each position $\vec{r}$ in D. Following the theory of potential, given $\phi(\vec{r},t)$ we can obtain a vector field of interactions

$$\vec{O}(\vec{r},t) = -\vec{\nabla}\phi(\vec{r},t) = -\int_{\vec{s}\in D} \vec{\nabla}W(|\,\vec{r}-\vec{s}\,|)\rho(\vec{s},t)d\vec{s}\,. \tag{2.4}$$

The field lines of $\vec{O}(\vec{r},t)$ define the orientation that the cytoneme should take. Note that in a certain way $\vec{O}(\vec{r},t)$ can also be seen as the force field felt by the cytoneme being in position $\vec{r}$ and at time t . This interpretation bears a direct similarity to the potentials generated by charges. $\vec{O}$ represents the 'force' that a 'point charge, with unit modulus and positive charge' would feel. This

is a direct consequence of how we have defined the constants β_X in (2.2). In this way, Ihog could be seen as a source of a positive field, while Dlp-Dally would play the role of a source of negative charge. It is important to note that, although proteins are effectively defined by a net charge, the 'charge' we are talking about in this model does not coincide with the electrostatic charge of these proteins. The sign that we are attributing to Ihog and Dlp-Dally only comes from their opposite properties in the orientation effect of cytonemes, and has nothing to do with their electrical properties. On the other hand, $W(r)$ measures the intensity with which the cytoneme is oriented towards the regions of maximum variation of $\rho(\vec{r}, t)$. Experimentally we have no indication of how the cytoneme accelerates or decelerates in the ECM. In fact, it appears that the orientation process is quite linear, with relatively constant speeds. For this reason, we will model the variation of $W(r)$ by the function

$$W'(r) := \begin{cases} \dfrac{K\delta_0^\alpha}{r^\alpha}, & r \in (0, \delta_0), \\ K, & r \in [\delta_0, \delta_1), \\ \dfrac{K}{1 + e^{\frac{2r-(\delta_1+\delta_2)}{(\delta_2-r)(r-\delta_1)}}}, & r \in [\delta_1, \delta_2), \\ 0, & r \in [\delta_2, \infty), \end{cases} \tag{2.5}$$

where K is the orientation intensity, $\alpha > 0$ a parameter that measures the singular interaction at the origin, and $0 < \delta_0 < \delta_1 < \delta_2$ values defining ranges of different cytoneme behaviors:

1. $0 \leq r < \delta_0$: **Collision/Repulsion range**. If the cytoneme is close enough to a source, you will feel a pull towards it (if the charge of the source is opposite to the effective charge of the cytoneme[1]), or in the opposite direction (if the source charge is the same as the effective charge carried by the cytoneme).

2. $\delta_0 \leq r < \delta_1$: **Orientation range**. At these distances, the cytoneme is oriented with the field at a constant speed.

3. $\delta_1 \leq r < \delta_2$: **Attenuation range**. Here the cytoneme feels less and less the contribution of ρ.

4. $\delta_2 \leq r$: **Non-interaction range**. Sources in this range will not be sensed by the cytoneme.

Please note that we are defining the variation of W, i.e., $W'(r)$. As we will see in the next spteps, we don't need to define $W(r)$, but if we were about to do it, we would need additionally an initial condition (for instance, $W(\delta_2) = 0$). The region belonging to the collision/repulsion has been added

[1] Recall that cytonemes also have concentrations of Ihog and Dlp-Dally

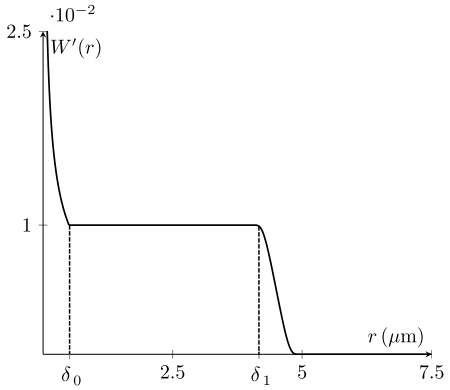

Figure 2.4: Interaction kernel variation $W'(r)$ for $K = 0.01$, $\alpha = \frac{1}{2}$, $\delta_0 = 0.5$, $\delta_1 = 4$ and $\delta_2 = 5$.

in the interaction nucleus in order to reproduce with the model contacts between opposite cytonemes (which attract each other if they carry opposite charges), and repulsion between cytonemes of the same type (equal charges). We want this interaction region to be as intense as possible, and therefore we have added a singular term proportional to $1/r^{\alpha}$. However, if we remember the definition of ρ in (2.1), this singular term would come into play for both concentrations in cytoneme membranes $[X]_{cyt}(\vec{r}, t)$, as concentrations in the membrane of any cell $[X](\vec{r})$. The latter is a continuous concentration, as we saw in Fig. 2.2. This would imply that, for every point $\vec{r}$ of the extracellular matrix, we would always find some contribution of ρ that is inside of the singular range of collision/repulsion. And this is an effect that we do not want in our model since the concentrations $[X](\vec{r})$ are there to orient only the cytonemes along with the extracellular matrix, and not to stick to it. Fortunately biology, in this regard, helps us to understand why this does not happen. Earlier we said that the imaginal disc can be approximated by a plane. This is essentially true, but the plane certainly has a biological limit on its thickness. This limit is precisely tied to the size of a cell, which we introduced in the introductory chapter that occupied an area of $10\,\mu\text{m}^2$. Approaching the cell to a sphere, and assuming that said area is the one projected in the imaginal disc plane, we then have that the minimum thickness of the plane on which the cytonemes move is about $4\,\mu$m. A certain number of cytonemes move through the tangle of cells within this thickness, but the vast majority of cytonemes tend to interact on the surface of the plane, where they have greater freedom to make contact. This surface is what is known as the Basal of the imaginal disc. For this reason, when in our model we speak of position $\vec{r}$ in the extracellular matrix, we are really referring to the position in the Basal plane, and not within the disc volume. Due to this, when we calculate the distance between the cytoneme and the concentrations $[X](\vec{r})$ given by the cells, we are intrinsically calculating the

projected distance in the plane. Therefore, we can choose $\delta_0 < 4\,\mu$m in (2.5) to always be outside the collision/repulsion region, and thus ensure that for this region we only consider concentrations $[X]_{cyt}(\vec{r},t)$ on the Basal plane.

2.1.2 Equation of motion of a cytoneme

Once $\vec{O}(\vec{r},t)$ was built, we would have completed most of our goal. That is, given the experimental concentration of $[Ihog]$, $[Dally]$ and $[Dlp]$ we would be able to predict the path that a cytoneme should take following the lines of the field $\vec{O}(\vec{r},t)$. However, we have yet to define mathematically how a cytoneme can follow these field lines. To do this, we will model a cytoneme as a parameterized curve $\vec{\gamma}(\xi,t)$ equipped by a typical load $\bar{\rho}$. We consider that the cytoneme does not elongate, imposing a restriction on $\vec{\gamma}$ of the kind

$$|\vec{\gamma}'(\xi,t)| = 1\,, \tag{2.6}$$

which means that the curve is parameterized by the arc, with $\vec{\gamma}' \equiv \frac{\partial\vec{\gamma}}{\partial\xi}$. Since cytonemes can grow and retract, the length $L(t)$ of the curve depends on time. Please, note that we are saying that the cytoneme is not able to elongate, but it can grow. Specifically, this means that we are not allowing the cytoneme to modify its actual inner structure, i.e., the actual actin fibers that configures it. But it can grow by creating new blocks of actin via polymerization. We will assume a linear polymerization growth at constant rate v, such that

$$\int_0^{L(t)} |\vec{\gamma}'(\xi,t)|d\xi = L(t) = L_0 + vt\,, \tag{2.7}$$

where we have taken into account the constraint (2.6) and L_0 is the initial length of the cytoneme. The sign of v will give us the state of growth or retraction of the cytoneme, where we will say that

- The cytoneme grows if $v > 0$,
- The cytoneme retracts if $v < 0$,
- The cytoneme does not change its length if $v = 0$.

Please note that v could be time dependent, or even depend on the interaction potential ϕ. However, due to the lack of experimental evidences in this work we have considered a constant grow/retraction velocity (i.e., $v = cte$), with its corresponding sign. On the other hand, the cytoneme will be influenced by the $\vec{O}$ field generated by (2.3). Since it is equipped with an effective charge $\bar{\rho}$, we can attribute to it a potential energy as $\bar{\rho}\phi(\vec{\gamma},t)$. We define the cytoneme Lagrangian as

$$\mathcal{L}(\xi,t,\vec{\gamma},\vec{\gamma}',\dot{\vec{\gamma}}) = \frac{1}{2}|\dot{\vec{\gamma}}|^2 - \bar{\rho}\phi(\vec{\gamma},t)\,, \tag{2.8}$$

with $\dot{\vec{\gamma}} \equiv \frac{\partial \vec{\gamma}}{\partial t}$. Please note that we have assumed unit mass in the kinetic energy term to simplify the notation. Once the Lagrangian is obtained, we will say that the cytoneme will be that curve $\vec{\gamma}(\xi, t)$ such that it minimizes the total action $\mathcal{S}$ defined by $\mathcal{L}$ and constrained to (2.6), that is

$$\begin{cases} \min \quad \mathcal{S}[\vec{\gamma}] = \iint_{\Omega_T} \mathcal{L}(\xi, t, \vec{\gamma}, \vec{\gamma}', \dot{\vec{\gamma}})\, dt\, d\xi, \\ \\ \text{s.t.} \quad C(\vec{\gamma}') = |\vec{\gamma}'| = 1\,, \end{cases} \tag{2.9}$$

for all (ξ, t) within a domain Ω_T (Fig. 2.5) such that

$$\Omega_T := \{(\xi, t) : \xi \in [0, L(t)]\,,\, t \in [0, T]\}\,. \tag{2.10}$$

To obtain the equation of motion of the cytoneme we have to impose boundary conditions on the boundary of Ω_T. A priori, we can set three conditions that come directly from the biology of the problem:

$$\begin{cases} \vec{\gamma}(\xi, 0) = \vec{\gamma}_0(\xi)\,, \text{ initial shape of the cytoneme, with initial length } L_0\,, \\ \dot{\vec{\gamma}}(\xi, 0) \equiv \frac{\partial \vec{\gamma}}{\partial t}(\xi, 0) = \dot{\vec{\gamma}}_0(\xi)\,, \\ \vec{\gamma}(0, t) = \vec{r}_0\,, \text{ base position of the cytoneme}\,. \end{cases} \tag{2.11}$$

However, we see that we still need a boundary condition that controls the evolution of the moving end of the cytoneme. This condition can be seen in Fig. 2.5, where the boundary delimiting the parameter ξ is variable with time according to $L(t) = L_0 + vt$. To deduce what the ideal condition for this boundary should be, we must first deduce the Euler-Lagrange equations attributed to our action functional. However, we have to keep in mind that we are dealing with a system restricted to (2.6). To solve this problem, we first introduce a new Lagrangian taking constrains into account:

$$\mathcal{L}^*(\xi, t, \vec{\gamma}, \vec{\gamma}', \dot{\vec{\gamma}}, \lambda) = \mathcal{L}(\xi, t, \vec{\gamma}, \vec{\gamma}', \dot{\vec{\gamma}}) - \frac{1}{2}\lambda(|\vec{\gamma}'|^2 - 1)\,. \tag{2.12}$$

By the method of Lagrange multipliers, we know that the curves $\vec{\gamma}$ minimizing (2.9) are also minimum of the functional

$$\mathcal{S}^*[\vec{\gamma}] = \iint_{\Omega_T} \mathcal{L}^*(\xi, t, \vec{\gamma}, \vec{\gamma}', \dot{\vec{\gamma}}, \lambda)\, dt\, d\xi\,. \tag{2.13}$$

To derive the Euler-Lagrange equations, we introduce a perturbation of the curve $\vec{\gamma}$ of the type

$$\vec{\beta}(\xi, t) = \vec{\gamma}(\xi, t) + \varepsilon\vec{\alpha}(\xi, t)\,, \tag{2.14}$$

where $\varepsilon \ll 1$ and $\vec{\alpha}(\xi, t)$ a generic perturbation. To make $\vec{\beta}$ compatible with the boundary conditions (2.11), we will force the perturbation $\vec{\alpha}$ to cancel in

$$\begin{cases} \vec{\alpha}(\xi, 0) = \vec{0}, \\ \vec{\alpha}(\xi, T) = \vec{0}, \\ \vec{\alpha}(0, t) = \vec{0}, \end{cases} \tag{2.15}$$

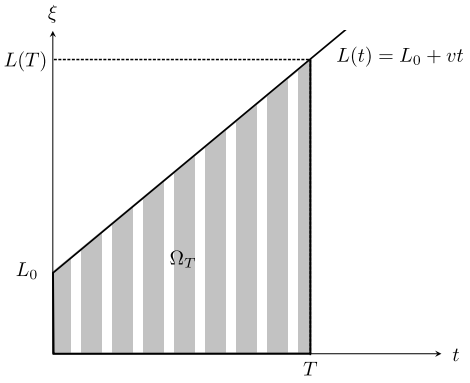

Figure 2.5: Integration domain for the functional (2.9).

for all $(\xi, t) \in \Omega_T$. On the other hand, since $\vec{\gamma}$ is a minimum of $\mathcal{S}^*$, automatically we get that

$$\frac{d\mathcal{S}^*}{d\varepsilon}[\vec{\beta}]\bigg|_{\varepsilon=0} = 0\,, \tag{2.16}$$

which holds if and only if[2]

$$\iint_{\Omega_T} \left(\frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}} \vec{\alpha} + \frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}'} \vec{\alpha}' + \frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} \dot{\vec{\alpha}} \right) d\xi dt = 0\,. \tag{2.17}$$

Our goal is to leave the integral as a perturbed function $\vec{\alpha}$. To do this, we first apply integration by parts in the addend with $\vec{\alpha}'$

$$\iint_{\Omega_T} \frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}'} \vec{\alpha}' = \int_0^T \frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}'} \vec{\alpha}(L(t), t) dt - \iint_{\Omega_T} \frac{d}{d\xi} \frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}'} \vec{\alpha} d\xi dt\,, \tag{2.18}$$

where we have taken into account the first boundary condition of (2.15). On the other hand, the term with $\dot{\vec{\alpha}}$ requires a bit more work since the integration limit at ξ is time dependent. To obtain the dependency with $\vec{\alpha}$, we rewrite the term as a function of the total derivative with respect to time

$$\iint_{\Omega_T} \frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} \dot{\vec{\alpha}} d\xi dt = \iint_{\Omega_T} \frac{d}{dt} \left(\frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} \vec{\alpha} \right) d\xi dt - \iint_{\Omega_T} \frac{d}{dt} \frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} \vec{\alpha} d\xi dt\,, \tag{2.19}$$

where the first term can be written using the fundamental theorem of calculus

$$\iint_{\Omega_T} \frac{d}{dt} \left(\frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} \vec{\alpha} \right) d\xi dt = \left[\int_0^{L(t)} \frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} \vec{\alpha} d\xi \right]_0^T - \int_0^T \frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} \vec{\alpha}(L(t), t) v dt\,. \tag{2.20}$$

[2]Please note that we are doing the formal deduction of the equations of motion of a system defined by the Lagrangian (2.12). However, it is important to remark that $\mathcal{L}$ is written in terms of ρ, which by (2.1), depends on delta functions on curves (i.e., all the existing cytonemes). Hence, the equations of motion given by this Lagrangian may need of a more rigorous study to justify the preceding steps. This falls outside of the scope of the present tesis.

The first term is null by the boundary conditions (2.15), and in this way the equation (2.17) is of the form

$$\iint_{\Omega_T} \left(\frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}} - \frac{d}{d\xi} \frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}'} - \frac{d}{dt} \frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} \right) \vec{\alpha} d\xi dt \tag{2.21}$$
$$+ \int_0^T \left(\frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}'} - v \frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} \right) \Bigg|_{\xi = L(t)} \vec{\alpha}(L(t), t) dt = 0 \,.$$

Since the perturbation is completely generic, the terms of the equation can be set to zero separately, where

$$\frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}} - \frac{d}{d\xi} \frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}'} - \frac{d}{dt} \frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} = \vec{0} \tag{2.22}$$

are the Euler-Lagrange equations and

$$\left(\frac{\partial \mathcal{L}^*}{\partial \vec{\gamma}'} - v \frac{\partial \mathcal{L}^*}{\partial \dot{\vec{\gamma}}} \right) \Bigg|_{\xi = L(t)} = \vec{0} \tag{2.23}$$

are the transversality conditions, for the boundary problem with free boundary $L(t) = L_0 + vt$. Deriving explicitly using (2.12), we are left with the final system of equations for the evolution of the cytoneme:

$$\begin{cases} \ddot{\vec{\gamma}} = -\bar{\rho} \vec{\nabla} \phi(\vec{\gamma}, t) - \dfrac{\partial}{\partial \xi} (\lambda \vec{\gamma}'), \\ \vec{\gamma}(0, t) = \vec{r}_0, \\ \vec{\gamma}(\xi, t) = \vec{\gamma}_0(\xi) \quad \text{y} \quad \dot{\vec{\gamma}}(\xi, 0) = \dot{\vec{\gamma}}_0(\xi), \\ |\vec{\gamma}'(\xi, t)| = 1, \\ \dot{\vec{\gamma}}(L(t), t) v = \lambda(L(t), t) \vec{\gamma}'(L(t), t) \,. \end{cases} \tag{2.24}$$

As we can see, the last equation represents the boundary condition at the tip of the cytoneme, which links growth with constrains via the Lagrange multiplier λ. On the other hand, the first equation would be equivalent to Newton's second law for the curve $\vec{\gamma}$, where the term $\frac{\partial}{\partial \xi}(\lambda \vec{\gamma}')$ would represent the binding force that maintains the constant tension of the cytoneme. Since the cytoneme is also subject to friction in the extracellular matrix, we can add a friction term of the type $\frac{1}{\tau} \dot{\vec{\gamma}}$, obtaining the final system

$$\begin{cases} \ddot{\vec{\gamma}} = -\bar{\rho} \vec{\nabla} \phi(\vec{\gamma}, t) - \dfrac{\partial}{\partial \xi} (\lambda \vec{\gamma}') - \dfrac{1}{\tau} \dot{\vec{\gamma}}, \\ \vec{\gamma}(0, t) = \vec{r}_0, \\ \vec{\gamma}(\xi, t) = \vec{\gamma}_0(\xi) \quad \text{y} \quad \dot{\vec{\gamma}}(\xi, 0) = \dot{\vec{\gamma}}_0(\xi), \\ |\vec{\gamma}'(\xi, t)| = 1, \\ \dot{\vec{\gamma}}(L(t), t) v = \lambda(L(t), t) \vec{\gamma}'(L(t), t) \,. \end{cases} \tag{2.25}$$

As can be seen, it is difficult to obtain a solution to this system because both the multiplier $\lambda(\xi, t)$ and the curve $\vec{\gamma}(\xi, t)$ are interrelated. However,

the transversality condition links the Lagrange multiplier with the growth speed at the tip of the cytoneme. This gives us the idea that, if we integrate the first equation of the system (2.25) between any position of the curve and the extreme, we could somehow eliminate the dependency on λ. Indeed, after integrating and taking into account the boundary condition we arrive at the integro-differential equation

$$\frac{d}{dt}\int_{\xi}^{L(t)} \dot{\vec{\gamma}}(\xi',t)d\xi' = \lambda(\xi,t)\vec{\gamma}'(\xi,t) - \int_{\xi}^{L(t)} \left(\frac{1}{\tau}\dot{\vec{\gamma}}(\xi',t) - \bar{\rho}\vec{\nabla}\phi(\xi',t)\right) d\xi' \,. \tag{2.26}$$

Lastly, the dependency on λ can be completely removed by multiplying on both sides by the normal vector to the curve at position ξ (ie, $\vec{\gamma}'^{\perp}(\xi,t)$), resulting

$$\begin{aligned} &\left(\frac{d}{dt}\int_{\xi}^{L(t)} \dot{\vec{\gamma}}(\xi',t)d\xi'\right)\vec{\gamma}'^{\perp}(\xi,t) \\ &= -\int_{\xi}^{L(t)} \left(\frac{1}{\tau}\dot{\vec{\gamma}}(\xi',t) + \bar{\rho}\vec{\nabla}\phi(\xi',t)\right)\vec{\gamma}'^{\perp}(\xi,t)d\xi' \,. \end{aligned} \tag{2.27}$$

2.1.3 Generalized coordinates of the cytoneme and discretization

Our next goal will be to discretize the system (2.27) for its subsequent numerical implementation. The first thing we are going to do is to introduce a change of coordinates that allows us to describe $\vec{\gamma}$ from the angle formed by its tangent vector to the curve, that is:

$$\vec{\gamma}'(\xi,t) = (\cos\theta(\xi,t), \sin\theta(\xi,t))\,, \tag{2.28}$$

where $\theta(\xi,t)$ is the tangent indicatrix [93] (i.e., the angle formed by the tangent vector to $\vec{\gamma}$) at position ξ and time t. Note that the binding condition (2.6) is still maintained. This allows us to rewrite the vector system (2.27) as the scalar integro-differential equation

$$\begin{aligned} &\frac{d}{dt}\int_{\xi}^{L(t)}\int_{0}^{\xi'} \cos(\theta(\xi'',t)-\theta(\xi,t))\dot{\theta}(\xi'',t)\,d\xi''\,d\xi' \\ &\qquad -\int_{\xi}^{L(t)}\int_{0}^{\xi'} \sin(\theta(\xi'',t)-\theta(\xi,t))\dot{\theta}(\xi'',t)\dot{\theta}(\xi,t)\,d\xi''\,d\xi' \\ &\qquad = -\frac{1}{\tau}\int_{\xi}^{L(t)}\int_{0}^{\xi'} \cos(\theta(\xi'',t)-\theta(\xi,t))\dot{\theta}(\xi'',t)\,d\xi''\,d\xi' \\ &-\bar{\rho}\int_{\xi}^{L(t)} \left((-\sin\theta(\xi,t))\partial_x\phi(t,\vec{\gamma}(\xi',t)) + (\cos\theta(\xi,t))\partial_y\phi(t,\vec{\gamma}(\xi',t))\right)\,d\xi'. \end{aligned} \tag{2.29}$$

To solve numerically the equation (2.29), we can discretize the curve $\vec{\gamma}$ by a chain of particles (nodes) that occupy positions $\vec{r}_i(t) \approx \vec{\gamma}(\xi_i, t)$ where

$$\begin{cases} \vec{r}_i(t) = \vec{r}_0 + \sum_{j=1}^{i} l_j(t)(\cos\theta_j(t), \sin\theta_j(t))\,, \ \forall i = 1, \dots, N(t) \\ \xi_i(t) = \sum_{j=1}^{i} l_j(t)\,, \ \forall i = 1, \dots, N(t)\,. \end{cases} \tag{2.30}$$

Please note that the number of nodes varies with time to account for the elongation or retraction of the cytoneme. By adding nodes every time the separation between the tip and the penultimate node reaches a certain value l, we achieve that the cytoneme increases or decreases in discrete time intervals

$$t_n = \frac{l}{|v|} n \ , \forall n \in \mathbb{Z}_0^+\,. \tag{2.31}$$

In this way we can describe $N(t)$ and the lengths $l_i(t)$ for each n and $t_{n-1} < t < t_n$ according to Table 2.2. That is, all the nodes are sepa-

Elongation dynamics ($v > 0$)
$N(t) = N_0 + n$
$l_i(t) = \begin{cases} l & \forall i = 1, \dots, N_0 + n - 1 \\ tv - (n-1)l & i = N_0 + n \end{cases}$
Retraction dynamics ($v < 0$)
$N(t) = N_0 - (n-1)$
$l_i(t) = \begin{cases} l & \forall i = 1, \dots, N_0 - n \\ tv - (n-1)l & i = N_0 - (n-1) \end{cases}$

Table 2.2: Law that describes the dynamics of growth and decrease of a cytoneme (discrete version)

rated by a fixed distance l except the tip, which will grow (or decrease) at a constant speed v until it reaches the typical length l. Upon reaching this separation, a new node is created, repeating the process over time. With this discretization we can approximate the integrals in (2.29) by sums, approximating the differentials by the distances between the nodes defined in (2.2). That is, for each ξ_i and $t_{n-1} < t < t_n$ we would make the change

$$\begin{cases} \xi' \approx \xi_j \to d\xi' \approx l_j(t)\,, \forall j = i, \dots, N(t)\,, \\ \xi'' \approx \xi_k \to d\xi'' \approx l_k(t)\,, \forall k = 1, \dots, j\,, \end{cases} \tag{2.32}$$

changing the integrals (2.29) into their discrete version:

$$\mathbf{M}(\vec{\theta}(t), t)\ddot{\vec{\theta}}(t)^T = -\frac{1}{\tau}\mathbf{M}(\vec{\theta}(t), t)\dot{\vec{\theta}}(t)^T + \vec{g}(\vec{\theta}(t), \dot{\vec{\theta}}(t), t)^T\,, \tag{2.33}$$

with

$$\vec{\theta}(t) = (\theta_1(t), \dots, \theta_{N(t)}(t)) \in \mathbb{R}^{N(t)}\,, \tag{2.34}$$

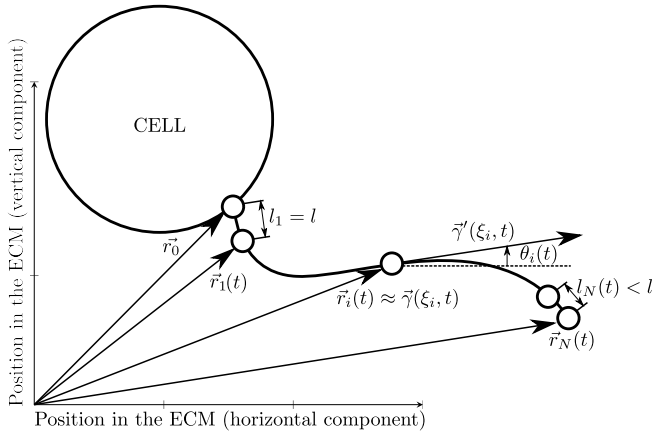

Figure 2.6: Discretization of $\vec{\gamma}(\xi, t)$

$\mathbf{M}(\vec{\theta}(t), t) = \Big(M_{i,j}(\vec{\theta}(t), t)\Big) \in \mathcal{M}_{N(t)\times N(t)}(\mathbb{R})$ tal que

$$M_{i,j}(\vec{\theta}(t), t) = \left(\sum_{k=\max\{i,j\}}^{N(t)} l_k(t) l_j(t)\right) \cos(\theta_i(t) - \theta_j(t))\,, \tag{2.35}$$

and $\vec{g}(\vec{\theta}(t), \dot{\vec{\theta}}(t), t) \in \mathbb{R}^{N(t)}$ such as

$$g_i(\vec{\theta}(t), \dot{\vec{\theta}}(t), t) = -v \sum_{j=1}^{N(t)} l_j(t)\dot{\theta}_j(t) \cos(\theta_j(t) - \theta_i(t)) \tag{2.36}$$

$$+ \sum_{j=i}^{N(t)} l_j(t)\bar{\rho}\Big(\sin\theta_i(t)(\partial_x\phi)(\vec{r}_j(t), t) - \cos\theta_i(t)(\partial_y\phi)(\vec{r}_j(t), t)\Big)$$

$$+ \sum_{j=i}^{N(t)} l_j(t) \sum_{k=1}^{j} l_k(t)\dot{\theta}_k^2(t) \sin(\theta_k(t) - \theta_i(t))\,. \tag{2.37}$$

The system (2.33) can be solved first by inverting the matrix $\mathbf{M}$ with some numerical method (for the simulations in the next section the L-U decomposition was used). Once the matrix is inverted, we would have a system of second order ordinary equations, which can be solved numerically in a simple way using the Runge-Kutta method (in our case, R-K4).

2.2 Drosophila aplication

In this last section we will test the results that we obtain with the model against experimental measurements. The numerical simulations have been

taken first in Wild-Type situations, where Isabel Guerrero's laboratory measured the stationary concentrations of Ihog, Dally and Dally-Like in the entire volume of different imaginal discs. Each measurement was projected in the basal plane (Z), in each imaginal disc. Afterwards, the mean concentration of each protein obtained was taken, generating a mean ellipse with the three concentrations. This ellipse was used as the input in our model, thus obtaining $\rho(\vec{r},t)$ according to (2.1). Since the concentrations are stationary, in this simulation the temporal variation of ρ will be zero at all time instants. However, this measure allows us to calculate the orientation field $\vec{O}(\vec{r},t)$ using (2.4). This would be the field that we would obtain in stationary situations, which we collect in Fig. 2.7. As we can see, we obtain a region where the field is most intense around the A/P border. This coincides with what was expected, and was experimentally measured in previous works such as [50]. Using this field, we have simulated the temporal evolution (2.33) for cytonemes emitted by six cells, three belonging each compartment. In the Fig. 2.7 we have collected several frames of this simulation. To check the extensibility of the model, we have also simulated situations outside the Wild-Type conditions. For this we have used the clones generated in Isabel Guerrero's laboratory as data, where there are situations of:

- Cellular regions with Ihog clones, interacting with cellular regions with clones of Ihog (Fig. 2.8 up).
- Cellular regions with Ihog clones, interacting with cellular regions with clones of Dally-Dlp (Fig. 2.8 down).

To simulate various cytonemes in these conditions, we first removed the cytonemes we want to simulate from the experimental data (marked in blue in the Fig. 2.8). In order not to create an artificial zero of concentration in these regions, we have added noise with a mean value of the background that existed in that region. These new modified data are what we have used to calculate $\rho(\vec{r},t)$, where depending on the simulation we have taken zero concentrations that do not show overexpression. This approach makes sense since overexpressing a protein causes the levels obtained to double the concentrations under conditions in WT, which makes the others negligible. Once ρ has been calculated, we have simulated curves $\vec{\gamma}(\vec{r},t)$ with the base located in the cytonemes deleted from the experimental data (Fig. 2.8 right). The results obtained with the model are capable of reproducing the shape of the erased cytonemes, which shows that the model is capable of performing well even in biologically artificial situations. With this we conclude this section on cytoneme orientation modeling. In the last chapter of the thesis we will return to discuss in a little more detail how this model could be adapted to take into account the transport of morphogens, a work that is still in progress (see Chapter 4).

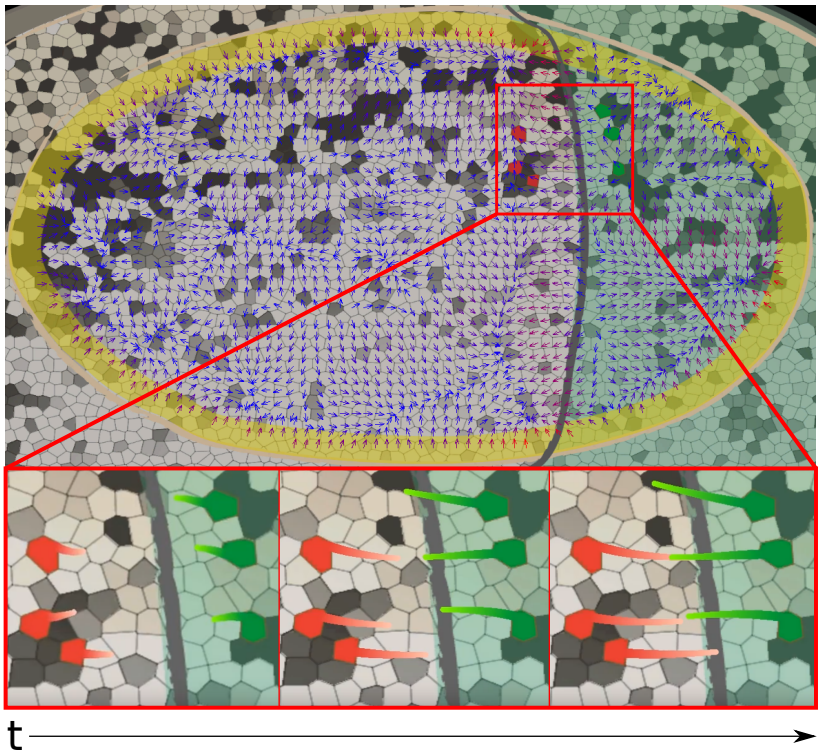

Figure 2.7: Simulation of the orientation field $\vec{O}(\vec{r}, t)$ for an imaginal disc with average Wild-Type concentrations of Ihog, Dally and Dlp. The color of the arrows indicates the intensity of the field, blue being the least intense and red being the most intense. The gray curve that cuts the disc in half marks the A/P boundary. At the bottom, 6 possible cells are shown emitting cytonemes that are oriented according to the calculated field, in different instants of time. Simulation parameters can be seen in Table B.1 in Appendix B.

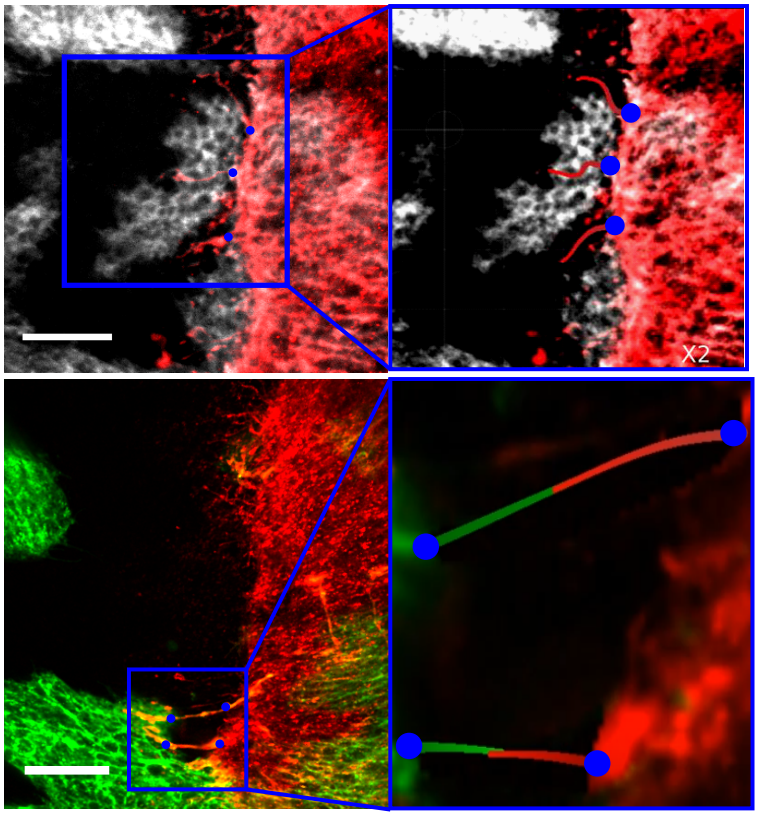

Figure 2.8: Simulation of cytonemes under overexpression conditions. Red and gray colors depicts Ihog overexpression clones, and in green Dlp-Dally overexpression clones. The images on the left correspond to the experimental image, while the images on the right show the reconstruction of the cytonemes marked in blue. Simulation parameters can be seen in Table B.1 in Appendix B.

Part II

Molecular scale

Chapter 3

Transcription dynamics

In previous chapters we have been studying the morphogenesis focusing on the signaling part. The morphogen, after being absorbed by the receptor cells, triggers complex signalling pathways that provoke cell responses [14, 65, 52, 72]. These responses are regulated by the transcription of different target genes of the morphogen signal, and typically ends up with the synthesis of protein which shape and function is encoded in the gen [72]. As we introduced in chapter 1, gene regulation is a complex process that follows several consecutive steps. The RNA polymerase (PolII) binds to DNA promoters, and its activity is directly activated or repressed by the transcription factors (TFs) that binds DNA enhancers [14, 72]. The transcription is hence controlled by the relative concentration of these proteins, the natural tendency of the proteins to bind the DNA (commonly referred as affinity), and the cleavage rules that drive their binding process. The cleavage rules in themselves are compounded, where the binding can involve competition for the same DNA binding sites, or assistance of other already bound elements (cooperativity) [85, 22, 9, 10]. Transcriptional control results in the gene (protein) production rates being controlled by the bound transcription factors which in turn use several biochemical mechanisms [88, 89, 87].

The numerous variables mentioned above makes difficult the understanding of the biochemical mechanisms that are really involved in a specific system. For this reason, experimental discussions in literature are frequently supported by theoretical models in order to decipher how these mechanisms interact [83, 62]. The modelling of transcriptional processes has been tackled from different mathematical perspectives [7, 61, 43, 20, 32], such as Bayesian [37, 36, 75, 28, 78] and Boolean [114], among others. In this work we will focus on models based on the statistical thermodynamic equilibrium approach [1, 97, 11, 17, 21]. The statistical thermodynamic approach, introduced in the pioneering works of Ackers–Shea and coworkers [1, 97], is also known as the BEWARE method (Binding Equilibrium Weighted Average Rate Expression) because transcription, and thus expression, is con-

sidered to be proportional to the probability of transcript initiation [47]. This modelling considers transcript initiation an average of all the possible micro-states where the system (proteins/binding sites) can be present. The micro-states are characterised by the equilibrium binding configurations.

As already mentioned in [96] the thermodynamic models follow one of two different biological control processes: *"recruitment"* or the *"stimulated transcription"* [89, 87, 67]. The use of one or the other will establish the role of the transcription factors in the PolII functionality later in the modelling. In the recruitment approach, all the configurations with a bound PolII have the same transcriptional efficiency, but the TFs control the PolII recruitment by TFs-PolII cooperative/anti-cooperative interactions. That is, the TFs are able to alter the PolII affinity for the promoter both positively (activators) and negatively (repressors) [11, 30, 54]. Alternatively, in the stimulated transcription approach, the PolII binding affinity is assumed to be fixed but the translational strength of any configuration is modulated in terms of the bound activators/inhibitors [62, 63, 69]. Although recruitment is considered to be the main way of controlling gene transcription by Ci/Gli factors [81, 79], Hh/Shh pathways have been modelled in recent publications either using the recruitment [83, 18, 46, 45, 8, 71, 123] or the stimulated [62, 69, 94, 111] approaches. See [98] for a comprehensive introduction to these kinds of modelling and [7, 96, 95, 34] and references therein for a review of case-study scenarios implemented in different areas. As we will see in next chapters, results obtained from different models are really sensitive to the modeling choice. Moreover, it is important to note that the averaging procedure performed in the thermodynamic methodology reveals one of its main drawbacks: the complexity of the deduced mathematical expressions when the model takes into consideration the wide variety of biochemical mechanisms involved in the original process: competition of multiple ligands for multiple binding sites [51], affinities, roles played by each ligand, signal strength and cooperativity [99, 101, 26], among others). The large number of micro-states produce entangled mathematical expressions, so the only way to extract valid information from the models is to go through a multiple-parameter non-trivial calibration process [7, 122]. In this thesis we propose an alternative use of this kind of model based on experimental results to understand the effect of the transcriptional control exerted by an enhancer, and how to use this information to connect with the tissular scale.

3.1 Thermo-statistical Modeling: BEWARE

In this section we will first deduce brand-new expressions of the BEWARE operator. We will work in a general framework, where we will assume that the gen expression is controlled by a number of M cooperative transcription factors. In the literature, both competition and cooperativity have been

determinant factors that tend to convolute greatly the mathematical expressions, even in the most simple case of two transcription factors. Hence, in order to get to the right expression of the BEWARE operator, we first need to declare briefly which are the model assumptions and, from there, how the competition and cooperativity are defined. Let us consider a promoter of a gene p controlled transcriptionally by the TFs $\mathcal{T} = \{T_1, ..., T_M\}$ by binding competitively to an enhancer with n binding sites. Let us mention that the transcription can be controlled by more than one enhancer. In this section we will deduce the model considering the transcription contribution of only one enhancer, but keep in mind that these expression can be generalized to any number of enhancers, with any number of binding sites.

We consider that some of the TFs will increase the production rates of the protein P (activators) while the rest try to repress the same rates (repressors). Thus $\mathcal{T}$ is divided in two different families, M_a activators $\mathcal{T}_A = \{T_1, ..., T_{M_a}\}$ and $M_r = M - M_a$ repressors $\mathcal{T}_R = \{T_{M_a+1}, ..., T_M\}$ such that $\mathcal{T} = \{\mathcal{T}_A, \mathcal{T}_R\}$. The goal of the statistical thermodynamic model is to describe the synthesis rate of P in terms of the TFs concentrations and their activator/repressor rol, that is,

$$\frac{d[P]}{dt} = \text{BEWARE}([T_1], ..., [T_M]) \tag{3.1}$$

where "BEWARE()" is the function specifying the dependence on the TFs, and is subjected to the biochemical mechanisms involved in all these processes. Our aim in this section is to show a methodology for deriving explicit simple analytical expressions for the BEWARE operator by using thermodynamic modelling. Although it is out of the main goals of this chapter, let us remark that the right hand side of model (3.1) is usually accompanied by a degradation contribution, $-\beta[P]$, being β a degradation rate [90]. As mentioned in the introductory section, here we will distinguish between models based on either recruitment or stimulated mechanisms. In order to clarify the difference between these two models, we enumerate in next paragraphs the main assumptions that are used in their development. We also outline in Fig. 3.1 the biochemical mechanisms, which are mainly related with affinity, cooperativity or the manner in which TFs control transcriptional activity.

3.1.1 Thermodynamic description: assumptions

H1) *Separated time scales:* The reactions driving transcriptional control are much more faster than the changes in TFs concentrations and the synthesis of the protein P. Thus, TFs/PolII binding in enhancers/promoter will be considered in thermodynamical equilibrium given by the Law of Mass Action [86]. For instance, in chick embryo neural tube Shh signalling, it has been pointed out that changes in Gli protein concentrations take place at a timescale of days compared to mRNA

variations in timescales of minutes or hours [62, 90].
The binding of a TF T_i to one of the n free binding sites, B, can be interpreted as the chemical reaction

$$T_i + B \underset{k_{-i}^{(1)}}{\overset{k_{+i}^{(1)}}{\rightleftharpoons}} BT_i,$$

where the Law of Mass Action establishes that the complex BT_i has concentration at equilibrium given by

$$[BT_i] = \frac{k_{+i}^{(1)}}{k_{-i}^{(1)}}[T_i][B] := \frac{[T_i]}{K_i^{(1)}}[B]\,.$$

In the notation proposed $K_i^{(1)} = k_{-i}^{(1)}/k_{+i}^{(1)}$ is the dissociation constant of the reaction, with units of concentration so that the quotient $\frac{[T_i]}{K_i^{(1)}}$ is dimensionless. The superscript (1) stands for the dissociation constant of a reaction that takes place in absence of another TF previously bound in the enhancer module. The dissociation constants quantify the affinity of the TFs for their binding sites, being more affine those TFs with lower dissociation constants. The binding of PolII in the promoter follows the same rule, which concentration at equilibrium is

$$[PolII] = \frac{k_{+RP}}{k_{-RP}}[PolII][B] := \frac{[PolII]}{K_{RP}}[B]\,.$$

Please note that K_{RP} doesn't need any superscript since we are working with biological modules that are controlled by n enhancers but only one promoter, hence the maximum number of bound PolII is reduced to one. In Fig. 3.1 these admisible bounds are indicated by black doubled sided arrows.

H2) *TFs binding sites, B, are constituted by n identical sites that can be occupied competitively by any TF.* The basic rule of this competition is that the dissociation constant of the free sites configuration does not depend on their position but might depend on other existing bound TFs in the same module.

Let us suppose the reaction of the binding between a molecule T_i and a free enhancer configuration B occurred. Then, if a second molecule T_j binds to this configuration we have

$$T_j + BT_i \underset{k_{-j}^{(2)}}{\overset{k_{+j}^{(2)}}{\rightleftharpoons}} BT_iT_j$$

with respective equilibrium concentration

$$[BT_iT_j] = \frac{[T_i][T_j]}{K_i^{(1)}K_j^{(2)}}[B]$$

where now the superscript (2) denotes the dissociation constant for a reaction of a TF that binds the operator with already one TF in some other site. Let us remark that previous expression is independent of the enhancers occupied since we are considering them identical. Note that the product could be also obtained by simply changing the order of linkage of the TFs, i.e.,

$$T_i + BT_j \underset{k_{-i}^{(2)}}{\overset{k_{+i}^{(2)}}{\rightleftharpoons}} BT_jT_i$$

although the corresponding concentration could not be the same as we explain in next paragraph.

In biological literature [9, 83, 11, 69] it has been observed that in this competition for the enhancers it could be involved a well known mechanism, called cooperativity, that assumes that one ligand supports the binding of others [53]. Thus, we will call non cooperative to TFs all those proteins whose enhancer's affinities are not modified because of any previously bound TFs, that is, they verify $K_l^{(2)} = K_l^{(1)}$ for all $l = 1, ..., M$. If that is the case, it is plausible to assume the same relation for later bindings, that is, $K_l^{(m)} = K_l^{(1)}$ for $m \geq 2$ and in consequence of this sequential independence we will denote the dissociation constant by K_l skipping the superscript. Then, if all the TFs under consideration are non cooperative we easily deduce that the concentration at equilibrium of a generic configuration with j_i proteins of the T_i specie is

$$[BT_1^{j_1}...T_M^{j_M}] = [B]\prod_{i=1}^{M}\left(\frac{[T_i]}{K_i}\right)^{j_i} \tag{3.2}$$

independently of the sequential order of binding and of the specific positions occupied for the TFs. Let's recall that, since n denotes the total number of free binding sites for TFs, then $\sum_{i=1}^{M} j_i \leq n$ has to be verified. $j_0 = n - \sum_{i=1}^{M} j_i (\geq 0)$, in the subsequent, will denote the number of free spaces in the configuration.

Cooperativity occurs when the existence of other previously bound protein, T_i, affects to the affinity of the new binding protein T_j (cooperativity is represented graphically by dotted arrows in Fig. 3.1). If the

binding process of a protein T_j is facilitated by other already bound T_i protein, TF-TF cooperativity, this can be modelled by considering:

$$K_j^{(2)} = K_j^{(1)}/c \quad \text{being } c > 1\,.$$

If cooperativity occurs in the presence of multiple TFs it would be necessary to know which TFs are affected by others TFs since the equilibrium concentration will depend on these relations. Regarding to this question in the literature several options have been considered. Partial cooperativity [83] would occur when the existence of a specific TF, T_i, modifies equally the affinity of any posterior transcription factor binding of the same family, that is $K_i^{(m)} = K_i^{(1)}/c_i$ for $m \geq 2$. Total cooperativity [69] would occur when the presence of any bound TF, T_i, modify the affinity of any posterior binding in the same manner, i.e. $K_j^{(m)} = K_j^{(1)}/c$ for $m \geq 2$ and $j = 1, ..., M$. These relations have been represented in Fig. 3.1 with red dotted arrows. Then, by direct adaptation of the previous considerations we have that

$$[BT_1^{j_1}...T_M^{j_M}] = [B]c^{\left(\sum_{i=1}^{M} j_i - 1\right)_+} \prod_{i=1}^{M} \left(\frac{[T_i]}{K_i}\right)^{j_i} \tag{3.3}$$

in the presence of total cooperativity while

$$[BT_1^{j_1}...T_M^{j_M}] = [B] \prod_{i=1}^{M} c_i^{(j_i-1)_+} \left(\frac{[T_i]}{K_i}\right)^{j_i} \tag{3.4}$$

if partial cooperativity for TFs occurs. Here $(\cdot)_+$ denotes the positive part function needed because cooperativity is not present unless two or more cooperative TFs are present in the configuration. In the subsequent we will denote by $\{\{T_1, ..., T_M\}_c\}$ and $\{\{T_1\}_{c_1}, ..., \{T_M\}_{c_M}\}$ the total and partial cooperativity respectively. Let us observe that this notation covers the case of non cooperativity since it would correspond to the case $\{\{T_1, ..., T_M\}_1\}$ or equivalently $\{\{T_1\}_1, ..., \{T_M\}_1\}$. Since cooperativity has been described to cause deep changes in transcriptional logic [18], in this work we present our results generalising both partial and total cases. The straightforward extension of total-partial cooperativity concepts is to consider that total cooperativity can occur only between some of the TFs, that is, between the elements of certain subsets of transcription factors. We will refer to this as mixed cooperativity, $\{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_N\}_{c_N}\}$, where $\mathcal{T}_i$ denotes each of the N disjoint subgroups of TFs cooperating with cooperativity constant c_i. By analogue arguments we obtain that, for mixed cooperativity,

$$[BT_1^{j_1}...T_M^{j_M}] = [B]\left(\prod_{i=1}^{N} c_i^{\left(\sum_{h\in\mathcal{I}_i} j_h - 1\right)_+}\right)\left[\prod_{i=1}^{M}\left(\frac{[T_i]}{K_i}\right)^{j_i}\right], \qquad (3.5)$$

where $\mathcal{I}_i = \{h; T_h \in \mathcal{T}_i\}$ is the set of subindexes of the TFs belonging to $\mathcal{T}_i$. As we will see in next paragraphs, this generalization will become very handy since we can deduce the expressions of different cooperative cases from this description.

H3) *The action of a bound TF is independent of the specific enhancer it is occupying, so the transcriptional contribution of configurations with the same number of TFs bound at different specific positions is the same.* Since the TFs compite for free enhancers, and there is not a predetermined binding order [115], multiple spatial configurations of occupied operators are allowed. So, in general, there is not an unique spatial distribution for a given configuration with a distribution of $(j_1, ..., j_M)$ bound transcription factors and j_0 free sites. For instance if we consider $n = 3$, $M = 3$ and $j_1 = j_2 = j_3 = 1$ there are six possible spatial distributions with the same elements ($T_1T_2T_3$, $T_2T_1T_3$, $T_1T_3T_2$, $T_2T_3T_1$, $T_3T_2T_1$, $T_3T_2T_1$). In our description spatial localisation of bound particles is not relevant, so for a concrete configuration $(j_1, ..., j_M)$ and j_0 free sites, we will identify the $\frac{n!}{j_0!\prod_{i=1}^{M} j_i!}$ spatial different plausible configurations.

Assumption H2) and H3) describe the possible configurations of TFs bound to the binding sites. Let us observe that these assumptions not only imply the spatial but also the sequential independence of the equilibrium concentrations. As mentioned in the introductory section, our deduction separates now in two modelling versions: the recruitment and the stimulated approaches [96, 87]. Our next hypotheses describe the PolIIs/promoter binding process in both versions.

HR4) *Recruitment assumption: TFs work by bringing the transcriptional machinery by TFs/PolII (anti-)cooperativity* [87, 88]. In [11], the synergy between a TF and PolII is interpreted in terms of a "glue-like" interaction that would give rise to a modification of the PolII binding affinity modelled analogously to a TFs/PolII cooperativity: each bound activator tries to pull the PolII in the promoter, modifying its affinity constant with a factor $\prod_{i=1}^{M_a} a_i^{j_i}$ where we denote $a_i > 1$ to the i-th *activator transcription intensity* for $i = 1, ..., M_a$. On the other hand, in

a symmetric manner we can model the effect of $M - M_a$ repressors in terms of a "repulsive-like" interaction by modifying the PolII binding affinity with a factor $\prod_{i=M_a+1}^{M} r_i^{j_i}$, where $r_i < 1$ is the i-th *repressor transcription intensity* for $i = M_a - 1, ..., M$ in this case. Then, in general, the PolII binding affinity will take the form

$$\frac{K_{RP}}{\prod_{i=1}^{M} t_i^{j_i}},$$

with $t_i = a_i > 1 \; \forall \, i = 1, ..., M_a$ and $t_i = r_i < 1 \; \forall \, i = M_a + 1, ..., M$. These TFs/PolII cooperative type interactions are indicated by a blue dotted arrow in Fig. 3.1. Please note that this description is in concordance with the affinity definition: If the denominator is larger than one, then PolII will be more affine. Since the denominator depends on the number of transcription factors and their activation/repression intensities, we are promoting or impeding the PolII cleavage, and hence, promoting or impeding the transcription itself.

HS4) *Stimulated assumption: Unaltered affinity of PolII at the promoter,* that is, the binding affinity of PolII to the promoter, K_{RP}, is invariant with respect to the bound TFs.

It is important to remark that only very few of the regulatory motifs shown in [11] match with the generality presented in this work. More concretely the cases of simple repressor and activator coincide with the case of a single binding site, $n = 1$, and glue-like interaction for activators and total repression, $r = 0$ for repressors. This comes from the specific character of the binding sites considered in [11], i.e. one binding site can be only occupied by an unique kind of molecule, which does not allow the competition we are describing in this work.

Let us consider the multi-index $\vec{j} = (j_1, ..., j_M) \in \mathbb{N}_0^M$ with j_i the number of bound TFs of the i-th specie in the set of enhancers, j_0 the number of free sites and $j_P = 1$ if there is a bound PolII and $j_P = 0$ otherwise. Sumarizing we have that all the possible ways of obtaining an equilibrium concentration with $(\vec{j}, j_P)$ TFs-PolII bound is given by the *microstates*

$$\begin{aligned} Z^{(n)}(\vec{j}, j_P = 1; \mathcal{C}) &= C(\mathcal{C}) \frac{n!}{\prod_{i=0}^{M} j_i!} [B] \left(\frac{[PolII]}{K_{RP}} \right) \prod_{i=1}^{M} \left(\frac{t_i [T_i]}{K_i} \right)^{j_i}, \quad (3.6) \\ Z^{(n)}(\vec{j}, j_P = 0; \mathcal{C}) &= C(\mathcal{C}) \frac{n!}{\prod_{i=0}^{M} j_i!} [B] \prod_{i=1}^{M} \left(\frac{[T_i]}{K_i} \right)^{j_i}, \end{aligned}$$

if the recruitment approach, HR4), is assumed and

$$Z^{(n)}(\vec{j}, j_P; \mathcal{C}) = C(\mathcal{C}) \frac{n!}{\prod_{i=0}^{M} j_i!} [B] \left(\frac{[PolII]}{K_{RP}} \right)^{j_P} \prod_{i=1}^{M} \left(\frac{[T_i]}{K_i} \right)^{j_i} \tag{3.7}$$

when stimulated assumption HS4) is considered. In both cases the variable $\mathcal{C}$ describes the relation of cooperativity, if there exist, between the TFs. More concretely, by using (3.3), (3.4) and (3.5), the cooperativity function C takes the value

$$C(\mathcal{C} = \{T_1, ..., T_M\}_c) = c^{\left(\sum_{i=1}^{M} j_i - 1 \right)_+}, \tag{3.8}$$

when total cooperativity holds,

$$C(\mathcal{C} = \{\{T_1\}_{c_1}, ..., \{T_M\}_{c_M}\}) = \prod_{i=1}^{M} c_i^{(j_i - 1)_+}, \tag{3.9}$$

if partial cooperativity is verified, and finally

$$C(\mathcal{C} = \{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_N\}_{c_N}\}) = \prod_{i=1}^{N} c_i^{\left(\sum_{h \in \mathcal{I}_i} j_h - 1 \right)_+} \tag{3.10}$$

when mixed cooperativity occurs. Let us observe that definitions (3.6)-(3.7) are absolutely consistent with the usual convention $0^j = 0$ when $j > 0$ because in the absence of a certain binding particles it is impossible to get any configuration with that kind of particles, and $0! = b^0 = 1$ with $b \geq 0$ since configurations with no bound particles of certain type are independent of that substance concentration.

3.1.2 Configurations probability

Now, the thermodynamic methodology proposes to describe from previous calculations a probability for any possible configuration [1, 97]. Let us notice that the sample space is determined by the multi-indices set

$$\Omega = \left\{ (\vec{j}, j_P) \, ; \vec{j} \in \mathbb{N}_0^M \, , \; |\vec{j}| \leq n \, , \; j_P = 0, 1 \right\}, \tag{3.11}$$

where the constraint on $|\vec{j}| = \sum_{i=1}^{M} j_i \leq n$ is due to the limit of capacity of the n enhancers for accepting bound TFs. Now, using the description of all the possible configurations in terms of the concentrations of TFs and PolII, we define the probability of finding the module in a particular configuration of j_P PolII and $j_1, ..., j_M$ TFs with cooperativity function $\mathcal{C}$ as

$$P^{(n)}(\vec{j}, j_P; \mathcal{C}) = \frac{Z^{(n)}(\vec{j}, j_P; \mathcal{C})}{\sum_{\{\vec{j}', j'_P\} \in \Omega} Z^{(n)}(\vec{j}', j'_P; \mathcal{C})}, \tag{3.12}$$

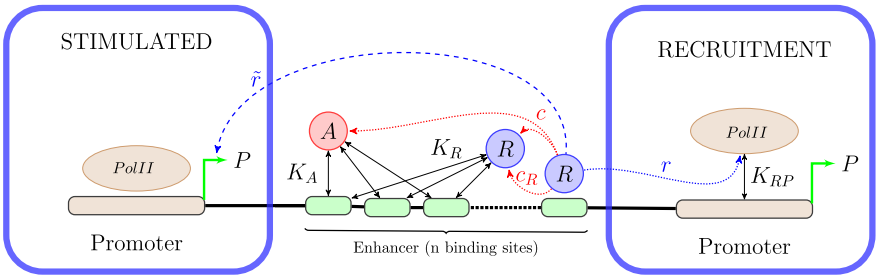

Figure 3.1: Representation of the biochemical mechanisms involved in transcriptional control. Competitive binding process of the TFs (red and purple circles) to an enhancer with n identical binding sites (green). RNA Polymerase (brown oval) binds to the promoter (brown rectangle). Affinities are indicated by black arrows. Alternative biochemical mechanisms are represented with discontinuous arrows: total *vs* partial binding cooperativity between TFs (red) and recruitment *vs* stimulated approaches for transcriptional control exerted by the TFs (blue). See assumption H2, HS4 and HR4 in section 3.1.1 for details.

for all $(\vec{j}, j_P) \in \Omega$.

In next subsections these probabilities, corresponding to each microstate, will be averaged (BEWARE operator) accordingly to the recruitment/stimulated transcription approaches.

3.1.3 Recruitment BEWARE operator

In the work of Shea et al [97], the BEWARE operator for the synthesis of a certain protein depends on the total probability of finding PolII in the promoter (i.e., proportional to the marginal distribution in the number of TFs evaluated at $j_P = 1$). This constitutes a new assumption, denoted by HR5, that has been widely used assuming recruitment assumption HR4 (see for instance [83, 11]). We will denote by the recruitment BEWARE operator the function

$$\text{BEWARE}_r(\vec{T}, [PolII]; \mathcal{C}) = C_B \sum_{|\vec{j}| \leq n} P^{(n)}(\vec{j}, j_P = 1; \mathcal{C})$$

where in definition (3.12) expression (3.6) is assumed and C_B is a proportionality constant. Let us note that C_B could depend on the control exerted in other enhancer modules or later stages of the whole genetic regulation process. Here $\vec{T}$ denotes a vector collecting the concentrations of all the TFs, that is, $\vec{T} = ([T_1], ..., [T_M])$, and $\mathcal{C}$ can be any of the cooperativity relations stablished in assumption *(H2)* whose contribution is determined by (3.8)-(3.10). Splitting the denominator in two sums, depending on the existence

of PolII bound to the configuration, this expression can be easily rewritten in terms of the regulation factor function F_{reg}:

$$\text{BEWARE}_r(\vec{T},[PolII];\mathcal{C}) = \frac{C_B}{1+\frac{\sum_{|\vec{j}|\leq n} Z^{(n)}(\vec{j},0;\mathcal{C})}{\sum_{|\vec{j}|\leq n} Z^{(n)}(\vec{j},1;\mathcal{C})}} = \frac{C_B}{1+\frac{K_{RP}}{[PolII]F_{reg}(\vec{T};\mathcal{C})}} \tag{3.13}$$

where

$$F_{reg}(\vec{T};\mathcal{C}) = \frac{\sum\limits_{|\vec{j}|\leq n} C(\mathcal{C})\frac{n!}{\prod\limits_{i=0}^{M} j_i!}\prod\limits_{i=1}^{M}\left(\frac{t_i[T_i]}{K_i}\right)^{j_i}}{\sum\limits_{|\vec{j}|\leq n} C(\mathcal{C})\frac{n!}{\prod\limits_{i=0}^{M} j_i!}\prod\limits_{i=1}^{M}\left(\frac{[T_i]}{K_i}\right)^{j_i}}\,. \tag{3.14}$$

Let us remark that activators and repressors play symmetric roles, mathematically speaking, since their activating/repressing nature is only reflected in the value of the parameter t_i being bigger than 1 for activators and smaller than 1 for repressors.

3.1.4 Stimulated BEWARE operator

On the other hand, the stimulated transcription approach associates to any configuration with a bound PolII a transcription density [63, 62, 69]. We will call this new assumption HS5. In this point we are going to follow the proposal of [69], adapted in later works [111, 94], where activators and repressors have to be distinguished. Let us consider that $\vec{T}_A = ([T_1],...,[T_{M_a}]) \in \mathbb{R}^{M_a}$, $\vec{T}_R = ([T_{M_a+1}],...,[T_M]) \in \mathbb{R}^{M_r}$ the activators and repressors vectors, and $\vec{j_A} = (j_1,...,j_{M_a}) \in \mathbb{Z}_+^{M_a}$, $\vec{j_R} = (j_{M_{a+1}},...,j_M) \in \mathbb{Z}_+^{M_r}$ the number of bound activators/repressors of the i-th specie, such that $M = M_a + M_r$. Considering the standard concatenation operator we have that $\vec{T} = (\vec{T}_A, \vec{T}_R) \in \mathbb{R}^M$ and $\vec{j} = (\vec{j}_A, \vec{j}_R) \in \mathbb{Z}_+^M$.

Furthermore, it is necessary to fix the basal and maximal/minimal transcription levels definition:

i) r_{bas} is basal transcription rate one would expect from a completely empty configuration with no stimulated transcription at all. If basal transcription is not assumed then $r_{bas} = 0$.

ii) $(\nu_{max}^{(n)} + r_{bas})$: is the level of maximal transcriptional rate of the system, given by a configuration filled with n of the most powerful activator. The dependence of $\nu_{max}^{(n)}$ with respect to the total number of enhancers can be justified from the experiments developed in [83] (see Fig. 4)

where it was observed that a diminishing in the number of enhancers produces a reduction in the maximal expression levels.

iii) $r_{bas}\tilde{r}_M^n$, with $\tilde{r}_M < 1$, the level of minimal transcriptional rate, that would correspond to the configuration completely bound to the most powerful repressors, assumed to be T_M.

From these basic levels and the probabilities $P^{(n)}$, given by (3.12) and (3.7), the activation or repression levels of all the possible configurations are determined by the following expressions:

(i) for states with PolII but no bound activators, that is $j_i = 0$ for any $i = 1, \ldots, M_a$,

$$r_{bas}\left(\prod_{i=M_a+1}^{M}\tilde{r}_i^{j_i}\right)P^{(n)}(\vec{j}, j_P = 1; \mathcal{C})\,,$$

where $\tilde{r}_M \leq \tilde{r}_i < 1$ is a constant that stands for the repression strength of the i-th repressor.

(ii) and in the opposite case, that is states with some bound activators ($|\vec{j}_A| > 0$ that is $j_i > 0$ for some $i = 1, \ldots, M_a$)

$$\left(\prod_{i=1}^{M_a}\tilde{a}_i^{j_i}\right)\left(\prod_{i=M_a+1}^{M}\tilde{r}_i^{j_i}\right)\left(r_{bas} + \nu_{max}^{(n)}\tilde{e}^{j_0+\sum_{i=M_a+1}^{M}j_i}\right)P^{(n)}(\vec{j}, j_P = 1; \mathcal{C})\,,$$

where $\tilde{e} < \min\{\tilde{a}_i; i = 1, \ldots, M_a\}$ is a constant of transcriptional efficiency proportional to free or repressor occupied enhancers, and $\tilde{a}_i \leq 1$ is a constant that stands for the activation strength of the i-th activator, being $\tilde{a}_i = 1$ in the case of the most powerful activator.

(iii) If we assume that the level of transcription of configurations with the same number of activators and repressors should coincide independently of the total number of sites in the configuration (n) this allow us to conclude a plausible expression for

$$\nu_{max}^{(n)} = \frac{\nu_{max}^{(1)}}{\tilde{e}^{n-1}} \tag{3.15}$$

where $\nu_{max}^{(1)}$ is defined in terms of the maximal transcription when only one enhancer is available.

The synthesis of the protein under consideration depends, therefore, on the addition of all these transcriptional efficiencies of states with a bound PolII,

which written in terms of the common factors r_{bas} and $\nu^{(n)}_{max}$ can be expressed as

$$\begin{aligned}&\mathrm{BEWARE}_s(\vec{T},[PolII];\mathcal{C})\\ &= \frac{r_{bas}}{1+\frac{K_{RP}}{[PolII]}}Basal(\vec{T};\mathcal{C})+\frac{\nu^{(n)}_{max}}{1+\frac{K_{RP}}{[PolII]}}Promoter(\vec{T};\mathcal{C})\end{aligned} \tag{3.16}$$

in terms of the *Basal* and *Promoter* functions

$$Basal(\vec{T};\mathcal{C})=\frac{\sum\limits_{|\vec{j}|\leq n} C(\mathcal{C})\frac{n!}{\prod\limits_{i=0}^{M} j_i!}\left[\prod\limits_{i=1}^{M_a}\left(\frac{\tilde{a}_i[T_i]}{K_i}\right)^{j_i}\right]\left[\prod\limits_{i=M_a+1}^{M}\left(\frac{\tilde{r}_i[T_i]}{K_i}\right)^{j_i}\right]}{\sum\limits_{|\vec{j}|\leq n} C(\mathcal{C})\frac{n!}{\prod\limits_{i=0}^{M} j_i!}\prod\limits_{i=1}^{M}\left(\frac{[T_i]}{K_i}\right)^{j_i}}, \tag{3.17}$$

$$Promoter(\vec{T};\mathcal{C})=\frac{\sum\limits_{\substack{|\vec{j}|\leq n\\ |\vec{j}_A|>0}} C(\mathcal{C})\frac{n!}{\prod\limits_{i=0}^{M} j_i!}\tilde{e}^{j_0}\left[\prod\limits_{i=1}^{M_a}\left(\frac{\tilde{a}_i[T_i]}{K_i}\right)^{j_i}\right]\left[\prod\limits_{i=M_a+1}^{M}\left(\frac{\tilde{r}_i\tilde{e}[T_i]}{K_i}\right)^{j_i}\right]}{\sum\limits_{|\vec{j}|\leq n} C(\mathcal{C})\frac{n!}{\prod\limits_{i=0}^{M} j_i!}\prod\limits_{i=1}^{M}\left(\frac{[T_i]}{K_i}\right)^{j_i}}. \tag{3.18}$$

Let us note that in [69] it was deduced the particular expression of these functionals for modelling the transcriptional rates of two Shh target genes, by acting on an enhancer module with $n=3$ sites, where $M=3$ TFs can bind, being $M_a=2$ of them activators and $M_r=1$ repressor. In that work the theoretical expressions assume the same affinity for activators and the repressor, and the activators have the same activation strength $\tilde{a}_1=\tilde{a}_2=1$. Those expressions also consider total cooperativity between TFs and the effects of PolII and its affinity is involved in the constants r_{bas} and $\nu^{(n)}_{max}$.

A remarkable fact that can be pointed out from expressions (3.13), (3.14), (3.16), (3.17) and (3.18) is that all the BEWARE operators, independently of the biochemical mechanisms involved, depend on the TFs concentrations and their binding sites affinities, $[T_i]$ and K_i respectively, but always through the quotients $[T_i]/K_i$. Despite of being a trivial observation, this will be the key ingredient for the appearance of the elasticity when the variability of the activation/repression thresholds is analysed in terms of affinity variations.

These two approaches focus on the transcription process, which is the first mechanism involved in the genetic activity control. However the whole process can be affected during the posterior RNA managing and interpretation. These process in this work are assumed to be linear and their effects are undercover in the value of the constants C_B, r_{bas} and $\nu^{(n)}_{max}$ appearing in (3.13) and (3.16).

3.1.5 Simplification of BEWARE operators expressions

One of the key points of the deduction developed in this tesis is the fact that the regulation factor (3.14), basal (3.17) and promoter (3.18) functions can be explicitly computed. This will give rise to simple rational and polynomical expressions, whose analysis may contribute to the understanding of the general biological process. These calculations exploit a classical strategy employed for obtaining the derivation of the General Binding Equation more than a century ago [53]. For instance in [18] we take advantage of particular cases of these simple expressions to deduce several transcription logics determined by the type of cooperativity between the TFs in the framework of Hh target genes. We start by remarking that the regulation factor, basal and promoter are rational functions where numerators and denominators correspond to polynomial expressions that can be expressed using next definition. Let $\vec{x} = (x_1, \ldots, x_M) \in \mathbb{R}^M$, $\mathcal{C}$ any of the cooperativity relations stablished in Subsection 3.1.1 *(H2)* and $C(\mathcal{C})$ determined by (3.8)-(3.10). Then we define the polynomial function

$$S_e^{(n)}(\vec{x};\mathcal{C}) = \sum_{|\vec{j}|\leq n} C(\mathcal{C}) \frac{n!}{\prod_{i=0}^{M} j_i!} e^{j_0} \prod_{i=1}^{M} x_i^{j_i}\,, \tag{3.19}$$

where the multi-index $\vec{j} \in \mathbb{N}_0^M$, $|\vec{j}| = \sum_{i=1}^{M} j_i$ and $j_0 = n - |\vec{j}|$.

Please recall the vectorial notation $\vec{T} = (\vec{T}_A, \vec{T}_R) \in \mathbb{R}^M$ where

- $\vec{T}_A = ([T_1], ..., [T_{M_a}]) \in \mathbb{R}^{M_a}$ is the vector of activators,
- $\vec{T}_R = ([T_{M_a+1}], ..., [T_M]) \in \mathbb{R}^{M_r}$ is the vector of repressors,

being the whole set of TFs concentrations $\vec{T} = (\vec{T}_A, \vec{T}_R) \in \mathbb{R}^M$. In concordance with previous vectors let us also consider

- $\vec{a}, \tilde{\vec{a}} \in \mathbb{R}^{M_a}$, the vector of the activation intensities such as $a_i \geq 1$ and $\tilde{a}_i \leq 1$, $\forall i = 1, ..., M_a$,
- $\vec{r}, \tilde{\vec{r}} \in \mathbb{R}^{M_r}$, the vector of the repression intensities such as $r_i, \tilde{r}_i \leq 1$ $\forall i = 1, ..., M_r$,
- $\vec{K}_A \in \mathbb{R}_{M_a}$, the vector of the activators binding affinities,
- $\vec{K}_R \in \mathbb{R}^{M_r}$, the vector of the repressor binding affinities.

Then, (3.14), (3.17), (3.18) can be equivalently written as

$$F_{reg}(\vec{T};\mathcal{C}) = \frac{S_1^{(n)}\Big((\vec{a}\circ\vec{T}_A/\vec{K}_A,\vec{r}\circ\vec{T}_R/\vec{K}_R);\mathcal{C}\Big)}{S_1^{(n)}\Big((\vec{T}_A/\vec{K}_A,\vec{T}_R/\vec{K}_R);\mathcal{C}\Big)}, \tag{3.20}$$

$$Basal(\vec{T};\mathcal{C}) = \frac{S_1^{(n)}\Big((\vec{\tilde{a}}\circ\vec{T}_A/\vec{K}_A,\vec{\tilde{r}}\circ\vec{T}_R/\vec{K}_R);\mathcal{C}\Big)}{S_1^{(n)}\Big((\vec{T}_A/\vec{K}_A,\vec{T}_R/\vec{K}_R);\mathcal{C}\Big)}, \tag{3.21}$$

$$\begin{aligned} Promoter(\vec{T};\mathcal{C}) &= \frac{S_{\tilde{e}}^{(n)}\Big((\vec{\tilde{a}}\circ\vec{T}_A/\vec{K}_A,\tilde{e}\vec{\tilde{r}}\circ\vec{T}_R/\vec{K}_R);\mathcal{C}\Big)}{S_1^{(n)}\Big((\vec{T}_A/\vec{K}_A,\vec{T}_R/\vec{K}_R);\mathcal{C}\Big)} \\ &- \frac{S_{\tilde{e}}^{(n)}\Big((\vec{0},\tilde{e}\vec{\tilde{r}}\circ\vec{T}_R/\vec{K}_R);\mathcal{C}\Big)}{S_1^{(n)}\Big((\vec{T}_A/\vec{K}_A,\vec{T}_R/\vec{K}_R);\mathcal{C}\Big)} \end{aligned} \tag{3.22}$$

where $\vec{x}\circ\vec{y}$ and $\vec{x}/\vec{y}$ denote the Hadamard (pointwise) product and division operators. Let us recall that the concatenation operator, $(\vec{x},\vec{y})$, is used to express the main argument of the previous expressions.

Our aim is to compute equivalent simplified expressions for (3.19). In the case of non-cooperativity, $C(\mathcal{C})=1$, expression (3.19) can be very easily computed,

$$S_e^{(n)}((x_1,\ldots,x_M);\{T_1,\ldots,T_d\}_1) = \left(e+\sum_{i=1}^{M}x_i\right)^n, \tag{3.23}$$

by direct application of the multinomial theorem. Accordingly to this result we will adopt the mathematical definition of (3.19) for empty vectors, that is when $M=0$, as $S_e^{(n)}(\cdot,\cdot)=e^n$.

On the other hand, in the presence of cooperativity the multinomial theorem can not be applied so straightforwardly, and in this case the expression of (3.19) is strictly determined by the cooperativity relations (3.8)-(3.9). However, we can use our definition of the general mixed cooperation (3.10) and resume all the possible cooperative cases in one general expression of the function $S_e^{(n)}$. This is, we will assume a general mixed cooperativity configuration $\mathcal{C}=\{\{\mathcal{T}_1\}_{c_1},...,\{\mathcal{T}_N\}_{c_N}\}$, and

- $\mathcal{I}_j$, the indices of the components of $\vec{x}$ corresponding to the TFs inside of the cooperative family with cooperativity constant c_j.

Then, it can be shown (see Appendix A.1) that

$$S_e^{(n)}\Big((x_1,\dots,x_M);\{\{\mathcal{T}_1\}_{c_1},...,\{\mathcal{T}_N\}_{c_N}\}\Big) = \sum_{\substack{|\vec{h}|_\infty\leq 1\\ \vec{h}\in\mathbb{N}_0^N}} \left(e+\sum_{j=1}^{N} h_j c_j \sum_{i\in\mathcal{I}_j} x_i\right)^n \prod_{j=1}^{N} \frac{\left(1-\frac{1}{c_j}\right)^{1-h_j}}{(c_j)^{h_j}} \tag{3.24}$$

where the addition on the multi-index $\vec{h}=(h_1,\dots,h_N)\in\mathbb{N}_0^N$ considers all the possible combinations where the components, h_j, can only be 0 or 1, that is $|\vec{h}|_\infty=\max_j\{h_j\}\leq 1$.

From this general result we can get, as particular cases, the value of (3.19) in the total and partial cooperativity cases. In the total cooperativity case all the TFs cooperate between them with the same cooperativity constant c, thus $N=1$ and the possible values of $\vec{h}$ are only $\vec{h}=(0)$ or $\vec{h}=(1)$ which leads to the expression

$$S_e^{(n)}((x_1,\dots,x_M);\{T_1,\dots,T_M\}_c) = \left(1-\frac{1}{c}\right)e^n + \frac{1}{c}\left(e+c\sum_{i=1}^{M}x_i\right)^n. \tag{3.25}$$

Let us observe that this expression coincides with (3.23) when $c=1$, that is, in absence of cooperativity.

On the other side, if TFs cooperate only between the proteins of the same specie we have that $N=M$, $\mathcal{I}_i=\{i\}$, and expression (3.24) reads,

$$S_e^{(n)}((x_1,\dots,x_M);\{T_1\}_{c_1},\dots,\{T_M\}_{c_M}) = \sum_{\substack{|\vec{h}|_\infty\leq 1\\ \vec{h}\in\mathbb{N}_0^M}} \left(e+\sum_{j=1}^{M} h_j c_j x_j\right)^n \prod_{j=1}^{M} \frac{\left(1-\frac{1}{c_j}\right)^{1-h_j}}{(c_j)^{h_j}} \tag{3.26}$$

for any M. Specially useful for next section will be the case of only two TFs, $M=2$. In this case $\vec{h}$ can be $(0,0)$, $(1,0)$, $(0,1)$ or $(1,1)$ giving rise to the expression:

$$S_e^{(n)}((x_1,x_2);\{T_1\}_{c_1},\{T_2\}_{c_2}) = \left(1-\frac{1}{c_1}\right)\left(1-\frac{1}{c_2}\right)e^n + \frac{(e+c_1x_1)^n}{c_1}\left(1-\frac{1}{c_2}\right) + \frac{(e+c_2x_2)^n}{c_2}\left(1-\frac{1}{c_1}\right) + \frac{(e+c_1x_1+c_2x_2)^n}{c_1c_2}. \tag{3.27}$$

3.2 Extreme cooperativity approach: Hill modules

In this section we are going to relate the developed thermodynamical modelling with a very frequently employed modelling approach in genetic control, the Hill modules. Using the analogy with Michaelis-Menten kinetic

equations some groups have proposed Michaelis-like functions for modelling genetic control (see [98, 43] and references therein). This approach has been used in order to capture cooperativity effects via adding Hill coefficients [55] to the Michaelis-like functions. It is frequently assumed in the literature, see for instance [17, 3, 80, 77], that the genetic transcription is controlled by combinations of Hill type functions. For example, the transcription obtained by a single activator/repressor corresponds to the expressions:

$$\frac{d[P]}{dt} = \alpha \frac{[A]^n}{K_d + [A]^n} + \eta \qquad \frac{d[P]}{dt} = \frac{\beta}{K_d + [R]^n} + \gamma \,. \tag{3.28}$$

where α and β are proportional constants, and η, γ correspond respectively to optional basal or minimal expression levels.

In [98] the authors show that the Hill functions can approximate the transcription rates under conditions of high cooperativity. In this section we will show that, indeed, the Hill functions (3.28) can be framed into the thermodynamic approach by assuming extreme cooperativity.

The stimulated BEWARE operators corresponding to a single activator/repressor transcription factor would be obtained from (3.16) and using

$$\begin{cases} Basal\big([A];\{A\}_c\big) = 1\,, \\ Promoter\big([A];\{A\}_c\big) = \dfrac{\frac{1}{c}\left(\tilde{e}+c\frac{[A]}{K_A}\right)^n - \frac{1}{c}\tilde{e}^n}{\left(1-\frac{1}{c}\right)+\frac{1}{c}\left(1+c\frac{[A]}{K_A}\right)^n}\,, \end{cases} \tag{3.29}$$

in the single activator case

$$\begin{cases} Basal\big([R];\{R\}_c\big) = \dfrac{\left(1-\frac{1}{c}\right)+\frac{1}{c}\left(1+c\tilde{r}[R]/K_R\right)^n}{\left(1-\frac{1}{c}\right)+\frac{1}{c}\left(1+c[R]/K_R\right)^n}\,, \\ Promoter\big([R];\{R\}_c\big) = 0\,, \end{cases} \tag{3.30}$$

for a single repressor. Let us remark that these expressions have been obtained from (3.21), (3.22).

The extreme cooperativity assumption implies that the cooperative binding constant c is large [98]. However, we cannot compute directly a limit in c without assuming some natural restrictions. By definition (see assumption H2 in section 3.1.1), the cooperativity interactions essentially modulate the binding affinity of the transcription factors in the binding sites. More specifically, if K_* is the affinity constant of a transcription factor, the affinity constant of a consecutive cleavage will be K_*/c with $c > 1$ (recall the inverse relation where, for lower affinity constants, the transcription factor will have higher binding affinity). Moreover, if we fill the n enhancers with the same transcription factor, it is clear to see that the "global" affinity constant of the whole process will be $K_d = K_*^n/c^{n-1}$. The extreme cooperativity assumption claims that, since we are interested on only modifying the cooperativity

between the transcription factors, we will compute a limit of $c \to \infty$ without modifying the global affinity constant in the process, keeping $K_d = cte$ (and consequently, $K_* \to 0$ [98]). These limits can be computed easily, so we get

$$\begin{aligned} \lim_{\substack{c\to\infty \\ K_d=cte}} Promoter\big([A];\{A\}_c\big) &= \frac{[A]^n}{K_d+[A]^n}\,, \\ \lim_{\substack{c\to\infty \\ K_d=cte}} Basal\big([R];\{R\}_c\big) &= \frac{K_d+\tilde{r}^n[R]^n}{K_d+[R]^n} \end{aligned}$$

which allow us to justify previous Hill modules because the limits of the corresponding BEWARE operators coincide with (3.28) by taking

$$\begin{cases} \alpha = \frac{r_{bas}}{1+K_{RP}[PolII]^{-1}}\,, \\ \eta = \frac{\nu_{max}^{(n)}}{1+K_{RP}[PolII]^{-1}}\,, \\ \beta = \frac{r_{bas}K_d(1-\tilde{r}^n)}{1+K_{RP}[PolII]^{-1}}\,, \\ \gamma = \frac{r_{bas}\tilde{r}^n}{1+K_{RP}[PolII]^{-1}}\,. \end{cases}$$

Let us notice that the optimal values η and γ vanish when $r_{bas} = 0$ or $\tilde{r} = 0$ respectively, which corresponds in the thermodynamical model to null basal level and a repressor executing a total repression once it is bound. Indeed, these are some extra hypothesis employed in [98] for justifying (3.28) (with $\eta = 0$ and $\gamma = 0$) that are not required in our approach.

It is also interesting to mention that the same reasoning can be used to prove that, under the extreme cooperativity assumption, the recruitment BEWARE operators also converge toward a generalised Hill type functions. Since

$$\begin{aligned} \lim_{\substack{c\to\infty \\ K_d=cte}} F_{reg}\big([A];\{A\}_c\big) &= \frac{K_d+a^n[A]^n}{K_d+[A]^n}\,, \\ \lim_{\substack{c\to\infty \\ K_d=cte}} F_{reg}\big([R];\{R\}_c\big) &= \frac{K_d+r^n[R]^n}{K_d+[R]^n}\,, \end{aligned}$$

the associated recruitment BEWARE operators verify

$$\begin{aligned} \lim_{\substack{c\to\infty \\ K_d=cte}} \mathrm{BEWARE}_r([A],[PolII];\{A\}_c) &= \alpha\frac{[A]^n}{\delta K_d+[A]^n}+\eta\,, \\ \lim_{\substack{c\to\infty \\ K_d=cte}} \mathrm{BEWARE}_r([R],[PolII];\{R\}_c) &= \frac{\beta}{\epsilon K_d+[R]^n}+\gamma\,, \end{aligned}$$

where now $\eta = \frac{C_B}{1+\frac{K_{RP}}{[PolII]}}$ and $\gamma = \frac{C_B r^n}{r^n+\frac{K_{RP}}{[PolII]}}$ are respectively a basal and minimal transcriptional rates, and $\alpha = \frac{C_B}{1+\frac{[PolII]}{K_{RP}}}\frac{a^n-1}{a^n+1}$, $\beta = \frac{K_{RP}}{[PolII]}\frac{C_B K_d(1-r^n)}{\left(r^n+\frac{K_{RP}}{[PolII]}\right)^2}$,

$\delta = \frac{1+\frac{K_{RP}}{[PolII]}}{a^n+\frac{K_{RP}}{[PolII]}}$ and $\epsilon = \frac{1+\frac{K_{RP}}{[PolII]}}{r^n+\frac{K_{RP}}{[PolII]}}$ are constants. Let us observe that the limits in this case are generalisations of the Hill classical modules (3.28) because of parameters $\delta < 1$ and $\epsilon > 1$.

When the modelling by Hill modules involve the effects of several transcription factors (see for instance [3]) it is not so clear which are the counterparts of (3.28). If the binding sites or any TF are independent and there are no cooperativity interactions, the Hill candidates can be computed straightforwardly [17, 11]. However, the same question is not clear when the TFs compite for the same binding sites or cooperate between them. Our modelling approach give us a clear strategy to propose Hill type modules in the presence of several TFs competing for the same enhancers. They might be deduced from the extreme cooperativity limit of the stimulated/recruitment BEWARE operators if the total cooperativity holds between the TFs:

$$\lim_{\substack{c\to\infty \\ K_d=cte}} F_{reg}\big(([T_1],...,[T_M]);\{\{T_1,...,T_M\}_c\}\big) = \frac{K_d + \left(\sum_{i=1}^{M} a_i[T_i]\right)^n}{K_d + \left(\sum_{i=1}^{M}[T_i]\right)^n}, \quad (3.31)$$

$$\lim_{\substack{c\to\infty \\ K_d=cte}} Basal\big(([T_1],...,[T_M]);\{\{T_1,...,T_M\}_c\}\big) = \frac{K_d + \left(\sum_{i=1}^{M} \tilde{a}_i[T_i]\right)^n}{K_d + \left(\sum_{i=1}^{M}[T_i]\right)^n}, \quad (3.32)$$

$$\lim_{\substack{c\to\infty \\ K_d=cte}} Promoter\big(([T_1],...,[T_M]);\{\{T_1,...,T_M\}_c\}\big) = \frac{\left(\sum_{i=1}^{M_a} \tilde{a}_i[T_i] + \sum_{j=M_a+1}^{M} \tilde{e}\tilde{r}_i[T_j]\right)^n - \left(\sum_{j=M_a+1}^{M} \tilde{e}\tilde{r}_i[T_j]\right)^n}{K_d + \left(\sum_{i=1}^{M}[T_i]\right)^n},$$

or

$$\lim_{\substack{c\to\infty \\ K_d=cte}} F_{reg}\big(([T_1],...,[T_M]);\{\{T_1\}_c,...,\{T_M\}_c\}\big) = \frac{K_d + \sum_{i=1}^{M} (a_i[T_i])^n}{K_d + \sum_{i=1}^{M}[T_i]^n}, \quad (3.33)$$

$$\lim_{\substack{c\to\infty \\ K_d=cte}} Basal\big(([T_1],...,[T_M]);\{\{T_1\}_c,...,\{T_M\}_c\}\big) = \frac{K_d + \sum_{i=1}^{M} (\tilde{a}_i[T_i])^n}{K_d + \sum_{i=1}^{M} [T_i]^n}\,, \tag{3.34}$$

$$\lim_{\substack{c\to\infty \\ K_d=cte}} Promoter\big(([T_1],...,[T_M]);\{\{T_1\}_c,...,\{T_M\}_c\}\big) = \frac{\sum_{i=1}^{M_a} (\tilde{a}_i[T_i])^n}{K_d + \sum_{i=1}^{M} [T_i]^n}\,,$$

when only partial cooperativity is present between TFs (see Appendix A.2 for a general proof).

3.3 Global Activator/Repressor variables reduction

Up until this point we have been able to obtain BEWARE operators that modelize situations with any number of transcription factors. However, even though these are different proteins, it is important to recall that the role they play in the transcription can be resumed as a binary function: activation of the signal, or repression. This makes us wonder if there is a possible way to group all the TFs in only two 'global' terms, one grouping all the activators, and the other all grouping all the repressors. Indeed, recalling the previous expressions, we can clearly see that this can be done if the following assumptions are fulfilled

- All the activators (repressors) have the same activation (repression) strength.
- All the activators (repressors) cooperate in the same family with the same cooperativity constant.

In this case the effect of all the activators and repressors in the BEWARE functional can be summarised in a global activator (repressor) variables:

$$[A] = \sum_{i=1}^{M_a} \frac{[T_i]}{\bar{K}_i}\ ;\ \bar{K}_i = \frac{K_i}{K_A}\ \text{ being }\ K_A = \frac{\sum_{i=1}^{M_a} K_i}{M_a}\,,$$

$$[R] = \sum_{i=M_a+1}^{M} \frac{[T_i]}{\bar{K}_i}\ ;\ \bar{K}_i = \frac{K_i}{K_R}\ \text{ being }\ K_R = \frac{\sum_{i=M_a+1}^{M} K_i}{M_r}\,.$$

The fact that the BEWARE operators only depend on these two global variables can be easily checked by replacing the new variables in the expressions of $S_{\tilde{e}}$ of (3.20), (3.21) and (3.22), and noting that the new functions obtained correspond to the expressions of the functions $S_{\tilde{e}}$ in the case of $M = 2$ transcription factors. More specifically, the regulation factor, basal and promoter functions read

$$F_{reg}\Big(([A],[R]);\{\{A,R\}_c\}\Big) = \frac{(c-1)+\bar{\alpha}_r^n\left(\frac{c[A]}{K_A},\frac{c[R]}{K_R}\right)}{(c-1)+\alpha_r^n\left(\frac{c[A]}{K_A},\frac{c[R]}{K_R}\right)}, \tag{3.35}$$

$$Basal\Big(([A],[R]);\{\{A,R\}_c\}\Big) = \frac{(c-1)+\bar{\alpha}_s^n\left(\frac{c[A]}{K_A},\frac{c[R]}{K_R}\right)}{(c-1)+\alpha_s^n\left(\frac{c[A]}{K_A},\frac{c[R]}{K_R}\right)}, \tag{3.36}$$

$$Promoter\Big(([A],[R]);\{\{A,R\}_c\}\Big) = \frac{\beta_s^n\left(\frac{c[A]}{K_A},\frac{c[R]}{K_R}\right)-\bar{\beta}_s^n\left(\frac{c[A]}{K_A},\frac{c[R]}{K_R}\right)}{(c-1)+\alpha_s^n\left(\frac{c[A]}{K_A},\frac{c[R]}{K_R}\right)}, \tag{3.37}$$

for the total cooperativity case and

$$F_{reg}\Big(([A],[R]);\{\{A\}_{c_A},\{R\}_{c_R}\}\Big) \tag{3.38}$$
$$= \frac{(c_A-1)(c_R-1)+(c_R-1)\gamma_r^n\left(\frac{c_A[A]}{K_A}\right)+(c_A-1)\bar{\beta}_r^n\left(\frac{c_R[R]}{K_R}\right)+\bar{\alpha}_r^n\left(\frac{c_A[A]}{K_A},\frac{c_R[R]}{K_R}\right)}{(c_A-1)(c_R-1)+(c_R-1)\bar{\gamma}_r^n\left(\frac{c_A[A]}{K_A}\right)+(c_A-1)\beta_r^n\left(\frac{c_R[R]}{K_R}\right)+\alpha_r^n\left(\frac{c_A[A]}{K_A},\frac{c_R[R]}{K_R}\right)},$$

$$Basal\Big(([A],[R]);\{\{A\}_{c_A},\{R\}_{c_R}\}\Big) \tag{3.39}$$
$$= \frac{(c_A-1)(c_R-1)+(c_R-1)\lambda_s^n\left(\frac{c_A[A]}{K_A}\right)+(c_A-1)\bar{\gamma}_s^n\left(\frac{c_R[R]}{K_R}\right)+\bar{\alpha}_s^n\left(\frac{c_A[A]}{K_A},\frac{c_R[R]}{K_R}\right)}{(c_A-1)(c_R-1)+(c_R-1)\lambda_s^n\left(\frac{c_A[A]}{K_A}\right)+(c_A-1)\gamma_s^n\left(\frac{c_R[R]}{K_R}\right)+\alpha_s^n\left(\frac{c_A[A]}{K_A},\frac{c_R[R]}{K_R}\right)},$$

$$Promoter\Big(([A],[R]);\{\{A\}_{c_A},\{R\}_{c_R}\}\Big) \tag{3.40}$$
$$= \frac{(c_R-1)\delta_s^n\left(\frac{c_A[A]}{K_A}\right)+\beta_s\left(\frac{c_A[A]}{K_A},\frac{c_R[R]}{K_R}\right)-(c_R-1)\bar{\delta}_s^n-\bar{\beta}_s^n\left(\frac{c_R[R]}{K_R}\right)}{(c_A-1)(c_R-1)+(c_R-1)\lambda_s^n\left(\frac{c_A[A]}{K_A}\right)+(c_A-1)\gamma_s^n\left(\frac{c_R[R]}{K_R}\right)+\alpha_s^n\left(\frac{c_A[A]}{K_A},\frac{c_R[R]}{K_R}\right)},$$

in the case of partial cooperativity, where we have written the Recruitment operators in terms of the polnomials

$$\begin{cases} \alpha_r(\tilde{x},\tilde{y}) = 1+\tilde{x}+\tilde{y}\,, \\ \bar{\alpha}_r(\tilde{x},\tilde{y}) = 1+a\tilde{x}+r\tilde{y}\,, \\ \beta_r(\tilde{y}) = 1+\tilde{y}\,, \\ \bar{\beta}_r(\tilde{y}) = 1+r\tilde{y}\,, \\ \gamma_r(\tilde{x}) = 1+a\tilde{x}\,, \\ \bar{\gamma}_r(\tilde{x}) = 1+\tilde{x}\,, \end{cases} \tag{3.41}$$

and the Stimulated operators in terms of

$$\begin{cases} \alpha_s(\tilde{x},\tilde{y}) = 1+\tilde{x}+\tilde{y}\,, \\ \bar{\alpha}_s(\tilde{x},\tilde{y}) = 1+\tilde{x}+\tilde{r}\tilde{y}\,, \\ \beta_s(\tilde{x},\tilde{y}) = \tilde{e}+\tilde{x}+\tilde{e}\tilde{r}\tilde{y}\,, \\ \bar{\beta}_s(\tilde{y}) = \tilde{e}+\tilde{e}\tilde{r}\tilde{y}\,, \\ \gamma_s(\tilde{y}) = 1+\tilde{y}\,, \\ \bar{\gamma}_s(\tilde{y}) = 1+\tilde{r}\tilde{y}\,, \\ \delta_s(\tilde{x}) = \tilde{e}+\tilde{x}\,, \\ \bar{\delta}_s = \tilde{e}\,, \\ \lambda_s(\tilde{x}) = 1+\tilde{x}\,. \end{cases} \tag{3.42}$$

This reduction is not trivial. In [62] it was addressed the problem of collecting the control executed by different TFs into two global Activation/Repression variables in the particular case of the Shh system. In this case the transcription is controlled by the concentration balance of $M = 3$ TFs: a pair of activators $GliA$, $Gli3A$ and a repressor $Gli3R$ competing for $n = 3$ binding sites. Gli-DNA enhancers binding affinities were described by dissociation constants K_1 for Gli1 and K_3 for both forms of Gli3. Since the model suggested total cooperativity between all the TFs and the activators, $GliA$ and $Gli3A$, it can be seen that the BEWARE operator deduced in [69, 94] depends on the global activator variable

$$[A] = \frac{[GliA]}{\bar{K}_1} + \frac{[Gli3A]}{\bar{K}_3}$$

with dissociation constant $K_A = \frac{K_1+K_3}{2}$ where the weights for defining A are given by $\bar{K}_1 = \frac{K_1}{K_A}$, $\bar{K}_3 = \frac{K_3}{K_A}$. Since only one repressor was taken into consideration, $Gli3R$, this plays the role of repressor variable with dissociation constant K_3.

Let us observe that these expressions allow us to relate the parameters in these models with measurable reference values as the minimal, basal and maximal expression levels, following the ideas presented in [62]. These values can be easily computed by letting $[A] \to 0\,,\ [R] \to \infty$ or $[A] \to 0\,,\ [R] \to 0$ or $[A] \to \infty\,,\ [R] \to 0$ respectively. As we resume in Table 3.1, these levels can be used to relate parameters of both Recruitment and Stimulated models.

3.4 Activation threshold and spatial genetic expression

Genetic experiments analyse the gene expression that results in the segregation of a protein P. The experiments work with either wild-type expression and/or expressions measured in mutants where some of the biochemical conditions have been modified. Examples are: Shh/Hh target genes

Table 3.1: Theoretical values for minimal, basal and maximal transcriptional rates for BEWARE operators in presence of global activator/represor variables.

	Minimal	Basal	Maximal
Recruitment	$\frac{C_B}{1+\frac{K_{RP}}{[PolII]r^n}}$	$\frac{C_B}{1+\frac{K_{RP}}{[PolII]}}$	$\frac{C_B}{1+\frac{K_{RP}}{[PolII]a^n}}$
Stimulated	$\frac{r_{bas}r^n}{1+\frac{K_{RP}}{[PolII]}}$	$\frac{r_{bas}}{1+\frac{K_{RP}}{[PolII]}}$	$\frac{r_{bas}+\nu^{(n)}_{max}}{1+\frac{K_{RP}}{[PolII]}}$

in [83, 62, 79, 3, 74, 91, 80], other Drosophila's target genes [6, 59] and prototypical biological systems, such as the λ phage [5, 112]. Of all these experimental approaches, of particular interest are those that compare transcription rates against basal levels. Correctly defining the basal transcription level is quite important because it depends on which part of the transcriptional control system you are considering. Since in our case we are focusing on the transcriptional effects due to signalling in the specific module of n binding sites, the basal level is the expression of P observed when these enhancers are disabled from receiving that signalling [83, 74]. These basal levels can still be signal-dependent since they could collect the signalling effects coming from other modules of enhancers or TFs. The deduced models allow us to predict when cells can be relatively activated or repressed with respect to the basal (net activated-repressed), in the presence of two opposing signals (activators *vs* repressors). This can be easily done by using a threshold level between the transcription factors. Each BEWARE operator defines a curve (function) in the $[A]-[R]$ plane which separates that plane in two regions. If the BEWARE is evaluated on concentrations of TFs that are below this curve, then it will predict an expression of P higher than the basal rate (and lower for concentrations above the curve). We have called these concentrations of transcription factors "Activation/Repression regions", and are depicted in Fig. 3.2 (C). It is important to note that the thresholds (i.e., the interphase between the Activation/Repression regions) depend strongly on the biochemical mechanisms involved in the transcriptional binding process by means of the BEWARE operator used in their determination. We can define the threshold for n binding sites as a function, $f_{m,l}([A];n)$, where the corresponding BEWARE operator fulfils the equation:

$$\text{BEWARE}_m([A], f_{m,l}([A];n), [RNAP];\mathcal{C}) = \text{BEWARE}_m(0, 0, [RNAP];\mathcal{C}) \tag{3.43}$$

where $\mathcal{C} = \{[A],[R]\}_c$ in the case of *t*otal cooperativity ($l = t$) and $\mathcal{C} = \{\{[A]\}_{c_A}, \{[R]\}_{c_R}\}$ in the case of *p*artial cooperativity ($l = p$). That is, the thresholds are the BEWARE operator level curves corresponding to the

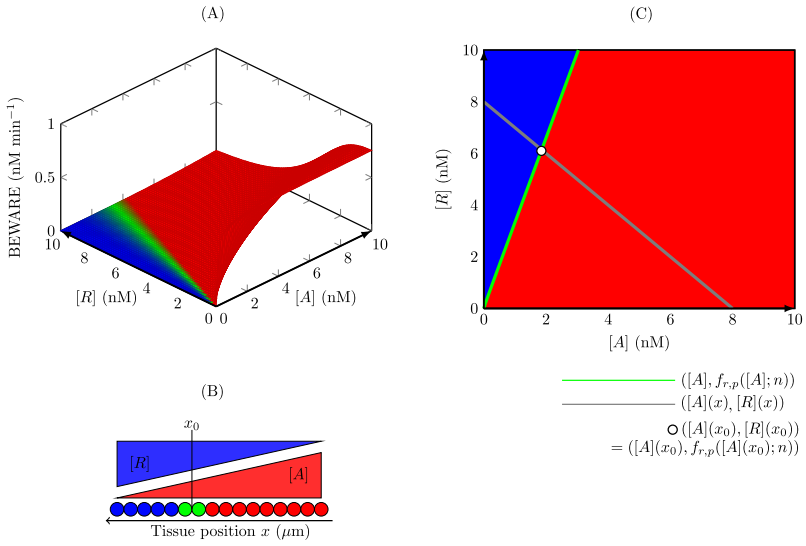

Figure 3.2: **Activation-repression threshold and net activated/repressed tissular regions**. **(A)** BEWARE operator representation. Blue, green and red correspond to values of the BEWARE operator, lower than, close to or higher than the basal level. **(B)** Triangles represent the tissular activator/repressor gradients governing gene transcription. Lower circles show net activated/repressed tissular regions determined by the upper gradients and the BEWARE operator in figure (A). The limiting position, x_0 corresponds to cells that have TF concentrations determined by the white circled in figure (C). **(C)** Representation of (A) and (B) in the [A]-[R] plane. The blue region is concentrations leading to BEWARE values below the basal level. The red region is concentrations leading to BEWARE values above the basal level. The green curve is the activation/repression threshold, that is, concentrations leading to the basal level. The grey curve shows the TF concentrations represented by the triangles in (B). The intersection between the green and grey curves, i.e. white circle, determines tissular regions where cells are net activated or repressed according to the BEWARE operator. Parameter values can be found in Table B.2. in Appendix B.

absence of signalling in the enhancers module, the basal level. It is important to note that $f_{m,l}([A]; n)$ is also a strictly increasing function (please, refer to Appendix A.4 for a proof of both existence and properties of this function). As we will see in the following, this curve can be used to extrapolate several pieces of transcriptional information from the molecular to the tissular level.

3.4.1 Net Activated Cell regions

Usually, morphogens are secreted from a specific part of the tissue and by spreading give rise to a gradient in the receptor cells which causes opposing gradients of activators and repressors of the corresponding transcription factors. The balance between the activator and repressor gradient gives a tissular region with higher levels of activators and lower levels of repressors depending on how close the cells are to the morphogenetic source. When all the enhancers are mutated to abolish TF binding we get basal expression levels. These levels take into account the control of other modules or other TFs. Then, the effect of the abolished enhancers can be measured by comparing the gene expression under the original enhancers distribution with the expressions in the mutant with removed enhancers. This comparison should determine the net effect of the removed enhancer, since in those tissular regions where the expression is higher/lower than the basal level the module causes relative activation/repression. Recalling the threshold function, $f_{m,l}$ predicts if a cell is activated or repressed by simply checking if the concentration of repressors inside of the cell is higher or lower than the threshold corresponding to its activators concentration. Or in other words, a cell with levels $[A]$ and $[R]$ of activators and repressors will be

- relatively activated if $[R] < f_{m,l}([A], n)$
- or relatively repressed if $[R] > f_{m,l}([A], n)$.

Hence, if we know the distribution of TF concentration across the developing tissue, we can deduce the position of cells that will be activated or repressed (Net Activated/Repressed Cell regions from now on). The information given by the activation and repression regions, added to the knowledge of the distribution of the transcription factors across the tissue, allows the analysis of how the activation profile is distributed spatially.

3.4.2 Opposing TF gradients

The pattern of activated-or-repressed cells is essential in the development of several biological systems. However, many morphogens, including Hedgehog, create opposing activator and repressor gradients that control the proper formation of these patterns. This is the case of for instance Drosophila's wing development, by the target genes *dpp* and *ptc* [18, 102], among others. The

study of the tissular activation/repression patterns due to opposing TF gradients has been extensively studied via modifications of several biochemical properties of the development system. Variations such as modification of TFs binding sites affinities, or even the number of binding sites, have been shown to have important effects in the pattern formation [83]. In order to model these opposing TFs concentrations, we are going to assume that they do not change over time and both are monotonic along the tissue,

$$[A] = [A](x) \text{ strictly decreasing and } [R] = [R](x) \text{ strictly increasing} \tag{3.44}$$

in terms of the position x in the tissue. Moreover, we will consider that total amount of TFs

$$h(x) = [R](x) + [A](x)\,, \tag{3.45}$$

is a non decreasing function on x. With this properties we obtain an spatially opposed TF gradient, where the concentration of repressors increases, and the concentration of activators decreases.

By (3.45), we have a curve in the $[A] - [R]$ plane of the form $\vec{\gamma}(x) = ([A](x), [R](x))$ (see figure 3.2C). This curve, from (3.44), can be reparametrized by $[A]$ such as $\vec{\gamma}([A]) = ([A], \phi([A]))$. It is important to note that $\phi([A])$ is a decreasing function since

$$\phi'([A]) = R'([A]^{-1}([A]))\frac{1}{[A]'([A])} < 0.$$

Now, please recall that the threshold function $f_m, l([A]; n)$ is an increasing function that divides the A-R plane in the Activation/Repression regions. Hence, there exists an intersection point $[A]_0$ such that $\phi([A]_0 = f_{m,l}([A]_0; n)$, and a position x_0 such that $\vec{\gamma}(x_0) = ([A]_0, f_{m,l}([A]_0; n))$. x_0 defines a spatial boundary that establish two simple (and connected) regions in the tissue: the net activated and repressed cells regions (NAC). This assures that the transcriptional control of an opposing-gradient is enough to establish activation patterns in the tissue, where the cells are either activated or repressed (see Fig. 3.2 B). Indeed, let us suppose that we have two threshold functions $f^{(1)}_{m,l}([A]; n)$ and $f^{(2)}_{m,l}([A]; n)$ such as the activation region of $f^{(1)}_{m,l}$ is higher than the activation region of $f^{(2)}_{m,l}$ (i.e., we have $f^{(1)}_{m,l} > f^{(2)}_{m,l}$ for all $[A]$). Then, it is easy to check that the intersection points $[A]^{(1)}_0$ (in the position $x^{(1)}_0$) and $[A]^{(2)}_0$ (in $x^{(2)}_0$) with $\phi([A])$ will maintain the order:

$$[A]^{(1)}_0 < [A]^{(2)}_0 \rightarrow x^{(1)}_0 > x^{(2)}_0\,. \tag{3.46}$$

This is, fixing a concentration of opposing transcription factors gradient, the higher the activation region is, the higher the number of activated cells (NAC) we should obtain in the tissue. This proves the potential of the

threshold function as an analytic tool in the modelling of transcription processes. The information, given in terms of concentrations of transcription factors inside of the cell, now can be decoded at tissular level outside the cell via the analysis of $f_{m,l}([A]; n)$.

As a matter of fact, it is important to remark that the monotonicity of $f_{m,l}$ does not necessarily correspond to the monotonicity of the BEWARE operator with respect to the TFs. Our analysis predicts concentration regions where the BEWARE operators involving total cooperation can exhibit behaviours of 'inverse logic' such as increases in the transcription rate where there is an increase in the concentration of repressors (or decreases where there is an increase in the concentration of activators). We refer to this effect as the "pull-effect", and it takes place when total cooperativity is very strong. Indeed, the analysis performed allows us to describe in great detail when these effects can be found (see A.3 for mathematical analysis, and Fig. 3.3 for a graphical explanation of these effects).

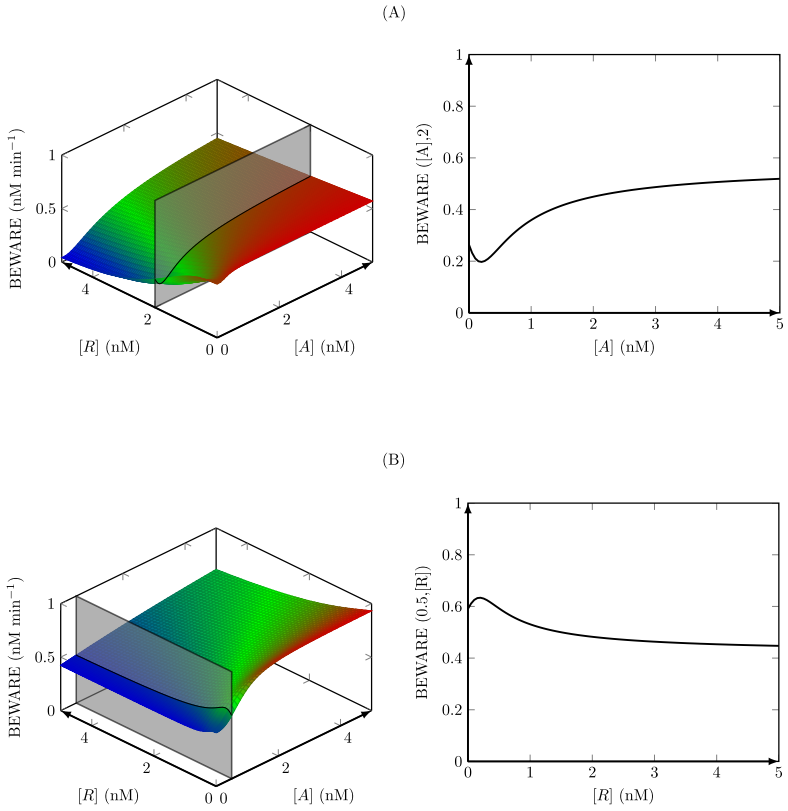

Figure 3.3: Inverse logic for a Recruitment BEWARE operator with total cooperativity. (A) Activators inverse logic. (B) Repressors inverse logic. Parameter values can be found in Table B.3 in Appendix B.

3.4.3 Sensitivity analysis of the threshold functions: Elasticity

In order to test out the change in the activation/repression regions, we have analysed the behaviour of the threshold function $f_{m,l}$ under the following perturbations:

1. *proportional reduction in affinity for the enhancers*:

$$K_A \to \eta K_A \quad \text{and} \quad K_R \to \eta K_R \quad \text{being} \quad \eta \geq 1\,, \tag{3.47}$$

2. *decrease in the number of available binding sites*:

$$n \to n - 1\,. \tag{3.48}$$

(3.47) represents the fact that an increase in the perturbation parameter η corresponds to lower affinities between the TFs and the enhancers, remembering that K_* are dissociation constants. Our study predicts that the response obtained by these perturbations are closely related in the case of stimulated and recruitment operators. Although both alterations interfere with the action of activators and repressors in the same way, it is surprising that in all these models the response to perturbations (3.47) and (3.48) is qualitatively predicted by the elasticity of the function $f_{m,l}$,

$$\epsilon_{m,l}([A];n) = \frac{[A]f'_{m,l}([A];n)}{f_{m,l}([A];n)} \approx \frac{\Delta f_{m,l}([A];n)/f_{m,l}([A];n)}{\Delta[A]/[A]}\,, \tag{3.49}$$

where $\Delta f_{m,l}([A];n) = f_{m,l}([A]+\Delta[A];n) - f_{m,l}([A];n)$. ϵ is a quantity usually used in Economics in order to measure a system's responses to proportional perturbations [58]. It is also known as a condition number in Numerical Analysis [44]. This index has also been introduced in biological contexts, for instance in Ecology [19]. Indeed, there exists a direct relationship between elasticity and the threshold response to perturbations (3.47) and (3.48). This is, if we annalyse the sign of the variation of the threshold with respect the perturbations (3.47) and (3.48), we can check that

$$\begin{aligned} sign\left\{ \frac{\delta f_{m,l}}{\delta \eta}([A];n)\Big|_{\eta=1} \right\} &= sign\left\{ f_{m,l}([A];n-1) - f_{m,l}([A];n) \right\} \\ &= sign\left\{ 1 - \epsilon_{m,l}([A];n) \right\}\,. \end{aligned} \tag{3.50}$$

This is applicable to recruitment and stimulated operators ($m = r, s$) in their total and partial cooperativity versions ($l = t, p$). Please note, both identities in (3.50) show that the Activation/Repression thresholds under signalling interferences (3.47) and (3.48) react in the same qualitative manner. The thresholds:

- will decrease in the *elastic* regime, that is, if $\epsilon_{m,l}([A]; n) > 1$,
- will not be modified in the *unit elastic* regime, that is, if $\epsilon_{m,l}([A]; n) = 1$,
- will increase in the *inelastic* regime, that is if $\epsilon_{m,l}([A]; n) < 1$.

We can interpret this result in terms of a loss of competitive advantage between the transcription factors. Binding cooperativity mechanisms between TFs are the clearest example of this competitive advantage. If TFs cooperate in their binding process this cooperation constitutes an advantage for such TFs, and this advantage is clearly amplified in the presence of

- high affinity enhancers, because the first binding required for cooperativity is more likely to occur,
- a high number of enhancers, because they allow improvement of TF's affinity by cooperativity.

So, perturbations (3.47) and (3.48) are clearly limiting the advantage given by cooperativity. This is more obvious in the case of (3.48) since in the limit case, $n = 1$, cooperativity can not operate at all. As we will see in next section, binding cooperativity is not the only advantage we can detect by means of the elasticity.

The inelastic case is interpreted as a situation where repressors "lose their advantage" over the activators, since a global decrease in affinity or number of enhancers gives rise to an increase in the threshold, as seen in (3.50), and consequently an increase of the activation region (see Fig. 3.2). Remembering that, given a fixed activator concentration, an increase of the threshold function implies that an increase of the repressor concentration is needed in order to get the same transcription rate as the basal. That is to say, the cells will enter in the activated state (i.e., transcription rates higher than the basal) "more easily" in the new situation because the repressors' advantage is not as effective as before. The same goes for the elastic case. In this case, the activators seem to lose the benefit of the advantage. Here (3.50) shows that the thresholds decrease, so the activation region decreases, because of perturbations (3.47) and (3.48). The unit elastic regime can be seen as a stable situation, where none of the perturbations modify the threshold (i.e., both activators and repressors have the same transcriptional advantage and perturbations affect both in the same way). Please note also that the same reasoning can be applied to proportional decrease in activators and repressors concentrations because of definition (3.49). In Fig. 3.4 these qualitative behaviours have been illustrated using a stimulated BEWARE operator where three sets of parameters have been chosen to represent all these situations.

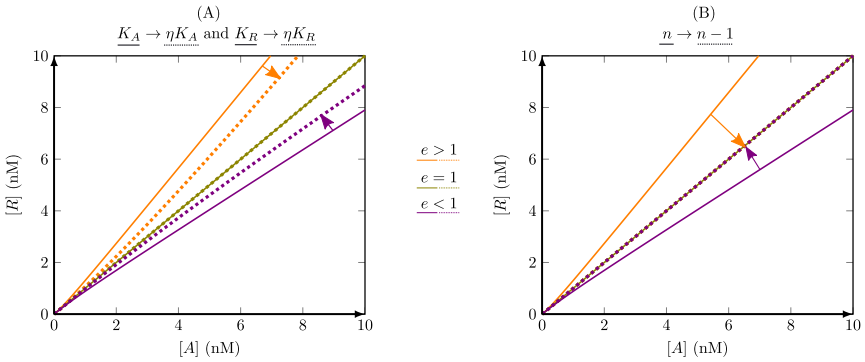

Figure 3.4: **Activation-repression threshold variations under TFs signalling interferences** (3.47) **and** (3.48). Thresholds determined by a stimulated BEWARE operator exhibiting different elasticities depending on the values of the model parameters. Orange, green and purple continuous lines are used to represent thresholds with elasticity greater than, equal to or less than one respectively. In **(A)** perturbation (3.47) has been applied with $\eta = 10$. In **(B)** perturbation (3.48) is where the number of enhancers changes from $n = 2$ to $n = 1$. The thresholds after perturbations are depicted in both cases as dotted lines. Orange and purple arrows show the threshold variations in the elastic and inelastic cases, since in the unit elastic case (green threshold) there is no change, in accordance with eq. (3.50). Parameter values can be found in Table B.4 in Appendix B, and the estimation of the elasticities for these values are in Table 3.2.

With the threshold definition (3.43), obviously $f_{m,l}([A]=0;n)=0$ holds true. So, the elasticity coefficient has a very simple geometrical interpretation: the comparison between the slope of the tangent line at the point $\big([A], f_{m,l}([A];n)\big)$ and the slope of the secant line intersecting the threshold curve at the origin $(0, f_{m,l}(0;n))$ and at the point $([A], f_{m,l}([A];n))$. That is,

$$\epsilon_{m,l}([A];n)>1 \iff f'_{m,l}([A];n) > \frac{f_{m,l}([A];n)-f_{m,l}(0;n)}{[A]-0}\,.$$

This expression also tells us that if $f_{m,l}$ is convex then it is elastic. In the same way, the concavity of $f_{m,l}$ implies being inelastic. Nevertheless, as we can see in Fig 3.4, this geometrical interpretation of ϵ can not always be easily recognised at first glance.

Although Hill models have been deduced in this work from the stimulated and recruitment BEWARE operators by *extreme cooperativity hypothesis*, these types of models do not inherit relationships (3.50). So this conceptual streamlining loses the original relationship between perturbations (3.47) and (3.48) and the elasticity function. Indeed, in Appendix A.6 it is proven that Hill threshold functions exist but they are always straight lines whose slope can change with the number of enhancers.

Lastly it is very important to remark that, knowing the behavior of the threshold under perturbations (3.47) and (3.48), we automatically can deduce the behaviour of the Net Activated Cells regions. I.e., if under (3.47) and (3.48) the threshold

- is elastic, then the NAC region decreases,
- is unit elastic, then the NAC region does not change,
- is inelastic, then the NAC region increases.

Which can be easily checked taking into account the relations (3.50) and (3.46). With this we have fulfilled our main goal: to connect fruitfully the molecular information that comes from the threshold function, with the tissular information described at celular level.

3.4.4 Recruitment, Stimulated and Hill BEWARE operators: Threshold comparative analysis

The end of this section is devoted to comparing the effects of the perturbations (3.47) and (3.48) on the activation/repression regions in terms of the thresholds predicted by the different BEWARE operators considered. This study shows the existence of two main factors that are seen to determine the model's response: the first one is the binding cooperativities involved in the TF binding process and the second is how the transcriptional effects of activators and repressors are modelled.

In order to understand the cooperativity effects between TFs in their binding process we have divided the partial cooperative case in two extreme scenarios: *activator cooperative* ($c_A > 1$, $c_R = 1$) and *repressor cooperative* ($c_A = 1$, $c_R > 1$) cases. Then, we have included the *null* and *total cooperative* case in the same scenario, since the only difference between the operators is the value of c (i.e., $c = 1$ for the null cooperative case and $c > 1$ for the total cooperative case). This allows us to identify more easily cases where the activators or repressors should lose "competitive advantages" or not. Particular cases with partial cooperativity between activators and repressors should be estimated numerically.

We analysed stimulated and recruitment BEWARE operators with each kind of cooperativity, obtaining interesting (but unexpected) results which are summarised in Table 3.2 (detailed proofs can be found in A.5.1 and A.5.2).

Table 3.2: Analytical estimations of elasticity for thresholds deduced from the BEWARE operators in the global activator/repressor framework. The values t_1, h_1, t_2, h_2 appearing in the case of the stimulated operator with two enhancers are defined in terms of the rest of the model parameters (see A.5.2 in Supplementary Material, for explicit definitions).

	Act coop.	null/total coop.	Rep coop.
Recr.	$\epsilon > 1$	$\epsilon = 1$	$\epsilon < 1$
Stim. (n=2)	$\epsilon < 1$ if $\tilde{e}>t_2$	$\epsilon < 1$ if $\tilde{e}>t_1$	$\epsilon < 1$ if $\tilde{e}>c_R t_1$ & $[A]>h_1$
	$\epsilon \leq 1$ if $\tilde{e}<t_2$ & $[A]\leq h_2$	$\epsilon = 1$ if $\tilde{e}=t_1$	$\epsilon \geq 1$ if $\tilde{e}>c_R t_1$ & $[A]\leq h_1$
	$\epsilon > 1$ if $\tilde{e}<t_2$ & $[A]>h_2$	$\epsilon > 1$ if $\tilde{e}<t_1$	$\epsilon > 1$ if $\tilde{e}\leq c_R t_1$
Hill	$\epsilon = 1$	$\epsilon = 1$	$\epsilon = 1$

From the modelling point of view (and following the intuition), the expected behaviour in relation to perturbations (3.47) and (3.48) should be:

- *Null/Total cooperative case*: The activators and repressors lose no competitive advantage in the null case or lose exactly the same amount of competitive advantage in the total cooperativity case. Hence the threshold function should not vary (unit elastic case).
- *Activator cooperative case*: The activators lose that competitive advantage over the repressors. Hence the threshold function should decrease (elastic case).
- *Repressor cooperative case*: The repressors now lose the competitive advantage over the activators. Hence the threshold function should increase (inelastic case).

Estimations exhibited in Table 3.2 show that the Recruitment BEWARE operator behaves as expected, that is, the elasticity of the thresholds deter-

mined by these operators are proven to be $\epsilon_{r,p} > 1$ in the Activator Cooperative case, $\epsilon_{r,p} < 1$ in the Repressor Cooperative case and $\epsilon_{r,t} = 1$ in the case of null/total cooperativity. In the case of the Stimulated BEWARE operators, the analysis of the elasticity variable (at $n = 2$) is more entangled. Here, the elasticity value is related not only to the cooperativities but also to other parameters that determine each micro-state transcription rate. In fact, we can observe that regardless of the (binding) cooperativity considered, we can get elastic, inelastic or unit-elastic situations depending on certain relationships between the parameters involved in the modelling. Let us explain this conclusion revealed by elasticity. Remembering that the basal transcriptional level, corresponding to the transcriptional rate of an empty micro-state, in this approach is $tr_{\emptyset,\emptyset} = r_{bas}$. Here $\emptyset$ represents an enhancer with non bound TFs. The transcriptional rates associated to micro-states with a single bound activator/repressor are $tr_{A,\emptyset} = r_{bas} + \tilde{e}\nu^{(2)}_{max} = r_{bas} + \nu^{(1)}_{max}$ and $tr_{R,\emptyset} = \tilde{r} r_{bas}$ respectively. If the linkage of a second TF of the same family occurs, then the new transcription rates become $tr_{A,A} = r_{bas} + \nu^{(2)}_{max} = r_{bas} + \nu^{(1)}_{max}/\tilde{e}$ and $tr_{R,R} = \tilde{r}^2 r_{bas}$.

Let us observe that the ratio in the variation of the transcriptional rate due to the binding of a repressor is constant regardless of the existence of other already bound repressor, that is,

$$\frac{tr_{R,\emptyset}}{tr_{\emptyset,\emptyset}} = \frac{tr_{R,\emptyset}}{tr_{R,R}} = \tilde{r}\,.$$

However, in the case of activators the analogous rates depend on the values of $\tilde{e}$, $\nu^{(1)}_{max}$ and r_{bas}

$$\frac{tr_{A,\emptyset}}{tr_{\emptyset,\emptyset}} = 1 + \frac{\nu^{(1)}_{max}}{r_{bas}} \quad \text{and} \quad \frac{tr_{A,\emptyset}}{tr_{A,A}} = \frac{r_{bas} + \frac{\nu^{(1)}_{max}}{\tilde{e}}}{r_{bas} + \nu^{(1)}_{max}}$$

and can vary from the first to the second binding protein. Then, the existence of the second binding site implies

1. a transcriptional advantage for repressors if $\frac{tr_{A,\emptyset}}{tr_{\emptyset,\emptyset}} > \frac{tr_{A,\emptyset}}{tr_{A,A}}$, because in that case comparatively the binding of a second activator is less effective than the binding of a second repressor,
2. a transcriptional advantage for activators if $\frac{tr_{A,\emptyset}}{tr_{\emptyset,\emptyset}} < \frac{tr_{A,\emptyset}}{tr_{A,A}}$, because in that case comparatively the binding of a second activator is more effective than the binding of a second repressor,
3. no advantage for any TF when $\frac{tr_{A,\emptyset}}{tr_{\emptyset,\emptyset}} = \frac{tr_{A,\emptyset}}{tr_{A,A}}$ because in that case the binding of a second TF is equally as effective as the first bound TF.

It is easy to check that (i), (ii) and (iii) directly correspond to the elastic, inelastic or unit-elastic cases under null/total cooperativity determined in

Table 3.2. That is, in the absence of binding cooperativity, the advantage that elasticity demonstrates is related to the possibility that the functioning of two adjacent activators can be more/less effective together than separately. Indeed, this "functional (anti-)cooperativity" mechanism is not new in literature, it was already proposed in [63]. However, with the elasticity we are able to analyse some specific cases of the models proposed in [63] and the values that the elasticity function takes reveal inconsistencies with the modelling guidelines proposed in that work (In Table B.4, we introduced some identifications/choices that can be made in order to fit the model proposed in [63] into the expression of a stimulated BEWARE operator). The last row of Table 3.2, that is, the estimates of the elasticity corresponding to the thresholds deduced from stimulated BEWARE operators with two enhancers ($n = 2$). This analysis shows that there are acceptable parameter values for the model that give us inelastic threshold functions that are fully compatible with the non existence of "functional cooperativity" ($\varepsilon_A = \varepsilon_R = 1$ according to notation in [63]). In Fig. 3.4, the inelastic threshold function (continuous purple line) has been obtained using some of these parameters (see values in Table B.4). In the same figure we can see how perturbed thresholds vary according to the elasticity being less than one. On the other hand, this model was deduced without considering binding or functional cooperativity.

So, our analysis shows that the model in [63], deduced from non "functional" and non "binding" affinity assumptions, can have an inelastic threshold. And it reacts as it has been proven to do under perturbations (3.47) and (3.48). This can be seen in Fig. 3.4. Our argument supports the notion that the asymmetric approach that has been adopted in stimulated operators for modelling the activators and repressors functionalities is causing this effect. The elasticity can also be estimated in the presence of partial binding cooperativity and we have seen that in both cases elasticity is able to balance the effects of both cooperativities.

This unexpected effect needs to be taken into account in analysis such as those developed in [18] because, if used, it can seriously alter the conclusions. This example is one reason why we were interested in performing sensitivity analysis in this work.

Regarding the Hill versions of the Recruitment and Stimulated operators, we can say that $\epsilon_h = 1$ for any kind of cooperativity and operator. However, it is important to note that identities in (3.50) are not valid in this framework. That is, the extreme cooperativity approach used to get the Hill type models is incompatible with the information that elasticity provides.

3.5 Drosophila aplication

Along the present tesis we have been working with Hedgehog (Hh). In *Drosophila*'s wing imaginal disc the secretion of Hh from the Posterior com-

partment cells induces the expression of several target genes inside the cells in the Anterior compartment. Among them are *decapentaplegic* (*dpp*) and *patched* (*ptc*). Both give rise to the synthesis of their corresponding proteins, Dpp and Ptc, which are essential for the wing central domain development [102, 105].In the embryonic ectoderm Hh also regulates *wingless* (*wg*) and *stripe* (*sr*) genes. However, it is known that the same signal of Hh produces different spatial expression of theses genes. That is to say, the expression of *ptc* is only limited to disc zones close to the *Anterior/Posterior* (A/P) border with high Hh concentrations, while *dpp* expresses in a broader disc range under low Hh concentrations (figure 3.5). This poses a question: Why does the same signal give rise to different spatial expressions for different genes? The answer to this question is still under debate. The current understanding is that both genes respond, basically, to the same principles that we list below.

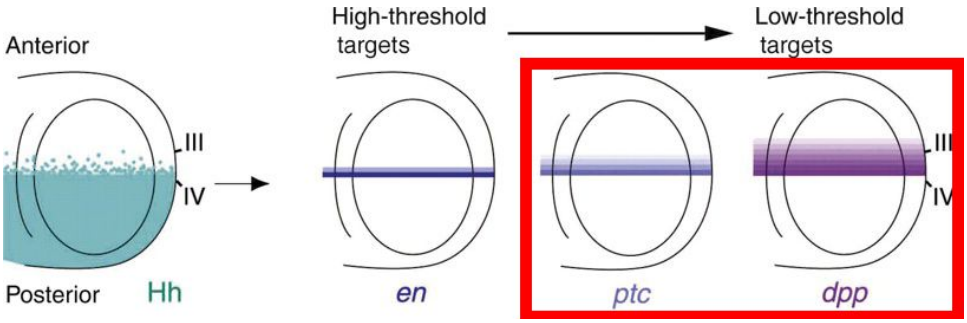

Figure 3.5: Spatial expression of the genes *patched* (*ptc*), *decapentaplegic* (*dpp*) and *engrailed* (*en*, not discussed in this thesis) in terms of the gradient of Hedgehog (Hh). Even under the same concentration of morphogen, *ptc* is expressed in a shorter region than *dpp*, depicted as High-Low (Hh) threshold target genes. Image obtained from Tabata's work [102].

Hh transcriptionally controls both Dpp and Ptc through the transcription factors Cubitus interruptus (Ci). As we already know, the transcription rate of the target genes is controlled by the TFs, in this case Cubitus Interruptus. Ci is present in two opposite forms: activator and repressor. The activators, CiA, attempt to promote the transcription rate while the repressors, CiR, attempt to decrease it. Please recall that Hh is secreted from the posterior into the anterior compartment (at around 60% of the DV axis). In the absence of Hh Ci appears in its repressed form but when the signal of Hh is absorbed by the cell, Cubitus changes it role presenting its activator form. So, the Hh gradient in the Anterior compartment creates opposing activator (CiA) and repressor (CiR) gradients of the wing imaginal disc. This means that we can model Cubitus as and opposing TF gradient, following what we have seen in section 3.4.2. This is, $[CiA](x)$ is a decreasing function with

the distance x from the A/P boundary. On the other hand, the experiments also shows that the total amount of Ci

$$h = [CiR](x) + [CiA](x)\,, \tag{3.51}$$

is constant. Hence, the concentration of CiR must increase with the distance from the A/P border. Following [83], we can model these concentrations with an exponential function

$$[CiA](x) = he^{-x/\sqrt{D}}\,, \quad [CiR](x) = h - [CiA](x)\,, \tag{3.52}$$

being h the TFs total concentration and D is the steepness of the gradient (see subplot in Fig. 3.6B).

Some recent works [83, 116, 91, 2] postulate that the reason for the proper spatial expression of these genes could be found in certain biochemical factors involved in the transcription process. Firstly, the binding of both PolIII and Ci in the promoter and enhancer is carried out by chemical reactions. As we have discussed in section 3.1, these require some free energy that is commonly characterised by the binding affinity. This affinity depends on several characteristics of the promoters and enhancers of each transcribed gene. In fact, in [91] it was observed that the enhancers with lower relative affinity seem to be necessary to obtain normal expression of *dpp* in regions of low signal. Secondly, it is possible that transcription factors that are already bound in some enhancers can modify the affinity of other binding elements. In this case, bound TFs may modify the free energy of a later binding reaction of either TF or RNAP via cooperativity.

The combination of all these biochemical factors (competition, cooperativity and binding affinities) gives rise to a very complex balance between the concentration of activators and repressors making it difficult to discern their interacting effects at tissular level. In [83, 91], the spatial expression of some of the Hh target genes was related to the respective binding affinity between Cubitus proteins and Hh/Ci module enhancers. The relative in vitro affinities of Ci sites in the *ptc* and *dpp* enhancers have been measured by electrophoretic mobility shift assays (see [83]). In [116], the Ci binding affinity of four Ci sites in the *wg* embryonic ectoderm enhancer was measured by using the same methodology. In [91], they predicted the binding affinity for Ci in the *Drosophila* genome for which Hh/Ci-regulated enhancers have been functionally characterised. It has been observed that *ptc* is activated by Hh/Ci in larval imaginal discs via a module with high-affinity Ci sites, by contrast, with the relative low-affinity of *dpp*, *wg* and *sr* enhancers located in their corresponding Hh/Ci modules. The experiments developed in [83] confirm that, under moderate Hh signal, the wild type low-affinity sites in dppD produce activation, whereas if they are substituted by high-affinity sites produce repression. Similar results were obtained in embryonic enhancers of *wg* and *sr* in [91, 116].

The large amount of biochemical variables that are present in the system calls for mathematical models that can shed some light on the origins of the differential spatial expression in the target genes of Hh, among others. In the next section we will use our new BEWARE operator expressions in order to have a better understanding of the transcriptional logic of target genes controlled by a Hh/Ci module of enhancers.

3.5.1 The case of Ptc and Dpp: Cooperativity between repressors

Since we have one gradient of activator (CiA) and one repressor (CiR), we can discuss in terms of the Global Activator/Repressor variables without loss of generality. Moreover, we can take advantage of this framework and use the threshold function in order to analyze the variations of the Net Activated Cells (NAC) regions introduced in section 3.4.1. We have discussed how the elasticity analysis of the threshold function could be used in order to deduce variations in the spatial expression of a gene. Hence we can use the same reasoning in this particular case for the expression of *ptc* and *dpp*. By the experiments, it is known that *ptc* shows a NAC region shorter than *dpp*. Since we have shown that the spatial expression variation is related to the threshold variations, the idea that we propose here is to analyze how the thresholds of *ptc* and *dpp* change with experimental perturbations, in a Recruitment BEWARE operator. In particular, the experimental evidences that motivate our analysis are mainly related with the affinity and the number n of binding sites. By electrophoretic mobility shift assays Parker and coauthors found in [83] that Ci binding sites in the *ptc* enhancer have considerably higher affinity than *dpp* sites. The same authors constructed transgenic fly lines that allow them to compare the transcriptional activity of reporter genes containing different variants of these sites modifying their affinity. Please recall that these modifications are the experimental realization of the theoretical perturbations (3.47) and (3.48) that we defined in section 3.4.3. Hence, we can use the results described in such section to deduce theoretically effects in the system that can be supported with the experimental evidences. Using the notation from [83, 91], we will work with different versions of the *dpp* enhancer, containing:

- n Low-affinity sites *dpp*D-nxCi^{WT}, with $n = 3, 1$, similar to the WT conditions of *dpp* enhancer.
- n High-affinity sites (*dpp*D-nxCi^{ptc}), with $n = 3, 1$, similar to the WT conditions of *ptc* enhancer and
- three null-affinity sites (*dpp*D-3xCi^{KO}).

The results obtained in their work can be resumed in the following three

experiments:

Exp. 0: *Measurement of the Basal transcription*
The reporter gene with null-affinity sites provided the basal expression, which by definition in the previous sections, can be used to define the NAC regions. Indeed, in [83, 91] the effects of Ci signalling with low- or high-affinity enhancers was measured comparing the gene activity *versus* the basal in any cell. Cells expressing a gene with higher expression rates than the basal level are called net activated cells, constituting the NAC range. Fig 3.6 shows how this range is determined by using our thermodynamic model in the same way as was done from measurements in [83].

Exp. 1: *Transcriptional effects of the reduction in binding sites*
In *dpp*D-1xCi^{ptc}, the range of net activated cells resulted wider than the range of net activated cells for *dpp*D-3xCi^{ptc}.

Essentially, the reduction from $n = 3$ to $n = 1$ binding sites implies the vanishing of any possible kind of cooperativity between Cubitus. If both activator and repressor interact with of total cooperativity, the threshold function for the Recruitment operator should be unit elastic. In such case, the interactions should be symmetric for CiA and CiR, and hence their disappearance should not modify the balance between net activated or repressed cellular ranges. However, the result obtained in the first experiment of Parker and coauthors differs from this. The NAC range changes with the reduction of the number of binding sites, which should imply some kind of asymmetry in the interactions. This could apply if partial cooperativity is present, either only between CiA or CiR. In this case the reduction of sites would provoke a modification in the threshold, due to its (in)elasticity. This is, when the cooperativity between activators is removed the NAC range should decrease (elastic) and increase when the repressor cooperativity is abolished (inelastic). In Fig 3.7 the reader can find the graphical representation of this reasoning for the first experiment.

Since the NAC region for CiA-CiR decreases with the number of enhancers, the theory deduces that there should be an asymmetric cooperativity that enhances more interactions between CiR than CiA. Moreover, by the relation (3.50), our model deduces that this behaviour should be also reproduced if perturbation (3.47) takes place. This can be corroborated with the second experiment of Parker et.al.

Exp. 2: *Differential affinity effects*
In *dpp*D-3xCi^{WT} the range of net activated cells resulted wider than the range of net activated cells for *dpp*D-3xCi^{ptc}.

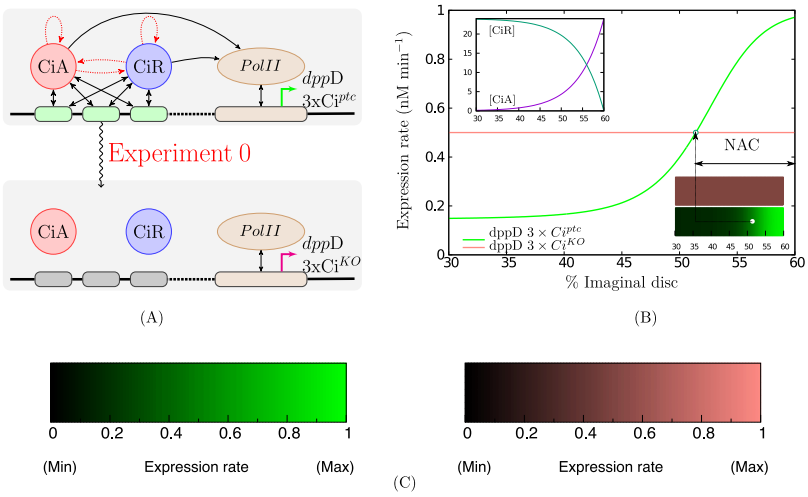

Figure 3.6: **Net activated cellular (NAC) range. A)** Schematic of the experiment for NAC range determination. The arrows represent all the possible interactions captured by a thermodynamic model determining the transcription rates: double-headed straight arrows show protein-DNA binding site affinities while single-headed black and red arrows are TFs-RNAP and TFs-TFs cooperativities respectively. The net activated cellular range of *dpp*D3xCi^{ptc}, a reporter gene with a version of the *dpp* enhancer with three high-affinity binding sites, is obtained by comparing its theoretical transcriptional activity with the activity of *dpp*D3xCi^{KO}, a gen containing different version of the *dpp* enhancer containing three null-affinity sites. Both cases are represented in the upper and lower schemes respectively. TFs binding sites are represented by rounded rectangles filled in green (high-affinity) or black (null affinity). **B)** Theoretical transcription rates predicted for both genes in cells of the Anterior compartment. This compartment occupies the 60% of the *Drosophila* imaginal disc and the Posterior compartment the rest (60% to 100%). The expression levels given by the BEWARE operators are between $0nM/min$ and $1nM/min$ being the basal level equal to $0.5nM/min$. These reference expression levels have been chosen for a proper appreciation of signal modulation. Since *dpp*D3xCi^{KO} has been modelled independent of external factors it is expressed at basal level anywhere. Cells expressing *dpp*D3xCi^{ptc} more than the basal level are in the NAC range. The expression of both genes in the wing imaginal disc is also indicated by using coloured bars. The blue circle inside the bar, indicate the position of a cell expressing *dpp*D3xCi^{ptc} at the basal level. The color scale used in these bars is shown in **C)** black meaning no expression ($0nM/min$), and full color meaning high expression ($1nM/min$). The inset in **B)** depicts the activator/repressors (CiA/CiR) gradients generated by Hh signalling: activator concentrations are higher close to the Anterior/Posterior border. Model parameters used in this plot can be found in Table B.5 of Appendix B. Image obtained from [18].

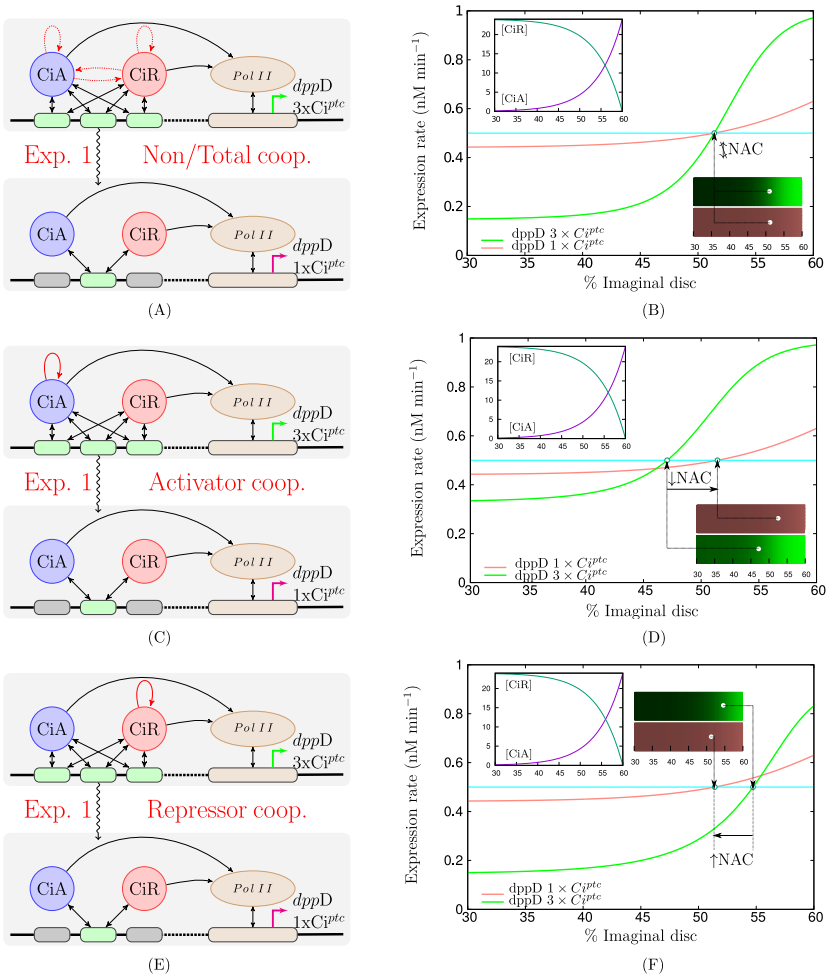

Figure 3.7: **Transcriptional responses to the experiment 1.** First column: schematic of the experiment 1: comparison of the expression ranges of reporter genes with 3 high-affinity sites (*dpp*D-3xCi^{ptc}) and a single high-affinity Ci site (*dpp*D-1xCi^{ptc}). **A)** corresponds to the non/total cooperativity case where, if cooperativity holds, all the TFs cooperate between them, **C)** to the activators cooperativity case, only activators cooperate, and finally **E)** to the repressors cooperativity case where only repressors cooperate. Second column shows the different transcriptional responses that can be theoretically described depending on the case of cooperativity considered. The schemes and plots employ the same keys explained in Fig 3.6, and perturbed parameters in Table B.6. Image obtained from [18].

On the one hand, this result was expected from the biology itself. After all, it is known that the WT expression of *ptc* is shorter than the WT expression of *dpp*. However, on the other hand, this result is tremendously interesting from the mathematical modeling point of view: It provides a direct experimental proof of the property (3.50), which we obtained theoretically and directly from the BEWARE model. Moreover, with this result we can deduce that the repressor cooperative model is the only in concordance with the results observed for *dpp*D-3xCi^{ptc} and *dpp*D-3xCi^{WT} in [83] (see figure 3.8).

Lastly, it is important to remark that the theory is also able to deduce the order (in size) of the NACs in each reporter gene. Let us consider partial cooperativity between CiR, and the thresholds functions $f_{r,p}^{1\times Ci^{Ptc}}$, $f_{r,p}^{1\times Ci^{WT}}$,$f_{r,p}^{3\times Ci^{WT}}$, $f_{r,p}^{3\times Ci^{Ptc}}$ for each reporter gene. Then, taking into account the relation (3.50) and the fact that the threshold in the repressor partial cooperative case is inelastic, we automatically get that

$$f_{r,p}^{1\times Ci^{Ptc}} > f_{r,p}^{1\times Ci^{WT}} > f_{r,p}^{3\times Ci^{WT}} > f_{r,p}^{3\times Ci^{Ptc}}$$

Which means that, by (3.46),

$$\text{NAC}^{1\times Ci^{Ptc}} > \text{NAC}^{1\times Ci^{WT}} > \text{NAC}^{3\times Ci^{WT}} > \text{NAC}^{3\times Ci^{Ptc}}$$

This is, the NAC range for *dpp*D-1xCi^{ptc} should contain the NAC range for *dpp*D-3xCi^{WT}, and the latest should contain the NAC range for *dpp*D-3xCi^{ptc}. This has also been corroborated experimentally, where in [83] it can be observed that the NAC ranges for *dpp*D-1xCi^{ptc}, *dpp*D-Ci^{WT} and *dpp*D-3xCi^{ptc} occupy from the 43%, 49% and 54% of the disc width, respectively, to the A/P border (which is around to the 60% of the disc width).
With this we get to the end of this chapter, and the main contents of the thesis. In next chapter we will discuss some on-going works, as well as possible modifications that can improve the models presented in previous chapters.

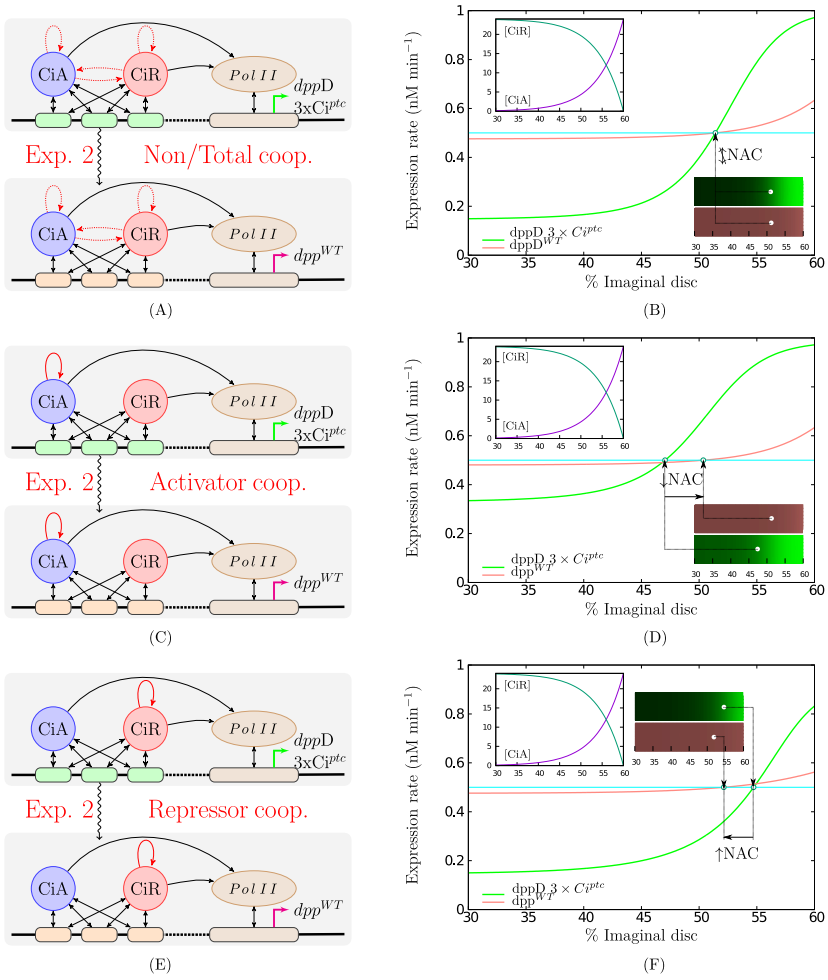

Figure 3.8: **Transcriptional responses to the experiment 2.** Figures in the first column: schematic of the experiment 2: comparison of the expression ranges of reporter genes with 3 high-affinity sites *dpp*D-3xCi^{ptc} or 3 low-affinity site *dpp*D-3xCi^{WT}. TFs-DNA binding site affinities are indicated by thicker or thinner double-headed straight arrows. **A)** correspond to the non or total cooperativity case where, if cooperativity holds, all the TFs cooperate between them, **C)** to the activators cooperativity case, only activators cooperate, and finally **E)** to the repressors cooperativity case where only repressors cooperate. Figures in the second column shows the transcriptional responses that can be theoretically described depending on the case of cooperativity considered. The schemes and plots employ the same keys explained in Fig 3.6, and perturbed parameters in Table B.6. Image obtained from [18].

Chapter 4

Results, conclusions and future work

Along chapters 2 (Cytonemes) and 3 (Transcription dynamics), we have introduced and developed different mathematical models in different biological scales. The application of these models have been tested in each chapter, using experimental evidences in different stages of *Drosophila*'s fly development. The results reveal that each model is able to, at some extent, capture the complexity of the systems. More specifically,

- **In Chapter 2:** we saw that a glypican-concentration based potential theory can deduce cytonemes orientations in different tissues. The results have been tested in *Drosophila*'s imaginal disc, in both Wild-Type and overexpression conditions, obtaining cytoneme trajectories that resemble the experimental ones.
- **In Chapter 3:** we saw that, using the correct hypotheses and general notation, the thermostatistical approach (i.e., BEWARE operators) can be reformulated as polynomical expressions. These allow novel mathematical and sensitivity analysis, that are able to connect different biochemical properties with elements that are defined at higher scales. Among these, we have been able to obtain the Threshold function. This is a strong tool that relates the tissue spatial information of an expressed protein, with cellular biochemical properties such as TF-TF;TF-PolII cooperativity, binding affinity and enhancers. The results have been tested in *Drosophila*'s Hh pathway, in different target genes (*ptc* and *dpp*) that are controlled by the same concentration of transcription factors (Ci). The model is able to predict what is the biochemical framework that better fits the experimental evidences. Moreover, the developed tools can be applied to perform straightforward sensitivity analysis and fitting of the model parameters.

However, it is important to remark that these model are in their initial states,

subjected to changes and possible improvements:

- **Cytonemes model:** In Chapter 2 we have deduced the trajectories of the cytonemes in terms of the vector field $\vec{O}(\vec{r}, t)$. However, note that we are not still saying anything about the effect that these trajectories have over the flux of morphogen. Recall that, in the case of *Drosophila*, Hh is produced by cells in the Posterior compartment. Then, it is traded via cytonemes to the cells in the Anterior compartment. It is in these cytoneme-cytoneme contacts (in biology terms, synaptic buttons), where the vesicles of Hh are interchanged. Hence, in order to describe mathematically how the gradient of Hh is formed in the imaginal disc, we need to determine an equation that describes the vesicles flux along the cytonemes, and the cytoneme-cytoneme contacts (some works has already proposed brand new mathematical models in this framework, such as in [66]). Please note that the vesicles flux should be described over the curve $\vec{\gamma}$ described in (2.25). Hence, the flux equations may be described in a dynamic 1-manifold, making the problem highly non-linear. This is a work that is still on process, in collaboration with Isabel Guerrero's laboratory (CBM, Madrid).

- **BEWARE operator model:** In Chapter 3 we have deduced the BEWARE operator in general for any number M of transcription factors and n binding sites. Remarkably, we have also deduced possible hypothesis that makes possible the reduction to an $M = 2$ system (what we called the Global Activator/Repressor variables, in Section 3.3). In this framework, we have deduced the threshold function, which has proven to be an essential tool for most of our analysis. It is important to recall that the threshold function is defined as the solution of the implicit equation (3.43). We have shown that the threshold exists, and it is unique in the case of $M = 2$. This defines a curve (i.e., 1-manifold) in the $[A]-[R]$ plane. The question now arises: is it possible to prove the existence of threshold for any number M of transcription factors? And if that is the case, what is the behavior of the $(M-1)$-manifold threshold in the space of $\vec{T}$? Answering these questions, we could repeat the questions that we did in the Global Activator/Repressor framework: relating the spatial expression of any gen with the biochemical properties for any number of TFs, obtaining sensitivity annalysis for any parameter, etc. This is a model extension that may me tackled in the future.

4.1 Future work: Burst dynamics

In this last section we are going to introduce briefly some of the current work developed during the research stay done by the author of the thesis.

The stay took place in the Institute of Quantitative Biosciences (University of Tokyo), at Dr. Takashi Fukaya's laboratory. There, the research team focused on the annalysis of the transcription dynamics at cellular level.

Recent advances in microscopic visualization technology have made possible to analyze events occurring in gene transcription with high precision and temporal resolution [41]. Let us remember that in Chapter 3 we already discussed the temporal relationship that existed between the evolution of the transcription to which a certain gene was copied with its protein product. The time scale in which proteins are produced is much longer than the scale in which the various events occur in DNA to result in transcription [62, 90]. Let us remember that these events are related to the anchoring / undocking of both RNA polymerase and the transcription factors in the promoter and number of enhancer binding sites. These, in balance, give rise to an average protein synthesis level that depends on both the position in the tissue (regulated by the concentrations of transcription factors), and the biochemical properties themselves in the cells that populate said region. However, recently the scientific community is making great efforts to follow this process on a much smaller scale, focusing on analyzing the problem in a way that is much more focused on the discrete transcription itself. Indeed, as shown in Fig. 4.1, the experiments reveal punctuated peaks and troughs in the number of active RNAP molecules (and in consequence, the number of copies of mRNA). These features have been related to the rate of RNAP

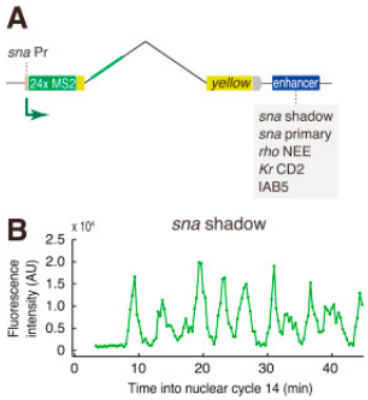

Figure 4.1: Transcription bursts in Drosophila embryo, in the gene *snail*. **A)** Scheme of *sna* promoter and enhancer (shadow enhancer in the pic). **B)** Transcription bursts of *sna*. The y-axis represents the number of transcripts in fluorescence intensity (see [38] for details of the experimental setup). Image also obtained from [38].

initiation is 'burst-like' [107], with the promoter rapidly loading multiple RNAP molecules onto the gene at a constant rate during discrete 'bursts' of

activity [13]. This and other evidence from live imaging [13, 39, 23], as well as data from fixed-tissue approaches [124, 82, 73, 118], support the idea that mathematical models should take into account the stochastic events of PolII binding in the promoter. In particular, recently the scientific community proposes what is called a 2-state Continuous-Time Markov Model (CTMM) of promoter switching [84, 119, 64].

4.1.1 Two-state model

In this model, promoters switch stochastically between ON and OFF states with rates k_{ON} and k_{OFF}. This is, k_{ON} and k_{OFF} are the rate parameters of two exponential random variables, each one representing the time between each OFF-ON and ON-OFF state jump such that

$$P_t(t; OFF \to ON) = k_{ON} e^{-k_{ON} t} \tag{4.1}$$

is the probability distribution of the time between OFF-ON events and

$$P_t(t; ON \to OFF) = k_{OFF} e^{-k_{OFF} t} \tag{4.2}$$

is the probability distribution of the time between ON-OFF events. Promoters in the ON state engage in the loading of PolII (and, correspondingly, mRNA production) at rate r [106, 42] (see Fig. 4.2)[1]. It is important to note that this process happens inside of any cell. As we saw in the previous chapter, the tissular position of the cells is important for the synthesis of the proteins, which was regulated by the spatial morphogen gradient and concentration of transcription factors. Hence, normally the ON/OFF ratios depend on the position $\vec{r}$ that the cell occupies in the tissue (i.e., $k_{ON} = k_{ON}(\vec{r})$ and $k_{OFF} = k_{OFF}(\vec{r})$). A two-state model is defined by a master equation of two

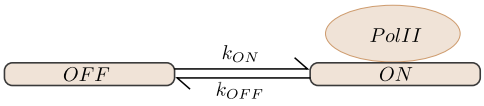

Figure 4.2: Two-state model of bursting of a single promoter. The promoter switches between the states OFF (PolII unbound) and ON (PolII bound) with rates k_{ON} and k_{OFF}, respectively.

ordinary differential equations:

$$\begin{cases} \frac{d}{dt} p(OFF; \vec{r}, t) = -k_{ON}(\vec{r}) p(OFF; \vec{r}, t) + k_{OFF}(\vec{r}) p(ON; \vec{r}, t) \,, \\ \frac{d}{dt} p(ON; \vec{r}, t) = k_{ON}(\vec{r}) p(OFF; \vec{r}, t) - k_{OFF}(\vec{r}) p(ON; \vec{r}, t) \,, \end{cases} \tag{4.3}$$

[1] This is just one (and the simplest) way of developing a two-state model. Please refer to [84] for a complementary discussion.

being $p(OFF;\vec{r},t)$ and $p(ON;\vec{r},t)$ the probability for the promoter of being in the OFF and ON states, inside of a cell positioned at position $\vec{r}$ in the development tissue and time t, respectively. The system (4.3) can be also rewritten in the matricial form

$$\frac{d}{dt}\vec{P}(\vec{r},t) = \vec{P}(\vec{r},t)Q(\vec{r}) \tag{4.4}$$

with $\vec{P}(\vec{r},t) = (p(OFF;\vec{r},t), p(ON;\vec{r},t))$ and

$$Q(\vec{r}) = \begin{pmatrix} -k_{ON}(\vec{r}) & k_{ON}(\vec{r}) \\ k_{OFF}(\vec{r}) & -k_{OFF}(\vec{r}) \end{pmatrix} \tag{4.5}$$

is called the Q-Matrix of the two-state CTMC, for each cell at position $\vec{r}$. Additionally, we could compute the approximate variation of the copies of produced mRNA as

$$\frac{d}{dt}[mRNA](\vec{r},t) = r\frac{d}{dt}p(ON;\vec{r},t)\,, \tag{4.6}$$

where for the sake of simplicity we are not taking into account the mRNA degradation term. Please also recall that this step is closely related to the Recruitment hypothesis that we took in Section 3.1.3. There, we assumed that the production of protein P was proportional to the number of copied mRNA, which at the same time was proportional to the probability of finding PolII in the promoter. The main difference here is that those probabilities were considered stationary (see Hypothesis H1 in Section 3.1.1), since the time scales for the evolution of protein P were bigger than the time scale of the binding events. Now we are trying to understand the transcriptional bursting dynamics, starting from a smaller time scale where the events are governed by the promoter states. This is what motivates the use of stochastic models, as for instance the two-state CTMM. However, as a matter of fact, normally (4.6) is not used. The main reason is that the study of discrete burst events can be estimated more straightforwardly with the parameters of the two-state model.

4.1.2 Burst properties in the two-state model

Following [124], from (4.4) several properties related to the bursts can be obtained. Firstly, by (4.1) it is straightforward to see that the expected value for the time in the OFF state is

$$t_{OFF} = \frac{1}{k_{ON}}\,. \tag{4.7}$$

By the same reason, using (4.2) we see that the expected the expected value for the time in the ON state (also known as burst size in the literature) is

$$t_{ON} = \frac{1}{k_{OFF}}\,. \tag{4.8}$$

This allows us to define the expected burst frequency in the promoter activity as

$$f = \frac{1}{t_{ON} + t_{OFF}} = \frac{k_{ON} k_{OFF}}{k_{ON} + k_{OFF}} = k_{OFF} \langle ON \rangle \,, \tag{4.9}$$

being

$$\langle ON \rangle = \frac{k_{ON}}{k_{ON} + k_{OFF}} \,. \tag{4.10}$$

Please note that (4.10) can be seen as the mean promoter occupancy. In fact, it easy to check that this magnitude is, indeed, the stationary solution of $p(ON; \vec{r}, t)$ in (4.3). Moreover, it is important to recall that (4.8)-(4.10) are defined at each cell position $\vec{r}$, since the rates k_{OFF} and k_{ON} are cell-dependent. This means that the burst dynamics could also depend on the tissular position, which is indeed the case (see next section). However, why these rates change with the tissular position is still under debate.

Some recent works [70] propose logistic regressions of these parameters with the transcription factors. However, the importance of considering the combinatory behaviour of the transcription factors in the enhancers has been already pointed out in several works [120, 100]. Moreover, the interaction between the enhancers and promoter is an important and emerging field [40]. This is, enhancers are thought to be responsible for driving transcriptional bursting from their target promoters [38]. Our plan is to study these two components (TFs and enhancer-promoter interactions) by using stochastic models similar to (4.4). However, in order to take into account the transcription factors in the burst dynamics, we would need to modify the 2-state model. I.e., we may need to take into account all the possible enhancer-promoter states (this is, all the states compatible to the state space Ω defined in (3.11)), with different transition rates that could be governed by the association (+) / dissociation (-) reaction dynamics of each transcription factor (see Fig. 4.3).

This approach opens new interesting questions, from both the theoretical and experimental point of view. As we saw in Fig. 4.1, the bursts appears to have some kind of oscillatory behavior. This could come from the balance between the binding events that occur in the enhancer, where ON states help the transitions to other ON states (cooperativity). Interestingly, there are some works that analyze the existence of cycles in multi-state Markov Chains [60]. Hence, there could exist a connection between TF-TF cooperativity, promoter-enhancer interactions, and the periodicity of the states. Moreover, the combination of experiments and stochastic numerical methods applied to Markov Chains (suchs as Viterbi, Forward-Backward Algorithms, etc) would be required in order to get to the point of the question. This is an exciting question that is still on preparation, done by the author of the thesis in collaboration with Takashi Fukaya's Laboratory (IQB, Tokyo).

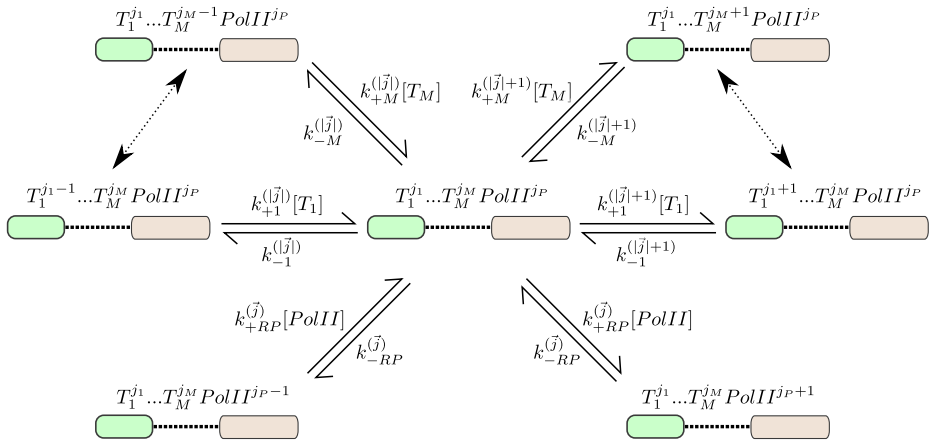

Figure 4.3: Scheme for the extension of the 2-state model of bursting, for single promoter controlled by M transcription factors. The scheme depicts the possible transitions for a state $[T_1^{j_1}...T_M^{j_M}PolII^{j_P}]$. The dotted arrows stand for the remaining reactions (not drawn in the scheme), for $j_2, ..., j_{M-1}$. The notation of the scheme is the same as the one used in Chapter 3.

Part III

Appendix

Appendix A

Proofs

A.1 Proof of equation (3.24)

In order to obtain equation (3.24) we first describe an iterative rule which will allow us to concrete the desired result for any mixed cooperativity configuration, covering as particular cases total and partial cooperativity.

Lemma 1.1 *Let us assume a general mixed cooperativity configuration* $\mathcal{C} = \{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_N\}_{c_N}\}$, *where* $\mathcal{I}_i$ *denotes the indices of* $\vec{x} \in \mathbb{R}^M$ *of the* N *subgroups of TFs cooperating with cooperativity constant* c_i. *Let us consider* $\vec{x}_i$ *a vector collecting all the values* x_h *with* $h \in \mathcal{I}_i$. *Then,*

i) the reordering of the values x_h *does not affects to expression (3.19), in particular,*

$$\begin{aligned} & S_e^{(n)}\Big((x_1, \ldots, x_M); \big\{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_N\}_{c_N}\big\}\Big) \\ = \;& S_e^{(n)}\Big((\vec{x}_1, \ldots, \vec{x}_N); \big\{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_N\}_{c_N}\big\}\Big). \end{aligned} \tag{A.1}$$

ii) The value of $S_e^{(n)}$ *evaluated on* N *cooperating subfamilies of TFs admits a a decomposition as the addition of two* $S_e^{(n)}$ *functions evaluated on* $N-1$ *cooperative families, accordingly to next iterative rule:*

$$\begin{aligned} & S_e^{(n)}\Big((\vec{x}_1, \ldots, \vec{x}_{N-1}, \vec{x}_N); \big\{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_{N-1}\}_{c_{N-1}}, \{\mathcal{T}_N\}_{c_N}\big\}\Big) \\ = \;& \frac{1}{c_N} S_e^{(n)}\Big((\vec{x}_1, \ldots, \vec{x}_{N-1}, c_N\vec{x}_N); \big\{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_{N-1}\}_{c_{N-1}}, \{\mathcal{T}_N\}_1\big\}\Big) \\ + \;& \left(1 - \frac{1}{c_N}\right) S_e^{(n)}\Big((\vec{x}_1, \ldots, \vec{x}_{N-1}, \vec{0}); \big\{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_{N-1}\}_{c_{N-1}}, \{\mathcal{T}_N\}_1\big\}\Big). \end{aligned}$$

Proof.
Assertion *i*) is obviously true since the value of $S_e^{(n)}$ depends multiplicatively of x_i.

In order to proof *ii*), let us reorder the vector of indexes $\vec{j} = (\vec{j}_1, \ldots, \vec{j}_N)$, being $\vec{j}_N = (j_{\alpha+1}, \ldots, j_M)$ the $M - \alpha$ TFs cooperating with cooperativity c_N. By splitting the addition in terms those terms where $\vec{j}_N$ is null and those where it is not we can rewrite

$$\begin{aligned} & S_e^{(n)}\Big((\vec{x}_1, \ldots, \vec{x}_N); \big\{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_N\}_{c_N}\big\}\Big) \\ &= \frac{1}{c_N} \sum_{\substack{|\vec{j}|\leq n \\ |\vec{j}_N|\geq 1}} \left(\prod_{i=1}^{N-1} c_i^{(|\vec{j}_i|-1)_+}\right) \frac{n!}{\prod_{i=0}^{M} j_i!} e^{j_0} \left(\prod_{i=1}^{\alpha} x_i^{j_i}\right) \left(\prod_{i=\alpha+1}^{M} (c_N x_i)^{j_i}\right) \\ &+ \sum_{\substack{|\vec{j}|\leq n \\ |\vec{j}_N|=0}} \left(\prod_{i=1}^{N-1} c_i^{(|\vec{j}_i|-1)_+}\right) \frac{n!}{\prod_{i=0}^{M} j_i!} e^{j_0} \left(\prod_{i=1}^{\alpha} x_i^{j_i}\right). \end{aligned} \tag{A.2}$$

Then, adding and subtracting the term

$$\frac{1}{c_N} \sum_{\substack{|\vec{j}|\leq n \\ |\vec{j}_N|=0}} \left(\prod_{i=1}^{N-1} c_i^{(|\vec{j}_i|-1)_+}\right) \frac{n!}{\prod_{i=0}^{M} j_i!} e^{j_0} \left(\prod_{i=1}^{\alpha} x_i^{j_i}\right) \left(\prod_{i=\alpha+1}^{M} (c_N x_i)^{j_i}\right)$$

in previous expression we get (A.2)

$$\begin{aligned} & S_e^{(n)}\Big((\vec{x}_1, \ldots, \vec{x}_N); \big\{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_N\}_{c_N}\big\}\Big) \\ &= \frac{1}{c_N} \sum_{|\vec{j}|\leq n} \left(\prod_{i=1}^{N-1} c_i^{(|\vec{j}_i|-1)_+}\right) \frac{n!}{\prod_{i=0}^{M} j_i!} e^{j_0} \left(\prod_{i=1}^{\alpha} x_i^{j_i}\right) \left(\prod_{i=\alpha+1}^{M} (c_N x_i)^{j_i}\right) \\ &+ \left(1 - \frac{1}{c_N}\right) \sum_{\substack{|\vec{j}|\leq n \\ |\vec{j}_N|=0}} \left(\prod_{i=1}^{N-1} c_i^{(|\vec{j}_i|-1)_+}\right) \frac{n!}{\prod_{i=0}^{\alpha} j_i!} e^{j_0} \left(\prod_{i=1}^{\alpha} x_i^{j_i}\right) \\ &= \frac{1}{c_N} S_e^{(n)}\Big((\vec{x}_1, \ldots, \vec{x}_{N-1}, c_N \vec{x}_N); \big\{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_{N-1}\}_{c_{N-1}}, \{\mathcal{T}_N\}_1\big\}\Big) \\ &+ \left(1 - \frac{1}{c_N}\right) S_e^{(n)}\Big((\vec{x}_1, \ldots, \vec{x}_{N-1}, \vec{0}); \big\{\{\mathcal{T}_1\}_{c_1}, ..., \{\mathcal{T}_{N-1}\}_{c_{N-1}}, \{\mathcal{T}_N\}_1\big\}\Big). \end{aligned}$$

□

Previous iterative property allow us to face up the proof of equation (3.24).

Proof of equation (3.24) The proof of the formula stated can be done by induction over the number of cooperating subfamilies. As mentioned in the main text, expression (3.24) reduces to (3.25) in the case of a single cooperating family, that is $N = 1$. (3.25) can be easily deduced by applying Lemma 1.1 and taking into consideration (3.23). Let us assume as

inductive hypothesis that (3.24) is valid when there exists $N-1$ cooperative subfamilies. Then, by using twice Lemma 1.1 we can write:

$$
\begin{aligned}
& S_e^{(n)}\Big((\vec{x}_1,...,\vec{x}_{N-1},\vec{x}_N);\Big\{\{\mathcal{T}_1\}_{c_1},...,\{\mathcal{T}_{N-1}\}_{c_{N-1}},\{\mathcal{T}_N\}_{c_N}\Big\}\Big) \\
&= \tfrac{1}{c_N} S_e^{(n)}\Big((\vec{x}_1,...,\vec{x}_{N-1},c_N\vec{x}_N);\Big\{\{\mathcal{T}_1\}_{c_1},...,\{\mathcal{T}_{N-1}\}_{c_{N-1}},\{\mathcal{T}_N\}_1\Big\}\Big) \\
&+ \Big(1-\tfrac{1}{c_N}\Big) S_e^{(n)}\Big((\vec{x}_1,...,\vec{x}_{N-1},\vec{0});\Big\{\{\mathcal{T}_1\}_{c_1},...,\{\mathcal{T}_{N-1}\}_{c_{N-1}},\{\mathcal{T}_N\}_1\Big\}\Big) \\
&= \frac{S_e^{(n)}\Big((\vec{x}_1,...,\vec{x}_{N-2},c_{N-1}\vec{x}_{N-1},c_N\vec{x}_N);\Big\{\{\mathcal{T}_1\}_{c_1},...,\{\mathcal{T}_{N-2}\}_{c_{N-2}},\{\mathcal{T}_{N-1}\cup\mathcal{T}_N\}_1\Big\}\Big)}{c_N c_{N-1}} \\
&+ \Big(1-\tfrac{1}{c_{N-1}}\Big)\frac{S_e^{(n)}\Big((\vec{x}_1,...,\vec{x}_{N-2},\vec{0},c_N\vec{x}_N);\Big\{\{\mathcal{T}_1\}_{c_1},...,\{\mathcal{T}_{N-2}\}_{c_{N-2}},\{\mathcal{T}_{N-1}\cup\mathcal{T}_N\}_1\Big\}\Big)}{c_N} \\
&+ \Big(1-\tfrac{1}{c_N}\Big) S_e^{(n)}\Big((\vec{x}_1,...,\vec{x}_{N-1},\vec{0});\Big\{\{\mathcal{T}_1\}_{c_1},...,\{\mathcal{T}_{N-1}\}_{c_{N-1}},\{\mathcal{T}_N\}_1\Big\}\Big) \qquad \text{(A.3)}
\end{aligned}
$$

Since all the $S_e^{(n)}$ operators are evaluated on data with only $N-1$ cooperating subfamilies we can apply them the inductive hypothesis and observe that:

- the first term on the right hand side of (A.3) would correspond to all the additive terms in the right hand side of (3.24) with multi-indices of the form $(h_1,\ldots,h_{N-2},1,1)$,
- the second term on the right hand side of (A.3) would correspond to all the additive terms in the right hand side (3.24) with multi-indices of the form $(h_1,\ldots,h_{N-2},0,1)$,
- the third term on the right hand side of (A.3) would correspond to all the additive terms in the right hand side (3.24) with multi-indices of the form $(h_1,\ldots,h_{N-2},h_{N-1},0)$,

where all the h_j can be 0 or 1. In consequence the addition of the three terms in (A.3) coincides with expression on the right hand side of (3.24) which concludes the proof. □

A.2 Deduction of the general Hill BEWARE operators

In order to deduce expressions (3.32)-(3.35) we need to recall first the general expressions of the Regulation Factor (3.20), Basal (3.21) and Promoter (3.22) functions defined in terms of the polynomial function (3.19). Since all these functions are written in terms of rational relations between S_e, we can define a general rational function

$$
R^{(n)} = \frac{S_{e_1}^{(n)}((x_1/K,...,x_M/K);\{\{\mathcal{X}_1\}_c,...,\{\mathcal{X}_N\}_c\})}{S_{e_2}^{(n)}((y_1/K,...,y_M/K);\{\{\mathcal{Y}_1\}_c,...,\{\mathcal{Y}_N\}_c\})} \qquad \text{(A.4)}
$$

with $S_e^{(n)}$ defined in the mixed cooperative case (3.24). Then, if we define the global affinity constant $K_d = K^n/c^{n-1}$, we can rewrite (A.4)

$$
\begin{aligned}
R^{(n)} &= \frac{\sum\limits_{\substack{|\vec{h}|_\infty \leq 1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(e_1 + \sum\limits_{j=1}^{N} h_j c \sum\limits_{i\in\mathcal{I}_j} \frac{x_i}{K} \right)^n \prod\limits_{j=1}^{N} \frac{\left(1-\frac{1}{c}\right)^{1-h_j}}{c^{h_j}}}{\sum\limits_{\substack{|\vec{h}|_\infty \leq 1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(e_2 + \sum\limits_{j=1}^{N} h_j c \sum\limits_{i\in\mathcal{I}_j} \frac{y_i}{K} \right)^n \prod\limits_{j=1}^{N} \frac{\left(1-\frac{1}{c}\right)^{1-h_j}}{c^{h_j}}} \\
&= \frac{\sum\limits_{\substack{|\vec{h}|_\infty \leq 1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(e_1 + \frac{c}{K} \sum\limits_{j=1}^{N} h_j \sum\limits_{i\in\mathcal{I}_j} x_i \right)^n \prod\limits_{j=1}^{N} (c-1)^{1-h_j}}{\sum\limits_{\substack{|\vec{h}|_\infty \leq 1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(e_2 + \frac{c}{K} \sum\limits_{j=1}^{N} h_j \sum\limits_{i\in\mathcal{I}_j} y_i \right)^n \prod\limits_{j=1}^{N} (c-1)^{1-h_j}} \\
&= \frac{\sum\limits_{\substack{|\vec{h}|_\infty \leq 1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(e_1 + \frac{c^{1/n}}{K_d^{1/n}} \sum\limits_{j=1}^{N} h_j \sum\limits_{i\in\mathcal{I}_j} x_i \right)^n \prod\limits_{j=1}^{N} (c-1)^{1-h_j}}{\sum\limits_{\substack{|\vec{h}|_\infty \leq 1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(e_2 + \frac{c^{1/n}}{K_d^{1/n}} \sum\limits_{j=1}^{N} h_j \sum\limits_{i\in\mathcal{I}_j} y_i \right)^n \prod\limits_{j=1}^{N} (c-1)^{1-h_j}}.
\end{aligned}
$$

By splitting the sums in those terms that have $|\vec{h}|_\infty = 0$ and $|\vec{h}|_\infty = 1$ we finally get

$$
R^{(n)} = \frac{(c-1)^N + \frac{c}{K_d} \sum\limits_{\substack{|\vec{h}|_\infty = 1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(\frac{K_d^{1/n}}{c^{1/n}} e_1 + \frac{c^{1/n}}{K_d^{1/n}} \sum\limits_{j=1}^{N} h_j \sum\limits_{i\in\mathcal{I}_j} x_i \right)^n \prod\limits_{j=1}^{N} (c-1)^{1-h_j}}{(c-1)^N + \frac{c}{K_d} \sum\limits_{\substack{|\vec{h}|_\infty = 1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(\frac{K_d^{1/n}}{c^{1/n}} e_2 + \frac{c^{1/n}}{K_d^{1/n}} \sum\limits_{j=1}^{N} h_j \sum\limits_{i\in\mathcal{I}_j} y_i \right)^n \prod\limits_{j=1}^{N} (c-1)^{1-h_j}}.
$$

Please note that both $(c-1)^N$ and

$$
\frac{c}{K_d} \sum_{\substack{|\vec{h}|_\infty = 1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(\frac{K_d^{1/n}}{c^{1/n}} e_1 + \frac{c^{1/n}}{K_d^{1/n}} \sum_{j=1}^{N} h_j \sum_{i\in\mathcal{I}_j} x_i \right)^n \prod_{j=1}^{N} (c-1)^{1-h_j}
$$

are polynomials of degree N in the c variable, since $|\vec{h}|_\infty = 1$ automatically implies that $\sum_{j=1}^{N} 1 - h_j \leq N-1$. We are interested on computing $\lim_{\substack{c\to\infty \\ K_d=cte}} R^{(n)}$, hence by l'Hôpital's rule the only terms that will remain in the limit will be those coefficients that multiply to c^N, this is, those terms such as $|\vec{h}| = 1$:

$$\begin{aligned}
&\lim_{\substack{c\to\infty \\ K_d=cte}} R^{(n)} \\
&= \frac{1 + \frac{1}{K_d} \sum\limits_{\substack{|\vec{h}|=1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(\sum\limits_{j=1}^{N} h_j \sum\limits_{i\in\mathcal{I}_j} x_i \right)^n}{1 + \frac{1}{K_d} \sum\limits_{\substack{|\vec{h}|=1 \\ \vec{h}\in\mathbb{N}_0^N}} \left(\sum\limits_{j=1}^{N} h_j \sum\limits_{i\in\mathcal{I}_j} y_i \right)^n} = \frac{K_d + \sum\limits_{j=1}^{N} \left(\sum\limits_{i\in\mathcal{I}_j} x_i \right)^n}{K_d + \sum\limits_{j=1}^{N} \left(\sum\limits_{i\in\mathcal{I}_j} y_i \right)^n} . \quad \text{(A.5)}
\end{aligned}$$

Please note that we can now directly deduce the expressions (3.32)-(3.35) assuming that $K_i = K$ and $c_i = c$ for all $i = 1, ..., M$.

A.3 Existence/non existence of inverse logic in the activator/repressor framework: pull effect

In this Appendix we are going to analyse the consistency of previous expressions with the fundamental notion of activators established in [33], this is, when activator concentrations are scaled up then the transcriptional activity increases. Analogously, it can be declared the fundamental notion of repressors saying that increases in repressor concentrations would diminish transcriptional activity. This basic idea is translated mathematically into:

$$\frac{\partial \text{BEWARE}\big(([A],[R]),[PolII];\mathcal{C}\big)}{\partial [A]} > 0\,, \quad \text{(A.6)}$$

$$\frac{\partial \text{BEWARE}\big(([A],[R]),[PolII];\mathcal{C}\big)}{\partial [R]} < 0\,. \quad \text{(A.7)}$$

However, as we will show this basic logic may not hold in some of the proposed models, and effects of 'inverse control logic' (i.e., an inversion on the inequalities (A.6) and (A.7)) can appear. In this case, both versions of modelling stimulated and recruitment, coincide predicting that this inverse logic would happen in the presence of strong enough cooperativity between activators and repressors. Furthermore, they coincide predicting that in the presence of cooperativity only between TFs of the same nature, that is partial cooperativity, the basic activator/repressor logic (A.6)-(A.7) always holds.

The results of the performed analysis are summarised in next Lemma.

Lemma 3.2 *BEWARE operators* (3.13)-(3.16) *depending only on two functional opposite TFs*

i) can exhibit inverse control logic when total cooperativity is considered. This occurs for large enough cooperativities for certain ranges of the parameters a, r, $\nu_{max}^{(n)}$ *and* r_{bas} *stablished in Table A.1.*

ii) However, this basic control logic always holds when partial cooperativity (3.9) *is considered.*

Table A.1: Parameter requirements for existence of inverse logic for bifunctional BEWARE operators in presence of total cooperativity.

	Act. Inv. Log. (A.6)	Rep Inv. Log. (A.7)
Recruitment	$r < a^{\frac{1}{1-n}}$ & $c > c_{th}^a$	$r > a^{1-n}$ & $c > c_{th}^r$
Stimulated	$\frac{\nu_{max}^{(n)}}{r_{bas}} < \frac{1-\tilde{r}^{n-1}}{(\tilde{e}\tilde{r})^{n-1}}$ & $c > \tilde{c}_{th}^a$	$\frac{r_{bas}}{\nu_{max}^{(n)}} < \frac{\tilde{e}\tilde{r}}{1-\tilde{r}}$ & $c > \tilde{c}_{th}^r$

The analysis developed in this Appendix for proving Lemma 3.2 also allows to describe the values of the concentrations that give rise to inverse logic. In the presence of relative low concentrations of activators (this is, low activators concentrations compared to the concentrations of repressors), the system could show an unnatural response to an increase of the activators, presenting a decrement in the transcription rate (BEWARE function). This 'pull effect' is a direct consequence of the total cooperativity. Since an increase of the activator concentration implies more activator bindings, these additional bindings improve the cleavage of any other transcription factor, included repressors. Taking into account that there is higher concentration of repressors in the system than activators, it is much more likely to 'pull' repressors in the enhancers, finding in consequence more states with more repressors than activators (i.e., more repression and hence less transcription). The delicate balances occurring in that case between activation, repression, affinities and binding cooperativity is reflected in Table A.1. Fig 3.3 illustrates these effects. Although behavioural tendencies (A.6)-(A.7) could appear to be very naive their relevance in our logical scheme is undoubtable. This can be seen in the proof of Proposition 4.1 where the existence of activation/repression thresholds is stated. In those cases where (A.6)-(A.7) does not hold, this requieres an extra effort in order to apply the Implicit Function Theorem. This kind of basic properties are also really relevant in specific applications, as for instance in repressilator type models [25].

In these arguments it will be helpful to use the positive symmetric func-

tions defined by:

$$H_n(X,Y)=\sum_{i=0}^{n-1}X^iY^{n-1-i}\ \text{ for any } X\,,Y\in(0,+\infty) \tag{A.8}$$

when $n>1$ and $H_1(X,Y)=1$ when $n=1$.

It can be easily checked that using this notation we have

$$X^{n-1}-Y^{n-1}=H_{n-1}(X,Y)(X-Y)=\frac{H_{n-1}(X,Y)}{H_n(X,Y)}\,(X^n-Y^n)\qquad n\geq 2. \tag{A.9}$$

We will employ in the subsequent several easy properties of these functions collected in next Lemma.

Lemma 3.3 *Let $X\,,Y\in(0,+\infty)$, then*

1. *$\frac{H_{n-1}(X,Y)}{H_n(X,Y)}$ is strictly decreasing with respect to X and Y.*
2. *If $X>Y$, the following relations are fulfilled*

$$\frac{1}{\frac{Y}{n-1}+X}\leq\frac{H_{n-1}(X,Y)}{H_n(X,Y)}\leq\frac{1}{\frac{X}{n-1}+Y}\,.$$

3. *If $Z\leq X$ then $\frac{H_{n-1}(X,Y)}{H_n(X,Y)}Z<1$.*

Proof

Properties *i)* and *ii)* can be trivially proven by using the identities:

$$\frac{H_{n-1}(X,Y)}{H_n(X,Y)}=\frac{1}{\frac{X}{\sum_{i=0}^{n-2}(Y/X)^i}+Y}=\frac{1}{\frac{Y}{\sum_{i=0}^{n-2}(X/Y)^i}+X}\,.$$

Estimate *iii)* can be also check by considering

$$\frac{H_{n-1}(X,Y)}{H_n(X,Y)}Z=\frac{\sum\limits_{i=0}^{n-2}Y^iX^{n-2-i}Z}{\sum\limits_{i=0}^{n-1}Y^iX^{n-1-i}}\leq\frac{\sum\limits_{i=0}^{n-2}Y^iX^{n-1-i}}{\sum\limits_{i=0}^{n-1}Y^iX^{n-1-i}}<1\,.$$

□

A.3.1 Existence of inverse logic in presence of high total cooperativity

This sub-appendix is devoted to prove statement *i)* in Lemma 3.2. That is, in presence of strong enough total cooperativity recruitment and stimulated approaches predict that the intuitive behavioural tendencies (A.6) or (A.7) can be violated. For the sake of simplicity, we will note $\tilde{x}=\frac{c[A]}{K_A}$ and $\tilde{y}=\frac{c[R]}{K_R}$.

- Activators inverse logic for the Recruitment BEWARE operator
  Since the recruitment BEWARE operator is increasing with respect to the regulation factor operator, we have

$$\begin{aligned}
&\text{sign}\left\{\frac{\partial \text{BEWARE}_r\big(([A],[R]),[PolII];\{A,R\}_c\big)}{\partial[A]}\right\}\\
&= \text{sign}\left\{\frac{\partial F_{reg}\big((\tilde{x},\tilde{y});\{A,R\}_c\big)}{\partial\tilde{x}}\right\}\\
&= \text{sign}\left\{(c-1)\big(a\bar{\alpha}_r^{n-1}-\alpha_r^{n-1}\big)+\bar{\alpha}_r^{n-1}\alpha_r^{n-1}(a\alpha_r-\bar{\alpha}_r)\right\}
\end{aligned} \tag{A.10}$$

  where $\bar{\alpha}_r = \bar{\alpha}_r(\tilde{x},\tilde{y}) = 1+a\tilde{x}+r\tilde{y}$, $\alpha_r = \alpha_r(\tilde{x},\tilde{y}) = 1+\tilde{x}+\tilde{y}$. Since we are assuming $r<1<a$ we have that $a\alpha_r - \bar{\alpha}_r > 0$. This proves that in absence of cooperativity ($c=1$) the sign of (A.10) will be always positive. Nevertheless, if $a\bar{\alpha}_r^{n-1}-\alpha_r^{n-1}$ takes a negative value and c is high enough, that it,

$$c > 1 + \frac{\bar{\alpha}_r^{n-1}\alpha_r^{n-1}(a\alpha_r-\bar{\alpha}_r)}{\alpha_r^{n-1}-a\bar{\alpha}_r^{n-1}} \tag{A.11}$$

  the sign of (A.10) can become negative. It can be checked that $\big(a\bar{\alpha}_r^{n-1}-\alpha_r^{n-1}\big)<0$ if and only if $a^{\frac{1}{n-1}}r<1$ and in that case it would occur for values in the cone

$$H_{r,a} = \left\{(\tilde{x},\tilde{y})\in(\mathbf{R}_0^+)^2; \tilde{y} > \frac{(a^{\frac{1}{n-1}}-1)+(a^{\frac{n}{n-1}}-1)\tilde{x}}{1-a^{\frac{1}{n-1}}r}\right\}.$$

  Thus, for any

$$c > c_{th}^a = \inf_{(\tilde{x},\tilde{y})\in H_{r,a}}\left\{1+\frac{\bar{\alpha}_r^{n-1}\alpha_r^{n-1}(a\alpha_r-\bar{\alpha}_r)}{\alpha_r^{n-1}-a\bar{\alpha}_r^{n-1}}\right\}$$

  we can assure that there exist some values $(\tilde{x},\tilde{y})\in H$ such that the sign computed in (A.10) will be negative.

- Repressors inverse logic for the Recruitment BEWARE operator
  By an analogous argument we get

$$\begin{aligned}
&\text{sign}\left\{\frac{\partial \text{BEWARE}_r([A],[R],[PolII];\{A,R\}_c)}{\partial[R]}\right\}\\
&= \text{sign}\left\{\frac{\partial F_{reg}(\tilde{x},\tilde{y})}{\partial\tilde{y}}\right\}\\
&= \text{sign}\left\{(c-1)\big(r\bar{\alpha}_r^{n-1}-\alpha_r^{n-1}\big)+\bar{\alpha}_r^{n-1}\alpha_r^{n-1}(r\alpha_r-\bar{\alpha}_r)\right\}
\end{aligned} \tag{A.12}$$

Since we are assuming $r<1<a$ we now have that $r\alpha_r-\bar{\alpha}_r<0$. Thus, the sign of (A.12) will be always negative in absence of cooperativity $c=1$. On the other hand, if $\left(r\bar{\alpha}_r^{n-1}-\alpha_r^{n-1}\right)$ takes a positive value and c is high enough, that it,

$$c>1+\frac{\bar{\alpha}_r^{n-1}\alpha_r^{n-1}(\bar{\alpha}_r-r\alpha_r)}{r\bar{\alpha}_r^{n-1}-\alpha_r^{n-1}} \tag{A.13}$$

the sign of (A.12) can become positive. It can be checked that $\left(r\bar{\alpha}_r^{n-1}-\alpha_r^{n-1}\right)>0$ if and only if $ar^{\frac{1}{n-1}}>1$ and in that case it would happen for values in the cone

$$H_{r,r}=\left\{(\tilde{x},\tilde{y})\in(\mathbf{R}_0^+)^2;\tilde{x}>\frac{(1-r^{\frac{1}{n-1}})+(1-r^{\frac{n}{n-1}})\tilde{y}}{ar^{\frac{1}{n-1}}-1}\right\}.$$

Thus, for any

$$c>c_{th}^r=\inf_{(\tilde{x},\tilde{y})\in H_{r,r}}\left\{1+\frac{\bar{\alpha}_r^{n-1}\alpha_r^{n-1}(\bar{\alpha}_r-r\alpha_r)}{r\bar{\alpha}_r^{n-1}-\alpha_r^{n-1}}\right\}$$

we can assure that there exist some values $(\tilde{x},\tilde{y})\in H$ such that the sign computed in (A.12) will be positive.

- Activators inverse logic for the Stimulated BEWARE operator

$$\begin{aligned}
&\text{sign}\left\{\frac{\partial \text{BEWARE}_s([A],[R],[PolII];\{A,R\}_c)}{\partial[A]}\right\}\\
&=\ \text{sign}\left\{\frac{r_{bas}}{1+\frac{K_{RP}}{[PolII]}}\frac{\partial Basal(\tilde{x},\tilde{y})}{\partial\tilde{x}}+\frac{\nu_{max}^{(n)}}{1+\frac{K_{RP}}{[PolII]}}\frac{\partial Promoter(\tilde{x},\tilde{y})}{\partial\tilde{x}}\right\}\\
&=\ \text{sign}\left\{(c-1)\Big((\bar{\alpha}_s^{n-1}-\alpha_s^{n-1})r_{bas}+\beta_s^{n-1}\nu_{max}^{(n)}\Big)\right.\\
&+\ \left.\alpha_s^{n-1}\Big(\bar{\alpha}_s^{n-1}(\alpha_s-\bar{\alpha}_s)r_{bas}+(\beta_s^{n-1}(\alpha_s-\beta_s)+\bar{\beta}_s)\nu_{max}^{(n)}\Big)\right\}
\end{aligned} \tag{A.14}$$

where adopting previous notation we have $\bar{\alpha}_s=\bar{\alpha}_s(\tilde{x},\tilde{y})=1+\tilde{x}+\tilde{r}\tilde{y}$, $\alpha_s=\alpha_s(\tilde{x},\tilde{y})=1+\tilde{x}+\tilde{y}$, $\bar{\beta}_s=\bar{\beta}_s(\tilde{y})=\tilde{e}+\tilde{e}\tilde{r}\tilde{y}$, $\beta_s=\beta_s(\tilde{x},\tilde{y})=\tilde{e}+\tilde{x}+\tilde{e}\tilde{r}\tilde{y}$. Please note also that thanks to $\tilde{r}<1$ and $X,Y\geq 0$ next relations are verified

$$\alpha_s\geq\bar{\alpha}_s\geq\beta_s\geq\bar{\beta}_s>0. \tag{A.15}$$

By using these estimates we can easily check

$$\alpha_s^{n-1}\Big(\bar{\alpha}_s^{n-1}(\alpha_s-\bar{\alpha}_s)r_{bas}+(\beta_s^{n-1}(\alpha_s-\beta_s)+\bar{\beta}_s)\nu_{max}^{(n)}\Big)>\tilde{e}\nu_{max}^{(n)}>0\,. \tag{A.16}$$

Hence if

$$\left((\bar{\alpha}_s^{n-1} - \alpha_s^{n-1})r_{bas} + \beta_s^{n-1}\nu_{max}^{(n)}\right) < 0 \text{ for some } \tilde{x},\ \tilde{y} \tag{A.17}$$

holds then the partial with respect to $\tilde{x}$ could change the sign for c large enough, more concretely,

$$c > 1 - \frac{\alpha_s^{n-1}\left(\bar{\alpha}_s^{n-1}(\alpha_s - \bar{\alpha}_s) + (\beta_s^{n-1}(\alpha_s - \beta_s) + \bar{\beta}_s)\frac{\nu_{max}^{(n)}}{r_{bas}}\right)}{(\bar{\alpha}_s^{n-1} - \alpha_s^{n-1})r_{bas} + \beta_s^{n-1}\frac{\nu_{max}^{(n)}}{r_{bas}}}. \tag{A.18}$$

Let us also observe that the inverse logic is not possible in the absence of total cooperativity, that is, $c > 1$ is required. Obviously, (A.17) will occurs whenever

$$\frac{\nu_{max}^{(n)}}{r_{bas}} < \sup_{\tilde{x},\tilde{y}} \frac{(\alpha_s^{n-1} - \bar{\alpha}_s^{n-1})}{\beta_s^{n-1}} = \frac{1 - \tilde{r}^{n-1}}{(\tilde{e}\tilde{r})^{n-1}}$$

where the supremum can be easily calculated thanks to

$$\begin{aligned} \frac{(\alpha_s^{n-1} - \bar{\alpha}_s^{n-1})}{\beta_s^{n-1}} &\leq \frac{(\alpha_s^{n-1} - (\tilde{r}\alpha_s)^{n-1})}{(\tilde{e}\tilde{r}\alpha_s)^{n-1}} \\ = \quad \frac{1 - \tilde{r}^{n-1}}{(\tilde{e}\tilde{r})^{n-1}} &= \lim_{\tilde{x}\to 0,\tilde{y}\to\infty} \frac{(\alpha_s^{n-1} - \bar{\alpha}_s^{n-1})}{\beta_s^{n-1}}. \end{aligned} \tag{A.19}$$

Thus, we conclude that, always that

$$\frac{\nu_{max}^{(n)}}{r_{bas}} < \frac{1 - \tilde{r}^{n-1}}{(\tilde{e}\tilde{r})^{n-1}}$$

the set

$$H_{s,a} = \left\{(\tilde{x},\tilde{y}) \in (\mathbb{R}_0^+)^2| \quad \left((\bar{\alpha}_s^{n-1} - \alpha_s^{n-1})r_{bas} + \beta_s^{n-1}\nu_{max}^{(n)}\right)|_{\tilde{x},\tilde{y})} < 0\right\}$$

is non empty and for any

$$c > \tilde{c}_{th}^a = \inf_{H_{s,a}} \left\{ 1 - \frac{\alpha_s^{n-1}\left(\bar{\alpha}_s^{n-1}(\alpha_s - \bar{\alpha}_s) + (\beta_s^{n-1}(\alpha_s - \beta_s) + \bar{\beta}_s)\frac{\nu_{max}^{(n)}}{r_{bas}}\right)}{(\bar{\alpha}_s^{n-1} - \alpha_s^{n-1})r_{bas} + \beta_s^{n-1}\frac{\nu_{max}^{(n)}}{r_{bas}}} \right\}$$

there exist points $(\tilde{x},\tilde{y})$ where the sign on the right hand side of (A.14) is negative. Obviously, the value of c will determine the final set of values where it occurs by condition (A.18).

- Repressors inverse logic for the Stimulated BEWARE operator
  The sign of the partial derivative of the BEWARE operator with respect to the repressor variable is

$$\begin{aligned}
&\operatorname{sign}\left\{\tfrac{\partial \mathrm{BEWARE}_s([A],[R],[PolII];\{A,R\}_c)}{\partial [R]}\right\}\\
&= \operatorname{sign}\left\{\frac{r_{bas}}{1+\frac{K_{RP}}{[PolII]}}\frac{\partial Basal(\tilde{x},\tilde{y})}{\partial \tilde{y}}+\frac{\nu^{(n)}_{max}}{1+\frac{K_{RP}}{[PolII]}}\frac{\partial Promoter(\tilde{x},\tilde{y})}{\partial \tilde{y}}\right\}\\
&= \operatorname{sign}\Big\{(c-1)\Big((\bar{\alpha}_s^{n-1}\tilde{r}-\alpha_s^{n-1})r_{bas}+\tilde{e}\tilde{r}(\beta_s^{n-1}-\bar{\beta}_s^{n-1})\nu^{(n)}_{max}\Big) \qquad (A.20)\\
&+ \alpha_s^{n-1}\Big(\bar{\alpha}_s^{n-1}(\tilde{r}\alpha_s-\bar{\alpha}_s)r_{bas}+(\tilde{e}\tilde{r}(\beta_s^{n-1}-\bar{\beta}_s^{n-1})-(\beta_s^n-\bar{\beta}_s^n))\nu^{(n)}_{max}\Big)\Big\}.
\end{aligned}$$

  Here

$$\alpha_s^{n-1}\Big(\bar{\alpha}_s^{n-1}(\tilde{r}\alpha_s-\bar{\alpha}_s)r_{bas}+(\tilde{e}\tilde{r}(\beta_s^{n-1}-\bar{\beta}_s^{n-1})-(\beta_s^n-\bar{\beta}_s^n))\nu^{(n)}_{max}\Big)<0\,, \qquad (A.21)$$

  since $\tilde{r}\alpha_s-\bar{\alpha}_s<0$ by definition and

$$\tilde{e}\tilde{r}(\beta_s^{n-1}-\bar{\beta}_s^{n-1})-(\beta_s^n-\bar{\beta}_s^n)=(\beta_s^n-\bar{\beta}_s^n)\left(\frac{H_{n-1}(\tilde{e}\tilde{r}\beta_s,\tilde{e}\tilde{r}\bar{\beta}_s)}{H_n(\beta_s,\bar{\beta}_s)}-1\right)<0 \qquad (A.22)$$

  by (A.9) and Lemma 3.3 *ii)*. By this reason, if

$$(\bar{\alpha}_s^{n-1}\tilde{r}-\alpha_s^{n-1})r_{bas}+\tilde{e}\tilde{r}(\beta_s^{n-1}-\bar{\beta}_s^{n-1})\nu^{(n)}_{max}>0 \qquad (A.23)$$

  then for c large enough, more concretely when

$$c>1-\frac{\alpha_s^{n-1}\Big(\bar{\alpha}_s^{n-1}(\tilde{r}\alpha_s-\bar{\alpha}_s)+(\tilde{e}\tilde{r}(\beta_s^{n-1}-\bar{\beta}_s^{n-1})-(\beta_s^n-\bar{\beta}_s^n))\frac{\nu^{(n)}_{max}}{r_{bas}}\Big)}{(\bar{\alpha}_s^{n-1}\tilde{r}-\alpha_s^{n-1})+\tilde{e}\tilde{r}(\beta_s^{n-1}-\bar{\beta}_s^{n-1})\frac{\nu^{(n)}_{max}}{r_{bas}}} \qquad (A.24)$$

  the sign computed in (A.20) can be positive. Again, this can not be longer true in absence of cooperativity, that is, when $c=1$. Now, (A.23) can occur if and only if

$$\frac{r_{bas}}{\tilde{e}\tilde{r}\nu^{(n)}_{max}}<\sup\frac{\beta_s^{n-1}-\bar{\beta}_s^{n-1}}{\alpha_s^{n-1}-\tilde{r}\bar{\alpha}_s^{n-1}}=\frac{1}{1-\tilde{r}} \qquad (A.25)$$

  where analogously as was done before, the supremum can be easily calculated from

$$\frac{\beta_s^{n-1}-\bar{\beta}_s^{n-1}}{\alpha_s^{n-1}-\tilde{r}\bar{\alpha}_s^{n-1}}\leq\frac{\bar{\alpha}_s^{n-1}}{\bar{\alpha}_s^{n-1}-\tilde{r}\bar{\alpha}_s^{n-1}}=\frac{1}{1-\tilde{r}}=\lim_{\tilde{x}\to\infty,\tilde{y}\to 0}\frac{\beta_s^{n-1}-\bar{\beta}_s^{n-1}}{\bar{\alpha}_s^{n-1}\tilde{r}-\alpha_s^{n-1}}\,.$$

  Then, arguing as before, we can conclude that (A.20) can be positive in the points determined by a set $H_{s,r}$ defined in terms of the condition defining by (A.23) and the for values c bigger than $\tilde{c}^r_{th}$ determined by the lower term of (A.24).

A.3.2 Direct logic in the presence of partial cooperativity

Now we prove that in the presence of partial cooperativity between activators and/or repressors the inverse logic can not occurs as stated in Lemma (3.2) *ii)*. In the rest of this proof we will denote $\tilde{x} = \frac{c_A[A]}{K_A}$ and $\tilde{y} = \frac{c_R[R]}{K_R}$.

- Activator direct logic for the partial cooperative recruitment BEWARE operator
  Using the increasing character of the recruitment BEWARE operator with respect to the regulation factor operator, we have

$$\begin{aligned}
\operatorname{sign}\Big\{\tfrac{\partial \mathrm{BEWARE}_r([A],[R],[PolII];\{A\}c_A,\{R\}c_R)}{\partial[A]}\Big\} &= \operatorname{sign}\Big\{\tfrac{\partial F_{reg}(\tilde{x},\tilde{y})}{\partial \tilde{x}}\Big\} \\
= \operatorname{sign}\Big\{&(c_A-1)(c_R-1)^2\Big(a\gamma_r^{n-1}-\bar{\gamma}_r^{n-1}\Big)+(c_R-1)^2\gamma_r^{n-1}\bar{\gamma}_r^{n-1}\Big(a\bar{\gamma}_r-\gamma_r\Big) \\
&+\bar{\alpha}_r^{n-1}\alpha_r^{n-1}(a\alpha_r-\bar{\alpha}_r)+(c_A-1)\Big(\tfrac{a}{\bar{\alpha}_r}\bar{\alpha}_r^n\beta_r^n-\tfrac{1}{\alpha_r}\alpha_r^n\bar{\beta}_r^n\Big) \\
&+(c_R-1)\Big(\bar{\alpha}_r^n\bar{\gamma}_r^n\Big(\tfrac{a}{\bar{\alpha}_r}-\tfrac{1}{\bar{\gamma}_r}\Big)+\gamma_r^n\alpha_r^n\Big(\tfrac{a}{\gamma_r}-\tfrac{1}{\alpha_r}\Big)\Big) \\
&+(c_A-1)(c_R-1)\Big(a\bar{\alpha}_r^{n-1}-\alpha_r^{n-1}+a\gamma_r^{n-1}\beta_r^n-\bar{\gamma}_r^{n-1}\bar{\beta}_r^n\Big)\Big\}
\end{aligned} \tag{A.26}$$

  where $\bar{\alpha}_r = \bar{\alpha}_r(\tilde{x},\tilde{y}) = 1 + a\tilde{x} + r\tilde{y}$, $\alpha_r = \alpha_r(\tilde{x},\tilde{y}) = 1 + \tilde{x} + \tilde{y}$, $\bar{\beta}_r = \bar{\beta}_r(\tilde{y}) = 1 + r\tilde{y}$, $\beta_r = \beta_r(\tilde{y}) = 1 + \tilde{y}$, $\bar{\gamma}_r = \bar{\gamma}_r(\tilde{x}) = 1 + \tilde{x}$, $\gamma_r = \gamma_r(\tilde{x}) = 1 + a\tilde{x}$. From these definitions we can easily check that almost all the terms inside the sign function are positive by using the estimates:

$$\begin{aligned}
&\bar{\alpha}_r \geq \gamma_r \geq \bar{\gamma}_r\,, \qquad a\bar{\gamma}_r \geq \gamma_r\,, \qquad \frac{a}{\bar{\alpha}_r} \geq \frac{1}{\alpha_r}\,, \qquad \frac{a}{\gamma_r} \geq \frac{1}{\alpha_r}\,, \\
&\bar{\alpha}_r\beta_r \geq \alpha_r\bar{\beta}_r\,, \qquad \alpha_r\gamma_r \geq \bar{\alpha}_r\bar{\gamma}_r\,, \qquad \frac{a}{\gamma_r} \geq \frac{1}{\bar{\gamma}_r}\,, \qquad \alpha_r \geq \beta_r \geq \bar{\beta}_r\,.
\end{aligned} \tag{A.27}$$

  The positivity of the last term can be checked by observing that the expression $\big(a\bar{\alpha}_r^{n-1} - \alpha_r^{n-1} + a\gamma_r^{n-1}\beta_r^n - \bar{\gamma}_r^{n-1}\bar{\beta}_r^n\big)$ is increasing with respect to the variable a. Then, this will be always positive if it is for $a = 1$. In this sense, we estimate

$$\begin{aligned}
&(1+\tilde{x}+r\tilde{y})^{n-1}-(1+\tilde{x}+\tilde{y})^{n-1}+(1+\tilde{x})^{n-1}(1+\tilde{y})^n-(1+\tilde{x})^{n-1}(1+r\tilde{y})^n \\
&\qquad \geq \big(1+r\tfrac{\tilde{y}}{1+\tilde{x}}\big)^{n-1}-\big(1+\tfrac{\tilde{y}}{1+\tilde{x}}\big)^{n-1}+(1+\tilde{y})^{n-1}-(1+r\tilde{y})^{n-1}
\end{aligned}$$

  where the lower term is positive since using (A.9) we get

$$\begin{aligned}
&(1+\tilde{y})^{n-1} - \left(1+\frac{\tilde{y}}{1+\tilde{x}}\right)^{n-1} \geq (1+r\tilde{y})^{n-1} - \left(1+r\frac{\tilde{y}}{1+\tilde{x}}\right)^{n-1} \\
\iff\ & H_{n-1}\left(1+\tilde{y}, 1+\frac{\tilde{y}}{1+\tilde{x}}\right) \geq \quad rH_{n-1}\left(1+r\tilde{y}, 1+r\frac{\tilde{y}}{1+\tilde{x}}\right)
\end{aligned} \tag{A.28}$$

  which holds because of the increasing character of the operator H_{n-1}.

- Repressor direct logic for the partial cooperative recruitment BEWARE operator

  The fact that the partial derivative of BEWARE_r with respect to the repressor variable is negative when partial cooperativity between activator and repressor can be verified in an analogous manner by the symmetric roles of activators and repressors in the recruitment operators.

- Activator direct logic for the partial cooperative stimulated BEWARE operator Since $Basal(([A],[R]);\{\{A\}_{c_A},\{R\}_{c_R}\})$ has the same expression as the Regulation Factor (3.14) with $a=1$, the sign of the derivatives can estimated following exactly the deduction for the recruitment operator. Hence, we need to check the sign only for the partial derivatives of the $Promoter\big(([A],[R]);\{\{A\}_{c_A},\{R\}_{c_R}\}\big)$ function

$$\frac{(c_R-1)(\delta_s^n-\bar{\delta}_s^n)+\beta_s^n-\bar{\beta}_s^n}{(c_A-1)(c_R-1)+\alpha_s^n+(c_R-1)\lambda_s^n+(c_A-1)\gamma_s^n}$$

  in order to conclude the proof of Lemma 3.2. In previous definition we have used $\alpha_s=\alpha_s(\tilde{x},\tilde{y})=1+\tilde{x}+\tilde{y}$, $\bar{\alpha}_s=\bar{\alpha}_s(\tilde{x},\tilde{y})=1+\tilde{x}+\tilde{r}\tilde{y}$, $\beta_s=\beta_s(\tilde{x},\tilde{y})=\tilde{e}+\tilde{x}+\tilde{e}\tilde{r}\tilde{y}$, $\bar{\beta}_s=\bar{\beta}_s(\tilde{y})=\tilde{e}+\tilde{e}\tilde{r}\tilde{y}$, $\gamma_s=\gamma_s(\tilde{y})=1+\tilde{y}$, $\bar{\gamma}_s=\bar{\gamma}_s(\tilde{y})=1+\tilde{r}\tilde{y}$, $\delta_s=\delta_s(\tilde{x})=\tilde{e}+\tilde{x}$, $\bar{\delta}_s=\bar{\delta}_s=\tilde{e}$, $\lambda_s=\lambda_s(\tilde{x})=1+\tilde{x}$ where, again, we are assuming that each function is evaluated on $\tilde{x}=\frac{c_A[A]}{K_A}$ and $\tilde{y}=\frac{c_R[R]}{K_R}$. The sign of the partial derivative with respect the activator variable will come from the sign of

$$\begin{aligned}
&\text{sign}\left\{\frac{\partial \text{BEWARE}_s\left(([A],[R]),[PolII];\{\{A\}_{c_A},\{R\}_{c_R}\}\right)}{\partial[A]}\right\}\\
&= \text{sign}\Big\{-\Big(\alpha^{n-1}+(c_R-1)\lambda_s^{n-1}\Big)\Big((c_R-1)(\delta_s^n-\bar{\delta}_s^n)+\beta_s^n-\bar{\beta}_s^n\Big)\\
&+ \Big((c_R-1)\delta_s^{n-1}+\beta_s^{n-1}\Big)\Big((c_A-1)(c_R-1)+\alpha_s^n+(c_R-1)\lambda_s^n+(c_A-1)\gamma_s^n\Big)\Big\}
\end{aligned}$$

  which, rearranging the terms, translates to check the sign of

$$\begin{aligned}
&\text{sign}\Big\{(c_R-1)^2\Big((c_A-1)\delta_s^{n-1}+\lambda_s^{n-1}(\lambda_s\delta_s^{n-1}-\delta_s^n+\bar{\delta}_s^n)\Big)\\
&+ \alpha_s^{n-1}(\beta_s^{n-1}\alpha_s-\beta_s^n-\bar{\beta}_s^n)\\
&+ (c_R-1)(c_A-1)\Big(\delta_s^{n-1}\gamma_s^n+\beta_s^{n-1}\Big)+(c_A-1)\gamma_s^n\beta_s^{n-1}\\
&+ (c_R-1)\Big(\delta_s^{n-1}\alpha_s^n+\lambda_s^n\beta_s^{n-1}-\alpha_s^{n-1}(\delta_s^n-\bar{\delta}_s^n)-\lambda_s^{n-1}(\beta_s^n-\bar{\beta}_s^n)\Big)\Big\}\,.
\end{aligned}$$

  The first terms are trivially positive since $\lambda_s>\delta_s$ and $\alpha_s>\beta_s$. Also it is positive the last term that can be rewritten, up to the multiplicative constant (c_R-1), as

$$\lambda_s^{n-1}\beta_s^{n-1}(\lambda_s-\beta_s)+\alpha_s^{n-1}\delta_s^{n-1}(\alpha_s-\delta_s)+\lambda_s^{n-1}\bar{\beta}_s^{n-1}+\alpha_s^{n-1}\bar{\delta}_s^n\geq 0$$

thanks to $|\lambda_s - \beta_s| \leq (\alpha_s - \delta_s)$ and $\alpha_s\delta_s \geq \lambda_s\beta_s$.

- Repressor direct logic for the partial cooperative stimulated BEWARE operator
  On the other hand,

$$\begin{aligned}
&\text{sign}\left\{\frac{\partial \text{BEWARE}_s\Big(([A],[R]),[PolII];\{\{A\}_{c_A},\{R\}_{c_R}\}\Big)}{\partial[R]}\right\}\\
&= \text{sign}\Big\{\Big(\tilde{e}\tilde{r}(\beta_s^{n-1}-\bar{\beta}_s^{n-1})\Big)\Big((c_A-1)(c_R-1)+\alpha_s^n+(c_R-1)\lambda_s^n+(c_A-1)\gamma_s^n\Big)\\
&- \Big(\alpha_s^{n-1}+(c_A-1)\gamma_s^{n-1}\Big)\Big((c_R-1)(\delta_s^n-\bar{\delta}_s^n)+\beta_s^n-\bar{\beta}_s^n\Big)\Big\}
\end{aligned}$$

and rearranging the terms, the same sign can be obtained from

$$\begin{aligned}
&\text{sign}\Big\{(c_A-1)(c_R-1)\Big(\tilde{e}\tilde{r}(\beta_s^{n-1}-\bar{\beta}_s^{n-1})-\gamma_s^{n-1}(\delta^n-\bar{\delta}_s^n)\Big)\\
&+ (c_R-1)\Big((\tilde{e}\tilde{r}\lambda_s)(\lambda_s\beta_s)^{n-1}\Big(1-\frac{\bar{\beta}_s^{n-1}}{\beta_s^{n-1}}\Big)+\delta_s(\delta_s\alpha_s)^{n-1}\Big(\frac{\bar{\delta}_s^n}{\delta_s^n}-1\Big)\Big)\\
&+ (c_A-1)\Big((\gamma_s\beta_s)^{n-1}(\tilde{e}\tilde{r}\gamma_s-\beta_s)+(\gamma_s\bar{\beta}_s)^{n-1}(\bar{\beta}_s-\tilde{e}\tilde{r}\gamma_s)\Big)\\
&+ (\alpha_s\beta_s)^{n-1}(\tilde{e}\tilde{r}\alpha_s-\beta_s)+(\alpha_s\bar{\beta}_s)^{n-1}(\bar{\beta}_s-\tilde{e}\tilde{r}\alpha_s)\Big\}. \qquad \text{(A.29)}
\end{aligned}$$

The first term in (A.29)

$$\begin{aligned}
&\tilde{e}\tilde{r}(\beta_s^{n-1}-\bar{\beta}_s^{n-1})-\gamma_s^{n-1}(\delta_s^n-\bar{\delta}_s^n)\\
&= (\delta_s-\bar{\delta}_s)(\tilde{e}\tilde{r}H_{n-1}(\beta_s,\bar{\beta}_s)-\gamma_s^{n-1}H_n(\delta_s,\bar{\delta}_s))\\
&= (\delta_s-\bar{\delta}_s)(\tilde{e}\tilde{r}H_{n-1}(\beta_s,\bar{\beta}_s)-H_n(\gamma_s\delta_s,\gamma_s\bar{\delta}_s))\\
&= (\delta_s-\bar{\delta}_s)H_n(\gamma_s\delta_s,\gamma_s\bar{\delta}_s)\left(\frac{\tilde{e}\tilde{r}H_{n-1}(\beta_s,\bar{\beta}_s)}{H_n(\gamma_s\delta_s,\gamma_s\bar{\delta}_s)}-1\right)\\
&\leq (\delta_s-\bar{\delta}_s)H_n(\gamma_s\delta_s,\gamma_s\bar{\delta}_s)\left(\frac{\tilde{e}\tilde{r}H_{n-1}(\gamma_s\delta_s,\gamma_s\bar{\delta}_s)}{H_n(\gamma_s\delta_s,\gamma_s\bar{\delta}_s)}-1\right)\leq 0
\end{aligned}$$

is negative because $\beta_s \leq \gamma_s\delta_s$, $\bar{\beta}_s \leq \gamma_s\bar{\delta}_s$ and Lemma A.9 *iii)* is used being $\tilde{e}\tilde{r} \leq \gamma_s\delta_s$. The rest of the terms in (A.29) can be proven to be negative following next estimates

$$\begin{aligned}
(\tilde{e}\tilde{r}\lambda_s)(\lambda_s\beta_s)^{n-1}\Big(1-\frac{\bar{\beta}_s^{n-1}}{\beta_s^{n-1}}\Big)+\delta_s(\delta_s\alpha_s)^{n-1}\Big(\frac{\bar{\delta}_s^n}{\delta_s^n}-1\Big) &\leq \delta_s(\delta_s\alpha_s)^{n-1}\Big(\frac{\bar{\delta}_s^n}{\delta_s^n}-\frac{\bar{\beta}_s^{n-1}}{\beta_s^{n-1}}\Big)\\
(\gamma_s\beta_s)^{n-1}(\tilde{e}\tilde{r}\gamma_s-\beta_s)+(\gamma_s\bar{\beta}_s)^{n-1}(\bar{\beta}_s-\tilde{e}\tilde{r}\gamma_s) &\leq (\gamma_s\bar{\beta}_s)^{n-1}(\bar{\beta}_s-\beta_s)\\
(\alpha_s\beta_s)^{n-1}(\tilde{e}\tilde{r}\alpha_s-\beta_s)+(\alpha_s\bar{\beta}_s)^{n-1}(\bar{\beta}_s-\tilde{e}\tilde{r}\alpha_s) &\leq (\alpha_s\bar{\beta}_s)^{n-1}(\bar{\beta}_s-\beta_s)
\end{aligned}$$

and thanks to

$$\begin{aligned}
&\alpha_s > \gamma_s \geq \bar{\gamma}_s\,, \qquad \lambda_s\beta_s \leq \alpha_s\delta_s\,, \qquad \tilde{e}\tilde{r}\lambda_s \leq \delta_s\,, \qquad 1 \geq \frac{\bar{\beta}_s}{\beta_s} \geq \frac{\bar{\delta}_s}{\delta_s}\,,\\
&\alpha_s > \delta_s \geq \bar{\delta}_s\,, \qquad \bar{\beta}_s \leq \beta_s\,, \qquad \tilde{e}\tilde{r}\gamma_s \leq \beta_s\,, \qquad \tilde{e}\tilde{r}\alpha_s \leq \beta_s\,. \qquad \text{(A.30)}
\end{aligned}$$

□

A.4 Existence of threshold

This appendix is devoted to proving the existence of activation/repression thresholds implicitly deduced from the BEWARE operators.

Proposition 4.1 *In the global activator/repressor framework each BEWARE operator, determines an unique, positive and increasing function*

$$[R] = f_{m,l}([A]; n)$$

fulfilling (3.43). *This function determines the threshold between two regions in the plane* $([A], [R])$. *Each region contains concentrations providing transcriptional levels either over the basal level, if* $[R] < f_{m,l}([A]; n)$, *or under the basal level, when* $[R] > f_{m,l}([A]; n)$.

These threshold functions depend on all the biochemical factors considered in the derivation of the BEWARE operator: the Recruitment $(m = r)$ (3.13) *or Stimulated* $(m = s)$(3.16) *(with* $r_{bas} > 0$*) approaches, the binding cooperativity mechanisms between the TFs (cooperative, total cooperative* $l = t$ *or partial cooperative* $l = p$*) or the number of enhancers* (n). *The particular case of the dependence with respect to the affinities coefficients* K_A-K_R *is given by*

$$f_{m,l}([A]; n) := K_R \tilde{f}_{m,l}\left(\frac{[A]}{K_A}; n\right) \tag{A.31}$$

where $\tilde{f}_{m,l}$ *is independent of both,* K_A *and* K_R.

Remark 4.1 *A very easy example can be shown in the case of the BEWARE operators with null/total cooperativity* (3.8). *Because of the increasing character of the recruitment operator* (3.13), *with respect to the regulation factor* F_{reg} *equation* (3.43) *translates directly to*

$$F_{reg}\big(([A], f_{r,t}([A]; n)); \mathcal{C}\big) = F_{reg}\big((0, 0); \mathcal{C}\big) = 1\,. \tag{A.32}$$

When we substitute the definition of the regulation factor (3.35) *in this expression we directly get that the threshold corresponds to the linear expression*

$$f_{r,t}([A]; n) = \frac{K_R}{K_A}\frac{a-1}{1-r}[A]\,.$$

Let us remark that, although in expression (3.35) *the cooperativity constant* c *and the number of enhancers* n *are present, they are not in this threshold expression. That implies that the thresholds for this model are the independent of the intensity of the total binding cooperativity between the species or the number of enhancers.*

Remark 4.2 *A similar argument can be performed for the stimulated BEWARE operator when $n = 1$. Obviously, this is independent of any kind of binding cooperativity because it can not occurs when only one binding site is available. In this case, direct computations gives rise to*

$$f_{s,t}([A];1) = \frac{K_R}{K_A}\frac{\nu_{max}^{(1)}}{r_{bas}}\frac{1}{1-\tilde{r}}[A]\,.$$

As we will mention in the subsequent the thresholds under the stimulated approach with total cooperativity are not in general independent of the number of enhancers, which will represent a remarkable difference between the stimulated and recruitment approach.

Remark 4.3 *Proposition 4.1 is no longer true for Stimulated operators when r_{bas} is null as can be trivially deduced from expressions* (3.37) *and* (3.40)*. In this case the threshold coincides with the axis $[A] = 0$, and any pair concentrations $([A],[R])$ with $[A] > 0$ leads to activation levels. Nevertheless, we have to remark that this threshold does not depends on cooperativity relations, number of enhancers nor affinities.*

Sketch of the Proof of Proposition 4.1

Here, we have adopted the notation $x = \frac{[A]}{K_A}$, $y = \frac{[R]}{K_R}$ for convenience. Let us mention that the threshold for the recruitment BEWARE operator with n enhancers and null/total cooperativity was already calculated in Remark 4.1 solving explicitly the equation

$$G_{r,t}(x,y;n) = (1 + acx + rcy)^n - (1 + cx + cy)^n = 0 \tag{A.33}$$

which is equivalent to (A.32). In current notation the solution to this equation is given by:

$$y = \tilde{f}_{r,t}(x;n) = \frac{a-1}{1-r}x\,, \tag{A.34}$$

where $\tilde{f}_{r,t}(x;n)$ is the function stated in (A.31). Now, undoing our original change of variable $x = \frac{[A]}{K_A}$, $y = \frac{[R]}{K_R}$ we recover the expression deduced in Remark 4.1 .

Although, in this case the the definition of these thresholds can be done explicitly, we would like to remark that the general existence result provided by the implicit function theorem provides very useful information for subsequents analysis. The argument we adopt follows the same scheme for all the BEWARE functionals considered, so we now introduce the outlines of the general proof and in sub-A.4.1 we check the validity of each particular model dependent requirements.

When we substitute the concrete expressions of the BEWARE operators (3.13)-(3.16) with their corresponding expressions for regulation factors, basal and promoter functions (see (3.35)-(3.40)) into equation (3.43) this is

equivalent to a equation $G_{m,l}\left(\frac{[A]}{K_A}, \frac{f([A])}{K_R}; n\right) = 0$. In the case of the recruitment BEWARE operators we get

$$\begin{aligned} G_{r,t}(x,y;n) &= \bar{\alpha}_r(cx,cy)^n - \alpha_r(cx,cy)^n\,, \\ G_{r,p}(x,y;n) &= \bar{\alpha}_r(c_A x, c_R y)^n - \alpha_r(c_A x, c_R y)^n + (c_A - 1)(\bar{\beta}_r(c_R y)^n - \beta_r(c_R y)^n) \\ &+ (c_R - 1)(\gamma_r(c_A x)^n - \bar{\gamma}_r(c_A x)^n)\,, \end{aligned} \tag{A.35}$$

being $\bar{\alpha}_r(X,Y) = (1 + aX + rY)$, $\alpha_r(X,Y) = (1 + X + Y)$, $\bar{\beta}_r(Y) = 1 + rY$, $\beta_r(Y) = 1 + Y$, $\bar{\gamma}_r(X) = 1 + X$ and $\gamma_r(X) = 1 + aX$. In the stimulated cases we obtain in the same way

$$\begin{aligned} G_{s,t}(x,y;n) &= r_{bas}(\bar{\alpha}_s^n(cx,cy) - \alpha_s^n(cx,cy)) + \nu_{max}^{(n)}\left(\beta_s^n(cx,cy) - \bar{\beta}_s^n(cy)\right), \\ G_{s,p}(x,y;n) &= r_{bas}\left(\bar{\alpha}_s^n(c_A x, c_R y) - \alpha_s^n(c_A x, c_R y) + (c_A - 1)(\bar{\gamma}_s^n(c_R y) - \gamma_s^n(c_R y))\right) \\ &+ \nu_{max}^{(n)}\left(\beta_s^n(c_A x, c_R y) - \bar{\beta}_s^n(c_R y) + (c_R - 1)\left(\delta_s^n(c_A x) - \bar{\delta}_s^n\right)\right) \end{aligned} \tag{A.36}$$

where now $\bar{\alpha}_s(X,Y) = 1 + X + \tilde{r}Y$, $\alpha_s(X,Y) = 1 + X + Y$, $\bar{\beta}_s(Y) = \tilde{e} + \tilde{e}\tilde{r}Y$, $\beta_s(X,Y) = \tilde{e} + X + \tilde{e}\tilde{r}Y$. $\bar{\gamma}_s(Y) = 1 + \tilde{r}Y$, $\gamma_s(Y) = 1 + Y$, $\bar{\delta}_s = \tilde{e}$ and $\delta_s(X) = \tilde{e} + X$. As it has been done in (A.33)-(A.34) G functions and the corresponding thresholds f will be denoted with subindices: s, r, t, p, corresponding to *s*timulated, *r*ecruitment, *t*otal/null cooperativity and *p*artial cooperativity respectively. Some other dependences can be included whenever necessary by using parameters, as for instance $G_{m,l}(\cdot,\cdot;n)$ determining that G a bivariate polynomial function of order n. In the subsequent, subindexes as well as the parameter will be skipped in all those cases where they are not relevant. Let us also notice that all the functions G have been defined such that

$$G\left(\frac{[A]}{K_A}, \frac{[R])}{K_R}\right) > 0 \iff BEWARE([A],[R]) > \text{basal level.} \tag{A.37}$$

In order to prove the existence of an implicit function determined by the equation $G(x,y) = 0$ for each of these problems, we propose to check next items:

a) for any $x > 0$, the equation

$$G(x,\cdot) = 0 \tag{A.38}$$

function has at least one root because $G(x,0) > 0$ and $\lim_{y\to\infty} G(x,y) = -\infty$,

b) for any (x,y) root of the equation (A.38) then $\frac{\partial G}{\partial y}(x,y) < 0$,

c) and for any (x,y) root of the equation (A.38) then $\frac{\partial G}{\partial x}(x,y) > 0$.

For instance, items a), b) and c) can be easily checked for $G_{r,t}$ defined in (A.33) thanks to $r < 1 < a$. These ingredients allow us to conclude the proof easily. Assertion b) implies the uniqueness of the roots stated in a) because $G(x,\cdot)$ is strictly decreasing at any of them. Since the partial derivative with respect to y is non-zero, the implicit function theorem assures that, given a point (x_1, y_1) such that $G(x_1, y_1) = 0$, then in some small enough neighbourhood of (x_1, y_1) there exist a parametrisation $(x, \tilde{f}(x))$ such that $G(x, \tilde{f}(x)) = 0$. This really justifies that, the function $\tilde{f}(x)$ is globally defined and unique. Finally, this function is also monotone increasing because of b) and c), since

$$\tilde{f}'(x) = -\frac{\frac{\partial G}{\partial x}}{\frac{\partial G}{\partial y}} > 0 \tag{A.39}$$

Let us see that these functions define the thresholds. Coming back to our original notation, the function f stated in Proposition 4.1 takes the value expression given by (A.31). These thresholds will depend as much as from the cooperative relations between TFs: not cooperative/total cooperative (t) or partial cooperative (p) as from the recruitment (r) or stimulated (s) approaches reason why we will add subindexes $t/p, r/s$ to f function denoting any threshold, as it was done in definition (A.34). It is also true that the thresholds will depend on the number of enhancers n which will be introduced as a parameter dependence $f_{m,l}([A]; n)$. □

A.4.1 Existence of thresholds for bifunctional beware operators in the activator/repressor framework

The rest of the Appendix is devoted to declaring the functions G determining the activation/repression thresholds and check the hypothesis $a) - c)$ they have to verify in order to conclude Proposition 4.1 for all the considered BEWARE models.

Proof of Proposition 4.1

- Stimulated BEWARE operator with null/total cooperativity ($r_{bas} > 0$)

  From definitions (3.16) and the corresponding basal level, see Table 3.1, we get that equation (3.43) in this case study translates into

  $$r_{bas}\big(Basal([A],[R]);\mathcal{C})-1\big)+\nu^{(n)}_{max}Promoter([A],[R]);\mathcal{C}) = 0\,. \tag{A.40}$$

  Replacing definition Basal and Promoter functions according to expressions (3.36) and (3.37), equation (A.40) leads to

  $$\begin{aligned} & G_{s,t}(x,y;n) \qquad\qquad (A.41)\\ &= r_{bas}\left(\bar{\alpha}^n_s(cx,cy) - \alpha^n_s(cx,cy)\right) + \nu^{(n)}_{max}\left(\beta^n_s(cx,cy) - \bar{\beta}^n_s(cy)\right) = 0\,, \end{aligned}$$

where $x = \frac{[A]}{K_A}$, $y = \frac{[R]}{K_R}$, $\bar{\alpha}_s(X,Y) = 1+X+\tilde{r}Y$, $\alpha_s(X,Y) = 1+X+Y$, $\bar{\beta}_s(Y) = \tilde{e} + \tilde{e}\tilde{r}Y$, $\beta_s(X,Y) = \tilde{e} + X + \tilde{e}\tilde{r}Y$. Let us assume in the subsequent that $n \geq 2$ since Proposition 4.1 is obviously true for $n = 1$. For simplicity we will note $\alpha_s = \alpha_s(cx, cy)$, $\bar{\alpha}_s = \bar{\alpha}_s(cx, cy)$, $\beta_s = \beta_s(cx, cy)$ and $\bar{\beta}_s = \bar{\beta}_s(cy)$. Using estimates (A.15) it is easy to prove a), that is, $G_{s,t}(x, 0; n) > 0$ and $\lim_{y\to+\infty} G_{s,t}(x, y; n) = -\infty$, so equation (A.41) has at least one real root.

A similar procedure can be used in order to prove b). In this case

$$\begin{aligned} &\text{sign}\left\{\frac{\partial G_{s,t}}{\partial y}\right\} \\ &\quad = \text{sign}\left\{\left(r_{bas}\left(\tilde{r}\bar{\alpha}_s^{n-1} - \alpha_s^{n-1}\right) + \nu_{max}^{(n)}\tilde{e}\tilde{r}\left(\beta_s^{n-1} - \bar{\beta}_s^{n-1}\right)\right)\right\}. \end{aligned} \tag{A.42}$$

Obviously b) holds when $n = 1$ because $\tilde{r} < 1$. When $n > 1$, Let us observe that we know that under certain circumstances the right hand side on (A.42) can be positive by (A.25). Nevertheless, we can prove that this does not occurs at the solutions of (A.41). From (A.25), we get by simple algebra

$$\begin{aligned} &y\left(r_{bas}\left(\tilde{r}\bar{\alpha}_s^{n-1} - \alpha_s^{n-1}\right) + \nu_{max}^{(n)}\tilde{e}\tilde{r}\left(\beta_s^{n-1} - \bar{\beta}_s^{n-1}\right)\right) \\ &\quad = -r_{bas}(1+x)\left(\tilde{r}\bar{\alpha}_s^{n-1} - \alpha_s^{n-1}\right) - \nu_{max}^{(n)}\tilde{e}\left(\beta_s^{n-1} - \bar{\beta}_s^{n-1}\right) - \nu_{max}^{(n)}x\beta_s^{n-1} \\ &\quad < -(1+x)\left(r_{bas}\left(\tilde{r}\bar{\alpha}_s^{n-1} - \alpha_s^{n-1}\right) + \nu_{max}^{(n)}\tilde{e}\tilde{r}\left(\beta_s^{n-1} - \bar{\beta}_s^{n-1}\right)\right). \end{aligned} \tag{A.43}$$

On the other hand, we also can directly overestimate $\tilde{r}\bar{\alpha}_s^{n-1}$ by taking $\tilde{r} = 1$ since $\tilde{r} < 1$ and $\bar{\alpha}_s > 0$

$$\begin{aligned} &\left(r_{bas}\left(\tilde{r}\bar{\alpha}_s^{n-1} - \alpha_s^{n-1}\right) + \nu_{max}^{(n)}\tilde{e}\tilde{r}\left(\beta_s^{n-1} - \bar{\beta}_s^{n-1}\right)\right) \\ &\quad \leq \left(r_{bas}\left(\bar{\alpha}_s^{n-1} - \alpha_s^{n-1}\right) + \nu_{max}^{(n)}\tilde{e}\tilde{r}\left(\beta_s^{n-1} - \bar{\beta}_s^{n-1}\right)\right). \end{aligned} \tag{A.44}$$

Since the upper bound in (A.43) coincides with the upper bound in (A.44) multiplied by a negative constant we can deduce that the right hand side on (A.42) is negative for $y > 0$. That is, b) is verified.

In order to check c) we have that

$$\text{sign}\left\{\frac{\partial G_{s,t}}{\partial x}\right\} = \text{sign}\left\{\left(r_{bas}\left(\bar{\alpha}_s^{n-1} - \alpha_s^{n-1}\right) + \nu_{max}^{(n)}\beta_s^{n-1}\right)\right\}. \tag{A.45}$$

Now, by using (A.9) and (A.41) as before we get

$$\begin{aligned}
&\left(\nu_{max}^{(n)}\beta_s^{n-1} + r_{bas}\left(\bar{\alpha}_s^{n-1} - \alpha_s^{n-1}\right)\right) \\
&= \nu_{max}^{(n)}\left(\beta_s^{n-1} - \frac{H_{n-1}(\alpha_s,\bar{\alpha}_s)}{H_n(\alpha_s,\bar{\alpha}_s)}\left(\beta_s^n - \bar{\beta}_s^n\right)\right) \\
&= \nu_{max}^{(n)}\beta_s^{n-1}\left(1 - \frac{H_{n-1}(\alpha_s,\bar{\alpha}_s)}{H_n(\alpha_s,\bar{\alpha}_s)}\beta_s\right) \\
&+ \nu_{max}^{(n)}\frac{H_{n-1}(\alpha_s,\bar{\alpha}_s)}{H_n(\alpha_s,\bar{\alpha}_s)}\bar{\beta}_s^n\,,
\end{aligned}$$

which is trivially positive, thanks to to Lemma 3.3 *iii)* using $\beta_s < \alpha_s$ from (A.15). Hence, $\left.\frac{\partial G_{s,t}}{\partial x}\right|_{G_{s,t}=0} > 0$. Thus, we can assert the existence of a unique increasing function $\tilde{f}_{s,t}$ such that

$$G_{s,t}\big(x, \tilde{f}_{s,t}(x;n);n\big) = 0 \quad x \in (0,\infty) \tag{A.46}$$

for any $c > 0$.

- Stimulated BEWARE operators with partial cooperativity ($r_{bas} > 0$)

  Now, in the case of BEWARE operators in presence of partial cooperativity the proofs are much more easy because we can take advantage of the direct activator/repressor logic verified by these models as was stated in Lemma 3.2.

  The expression defining the threshold condition (3.43) leads to:

$$\text{BEWARE}_s([A],[R],[PolII];\{A\}_{c_A},\{R\}_{c_R}) - \frac{r_{bas}}{1+\frac{K_{RP}}{[PolII]}} = 0\,.$$

  Then, multiplying this equation by $1+\frac{K_{RP}}{[PolII]}$ times the denominator of the basal and promoter expressions, $Den_s([A],[R])$, and substituting the corresponding expressions of the basal and promoter functionals,(3.39) and (3.40), we get

$$\begin{aligned}
0 = &Den_s([A],[R])\times \\
&\times \left(\left(1+\tfrac{K_{RP}}{[PolII]}\right)\text{BEWARE}_s([A],[R],[PolII];\{A\}_{c_A},\{R\}_{c_R}) - r_{bas}\right) \\
&= r_{bas}(\bar{\alpha}_s^n(c_Ax,c_Ry) - \alpha_s^n(c_Ax,c_Ry) + (c_A-1)(\bar{\gamma}_s^n(c_Ry) - \gamma_s^n(c_Ry))) \\
&+ \nu_{max}^{(n)}\big(\beta_s^n(c_Ax,c_Ry) - \bar{\beta}_s^n(c_Ry) + (c_R-1)\big(\delta_s^n(c_Ax) - \bar{\delta}_s^n\big)\big) \\
&= G_{s,p}(x,y;n)\,,
\end{aligned} \tag{A.47}$$

  where, in addition to the functions α_s, $\bar{\alpha}_s$, β_s and $\bar{\beta}_s$ appearing in (A.41), we now also have $\bar{\gamma}_s(Y) = 1 + \tilde{r}Y$, $\gamma_s(Y) = 1 + Y$, $\bar{\delta}_s = \tilde{e}$ and

$\delta_s(X) = \tilde{e} + X$. Equation (A.47) has at least one real root for any $x > 0$, since $G_{s,p}(x, 0; n) > 0$ and $\lim_{y\to+\infty} G_{s,p}(x, y; n) = -\infty$, that is, a) holds.

We can now check b) and c) very easily by observing that

$$\operatorname{sign}\left\{\left.\frac{\partial G_{s,p}}{\partial y}\right|_{G_{s,p}=0}\right\} = \operatorname{sign}\left\{\frac{\partial \text{BEWARE}_s}{\partial [R]}([A],[R];\{PolII\};\{A\}_{c_A},\{R\}_{c_R})\right\} < 0\,,$$

$$\operatorname{sign}\left\{\left.\frac{\partial G_{s,p}}{\partial x}\right|_{G_{s,p}=0}\right\} = \operatorname{sign}\left\{\frac{\partial \text{BEWARE}_s}{\partial [A]}([A],[R];\{PolII\};\{A\}_{c_A},\{R\}_{c_R})\right\} > 0\,,$$

thanks to identities (A.47) and Lemma 3.2 *ii)* proved in A.3.2.

- Recruitment BEWARE operators with partial cooperativity

  As it was introduced in Remark 4.1 the threshold equation, (3.43), translates directly to (A.32) for Recruitment BEWARE operators.

  Then, by multiplying this equation by the denominator of the regulation factor (3.38), $Den_r([A],[R])$, we get the equivalent expression

$$\begin{aligned} 0 = Den_r([A],[R])\Big(F_{reg}\big(([A],[R]);\{\{A\}_{c_A},\{R\}_{c_R}\}\big) - 1\Big) \\ = \bar{\alpha}_r(c_A x, c_R y)^n - \alpha_r(c_A x, c_R y)^n + (c_A - 1)(\bar{\beta}_r(c_R y)^n - \beta_r(c_R y)^n) \\ + (c_R - 1)(\gamma_r(c_A x)^n - \bar{\gamma}_r(c_A x)^n) = G_{r,p}(x, y; n) \end{aligned} \tag{A.48}$$

  with $\bar{\alpha}_r(X,Y) = (1 + aX + rY)$, $\alpha_r(X,Y) = (1 + X + Y)$, $\bar{\beta}_r(Y) = 1 + rY$, $\beta_r(Y) = 1 + Y$, $\bar{\gamma}_r(X) = 1 + X$ and $\gamma_r(X) = 1 + aX$. Note that equation (A.48) verifies a) since $G_{r,p}(x, 0; n) > 0$ and $\lim_{y\to+\infty} G_{r,p}(x, y; n) = -\infty$ for any $x > 0$ thanks to $a > 1$ and $r < 1$ are assumed.

  In this case, the testing $b)-c)$ is again trivial by using identities (A.48) and Lemma 3.2 *ii)* as indicated

$$\operatorname{sign}\left\{\left.\frac{\partial G_{r,p}}{\partial y}\right|_{G_{r,p}=0}\right\} = \operatorname{sign}\left\{\frac{\partial F_{reg}}{\partial [R]}\big(([A],[R]);\{\{A\}_{c_A},\{R\}_{c_R}\}\big)\right\} < 0\,,$$

$$\operatorname{sign}\left\{\left.\frac{\partial G_{r,p}}{\partial x}\right|_{G_{r,p}=0}\right\} = \operatorname{sign}\left\{\frac{\partial F_{reg}}{\partial [A]}\big(([A],[R]);\{\{A\}_{c_A},\{R\}_{c_R}\}\big)\right\} > 0\,.$$

□

A.5 Proof of equation (3.50)

In this Appendix it can be found the results linking the elasticity functions with perturbations (3.47) and (3.48) as well as the elasticity estimates summarised in Table 3.2.

Lemma 5.4 *The sense of variation of the threshold function under the affinity perturbation* (3.47) *is determined by the elasticity function* (3.49)*, by the expression*

$$\frac{\delta f_{m,l}}{\delta \eta}([A];n)\Big|_{\eta=1} = f_{m,l}([A];n)\Big(1-\epsilon_{m,l}([A];n)\Big)\,.$$

Proof. This result is a trivial consequence of equation (A.31) because that expression allow us to directly compute the variability of the threshold under perturbation (3.47)

$$\begin{aligned}\frac{\partial f_{m,l}}{\partial \eta}([A];n) &= \frac{\partial}{\partial \eta}\left(\eta K_R \tilde{f}_{m,l}\left(\frac{[A]}{\eta K_A};n\right)\right)\\ &= K_R\tilde{f}_{m,l}\left(\frac{[A]}{\eta K_A};n\right) - \eta K_R \frac{[A]}{\eta^2 K_A}\tilde{f}'_{m,l}\left(\frac{[A]}{\eta K_A};n\right)\end{aligned}$$

and in consequence

$$\frac{\delta f_{m,l}}{\delta \eta}([A];n)\Big|_{\eta=1} = f_{m,l}([A];n) - [A]f'_{m,l}([A];n) = f_{m,l}([A];n)\big(1-\epsilon_{m,l}([A];n)\big)\,,$$

where we have used $f'_{m,l}([A];n) = \frac{K_R}{K_A}\tilde{f}'_{m,l}\left(\frac{[A]}{K_A};n\right)$. □

Corollary 5.1 *The thresholds determined by stimulated and recruitment BE-WARE operators* ($m = s/r$)*,* (3.16) *and* (3.13) *respectively, considering* (3.35)-(3.40) ($l = t/p$) *and assumption* (3.15)*, change in the same manner under affinity and enhancers number reductions, that is* (3.47) *and* (3.48) *since:*

$$\begin{aligned}sign\frac{\delta f_{m,l}}{\delta \eta}([A];n) &= sign\left(G_{m,l}\left(x,y;n-1\right)\big|_{x=\frac{[A]}{K_A},y=\frac{f_{m,l}([A];n)}{K_R}}\right),\\ &= sign\Big(f_{m,l}([A];n-1) - f_{m,l}([A];n)\Big).\end{aligned}$$

Furthermore, in both cases, the limit behaviours they would tend to in the case of very low affinities ($\eta \to \infty$) *and on the case of one only available enhancer coincides, that is,*

$$\lim_{\eta\to\infty} f_{m,l}([A];n) = \frac{\partial f_{m,l}}{\partial [A]}([A]=0;n)[A] = f_{m,l}([A];1).$$

Proof.

Combining Lemma 5.4 with expression (A.39) we have that for any BE-WARE operator it is verified

$$\begin{aligned}\tfrac{\delta f_{m,l}}{\delta \eta}([A];n) &= \tfrac{1}{\eta}\Big(\eta K_R\tilde{f}_{m,l}\Big(\tfrac{[A]}{\eta K_A};n\Big) - [A]\tfrac{K_R}{K_A}\tilde{f}'_{m,l}\Big(\tfrac{[A]}{\eta K_A};n\Big)\Big)\\ &= \tfrac{1}{\eta\frac{\partial G_{m,l}}{\partial y}}\Big(y\tfrac{\partial G_{m,l}}{\partial y}(x,y;n) + x\tfrac{\partial G_{m,l}}{\partial x}(x,y;n)\Big)\Big|_{x=\frac{[A]}{\eta K_A},\, y=\eta K_R\tilde{f}_{m,l}\left(\frac{[A]}{\eta K_A};n\right)}\end{aligned}$$

which gives

$$\frac{\delta f_{m,l}}{\delta \eta}([A];n)\Big|_{\eta=1} = \frac{1}{\frac{\partial G_{m,l}}{\partial y}} \left(y\frac{\partial G_{m,l}}{\partial y}(x,y;n) + x\frac{\partial G_{m,l}}{\partial x}(x,y;n) \right)\Big|_{x=\frac{[A]}{K_A}, y=\frac{f_{m,l}([A];n)}{K_R}} \quad \text{(A.49)}$$

Now, simple algebraic computations allow to rewrite this expression since

$$y\frac{\partial G_{m,l}}{\partial y}(x,y;n) + x\frac{\partial G_{m,l}}{\partial x}(x,y;n) \quad = \quad nG_{m,l}(x,y;n) - nG_{m,l}(x,y;n-1)\,.$$

Let us remark that this expression is only valid in the stimulated approach if hypothesis (3.15) is assumed.

The first identity of the statement of this Corollary can be trivially deduced from these expressions because

$$G_{m,l}\,(x,y;n)\big|_{x=\frac{[A]}{K_A}, y=\frac{f_{m,l}([A];n)}{K_R}} = 0$$

holds by definition and, as mentioned in the proof of Proposition 4.1, it is also verified that:

$$\frac{\partial G_{m,l}}{\partial y}\,(x,y;n)\Big|_{x=\frac{[A]}{K_A}, y=\frac{f_{m,l}([A];n)}{K_R}} < 0\,.$$

The second identity is also obvious since

$$G_{m,l}\,(x,y;n-1)\big|_{x=\frac{[A]}{K_A}, y=\frac{f_{m,l}([A];n)}{K_R}} > 0 \iff f_{m,l}([A];n) < f_{m,l}([A];n-1)\,.$$

Let us now justify the thresholds limit in the case of very low affinity, that is, $\eta \to \infty$

$$\begin{aligned} \lim_{\eta\to\infty} f_{m,l}([A];n) &= \lim_{\eta\to\infty} \eta\tilde{f}_{m,l}\left(\frac{[A]}{\eta K_A};n\right) \\ &= \frac{K_R}{K_A}[A] \lim_{\eta\to\infty} \frac{\tilde{f}_{m,l}\left(\frac{[A]}{\eta K_A};n\right) - \tilde{f}_{m,l}\,(0;n)}{\frac{[A]}{\eta K_A}} = \frac{K_R}{K_A}\tilde{f}'_{m,l}(0;n)[A]\,. \end{aligned} \quad \text{(A.50)}$$

From (A.39) we directly get $\tilde{f}'(0) = -\frac{\frac{\partial G}{\partial x}}{\frac{\partial G}{\partial y}}(0,0)$. By a direct substitution of G functions expressions for any BEWARE operator we get

$$\tilde{f}'_{r,*}(0;n) = \frac{a-1}{1-r} \qquad \tilde{f}'_{s,*}(0;n) = \frac{\nu^{(n)}_{max}\tilde{e}^{n-1}}{r_{bas}(1-\tilde{r})} = \frac{\nu^{(1)}_{max}}{\tilde{r}_{bas}(1-r)}$$

respectively for the recruitment and stimulated approaches. Substituting these expressions in (A.50) we conclude the proof thanks to Remarks 4.1 and 4.2 in A.4. □

These formulas stated in this Corollary confirms that the reaction to perturbation (3.47) is strongly related with the variation of the same threshold under perturbation (3.48). Indeed, in all these expressions

$$G_{m,l}\left(x,y;n-1\right)|_{x=\frac{[A]}{K_A},y=\frac{f_{m,l}([A];n)}{K_R}} \tag{A.51}$$

gives the value of the function G when the BEWARE operator has only $n-1$ enhancers evaluated on the threshold obtained for the same BEWARE operator but with n enhancers. The sign of this computation is directly related to the relative positions of the thresholds for n or $n-1$ enhacers because of (A.37).

Our last Lemma is devoted to showing how different are the predicted behaviours to perturbations in affinity or the number of enhancers, that is (3.47) or (3.48), depending on the stimulated or recruitment approach adopted to deduce the BEWARE operator. Our results suggests that under the recruitment approach the behaviour of the thresholds is determined basically by the kind of binding cooperativity relationships between the TFs, in the sense that if some TFs specie cooperate only between them then this is a competitive advantage that perturbations (3.47)-(3.48) interrupt. On the other side, the thresholds deduced from stimulated BEWARE operators are not only dependent on the cooperativity relations but also on the value of the parameter $\tilde{e}$. This, a priory harmless parameter, has been proven to change the threshold elasticities provoking qualitative different responses, that is, it is able to alter the competitive advantages that cooperativities between TFs provide in an unclear way. The elasticity can be analitically estimated for thresholds deduced from BEWARE models for both stimulated and recruitment approaches in some cooperativity specific regimes: null/-total cooperativity and in those cases when either only activators or only repressors can cooperate. All these estimates are collected in Table 3.2.

Now we focus on checking the elasticity estimates in Table 3.2 by considering separately the recruitment and stimulated approaches.

A.5.1 Deduction of elasticity estimates in Table 3.2: recruitment case.

As we have shown in (A.32), the threshold functions for the recruitment operator with total cooperativity are straight lines independent on the number of enhancers n. Both facts allow us to deduce the unit elastic character of those thresholds.

Imposing $c_A = 1$ and $c_R > 1$ in definition (A.35) we get that the the implicit equation for the threshold $f_{r,p}(x;n)$ in the repressor cooperative case is

$$G_{r,p}(x,y;n)|_{x=\frac{[A]}{K_A},y=\frac{f_{r,p}([A];n)}{K_R}} = \bar{\alpha}_r^n - \alpha_r^n + (c_R-1)(\gamma_r^n - \bar{\gamma}_r^n) = 0 \tag{A.52}$$

where $\bar{\alpha}_r$, α_r, γ_r and $\bar{\gamma}_r$ are evaluated at the points $\left(\frac{[A]}{K_A}, c_R\frac{f_{r,p}([A];n)}{K_R}\right)$. Please note that, since $\gamma_r \geq \bar{\gamma}_r$ by definition, then $\alpha_r \geq \bar{\alpha}_r$ and we can order the functions such as

$$\alpha_r \geq \bar{\alpha}_r \geq \gamma_r \geq \bar{\gamma}_r\,. \tag{A.53}$$

The idea of the proof is to relate $G_{r,p}(\cdot,\cdot;n)$ with $G_{r,p}(\cdot,\cdot;n-1)$ by considering the function

$$g(z) = z^{\frac{n}{n-1}}\,. \tag{A.54}$$

It is trivial to see that, by the mean value theorem,

$$\alpha_r^n - \bar{\alpha}_r^n = g(\alpha_r^{n-1}) - g(\bar{\alpha}_r^{n-1}) = g'(c)(\alpha_r^{n-1} - \bar{\alpha}_r^{n-1}) = \frac{n}{n-1}c^{\frac{1}{n-1}}(\alpha_r^{n-1} - \bar{\alpha}_r^{n-1}) \tag{A.55}$$

for a $c \in \mathbb{R}$ such as $\alpha_r^{n-1} > c > \bar{\alpha}_r^{n-1}$. The same goes for $\gamma_r^n - \bar{\gamma}_r^n$, where

$$\gamma_r^n - \bar{\gamma}_r^n = \frac{n}{n-1}\tilde{c}^{\frac{1}{n-1}}(\gamma_r^{n-1} - \bar{\gamma}_r^{n-1}) \tag{A.56}$$

for a $\tilde{c} \in \mathbb{R}$ such as $\gamma_r^{n-1} > \tilde{c} > \bar{\gamma}_r^{n-1}$. Obviously, inequalities (A.53) imply $c > \tilde{c}$. Taking this relation into account, if we replace (A.55)-(A.56) in (A.52), we obtain the following inequality

$$\frac{n}{n-1}c^{\frac{1}{n-1}}\left(\bar{\alpha}_r^{n-1} - \alpha_r^{n-1} + (c_R-1)(\gamma_r^{n-1} - \bar{\gamma}_r^{n-1})\right) > 0\,,$$

which implies

$$G_{r,p}(x,y;n-1)\big|_{x=\frac{[A]}{K_A}, y=\frac{f_{r,p}([A];n)}{K_R}} > 0$$

and hence the inelastic character ($\epsilon_{r,p} < 1$) of the threshold $f_{r,p}([A];n)$ in the repressor cooperativity case by Lemma 5.4 and Corollary 5.1.

Analogously, we get in the opposite regime, $c_R = 1$ and $c_A > 1$, that

$$G_{r,p}(x,y;n)\big|_{x=\frac{[A]}{K_A}, y=\frac{f_{r,p}([A];n)}{K_R}} = \bar{\alpha}_r^n - \alpha_r^n + (c_A-1)(\bar{\beta}_r^n - \beta_r^n) = 0$$

where now $\bar{\beta}_r < \beta_r < \alpha_r < \bar{\alpha}$ holds. Then, by using the function g in the same way we can prove:

$$G_{r,p}(x,y;n-1)\big|_{x=\frac{[A]}{K_A}, y=\frac{f_{r,p}([A];n)}{K_R}} = \bar{\alpha}_r^{n-1} - \alpha_r^{n-1} + (c_R-1)(\gamma_r^{n-1} - \bar{\gamma}_r^{n-1}) < 0\,,$$

which, using Lemma 5.4 and Corollary 5.1, proves the elastic character ($\epsilon_{r,p} > 1$) of the threshold $f_{r,p}([A];n)$ whenever $c_R = 1$ and $c_A > 1$.

A.5.2 Deduction of elasticity estimates in Table 3.2, Stimulated case ($n = 2$)

Recalling from (4.2) that the threshold $f_{s,p}(x;1)$ is a straight line of slope $m = \frac{\nu^{(1)}_{max}}{r_{bas}}\frac{1}{1-\tilde{r}}\frac{K_R}{K_A}$, and evaluating the implicit threshold function (A.47) on it, it is easy to check that

$$G_{s,p}(x, mx) = K_0 + K_1 x + K_2 x^2 \tag{A.57}$$

with

$$K_i = \frac{2!}{i!(2-i)!}\Bigg(r_{bas}\left((1+m\tilde{r})^i-(1+m)^i+(c_A-1)\left((\tilde{r}m)^i-m^i\right)\right) \\ +\nu^{(2)}_{max}\tilde{e}^2\left(\left(\frac{1}{\tilde{e}}+m\tilde{r}\right)^i-(m\tilde{r})^i+(c_R-1)\left(\frac{1}{\tilde{e}}\right)^i\right)\Bigg). \tag{A.58}$$

In the null cooperativity case, $c_A = c_R = 1$, we get $K_0 = K_1 = 0$ and in consequence

- $G_{s,p}(x, mx) = 0$ and $\epsilon = 1\ \forall x \geq 0 \iff K_2 = 0 \iff \frac{1}{\tilde{e}} = \left(2+\frac{\nu^{(1)}_{max}}{r_{bas}}\right) := t_1^{-1}$,
- $G_{s,p}(x, mx) < 0$ and $\epsilon < 1\ \forall x \geq 0 \iff K_2 < 0 \iff \frac{1}{\tilde{e}} < t_1^{-1}$,
- $G_{s,p}(x, mx) > 0$ and $\epsilon > 1\ \forall x \geq 0 \iff K_2 > 0 \iff \frac{1}{\tilde{e}} > t_1^{-1}$.

In the repressor cooperativity regime, $c_A = 1$ and $c_R > 1$, it can be estimated that $K_0 = 0$, $K_1 > 0$ and in consequence

- $G_{s,p}(x, mx) > 0$ and $\epsilon > 1\ \forall x \geq 0 \iff K_2 > 0 \iff \frac{c_R}{\tilde{e}} >= t_1^{-1}$.
- When $K_2 \leq 0$, which happens if and only if $\frac{c_R}{\tilde{e}} \leq t_1^{-1}$,we get
  - $G_{s,p}(x, mx) \geq 0$ and $\epsilon \geq 1$ for $0 \leq x \leq 2\frac{c_R-1}{\frac{1}{t_1}-\frac{c_R}{\tilde{e}}} := \frac{h_1}{K_A}$,
  - $G_{s,p}(x, mx) < 0$ and $\epsilon < 1\ \ \forall x > \frac{h_1}{K_A}$,

being $t_1 = \left(2+\frac{\nu^{(1)}_{max}}{r_{bas}}\right)^{-1}$ and $t_2 = \left(d2+\frac{\nu^{(1)}_{max}}{r_{bas}}\frac{c_A(1+\tilde{r})+2\tilde{r}}{1-\tilde{r}}\right)^{-1}$ described in Table 3.2.

Finally, in the case activators cooperativity regime, $c_A > 1$ and $c_R = 1$, we get $K_0 = 0$, $K_1 < 0$ and in consequence

- $G_{s,p}(x, mx) < 0$ and $\epsilon < 1\ \forall x \geq 0 \iff K_2 < 0 \iff \frac{1}{\tilde{e}} < 2+\frac{c_A(\tilde{r}+1)+2\tilde{r}}{1-\tilde{r}}\frac{\nu^{(1)}_{max}}{r_{bas}} := t_2^{-1}$.
- When $K_2 \geq 0$, which occurs if and only if $\frac{1}{\tilde{e}} \geq t_2^{-1}$ then we have that
  - $G_{s,p}(x, mx) \leq 0$ and $\epsilon \leq 1$ for $0 \leq x \leq 2\frac{c_A-1}{\frac{1}{\tilde{e}}-\frac{1}{t_2}} := \frac{h_2}{K_A}$,
  - $G_{s,p}(x, mx) > 0$ and $\epsilon > 1\ \ \forall x > \frac{h_2}{K_A}$.

Let us mention that the proof of the total cooperativity case is analog to the proof of the null cooperativity case.

In summary, the elasticities of the thresholds of the stimulated operator with two enhancers depends not only on the cooperativity considered, but also on the rest of the model parameters in terms of the quantities: $t_1 = \left(2 + \frac{\nu^{(1)}_{max}}{r_{bas}}\right)^{-1}$, $h_1 = 2K_A \frac{c_R - 1}{\frac{1}{t_1} - \frac{c_R}{\tilde{e}}}$, $t_2 = \left(2 + \frac{\nu^{(1)}_{max}}{r_{bas}} \frac{c_A(1+\tilde{r})+2\tilde{r}}{1-\tilde{r}}\right)^{-1}$, $h_2 = 2K_A \frac{c_A - 1}{\frac{1}{\tilde{e}} - \frac{1}{t_2}}$.

A.6 Threshold sensitivity analysis with Hill type operators

Please note that the previous sensitivity analysis can be also applied in the Hill modelling framework. This is, making the reduction to the global Activation and Repressor variables,

$$\lim_{\substack{c\to\infty \\ K_d=cte}} F_{reg}\big(([A],[R]);\{\{A,R\}_c\}\big) = \frac{1 + \left(a\frac{[A]}{K_d} + r\frac{[R]}{K_d}\right)^n}{1 + \left(\frac{[A]}{K_d} + \frac{[R]}{K_d}\right)^n}, \tag{A.59}$$

$$\lim_{\substack{c\to\infty \\ K_d=cte}} Basal\big(([A],[R]);\{\{A,R\}_c\}\big) = \frac{1 + \left(\frac{[A]}{K_d} + \frac{[R]}{K_d}\right)^n}{1 + \left(\frac{[A]}{K_d} + \frac{[R]}{K_d}\right)^n}, \tag{A.60}$$

$$\lim_{\substack{c\to\infty \\ K_d=cte}} Promoter\big(([A],[R]);\{\{A,R\}_c\}\big) = \frac{\left(\frac{[A]}{K_d} + \tilde{e}\tilde{r}\frac{[R]}{K_d}\right)^n - \left(\tilde{e}\tilde{r}\frac{[R]}{K_d}\right)^n}{1 + \left(\frac{[A]}{K_d} + \frac{[R]}{K_d}\right)^n},$$

and

$$\lim_{\substack{c\to\infty \\ K_d=cte}} F_{reg}\big(([A],[R]);\{\{A\}_c,\{R\}_c\}\big) = \frac{1 + \left(a\frac{[A]}{K_d}\right)^n + \left(r\frac{[R]}{K_d}\right)^n}{1 + \left(\frac{[A]}{K_d}\right)^n + \left(\frac{[R]}{K_d}\right)^n}, \tag{A.61}$$

$$\lim_{\substack{c\to\infty \\ K_d=cte}} Basal\big(([A],[R]);\{\{A\}_c,\{R\}_c\}\big) = \frac{1 + \left(\frac{[A]}{K_d}\right)^n + \left(\tilde{r}\frac{[R]}{K_d}\right)^n}{1 + \left(\frac{[A]}{K_d}\right)^n + \left(\frac{[R]}{K_d}\right)^n}, \tag{A.62}$$

$$\lim_{\substack{c\to\infty \\ K_d=cte}} Promoter\big(([A],[R]);\{\{A\}_c,\{R\}_c\}\big) = \frac{\left(\tilde{e}\tilde{r}\frac{[R]}{K_d}\right)^n}{1 + \left(\frac{[A]}{K_d}\right)^n + \left(\frac{[R]}{K_d}\right)^n}, \tag{A.63}$$

where we can deduce the threshold implicit function

$$G^H_{*,*}(x,y;n) \mid_{x=\frac{[A]}{K_d}, y=\frac{f^H_{*,*}([A];n)}{K_d}} = 0\,, \tag{A.64}$$

with $f^H_{*,*}([A];n)$ the threshold function for the Hill modules. Moreover, Lemma 5.4 is also fullfilled in this case, where it is easy to check that all thresholds functions are straight lines in the $[A]-[R]$ plane, and hence $\epsilon = 1$. In order to proof that $f^H_{*,*}([A])$ are linear for all the Hill BEWARE operators (A.59)-(A.63), we need to first obtain the implicit equation for each threshold. Following the same procedure done in the non-extreme cooperativity case, we get

$$G^H_{r,t}(x,y;n) = (ax+ry)^n - (x+y)^n\,, \tag{A.65}$$

$$G^H_{s,t}(x,y;n) = r_{bas}\big((x+\tilde{r}y)^n-(x+y)^n\big)+\nu^{(n)}_{max}\big((x+\tilde{e}\tilde{r}y)^n-(\tilde{e}\tilde{r}y)^n\big)\,, \tag{A.66}$$

$$G^H_{r,p}(x,y;n) = (r^n-1)y^n + (a^n-1)x^n \tag{A.67}$$

and

$$G^H_{s,p}(x,y;n) = r_{bas}(\tilde{r}^n-1)y^n + \nu^{(n)}_{max}x^n\,. \tag{A.68}$$

Since, all these functions are homogeneous, that is, $G^H(x,y;n) = x^n G^H(1,y/x;n)$ it is straightforward to check that

$$\left(y\frac{\partial G^H_{*,*}}{\partial y}(x,y;n) + x\frac{\partial G^H_{*,*}}{\partial x}(x,y;n)\right) = nG^H_{*,*}(x,y;n)$$

which means, from equation (A.49), that the threshold is a straight line in the $[A]-[R]$ plane, and hence $\epsilon = 1$ from Lemma 5.4.

However, it is important to note that Collorary 5.1 is not fulfilled in general, and the dependence of $f^H_{*,*}([A];n)$ with the number of enhancers varies depending on different parameter relations in the Hill versions of the Recruitment and Stimulated BEWARE operators.

Appendix B

Parameters

Parameter	Value
α $(a.u)$	1
β $(a.u)$	-1
γ $(a.u)$	-1
k $(\mu \mathrm{m}^2 st^{-2})$	0.01
δ_0 $(\mu \mathrm{m})$	0.5
δ_1 $(\mu \mathrm{m})$	4
δ_2 $(\mu \mathrm{m})$	5
l $(\mu \mathrm{m})$	0.05
v $(\mu \mathrm{m} st^{-1})$	1
τ (st)	0.01

Table B.1: Parameters used for all the simulations of cytonemes in Chapter 2. $a.u$ stands for arbitrary units; st stands for simulation timescale.

Parameter	Value
C_B $(nM \mathrm{min}^{-1})$	1
$K_{RP}/[PolII]$ $(a.u)$	1
c_A $(a.u)$	25
c_R $(a.u)$	100
a $(a.u)$	4.7

Parameter	Value
r $(a.u)$	0.2
K_A (nM)	0.8
K_R (nM)	0.8
n $(a.u)$	3

Table B.2: Parameters used in Fig. 3.2 for the Recruitment BEWARE operator with partial cooperativity

Parameter	(A)	(B)
C_B $(nM\text{min}^{-1})$	1	1
$K_{RP}/[PolII]$ $(a.u)$	1	1
c $(a.u)$	400	400
a $(a.u)$	1.1	2.5
r $(a.u)$	0.1	0.9
K_A (nM)	20	100
K_R (nM)	100	20
n $(a.u)$	3	3

Table B.3: Parameters used in Figure 3.3 for the Recruitment BEWARE operator with total cooperativity (inverse logic case)

Parameter (present work)	Parameter (in [63])	$\epsilon > 1$	$\epsilon = 1$	$\epsilon < 1$
$\tilde{e}$ $(a.u)$	$0.5\epsilon_A$	0.1	0.25	0.5
$\tilde{r}$ $(a.u)$	$\rho = 0.5\epsilon_R$	0.5	0.5	0.5
r_{bas} $(nM\text{min}^{-1})$	S_{XB}	50	50	50
$\nu^{(1)}_{max}$ $(nM\text{min}^{-1})$	S_{XA}	100	100	100
K_A (nM)	$K_{A1} = K_{A2}$	4	4	4
K_R (nM)	$K_{R1} = K_{R2}$	1	1	1
c $(a.u)$	$c_A = c_R = c_{AR}$	1	1	1
$K_{RP}/[PolII]$ $(a.u)$		1	1	1

Table B.4: Parameters used in Figure 3.4 for the Stimulated BEWARE operator with null cooperativity under perturbations (3.47) and (3.48). These parameters have been adopted from [63]

BEWARE constant	C_B	$1\,nM\text{min}^{-1}$
Transcriptional activation intensity	a	4.35
Transcriptional repression intensity	r	5×10^{-5}
Dissociation constant of activators for gene enhancers	K_A	$9 \times 10^{1}\,nM$
Dissociation constant of repressors for gene enhancers	K_R	$9 \times 10^{1}\,nM$
RNA polymerase binding affinity	K_{RP}	$[PolII]$
RNA polymerase concentration	$[PolII]$	K_{RP}
Total cooperativity constant	c	1
Activator partial cooperativity constant (cooperative activators)	c_A	10
Repressor partial cooperativity constant (cooperative activators)	c_R	1
Activator partial cooperativity constant (cooperative repressors)	c_A	1
Repressor partial cooperativity constant (cooperative repressors)	c_R	10
TFs total concentration	h	$24\,nM$
TFs gradient steepness	D	593

Table B.5: **Parameters used for the green curves in figures 3.6, 3.7 and 3.8**. The values for C_B, $[PolII]$ and K_{RP} imply that the expression levels given by the BEWARE operators are bounded by $1nM/min$ being the basal level equal to $0.5nM/min$. In this way signal modulations can be properly appreciated.

Perturbation	Value
Higher activator affinity (perturbation (3.47))	$K_A \to K_A \times 10^{-1}$
Higher repressor affinity (perturbation (3.47))	$K_R \to K_R \times 10^{-1}$
Lower number of binding sites (perturbation (3.48))	$n \to 1$

Table B.6: **Parameters used for the magenta curves in figures 3.6, 3.7 and 3.8**, using the non-perturbed parameters in Table B.5.

Bibliography

[1] G K Ackers, A D Johnson, and M A Shea. Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the National Academy of Sciences*, 79(4):1129–1133, 1982.

[2] Daniel Aguilar-Hidalgo, David Ba, Diana García-Morales, and Fernando Casares. Toward a study of gene regulatory constraints to morphological evolution of the drosophila ocellar region. *Development Genes and Evolution*, 226, 04 2016.

[3] Daniel Aguilar-Hidalgo, Maria Dominguez-Cejudo, Gabriele Amore, Anette Brockmann, María Lemos, Antonio Cordoba, and Fernando Casares. A hh-driven gene network controls specification, pattern and size of the drosophila simple eyes. *Development (Cambridge, England)*, 140, 11 2012.

[4] Ariel Altaba, Christophe Mas, and Barbara Stecca. The gli code: an information nexus regulating cell fate, stemness and cancer. trends cell biol 17:438-447. *Trends in cell biology*, 17:438–47, 10 2007.

[5] Adam Arkin, John Ross, and Harley McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage. *Genetics*, 149:1633–48, 09 1998.

[6] David Arnosti, S Gray, Scott Barolo, Jumin Zhou, and M Levine. The gap protein knirps mediates both quenching and direct repression in the. *The EMBO journal*, 15:3659–66, 08 1996.

[7] Ahmet Ay and David Arnosti. Mathematical modeling of gene expression: A guide for the perplexed biologist. *Critical reviews in biochemistry and molecular biology*, 46:137–51, 04 2011.

[8] Audun Bakk, Ralf Metzler, and Kim Sneppen. Sensitivity of or in phage Î». *Biophysical journal*, 86:58–66, 02 2004.

[9] Arieh Ben-Naim. Cooperativity in binding of proteins to dna. *The Journal of Chemical Physics*, 107:10242–10252, 12 1997.

[10] Arieh Ben-Naim. Cooperativity in binding of proteins to dna. ii. binding of bacteriophage lambda repressor to the left and right operators. *The Journal of Chemical Physics*, 108:6937–6946, 04 1998.

[11] Lacramioara Bintu, Nicolas Buchler, Hernan Garcia, Ulrich Gerland, Terence Hwa, Jane Kondev, and Rob Phillips. Transcriptional regulation by the numbers: Models. *Current opinion in genetics & development*, 15:116–24, 05 2005.

[12] Marcus Bischoff, Ana Gradilla, Irene Seijo, Germán Andrés, Carmen Rodríguez-Navas, Laura González-Méndez, and Isabel Guerrero. Cytonemes are required for the establishment of a normal hedgehog morphogen gradient in drosophila epithelia. *Nature cell biology*, 15, 10 2013.

[13] Jacques P. Bothma, Hernan G. Garcia, Emilia Esposito, Gavin Schlissel, Thomas Gregor, and Michael Levine. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living drosophila embryos. *Proceedings of the National Academy of Sciences*, 111(29):10598–10603, 2014.

[14] Ralph Bradshaw and E. Dennis. *Handbook of cell signaling, 2/e.* 01 2010.

[15] Paul Bressloff and Hyunjoong Kim. Bidirectional transport model of morphogen gradient formation via cytonemes. *Physical Biology*, 15, 01 2018.

[16] Paul Bressloff and Hyunjoong Kim. Search-and-capture model of cytoneme-mediated morphogen gradient formation. *Physical Review E*, 99, 05 2019.

[17] Nicolas Buchler, Ulrich Gerland, and Terence Hwa. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A*, 100:5136 – 5141, 11 2002.

[18] Manuel Cambón and Óscar Sánchez. Analysis of the transcriptional logic governing differential spatial expression in hh target genes. *PLOS ONE*, 14(1):1–25, 01 2019.

[19] Hal Caswell. Matrix population models: Construction, analysis, and interpretation. sinauer associates. *SERBIULA (sistema Librum 2.0)*, 01 2001.

[20] William Chen, Mario Niepel, and Peter Sorger. Classic and contemporary approaches to modeling biochemical reactions. *Genes & development*, 24:1861–75, 09 2010.

[21] Michael Cohen, Karen Page, Ruben Perez-Carrasco, Chris Barnes, and James Briscoe. A theoretical framework for the regulation of shh morphogen-controlled gene expression. *Development (Cambridge, England)*, 141:3868–78, 10 2014.

[22] Javier De Las Rivas and Celia Fontanillo. Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks. *PLoS computational biology*, 6:e1000807, 06 2010.

[23] Jonathan Desponds, Huy Tran, Teresa Ferraro, Tanguy Lucas, Carmina Perez Romero, Aurelien Guillou, Cecile Fradin, Mathieu Coppey, Nathalie Dostatni, and Aleksandra M. Walczak. Precision of readout at the hunchback gene: Analyzing short transcription time traces in living fly embryos. *PLOS Computational Biology*, 12(12):1–31, 12 2016.

[24] Creative Diagnostics. Creative diagnostics, 2009.

[25] Rui Dilao. The regulation of gene expression in eukaryotes: Bistability and oscillations in repressilator models. *Journal of Theoretical Biology*, 340:199 – 208, 2014.

[26] Ian Dodd, Keith Shearwin, Alison Perkins, Tom Burr, Ann Hochschild, and John Egan. Cooperativity in long-range gene regulation by the lambda ci repressor. *Genes & development*, 18:344–54, 03 2004.

[27] Ian Dunham, Anshul Kundaje, Shelley Aldred, Patrick Collins, Carrie Davis, Francis Doyle, Charles Epstein, Seth Frietze, Jennifer Harrow, Rajinder Kaul, Jainab Khatun, Bryan Lajoie, Stephen Landt, bumkyu Lee, Florencia Pauli Behn, Kate Rosenbloom, Peter Sabo, Alexias Safi, Amartya Sanyal, and Ewan Birney. The encode project consortium: An integrated encyclopedia of dna elements in the human genome. *Nature*, 489:57–74, 09 2012.

[28] Michael Elowitz, Arnold Levine, Eric Siggia, and Peter Swain. Stochastic gene expression in a single cell. *Science (New York, N.Y.)*, 297:1183–6, 09 2002.

[29] Harold Erickson. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biological procedures online*, 11:32–51, 06 2009.

[30] Walid Fakhouri, Ahmet Ay, Rupinder Sayal, Jacqueline Dresch, Evan Dayringer, and David Arnosti. Deciphering a transcriptional regulatory code: Modeling short-range repression in the drosophila embryo. *Molecular systems biology*, 6:341, 01 2010.

[31] Miguel Angel Fernández-Moreno, Carol L. Farr, Laurie S. Kaguni, and Rafael Garesse. *Drosophila melanogaster as a Model System to Study Mitochondrial Biology*. Humana Press, Totowa, NJ, 2007.

[32] Vladimir Filkov. Identifying gene regulatory networks from gene expression data. *Handbook of Computational Molecular Biology*, 01 2006.

[33] Till Frank, Aimée Carmody, and Boris Kholodenko. Versatility of cooperative transcriptional activation: A thermodynamical modeling analysis for greater-than-additive and less-than-additive effects. *PloS one*, 7:e34439, 04 2012.

[34] Till Frank, Miguel Cavadas, Lan Nguyen, and Alex Cheong. *Nonlinear Dynamics in Transcriptional Regulation: Biological Logic Gates*, volume 7, pages 43–62. 07 2016.

[35] Vernon French, Marieke Feast, and Linda Partridge. Body size and cell size in drosophila: the developmental response to temperature. *Journal of Insect Physiology*, 44(11):1081 – 1089, 1998.

[36] Nir Friedman. Inferring cellular networks using probabilistic graphical models. *Science (New York, N.Y.)*, 303:799–805, 03 2004.

[37] Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. Using bayesian networks to analyze expression data. *Journal of computational biology : a journal of computational molecular cell biology*, 7:601–20, 02 2000.

[38] Takashi Fukaya, Bomyi Lim, and Michael Levine. Enhancer control of transcriptional bursting. *Cell*, 166, 06 2016.

[39] Takashi Fukaya, Bomyi Lim, and Michael Levine. Enhancer control of transcriptional bursting. *Cell*, 166(2):358 – 368, 2016.

[40] Eileen E. M. Furlong and Michael Levine. Developmental enhancers and chromosome topology. *Science*, 361(6409):1341–1345, 2018.

[41] Hernan Garcia and Thomas Gregor. *Live Imaging of mRNA Synthesis in Drosophila*, volume 1649, pages 349–357. 01 2018.

[42] Diana García-Morales, Tomás Navarro, Antonella Iannini, David Míguez, and Fernando Casares. Dynamic hh signaling can generate temporal information during tissue patterning. *Development*, 146:dev.176933, 03 2019.

[43] Jordi Garcia-Ojalvo. Physical approaches to the dynamics of genetic circuits: A tutorial. *Contemporary Physics - CONTEMP PHYS*, 52, 05 2011.

[44] Walter Gautschi. *Numerical Analysis*. Birkhäuser Basel, 2011.

[45] Jason Gertz and Barak Cohen. Environment-specific combinatorial cis-regulation in synthetic promoters. *Molecular systems biology*, 5:244, 02 2009.

[46] Jason Gertz, Eric Siggia, and Barak Cohen. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*, 457:215–8, 12 2008.

[47] Alex Gilman and Adam Arkin. Genetic c̈ode¨: Representations and dynamical models of genetic components and networks. *Annual review of genomics and human genetics*, 3:341–69, 02 2002.

[48] Etienne Giroux Leprieur, Adrien Costantini, Vivianne Ding, and Biao He. Hedgehog signaling in lung cancer: From oncogenesis to cancer treatment resistance. *International Journal of Molecular Sciences*, 19:2835, 09 2018.

[49] Laura González-Méndez, Ana-Citlali Gradilla, David Sánchez-Hernández, Esperanza González, Adrián Aguirre-Tamaral, Carlos Jiménez-Jiménez, Milagros Guerra, Gustavo Aguilar, Germán Andrés, Juan M Falcón-Pérez, and Isabel Guerrero. Polarized sorting of patched enables cytoneme-mediated hedgehog reception in the drosophila wing disc. *The EMBO Journal*, n/a(n/a):e103629.

[50] Laura González-Méndez, Irene Seijo-Barandiarán, and Isabel Guerrero. Cytoneme-mediated cell-cell contacts for hedgehog reception. *eLife*, 6:e24045, aug 2017.

[51] Joshua Granek and Neil Clarke. Explicit equilibrium modeling of transcription-factor binding and gene regulation. *Genome biology*, 6:R87, 02 2005.

[52] P. Griffiths and Karola Stotz. Genetics and philosophy: An introduction. *Genetics and Philosophy: An Introduction*, pages 1–270, 01 2011.

[53] Bisswanger H. *Enzyme Kinetics: Principles and methods*. WILEY-VCH, 2008.

[54] Xin He, Abul Hassan Samee, Charles Blatti, and Saurabh Sinha. Thermodynamics-based models of transcriptional regulation by enhancers: The roles of synergistic activation, cooperative binding and short-range repression. *PLoS computational biology*, 6, 09 2010.

[55] AV Hill. The combinations of haemoglobin with oxygen and with carbon monoxide. i. *The Biochemical journal*, 7:471–80, 11 1913.

[56] Tom W Hiscock and Sean G Megason. Orientation of turing-like patterns by morphogen gradients and tissue anisotropies. *Cell Systems*, 1:408–416, 12 2015.

[57] Brandon Ho, Anastasia Baryshnikova, and Grant Brown. Unification of protein abundance datasets yields a quantitative saccharomyces cerevisiae proteome. *Cell Systems*, 6, 01 2018.

[58] A. Wride J. Sloman. *Economics, 7th Ed.* 01 2009.

[59] Hilde Janssens, Shuling Hou, Johannes Jaeger, Ah-Ram Kim, Ekaterina Myasnikova, David Sharp, and John Reinitz. Quantitative and predictive model of transcriptional control of the drosophila melanogaster even skipped gene. *Nature genetics*, 38:1159–65, 11 2006.

[60] Chen Jia and Daquan Jiang. Cycle symmetries and circulation fluctuations for discrete-time and continuous-time markov chains. *The Annals of Applied Probability*, 26:2016, 09 2016.

[61] Hidde Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of computational biology : a journal of computational molecular cell biology*, 9:67–103, 02 2002.

[62] Jan Junker, Kevin Peterson, Yuichi Nishi, Junhao Mao, Andrew McMahon, and Alexander Oudenaarden. A predictive model of bifunctional transcription factor signaling during embryonic tissue patterning. *Developmental Cell*, 31:448–460, 11 2014.

[63] Andrew Keller. Model genetic circuits encoding autoregulatory transcription factors. *Journal of theoretical biology*, 172:169–85, 02 1995.

[64] Thomas B. Kepler and Timothy C. Elston. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophysical Journal*, 81(6):3116 – 3136, 2001.

[65] Anna Kicheva, Michael Cohen, and James Briscoe. Developmental pattern formation: Insights from physics and biology. *Science (New York, N.Y.)*, 338:210–2, 10 2012.

[66] Hyunjoong Kim and Paul Bressloff. Direct vs. synaptic coupling in a mathematical model of cytoneme-based morphogen gradient formation. *SIAM Journal on Applied Mathematics*, 78:2323–2347, 09 2018.

[67] Miltiadis Kininis, Gary Isaacs, Leighton Core, Nasun Hah, and W Kraus. Postrecruitment regulation of rna polymerase ii directs rapid signaling responses at the promoters of estrogen target genes. *Molecular and cellular biology*, 29:1123–33, 03 2009.

[68] Shigeru Kondo and Takashi Miura. Reaction-diffusion model as a framework for understanding biological pattern formation. *Science (New York, N.Y.)*, 329:1616–20, 09 2010.

[69] Karen Lai, Matthew Robertson, and David Schaffer. The sonic hedgehog signaling system as a bistable genetic switch. *Biophysical journal*, 86:2748–57, 06 2004.

[70] Nicholas C. Lammers, Vahe Galstyan, Armando Reimer, Sean A. Medin, Chris H. Wiggins, and Hernan G. Garcia. Multimodal transcriptional control of pattern formation in embryonic development. *Proceedings of the National Academy of Sciences*, 117(2):836–847, 2020.

[71] S Law, G Bellomy, P Schlax, and M Record. In vivo thermodynamic analysis of repression with and without looping in lac constructs. estimates of free and local lac repressor concentrations and of physical properties of a region of supercoiled plasmid dna in vivo. *Journal of molecular biology*, 230:161–73, 04 1993.

[72] Tong Lee and Richard Young. Transcriptional regulation and its misregulation in disease. *Cell*, 152:1237–51, 03 2013.

[73] Shawn Little, Mikhail Tikhonov, and Thomas Gregor. Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell*, 154:789–800, 08 2013.

[74] David S Lorberbaum, Andrea I Ramos, Kevin A Peterson, Brandon S Carpenter, David S Parker, Sandip De, Lauren E Hillers, Victoria M Blake, Yuichi Nishi, Matthew R McFarlane, Ason CY Chiang, Judith A Kassis, Benjamin L Allen, Andrew P McMahon, and Scott Barolo. An ancient yet flexible *cis*-regulatory architecture allows localized hedgehog tuning by *patched/Ptch1*. *eLife*, 5:e13550, may 2016.

[75] Florian Markowetz and Rainer Spang. Inferring cellular networks: A review. *BMC bioinformatics*, 8 Suppl 6:S5, 02 2007.

[76] Paul Martin and Susan M. Parkhurst. Parallels between tissue repair and embryo morphogenesis. *Development*, 131(13):3021–3034, 2004.

[77] G Meir, E. Munro, and Garrett Odell. The segment polarity network is a robust development module. *Nature*, 406, 01 2004.

[78] Andreas Milias-Argeitis, Ana Paula Oliveira, Luca Gerosa, Laura Falter, Uwe Sauer, and John Lygeros. Elucidation of genetic interactions in the yeast gata-factor network using bayesian model selection. *PLOS Computational Biology*, 12:e1004784, 03 2016.

[79] Bruno Muller and Konrad Basler. The repressor and activator forms of cubitus interruptus control hedgehog target genes through common generic gli-binding sites. *Development (Cambridge, England)*, 127:2999–3007, 08 2000.

[80] Marcos Nahmad and Angelike Stathopoulos. Dynamic interpretation of hedgehog signaling in the drosophila wing disc. *PLoS biology*, 7:e1000202, 09 2009.

[81] Pawel Niewiadomski, Sylwia Niedziolka, Lukasz Markiewicz, Tomasz Uspienski, Brygida Baran, and Katarzyna Chojnowska. Gli proteins: Regulation in development and cancer. *Cells*, 8:147, 02 2019.

[82] Adam Paré, Derek Lemons, Dave Kosman, William Beaver, Yoav Freund, and William McGinnis. Visualization of individual scr mrnas during drosophila embryogenesis yields evidence for transcriptional bursting. *Current Biology*, 19(23):2037 – 2042, 2009.

[83] David Parker, Michael White, Andrea Ramos, Barak Cohen, and Scott Barolo. The cis-regulatory logic of hedgehog gradient responses: Key roles for gli binding affinity, competition, and cooperativity. *Science signaling*, 4:ra38, 06 2011.

[84] J. Peccoud and B. Ycart. Markovian modeling of gene-product synthesis. *Theoretical Population Biology*, 48(2):222 – 234, 1995.

[85] M. Perutz. Mechanisms of cooperative and allosteric regulation in proteins. *Quarterly reviews of biophysics*, 22:139–237, 06 1989.

[86] Michael Pilling. Mathematical models of chemical reactions. theory and applications of deterministic and stochastic models. *Journal of Photochemistry and Photobiology A-chemistry - J PHOTOCHEM PHOTOBIOL A-CHEM*, 49:409–410, 10 1989.

[87] Mark Ptashne. Regulation of transcription: From lambda to eukaryotes. *Trends in biochemical sciences*, 30:275–9, 07 2005.

[88] Mark Ptashne and Alexander Gann. Ptashne, m. & gann, a. transcriptional activation by recruitment. nature 386, 569-577. *Nature*, 386:569–77, 05 1997.

[89] Mark Ptashne and Alexander Gann. *Genes and Signals*. 01 2001.

[90] Michal Rabani, Joshua Levin, Lin Fan, Xian Adiconis, Raktima Raychowdhury, Manuel Garber, Andreas Gnirke, Chad Nusbaum, Nir Hacohen, Nir Friedman, Ido Amit, and Aviv Regev. Metabolic labeling of rna uncovers principles of rna production and degradation dynamics in mammalian cells. *Nature biotechnology*, 29:436–42, 05 2011.

[91] Andrea Ramos and Scott Barolo. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368:20130018, 11 2013.

[92] P.D. Robert J. Brooker. *Loose Leaf Version for Genetics: Analysis and Principles*. McGraw-Hill Education, 2014.

[93] Daniel D. Rocker. *Variational Methods on Elastic Curves*. Electronic Theses and Dissertations., 2013.

[94] Krishanu Saha and David Schaffer. Signal dynamics in sonic hedgehog tissue patterning. *Development (Cambridge, England)*, 133:889–900, 04 2006.

[95] Eran Segal, Tali Raveh-Sadka, Mark Schroeder, Ulrich Unnerstall, and Ulrike Gaul. Predicting expression patterns from regulatory sequence in drosophila segmentation. *Nature*, 451:535–40, 02 2008.

[96] Eran Segal and Jonathan Widom. From dna sequence to transcriptional behavior: A quantitative approach. *Nature reviews. Genetics*, 10:443–56, 08 2009.

[97] M. A. Shea and G. Ackers. The or control system of bacteriophage lambda. a physical-chemical model for gene regulation. *Journal of molecular biology*, 181 2:211–30, 1985.

[98] Marc Sherman and Barak Cohen. Thermodynamic state ensemble models of cis-regulation. *PLoS computational biology*, 8:e1002407, 03 2012.

[99] Marcia Simpson-Brose, Jessica Treisman, and Claude Desplan. Synergy between the hunchback and bicoid morphogens is required for anterior patterning in drosophila. *Cell*, 78:855–65, 10 1994.

[100] Stephen Small and David N. Arnosti. Transcriptional enhancers in drosophila. *Genetics*, 216(1):1–26, 2020.

[101] P Szymanski and M Levine. Multiple modes of dorsal - bhlh transcriptional synergy in the drosophila embryo. *The EMBO journal*, 14:2229–38, 06 1995.

[102] Tetsuya Tabata and Yuki Takei. Morphogens, their identification and regulation. *Development*, 131(4):703–712, 2004.

[103] Kiyofumi Takabatake, Tsuyoshi Shimo, Jun Murakami, Chang Anqi, Hotaka Kawai, Saori Yoshida, May Wathone Oo, Omori Haruka, Sukegawa Shintaro, Hidetsugu Tsujigiwa, Keisuke Nakano, and Hitoshi

Nagatsuka. The role of sonic hedgehog signaling in the tumor microenvironment of oral squamous cell carcinoma. *International Journal of Molecular Sciences*, 20:5779, 11 2019.

[104] D'Arcy Wentworth Thompson. *On Growth and Form*. Canto Classics. Cambridge University Press, 2014.

[105] Carlos Torroja, Nicole Gorfinkiel, and Isabel Guerrero. Mechanisms of hedgehog gradient formation. *Journal of neurobiology*, 64:334–56, 09 2005.

[106] Chung Jung Tsai and Ruth Nussinov. A unified view of "how allostery works". *PLoS computational biology*, 10:e1003394, 02 2014.

[107] Ed Tunnacliffe and Jonathan Chubb. What is a transcriptional burst? *Trends in Genetics*, 36, 02 2020.

[108] Alan Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society B*, 237:37–72, 1952.

[109] Graf Ulrich, van Schaik Nancy, and E. WÃ$\frac{1}{4}$rgler Friedrich. *Drosophila genetics*. Springer, 1991.

[110] Uniprot. Uniprot, 2002.

[111] Michela Verbeni, Óscar Sánchez, E Mollica, I Siegl-Cachedenier, Alan Carleton, Isabel Guerrero, A Altaba, and J Soler. Morphogenetic action through flux-limited spreading. *Physics of life reviews*, 10, 06 2013.

[112] Peter von Hippel, Arnold Revzin, Carol Gross, and Amy Wang. Nonspecific dna binding of genome regulating proteins as a biological control mechanism: 1. the lac operon: Equilibrium aspects. *Proceedings of the National Academy of Sciences of the United States of America*, 71:4808–12, 01 1975.

[113] Mong-Lien Wang, Shih-Hwa Chiou, and Cheng-Wen Wu. Targeting cancer stem cells: Emerging role of nanog transcription factor. *OncoTargets and therapy*, 6:1207–1220, 09 2013.

[114] Rui-Sheng Wang, Assieh Saadatpour, and Réka Albert. Boolean modeling in systems biology: An overview of methodology and applications. *Physical biology*, 9:055001, 09 2012.

[115] J Weiss. The hill equation revisited: uses and misuses. *The FASEB Journal*, 11:835–841, 09 1997.

[116] Michael White, Davis Parker, Scott Barolo, and Barak Cohen. A model of spatially restricted transcription in opposing gradients of activators and repressors. *Molecular systems biology*, 8:614, 09 2012.

[117] Wikipedia. Wikipedia, 2020.

[118] Heng Xu, Leonardo Sepúlveda, Lauren Figard, Anna Sokac, and Ido Golding. Combining protein and mrna quantification to decipher transcriptional regulation. *Nature methods*, 12, 06 2015.

[119] Heng Xu, Samuel Skinner, Anna Sokac, and Ido Golding. Stochastic kinetics of nascent rna. *Physical Review Letters*, 117, 09 2016.

[120] Julia Zeitlinger. Seven myths of how transcription factors read the cis-regulatory code. *Current Opinion in Systems Biology*, 2020.

[121] Chengting Zhang and Steffen Scholpp. Cytonemes in development. *Current Opinion in Genetics and Development*, 57:25–30, 08 2019.

[122] Xin Zhou and Zhen Su. tcal: Transcriptional probability calculator using thermodynamic model. *Bioinformatics (Oxford, England)*, 24:2639–40, 10 2008.

[123] Robert Zinzen, Kate Senger, Mike Levine, and Dmitri Papatsenko. Computational models for neurogenic gene expression in the drosophila embryo. *Current biology : CB*, 16:1358–65, 08 2006.

[124] Benjamin Zoller, Shawn C. Little, and Thomas Gregor. Diverse spatial expression patterns emerge from unified kinetics of transcriptional bursting. *Cell*, 175(3):835 – 847.e25, 2018.

–¿Qué te parece desto, Sancho? – Dijo Don Quijote –
Bien podrán los encantadores quitarme la ventura,
pero el esfuerzo y el ánimo, será imposible.

Segunda parte del Ingenioso Caballero
Don Quijote de la Mancha
Miguel de Cervantes