Universitat Autònoma de Barcelona

Studying the genotype-phenotype map through general mathematical models of embryonic development and its application on tooth morphogenesis and evolution

Miquel MARIN RIERA

UNIVERSITAT AUTÒNOMA DE BARCELONA

TESI DOCTORAL

Studying the genotype-phenotype map through general mathematical models of embryonic development and its application on tooth morphogenesis and evolution

Estudi del mapa genotip-fenotip per mitja de models matematics de desenvolupament embrionari i la seva aplicacio a la morfogenesis i evolucio de les dents

Autor:
Miquel MARIN RIERA

Supervisor:
Dr. Isaac SALAZAR CIUDAD

PROGRAMA DE DOCTORAT EN GENÈTICA

Grup de Genòmica, Bioinformàtica i Evolució
Departament de Genètica i Microbiologia

2016

UNIVERSITAT AUTÒNOMA DE BARCELONA

Abstract

Facultat de Biociències
Departament de Genètica i Microbiologia

Doctor of Philosophy

Studying the genotype-phenotype map through general mathematical models of embryonic development and its application on tooth morphogenesis and evolution

by Miquel MARIN RIERA

One of the main challenges of evolutionary biology is to understand how species change over generations. Phenotypic evolutionary change can only be understood by knowing how natural selection acts on the phenotype and which heritable phenotypic variation arises in each generation within populations. Heritable genetic variation is associated to the correspondent phenotypic variation by means of the genotype-phenotype map (GPM). In the case of morphological evolution, the GPM is determined by embryonic development which consists on complex and dynamic networks of interaction between genes, cells and tissues. The nature of the GPM tells us the ensemble of phenotypes that can be produced by development and also the specific genetic mutations that are required to reach a certain phenotype. In this thesis we use the computational approach to generate theoretical predictions on how the complexity of the GPM influences morphological variation arising from development and what effect has that on the evolutionary dynamics of populations. We use the mammalian tooth as a model system due to its high morphological complexity and ecological and evolutionary relevance. By performing *in silico* evolution we find that selective regimes that focus on all the details of the phenotype fail to drive populations to phenotypic optimality. We also develop a general model of animal development that is able to reproduce a wide range of developmental mechanisms. We then use this general model to extend the tooth development model and we explore the morphological variation arising from alterations in cell-cell adhesion and biomechanics. By simulating tooth development and evolution by means of mathematical models we have been able to make predictions on how a complex GPM arising from development affects the distribution of phenotypic variation and the effective of natural selection during adaptive evolution.

UNIVERSITAT AUTÒNOMA DE BARCELONA

Resum

Facultat de Biociències
Departament de Genètica i Microbiologia

Doctor

Estudi del mapa genotip-fenotip per mitja de models matematics de desenvolupament embrionari i la seva aplicacio a la morfogenesis i evolucio de les dents

de Miquel MARIN RIERA

Entendre com les especies canvien al llarg de les generacions es un dels principals desafiaments de la biologia evolutiva. Per tal d'entendre el canvi evolutiu a nivell del fenotip és necessari conèixer quina variació fenotípica heretable és present a cada generació i com la seleccio natural actua sobre el fenotip. La variació genètica heretable s'associa a la variació fenotípica corresponent a través del mapa genotip-fenotip (MGF). En el cas de l'evolució morfològica, el MGF ve donat pel desenvolupament embrionari, que consisteix en xarxes d'interacció complexes i dinàmiques entre gens, cèl·lules i teixits. El MGF ens indica el conjunt de fenotips que el desenvolupament pot produir i també les mutacions genètiques especiífiques que es requereixen per aconseguir un determinat fenotip. En aquesta tesi utilitzem l'aproximació computacional per tal de generar prediccions teòriques sobre com la complexitat del MGF influeix en la variació morfològica resultant del desenvolupament i quin efecte té en les dinàmiques evolutives de les poblacions. Utilitzem la dent de mamífer com a sistema model a causa de la seva alta complexitat morfològica i la seva rellevància ecològica i evolutiva. Per mitjà de la simulació d'evolució *in silico* trobem que els règims selectius que es centren en tots els detalls del fenotip no aconsegueixen portar les poblacions a l'optimalitat fenotípica. També desenvolupem un model general de desenvolupament animal que és capaç de reproduir un ampli ventall de mecanismes de desenvolupament. Després utilitzem aquest model per extendre el model de desenvolupament de dents i explorem la variacio morfològica provinent d'alteracions en l'adhesió i biomecànica cel·lulars. Per mitjà de la simulació del desenvolupament de dents amb models matemàtics hem pogut fer prediccions sobre com un MGF complex afecta la distribució de la variació fenotípica i l'efecte de la selecció natural durant l'evolució adaptativa.

Acknowledgements

I thank my supervisor Isaac Salazar-Ciudad for his advice and guidance. I also thank Miguel Brun-Usan, Roland Zimm, Tommi Välikangas and Isaac Salazar-Ciudad for their contribution to developing the EmbryoMaker software. I want to thank Miguel Brun-Usan, Roland Zimm, Irepan Salvador-Martinez, Pascal Hagolani, Alexis Matamoro-Vidal, Jukka Jernvall and Stuart Newman for helpful comments on the multiple manuscripts I wrote during this thesis, and also CSC of Finland (IT Centre for Science) which provided computational resources for some of the work done during this thesis. This work was funded by Generalitat de Catalunya (2013FI-B00439) and Finnish Academy (WBS 1250271).

Contents

List of Figures

List of Abbreviations

GPM	Genotype-Phenotype MAP
PEK	Primary Enamel Knot
SEK	Secondary Enamel Knot
FGF	Fibroblast Growth Factor
SHH	Sonic HedgeHog
BMP	Bone Morphogenetic Protein
PFM	Phenotype-Fitness Map
IEE	Inner Enamel Epithelium
OEE	Outer Enamel Epithelium
RNA	RiboNucleic Acid
DNA	DeoxiRiboNucleic Acid
EMD	Euclidean Morphological Distance
OPC	Orientation Patch Count
PCA	Principal Component Analysis
CPM	Cellular Potts Model
IBCell	Immersed Boundary Cell
SEM	Subcellular Elements Model
ECM	Extracellular Matrix

Dedicated to my family and friends

Chapter 1

Introduction

One of the main challenges of evolutionary biology is to understand how organisms change over generations. Classically, natural selection acting on phenotypic variation has been proposed to be the main driver of phenotypic change (Darwin, 1859). Thus, in order to predict phenotypic evolutionary change one needs to understand both natural selection and which heritable phenotypic variation arises in each generation in populations. Most of this heritable phenotypic variation arises from genetic variation, but we do not completely understand by which mechanisms specific genetic variants give rise to specific phenotypic variants. This is because an organism's phenotype, especially its morphology, arises through, and because of, a complex process of embryonic development based on complex networks of interaction between genes, cells and tissues.

Development is a process in which a single cell transforms, over time, into an organism composed of multiple cells and cell types arranged in specific and complex spatial patterns. This process can be understood as a series of transformations of specific spatial distribution of cell types, or patterns, into other, usually more complex, distributions of cell types over space. Any morphological difference between two individual organisms in a species arises first as a difference in these transformations at some stage during development. Each such pattern transformations in development involves complex networks of genetic, cellular and tissue interactions over time. Consequently, any morphological difference between two individuals arises as a difference in some of these interactions. Thus, in order to understand, for example, why some change in a specific gene leads to a specific set of morphological changes we need to understand this genetic variant in the context of the dynamic networks of interactions between genes, cells and tissues in a specific pattern transformation in development. We call "developmental mechanism" any gene network that regulates at least one cell behaviour and is involved in a pattern transformation.

Generally, in order to understand how morphology in a population changes over evolutionary time one needs to address two questions: 1) which phenotypic variation appears at each generation and 2) which phenotypic variation is filtered out by ecological factors. To address question 1 it is necessary to understand development. A good understanding of the development of a species or lineage at a certain point during evolution significantly enhances the capacity to understand its evolutionary trajectory (especially if something is known about question 2).

To address question 1 two aspects of development should be studied:

1. Variational properties (Salazar-Ciudad, 2006). They are defined as the range of morphologies that can be produced by the development of

an organism, or in a specific developmental mechanism when there are changes in the environment or small genetic mutations. By small we mean mutations that do not change the topology of the genetic network of a developmental mechanism.

2. The genotype-phenotype map (GPM). It describes which specific genetic changes are associated with which morphologies within the variational properties (Waddington, 2012).

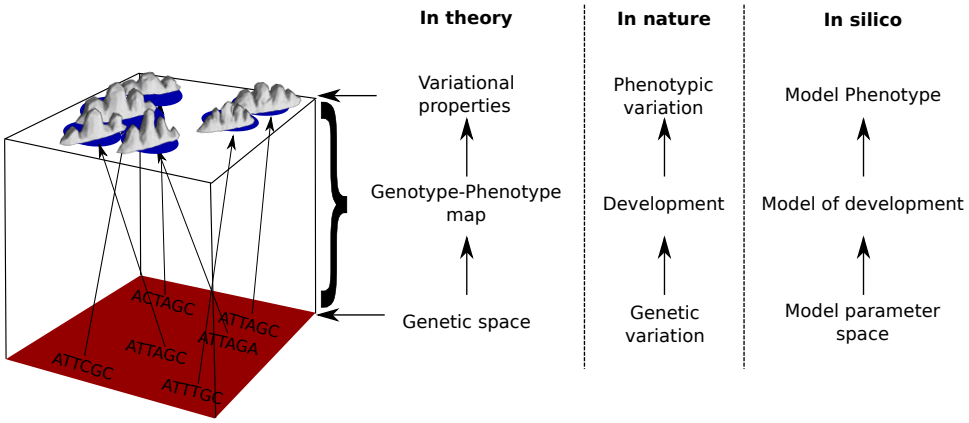

FIGURE 1.1: Given a certain developmental process in an individual, a population or a species, there is a limited amount of different phenotypes that can arise through changes in the genotype. The ensemble of phenotypes that can arise through a developmental process is called variational properties and the association between each genetic change and a phenotypic change is the genotype-phenotype map. In nature, specific genetic changes are associated to specific phenotypic changes through development. The GPM can be reproduced *in silico* by integrating our knowledge of development into a mathematical model. Variation in the model parameters will result in different model phenotypes. By systematically exploring the model parameter space one can infer the variational properties of that specific developmental system.

Variational properties and GPM are not independent of one another. The range of morphologies that compose the variational properties entirely depend on the nature of the GPM, since the latter determines which genotypes will arise from each of the genotypes. In other words, the variational properties describe what development, or a developmental mechanism, can produce at the phenotypic level, whereas the GPM describes the distribution of those phenotypic variants with respect to all genetic variants (fig. 1.1). Both concepts can be used in the context of the development of a whole organism, a part of an organism or a specific developmental mechanism in a specific pattern transformation (fig. 1.1).

It is possible to estimate the GPM by statistical methods without understanding the underlying developmental dynamics (Lande and Arnold, 1983). One can empirically obtain variance-covariance matrices of trait values (P) and trait heritability (G) for a set of phenotypic traits in a population. It has been shown that given a certain selection pressure (represented as vector s on the set of traits (represented by vector z), one can predict the

mean phenotype of the population after a selection event using the G and P matrices as follows (Lande and Arnold, 1983),

$$\Delta z = GP^{-1}s \tag{1.1}$$

where Δz represents the difference of the mean phenotype before and after the selection event. However, this method can only predict pheotypic evolutionary change in short time scales since it cannot predict how the G matrix changes over time. In order to be able to predict the latter, one needs to mechanistically understand the underlying development.

There is usually abundant but incomplete experimental information about those. Most morphologies are complex, that is composed of many measurable traits, and variation on each trait is highly interdependent on the others. Complex morphologies tend to have complex GPMs stemming from development (Salazar-Ciudad and Jernvall, 2005; Salazar-Ciudad and Marín-Riera, 2013), which means a single trait seldom shows gradual variation and that will covariate with other traits. This makes morphological variation in complex structures very hard to predict unless one understands the underlying GPM, which is the process of development. Most developmental processes involve complex dynamics consisting of extensive cell movements and cell-cell communication happening simultaneously (Salazar-Ciudad, 2010b).

Thus in order to be able to predict the variational properties and GPM of development it is necessary to implement the understanding, or hypotheses, about how genes, cells and tissues interact during development into mathematical models. These can then be used to understand the arising of subtle and complex morphological variation. Thus, mathematical models can make quantitative predictions on the variational properties and the GPM based on the causal links between each genotype and the corresponding phenotype arising from development (Salazar-Ciudad and Jernvall, 2010). In addition, using mathematical models it is easier and cheaper to explore the variational properties and inferring the underlying causes of phenotypic variation because it can be done systematically.

The aim of this thesis is to provide theoretical predictions on how the complexity of the GPM influences the morphological variation arising from development and what effect that complexity has on the evolutionary dynamics of populations. We use the mammalian tooth as a model system due to its high morphological complexity and ecological and evolutionary relevance. Chapter 3 of this thesis describes a model of tooth evolution and development that is used to infer which types of selective pressures facilitate or prevent adaptive change in complex tooth morphologies (question 2). Chapter 4 presents a general modelling framework for animal development that is later used in Chapter 5 to design a new and extended model of tooth development, which is used to study the phenotypic variation arising from tooth development (question 1) and provide predictions on which aspects of morphogenesis account for variation in specific morphological features of teeth.

1.1 Mammalian tooth as model system for evolutionary developmental biology

The mammalian tooth has been and it is being extensively studied both in the fields of development and evolution (Butler, 1983; Butler, 1995). From the developmental point of view, it is an interesting system due to the complexity of its dynamics, which combine cell signalling and extensive cell movements (Salazar-Ciudad, 2008; Salazar-Ciudad, 2012). Teeth develop in relative isolation from the rest of the body. That allows experimental manipulation of the developing tooth germ with minimal interference on the morphogenesis of the whole embryo. From the ecological and evolutionary point of view, teeth are required for the breakdown of food in mammals and thus tooth morphology is closely related with the animal's diet (Evans et al., 2007). Moreover, teeth are often the only traces left by extinct mammalian species in the fossil record and so, tooth morphology is very important in the study of mammalian evolution. For these reasons teeth are an especially relevant organ system at the crossroads of the fields of developmental and evolutionary biology (evo-devo) (Salazar-Ciudad, 2012). From the perspective of evo-devo, the question is not so much how a wild type tooth is achieved, but rather how variation in its development will result in different tooth morphologies (Harjunmaa et al., 2014).

Tooth development consists of the growth and folding of an epithelial invagination, on the growth and condensation of an underlying mesenchyme (the dental papilla) and the differentiation of these tissues into the different cell types of the tooth crown. Such a process involves inductive interactions mediated by signalling in the epithelium, the mesenchyme and between them. Different molecular signals diffusing in the complex spatial context of the tooth germ determine the differentiation of non-proliferative signalling centres called the enamel knots (EK) in specific positions, which in turn regulate the patterning and positioning of tooth cusps (Jernvall, 2000). The most studied case of tooth development is that of the mouse first molar, but most other molars seem to develop similarly (Keränen et al., 1998).

At embryonic day 13 (E13), the prospective molar consists of a round epithelial invagination surrounded by a mesenchymal condensation called the tooth bud (fig. 1.2A). A non-proliferative signalling centre called the primary enamel knot (PEK) appears at the distal tip of the bud (fig. 1.2B) (Jernvall and Thesleff, 2000). At this stage the epithelial bud consists of two cell populations, a basal layer of epithelial cells contacting the basement membrane and expressing P-cadherin, and a suprabasal cell population composing the bulk of the bud and expressing E-cadherin (Jussila et al., 2015).

By E15 two epithelial folds, called the cervical loops, have emerged at the buccal and lingual sides of the PEK, and the morphology is reminiscent of a cap (fig. 1.2C). The cervical loops, which are composed of both basal and suprabasal cells, first grow radically (i.e. in the direction where the roots of the developed tooth will be, as opposed to occlusal) and later towards the mid line of the tooth germ, effectively enveloping the mesenchymal condensate. During the late cap stage, the PEK disappears and two

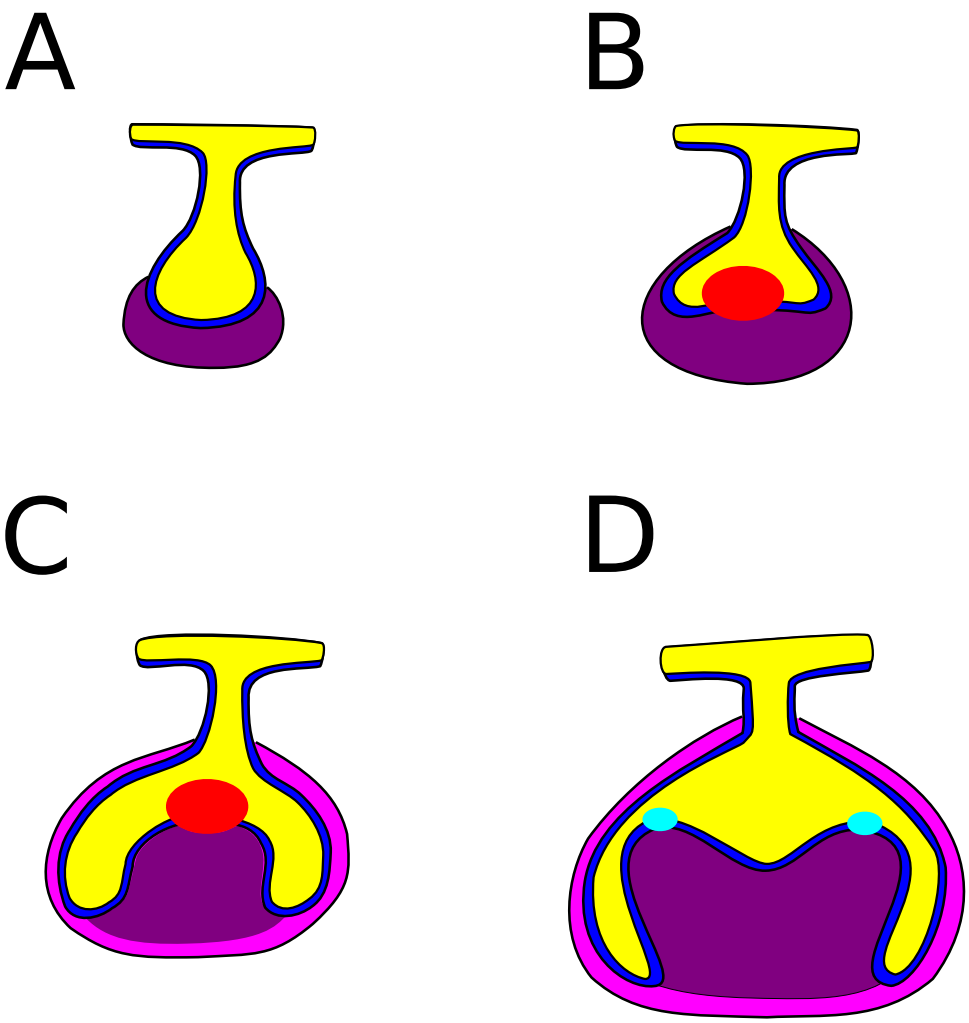

FIGURE 1.2: A, by E13, an epithelial bud (blue and yellow) has invaginated from the oral epithelium into the jaw mesenchyme and a mesenchymal condensate has started to form (pink). B, soon after that, a signalling centre called the primary enamel knot (PEK, red) appears at the distal tip of the bud and two epithelial folds start to grow in the bucal and lingual sides of the tooth germ. C, by E15 the growing cervical loops have eventually surrounded part of the mesenchyme. D, Between E15 and E16, the PEK disappears and new signalling centres called secondary enamel knots (SEK, cyan) are induced while the cervical loops continue to elongate.

secondary enamel knots (SEK) appear at its lingual and buccal sides (Jernvall and Thesleff, 2000). At the SEKs cell proliferation is downregulated while it is upregulated in the surrounding epithelia through the action of diffusible growth factors secreted by the SEKs (such as FGFs and SHH; Laurikkala et al., 2001; Jernvall et al., 1994). As a consequence the epithelium between SEKs grows and bends towards the dental mesenchyme forming a valley (fig. 1.2D) (Jernvall, 1995). Since there is no downward growth in the SEKs these are left behind and, consequently, form the cusps (with each SEK at the tip of each cusp) (Jernvall and Thesleff, 2000).

By E16 the position and shape of the first two cusps starts to become evident and two more SEKs appear at a certain distance posterior to the developing cusps. At this stage the cervical loops separate the mesenchyme

into two populations: the dental mesenchyme enclosed by the loops and the follicular mesenchyme surrounding the whole of the tooth germ (Rothová, Peterková, and Tucker, 2012). The former will give rise to the dentin forming odontoblasts and to the tooth pulp while the latter will not become part of the adult tooth. The suprabasal cells composing the bulk of the tooth germ start to vacuolate and secrete extracellular matrix, and at this stage they are called stellate reticulum (Butler, 1956). Two different epithelial populations can also be distinguished; the inner enamel epithelium (IEE) consisting of the epithelium enclosed between the two cervical loops, and the outer enamel epithelium (OEE) consisting of the epithelium facing outwards of the tooth germ. The final shape of the tooth crown and cusps corresponds to the shape of the IEE surface , since this is where the enamel secreting ameloblasts will differentiate.

By E17 two additional SEKs appear anterior to the first two cusps and two days later, at postnatal day 0 (P0), the shape of the tooth crown is already determined and the ameloblast and odontoblast layers can be distinguished.

1.2 Mathematical models of development as tools to integrate experimental evidence and predict the outcome of new experiments

There is a long list of realistic mathematical models of development involving both cell communication and morphogenesis by mechanical interactions such as in teeth (Salazar-Ciudad and Jernvall, 2002; Salazar-Ciudad and Jernvall, 2010), limbs (Hentschel et al., 2004; Raspopovic et al., 2014), turtle shell formation (Moustakas-Verho et al., 2014), fly egg cover (Osterfield et al., 2013), blood vessel development (Merks et al., 2008), pancreas development (Setty et al., 2008), lung development (Guo et al., 2014 ; Menshykau, Kraemer, and Iber, 2012) to cite a few. Mathematical models of development are based on experiments on specific developmental systems (such as an organ or part of the embryo at a specific stage). These experiments should allow to raise one or several hypotheses on the mechanism driving development in the respective system and in the main morphogenetic (movement of cells and tissues) and inductive (spatial changes in gene expression) events taking place during development.

The model is then built by formulating the mechanistic hypotheses in mathematical terms. This involves determining the spatial and temporal boundaries of the system, that is the cell layers that will be included and the developmental stages that will be modelled. It is also necessary to propose a gene network that specifies the interactions between genes, but also how certain genes regulate the cell behaviours involved in the development of the system (such as cell proliferation, migration, cell adhesion, etc.). In order to validate the model, and thus to accept or reject a mechanistic hypothesis, the model should be used to reproduce the wild-type phenotype of the system being studied and each of the developmental stages being studied. For a more stringent and informative test, however, the model should also be able to reproduce the phenotypic variation observed in the system (which is part of its variational properties), either in mutants, experimental manipulations of development or as natural phenotypic variation

between individuals within a species or between species. If the outcome of the model (the phenotype) matches the observed variation then the mechanistic hypothesis cannot be rejected. Otherwise, the hypothesis has to be rejected and another one needs to be devised. Failure, however, is informative about which aspects of the real dynamics are more poorly understood and, thus, are helpful for orienting future experiments. Furthermore, a validated model of development may be able to predict the outcome of experiments that have not been carried out yet. Thus, good models of development can become powerful tools to assist and guide experimental research programs.

Mathematical models of development usually include:

1. Initial conditions. Usually studies in development focus on a specific time range limited by two developmental stages. The initial conditions should reflect the developmental pattern observed in the real system at the initial stage considered.

2. The basic mechanics of cells and tissues. The mechanical interactions between different cells and between cells and the extracellular matrix that take place in the developmental system under study should be implemented in the model (e.g. regarding cell-cell adhesion, cell volume conservation and cell migration, to cite a few). Cell mechanics can be ignored in systems where cells do not move.

3. Intracellular gene networks and extracellular signals diffusing between cells (that can also be ignored in models where there is no cell-cell communication). 4) Cell behaviours (such as cell division, cell death, cell adhesion, cell contraction, etc.) and their regulation by the gene network. In order to illustrate the structure and implementation of a development model including all the previous elements, we will describe the tooth development model (Salazar-Ciudad and Jernvall, 2010) (fig. 1.3).

1.2.1 The tooth development model

Tooth development consists of the growth and folding of the dental epithelium over a mesenchymal condensate. During this process, epithelial signalling centres called secondary enamel knots are induced in specific positions and regulate cell proliferation and differentiation through cell signalling to neighbouring cells. Epithelial proliferation is downregulated in the enamel knots, but not in the space between them. As a result the epithelium grows and folds down between the knots, thus creating valleys between them and the knots end up at the tips of what would become the tooth cusps. The first knot is induced by the underlying mesenchymal condensate but, once formed, all knots secrete a second inhibitory signal that will preclude the formation of new knots at a certain distance from the existing knots. The tooth model summarizes those processes by defining a spatial context for cells, their mechanics, and a gene network that regulates the cell behaviours of proliferation and differentiation (Salazar-Ciudad and Jernvall, 2010).

In the tooth model, molecules are produced by cells and diffuse in the extracellular space between them (from now on extracellular signals). The model considers only three of those. 1) An activator signal that comes in

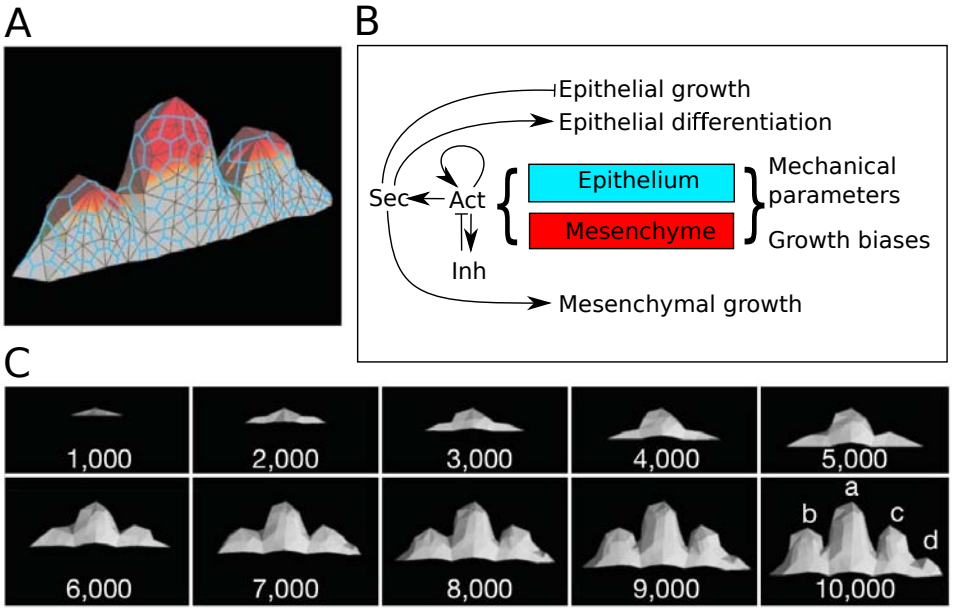

FIGURE 1.3: A, depiction of cells and their mechanical interactions in the developing tooth epithelium. Epithelial surface depicts the concentration of activator signal (in red), blue lines indicate cell borders, black lines depict mechanical interactions between neighboring cells. B, schematization of the gene network interactions, cell behaviours and mechanical parameters included in the model. Cell growth and differentiation are regulated by a signalling network. Activator signal Act promotes its own expression, the expression of Inh and Sec. Inhibitor signal Inh inhibits the expression of Act. Secondary signal Sec promotes cell differentiation in the epithelium and inhibits its growth (in absence of Sec epithelial cells grow at a default rate) and also promotes growth in the mesenchyme. C, time sequence of a model simulation from a lateral view. A small flat epithelium starts growing over the underlying mesenchyme. When some cells receive enough concentration of Act they differentiate into enamel knots, they stop growing and start forming the tips of the cusps. New enamel knots arise at a certain distance of existing knots because of Inh being secreted by the differentiated knots. Inh prevents the differentiation of cells close to the knots. The undiferentiated epithelium between existing knots continues growing downwards and form the valleys between the cusps.

through the borders of the system and promotes its own synthesis and secretion on epithelial cells. It also induces the differentiation of epithelial cells into enamel knots and thus inhibits cell proliferation (fig. 1.3B). 2) An inhibitor signal is only produced at the enamel knot cells and inhibits the secretion of the activator in the cells that receive it (fig. 1.3B). 3) A third extracellular signal is secreted from the enamel knots and induces cell differentiation (fig. 1.3B). Although there are a several extracellular signals expressed during tooth development (Salazar-Ciudad, 2008) they all have one of the following roles: promotion of cell differentiation (BMP2), promotion of knot differentiation (BMP4, Activin), inhibition of knot differentiation (Shh) and cell proliferation (FGFs). Thus, for the sake of simplicity the model only includes one signal for each role.

The model allows to simulate the real time dynamics of tooth development (fig. 1.3C). The simulation starts with a small, flat epithelium above

a block of mesenchyme (closely resembling the earliest stages in tooth development). As a result of model dynamics the epithelium starts growing downwards and soon the first enamel knot appears. As the tooth epithelium keeps growing, new knots appear separated from the existing ones and the growing epithelium between them becomes the valleys between the cusps (fig. 1.3C). Thus, signalling and induction are taking place while the shape of the tissues (where diffusion is taking place) is constantly changing. Only the initial conditions and the model parameters are specified before the simulation. It is important to note that, the final tooth phenotype and its change over developmental time arise as a result of the intrinsic dynamics between cells and gene products during the course of the developmental simulations. The outcome or phenotype of the model is, at each time in development, the morphology of the developing tooth, which is the distribution of cells in three-dimensional space, and the concentration of each molecule in them.

The model defines a number of parameters that quantify certain aspects of the cellular and signalling interactions (fig. 1.3C), such as the growth rate of cells, mechanical properties of cells, diffusivity constants of molecular signals and strength of regulatory interactions between genes. Different values of these parameters will change the dynamics of development during the simulation and thus alter its final outcome. For example, decreasing the parameter value specifying the secretion rate of the inhibitor in the enamel knots will reduce the distance between cusps in the final phenotype, and that happens because during the model simulation the amount of inhibitor secreted around knots is smaller which allows for new knots to differentiate closer to existing ones.

The model does not include all known genes and gene products involved in tooth development (for simplicity, the only gene products considered are the extracellular signals themselves). In many cases we do not know where are they located in the gene network of tooth development and thus cannot be included in the model. In other cases, variation in these genes produces variation in the tooth that is not interesting from the point of view of evolution and the GPM. That is the case when, for example, the mutation produces either no effect on tooth morphology or prevents teeth from forming at all. Any gene relevant for the production of morphology and its variation does so, necessarily, because it affects one or more of the cell behaviours involved in tooth development (that is cell signalling, cell division, cell adhesion and cell differentiation). For simplicity then the effect of these not-included genes can be encapsulated within the model parameters. For example the parameter specifying how much the activator promotes its own secretion on cells is in fact dependent on large number of gene products (this goes from this signal extracellular receptor to all the genes involved in this signal transduction and in the synthesis and secretion of the activator). Many of these developmental parameters are dependent on specific sets of gene products and then are, in a way, genetically encoded. Thus, instead of considering all these genes we can consider, for simplicity and in a first step, that parameters can change gradually to take any continuous value within a range. For the purpose of the GPM a genetic change in these genes would manifest itself as mere quantitative changes in a specific parameter. This means, in the context of GPMs, that these models consider a developmental parameter space rather than a genetic parameter

space *sensu stricto*.

The tooth model is a useful tool to predict evolutionary transitions from the fossil record (Salazar-Ciudad and Jernvall, 2002) and to assist empirical research on tooth development (Harjunmaa et al., 2014). Perturbation experiments or mutants that affect tooth development can be reproduced in the model by changing parameter values from the wild type phenotype (which could be seen as *in silico* mutations). For example, the inhibition of the Shh signalling pathway *in vivo* and *in vitro* causes a shortening in the separation distance between cusps, which results in an increase in their number due to a tighter packing of the cusps (Harjunmaa et al., 2012). That phenotype is reproduced in the model by decreasing the parameters affecting the rate of inhibitor secretion at the knots, which suggests that Shh is most likely acting as, or contributing to the pathway that inhibits the differentiation of the enamel knots. It is important to note that in this case the tooth model does not need to integrate any information from the Shh inhibition experiment in order to predict its outcome. Given a solid mechanistic hypothesis of the developmental process under study, theoretical models of development can provide predictions to the outcome of an experiment before it is carried out.

1.3 Predicting the effectiveness of natural selection on evolving populations *in silico* using models of development

Evolutionary theory states that the phenotypes of individuals better adapted to the environment will be more likely to be perpetuated due to a larger descendence compared to other individuals. This can be imagined using the metaphor of the adaptive landscape (Wright, 1932). Adaptive landscapes are relief maps in which each point in the surface is a genotype, and the height of the surface at that point is equal to its fitness. Populations at a certain point of the landscape can only move to the closest points in the landscape through mutation, but will only do so as long as it does not lead to a decrease in fitness. Thus, in the landscape metaphor populations always climb uphill and almost never downhill (stochastic drift may cause small downhill movements). It has been theoretically argued that the shape of an adaptive landscape has important implications for the ability of populations to adaptively evolve. Populations evolving on smooth landscapes will easily reach maximum fitness (the highest adaptive peak), whereas in rugged landscapes (i.e. with a rough surface and many fitness peaks) populations will often get trapped in sub-optimal local peaks (fig. 1.4, Kaufmann, 1993a).

The fitness of an individual is determined by both its phenotype and the environment (i.e. the ecological factors determining whether the phenotype is suitable for survival and/or reproduction). Then the adaptive landscape can in fact be decomposed into two different landscapes (fig. 1.4). A phenotype-fitness landscape mapping all possible phenotypes and their corresponding fitness (also called phenotype-fitness map or PFM) and a genotype-phenotype landscape mapping all possible genotypes and their corresponding phenotypes (the GPM). The structures of both landscapes have important implications for evolutionary dynamics. Complex GPMs

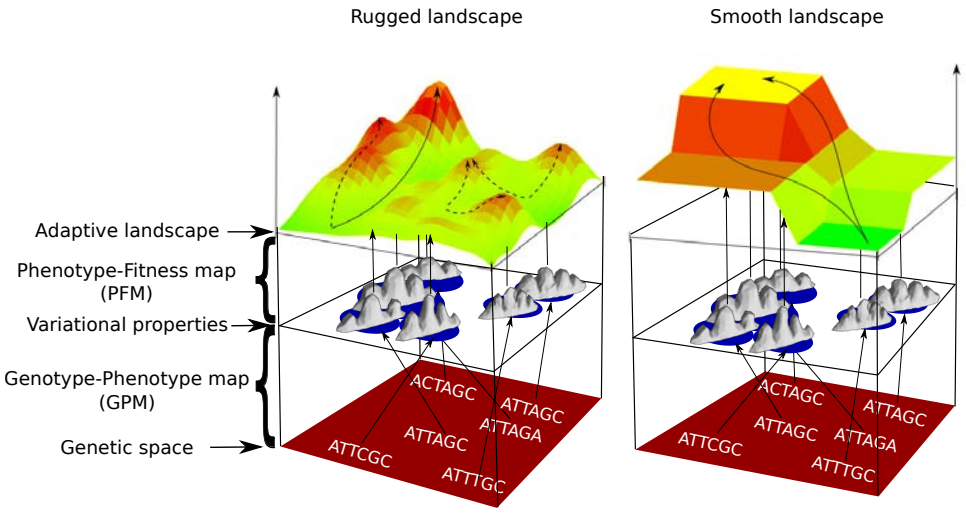

FIGURE 1.4: Two adaptive landscapes with their respective GPM and PFM are depicted. An adaptive landscape is an abstract space, depicted here as a relief surface, that determines the possible evolutionary trajectories of a population. The height of the surface depends on the fitness of the phenotype, thus populations will tend to climb uphill and seldom downhill. Once a peak is reached, adaptive change is precluded since it is not possible to further increase fitness. Rugged adaptive landscapes (left) consist in many peaks of different fitness values and so populations are more likely to get trapped in peaks of low fitness. Smooth, landscapes usually have fewer peaks, thus populations will often reach the peak of maximum fitness (right).

have been suggested to be the most prevalent, since they are present from the level of secondary RNA structure (Stadler, 2006) to the level of developing organisms (Salazar-Ciudad and Jernvall, 2005). It has also been shown that complex GPMs may increase the complexity of the adaptive landscape by increasing the disparity of morphological variation (Salazar-Ciudad and Jernvall, 2005).

The complexity of the GPM depends on how disparate phenotypes are when they are genetically similar. Let's consider an arbitrary phenotype within the variational properties and then consider a series of mutants that differ from the "wild type" in one genetic mutation. If the mutant phenotypes are similar to the wild type, that is the phenotypic change accompanying the mutations is small, then the GPM is simple. On the contrary if the mutant phenotypes show a large morphological disparity with the wild type, meaning that the small mutations have led to large phenotypic changes, then the GPM is complex. To put it in other words, a high complexity of the GPM is characterised by a low correlation between genetic change and phenotypic change. Hence, a complex or uncorrelated GPM may lead to an uncorrelated adaptive landscape and thus prevent adaptive evolution (Salazar-Ciudad and Jernvall, 2005).

Even if most morphological structures have a complex GPM, they often exhibit millions of years of adaptive evolution. Thus, there have to be, in theory, other factors that counteract the effect of GPM complexity on the ruggedness of the adaptive landscapes on the adaptability of populations. This question has been addressed by simulating the evolution of populations in which the individual's phenotype arises through a realistic model

of development. Chapter 3 of this thesis presents the results of this study, in which tooth evolution was simulated using the tooth development model and the shape of the adaptive landscape was estimated in different selective regimes (Salazar-Ciudad and Marín-Riera, 2013).

During the evolutionary simulations, individual fitness is determined by a function assessing how much an individual phenotype resembles an arbitrary target phenotype. In this evolutionary model, populations "walk" on the adaptive landscape (technically they walk at the same time on the phenotype-fitness landscape and the GPM), thus the evolutionary process and the dynamics of adaptation can be monitored. The degree of ruggedness of the adaptive landscape can be inferred by measuring how often populations reach the optimum (that is the target phenotype) in a series of evolutionary simulations. Different functions can be used to calculate individual fitness based on the phenotype. By assessing the frequency with which populations reach the optima under different fitness functions, we can infer the shape of the adaptive landscape in different selective regimes.

In the same study we also explore the variational properties of the tooth development model and analyse the distribution of complex phenotypes within the parameter space.

1.4 Expanding models of animal morphogenesis in order to analyse the properties of development across systems

Mathematical models of development allow to systematically explore the variational properties and the GPM, thus providing an understanding of which phenotypic variation natural selection can act on at each generation. Most models of development used in earlier studies either focus on one specific organ, or simply do not include all the cell behaviours known to be involved in animal development. Organ-specific models of development have been shown to reproduce quite accurately the relationship between genotype and phenotype of its real counterpart (Salazar-Ciudad and Jernvall, 2010; Zhu et al., 2010; Hentschel et al., 2004; Raspopovic et al., 2014; Moustakas-Verho et al., 2014) which has allowed to make predictions about evolutionary transitions. Thus, their application is restricted to a specific system and cannot provide general predictions on the properties of development in general. Moreover, the design of those models constrains them to only reproduce the GPM of a single organ, so it is not possible to infer evolutionary transitions between different organs (e.g. the transition between reptilian scales and avian feathers).

The next logical step in the modelling of developmental mechanisms should be the design of models that are able to reproduce the development of a variety of organs and structures. By reproducing the development of different organs with the same model we could quantitatively compare their differences at the level of development and infer which changes would be required to evolve from one organ to the other. It is necessary to devise more general and detailed models that include a wider range of cell dynamics and cell behaviours in order to study the general properties of development and the variational properties and GPM of a particular group of organs. Recently, a number of modelling frameworks have been published

that implement some of these cell behaviours (Swat et al., 2012; Starruß et al., 2014b; Izaguirre et al., 2004; Farhadifar et al., 2007; Smith et al., 2012; Pitt-Francis et al., 2009).

In order for a general model to predict the variational properties of development it requires not only the inclusion of all cell behaviours, but also to account for the different properties of epithelial and mesenchymal cells and extracellular matrix. Epithelial and mesenchymal cells differ greatly on the way they interact with other cells to form tissues and on how they transmit mechanical forces. There are a number of models that are good at representing epithelial cells (e.g. vertex model; Honda, Tanemura, and Nagai, 2004) and mesenchymal cells (e.g. subcellular elements model; Newman, 2005, Cellular Potts model; Graner and Glazier, 1992), but very few are good at representing both.

The Cellular Potts Model (CPM) is a regular lattice based model in which each cell occupies several contiguous lattice positions (Graner and Glazier, 1992) (fig. 1.5). The model determines whether a cell spreads to, or retracts from a specific lattice position by changing the identity of that specific position, which can then belong to a different cell or to the extracellular space, and thus change cell and embryo morphology. A Monte Carlo method is used to solve the temporal progression of the system. In this method, potential energies are calculated for each lattice position (that is for each part of the cell and extracellular space) based on the local stability of the system. In order to calculate the potential energy at each position, a set of rules have to be defined that can vary in different instantiations of the model. The most common rules are based on cell volume and area conservation and cell-cell adhesion. Changes in lattice positions are made asynchronously, that is one position is chosen at random each time (fig. 1.5A). Then a change in the identity of the lattice position is made that minimizes the potential energy of the system. Energy minimization is also stochastic, that is, an identity change that increases the energy of the system is less likely than one that decreases it, but it is still possible. Usually a "temperature parameter" is defined that determines the likelihood of positive energy changes in the system. Several additional cell behaviours have been incorporated into this model (Hogeweg, 2000). Recent versions of the model include cell division, cell migration (by chemotaxis), apoptosis, and some form of extra-cellular signaling in 2D and 3D (Cickovski et al., 2005; Marée, Grieneisen, and Hogeweg, 2007). There are some user friendly interfaces that implement this model (Swat et al., 2012; Starruß et al., 2014a). The CPM is specially suitable for developmental systems in which cells change their neighbours and shape in a rather dynamic way (such as in cell sorting and migration by chemotaxis and haptotaxis, Steinberg, 1996; Cai and Montell, 2014), but it is not good at predicting the transmission of mechanical forces across cells and tissues.

In the Vertex model (Honda, Tanemura, and Nagai, 2004; Honda et al., 2008) space is not discrete but continuous. Cells are defined by the contact interfaces with their neighbours (fig. 1.5). This means cells are represented as a mesh of polygons in the 2D case and of polyhedra in the 3D case. The

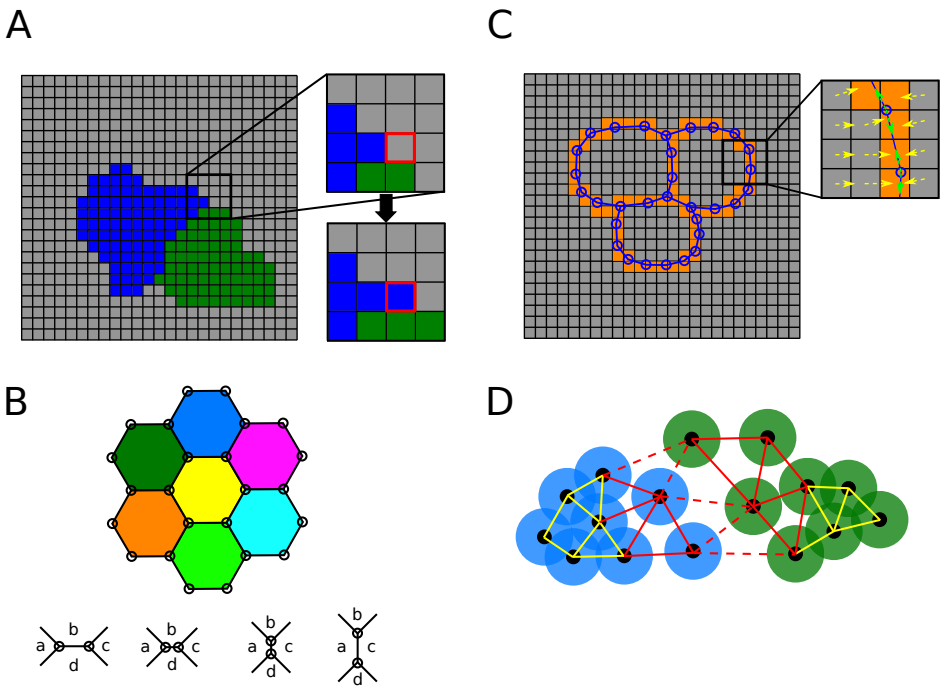

FIGURE 1.5: A, the Cellular Potts is a lattice based model in which each lattice position can take the identity of any cell in the system (blue or green in this case), or of the extracellular space (grey). Each cell is represented by the ensemble of lattice positions with their specific identity. Cell dynamics take place by changing the identity of random lattice positions based on a set of specific rules (inset, red square). B, the Vertex Model considers a continuous space and cells are defined by the contact interfaces with other cells (i.e. vertices and edges). Each cell is depicted in a different color, but note that vertices and edges are usually shared by multiple cells. Model dynamics can lead to cell rearrangements by shifting the relative position of vertices. The sequence at the bottom of B shows the process of rearrangement in which cells b and d lose contact and a and c come into contact. C, the Immersed Boundaries Model uses a continuous space to model cell membranes as articulated elastic bodies (blue) and a discretised lattice space to model the rheology of cell cytoplasm and extracellular medium (grey). The two layers are connected by an interpolation function (orange squares) that determines how the flow velocity field in the lattice will exert a mechanical force in the cell membrane (inset, yellow arrows) and how the forces arising from the elastic deformation of the cell membrane (inset, green arrows) will affect the flow velocity field in the lattice. D, the Subcellular Elements Model consider each cell composed of several spherical elements in a continuous space. Each element has a volume defined by an equilibrium radius (green and blue circles). When two elements are closer than the equilibrium distance (overlapping circles), repulsive forces arise (yellow lines), whereas if they are farther than that, attractive forces arise (red lines). Blue and green represent elements from different cells, solid lines indicate interactions within the same cell, dashed lines indicate intercellular interactions.

temporal progression of the system is determined by changes in vertex positions. These can be solved either by means of a system of equations of motion or with the Monte Carlo method. In both cases, the factors that determine the direction movement of vertices are cell surface area/volume conservation, cell membrane contractility, line tension at the contact interface

between cells and cell-cell adhesion (Osterfield et al., 2013). Cell rearrangements (i.e. changes in the neighbouring connections between cells) are possible in this model through the interchange of vertices between groups of cells (Honda, Tanemura, and Nagai, 2004; Honda and Nagai, 2015, see fig. 1.5B). This model is especially suitable for modelling flat two dimensional epithelia (i.e. that do not bend or fold in 3D) and highly packed three-dimensional tissues with little extracellular space between cells (Honda et al., 2008). Since all calculations, and cell shape itself, are defined at the contact interface between cells this model does not easily accommodate processes involving very dynamical cell contact changes (as in cell sorting and convergent extension, Tada and Heisenberg, 2012), ECM and the interactions between cells and ECM or with environment (such as in migration by haptotaxis and chemotaxis, Paluch and Heisenberg, 2009).

The viscoelastic model or Immersed Boundaries Cell (IBCell) Model (Rejniak, Kliman, and Fauci, 2004; Rejniak, 2007) combines an off-grid representation of cells as elastic bodies with a lattice Boltzmann Method to simulate cell cytoplasm and intercellular fluids (fig. 1.5C). Thus, each cell is delimited in a continuous 2D space by a set of points connected by elastic springs, conforming the cell membrane, whereas fluid rheology is calculated on a regular grid in discrete space. Cell membranes separate the medium inside the cells (the cytoplasm) from the one outside them (the extracellular environment). The model uses a Dirac delta function (Peskin, 1972) to link the mechanics of the elastic elements and the fluid medium. Cell membrane movement and deformation is governed by the elastic forces generated by the springs in the membrane and by the pressures coming from the medium, which are interpolated based on the velocity field near the cell membranes. Deformations originated at the cell membrane also feed back to the fluid grid by affecting the fluid velocity field near the cell membranes (fig. 1.5C). This allows to model not only the diffusion of molecules in a context where cell membranes, cell cytoplasm and the extracellular space are taken into account, but also the advection of fluids due to cells moving and displacing the surrounding media. This implementation and level of detail makes the model very heavy computationally which precludes, according to the authors (Tanaka, Sichau, and Iber, 2015), its extension to 3D problems. Several cell behaviours have been implemented, such as cell proliferation (Rejniak, 2007), cell growth, differentiation and cell-cell signalling (Merks et al., 2011). This model has been implemented in several modelling frameworks (Merks et al., 2011; Tanaka, Sichau, and Iber, 2015; Shapiro, Meyerowitz, and Mjolsness, 2013).

In the subcellular elements model (SEM) (Newman, 2005; Sandersius and Newman, 2008) each cell is composed of several subcellular elements that mechanically interact between each other as well as with elements belonging to other cells (fig. 1.5D). This model has the advantage over the CPM that cells move in a continuous space and over the vertex model that cells are defined by their volumes and not by their contact interfaces with other cells. Each element occupies a certain volume in space, which is determined by an equilibrium radius. If the distance between two adjacent elements is shorter than the sum of their equilibrium radii, a repulsive force will tend to separate them, whereas when distance is longer than that an attractive force will tend to drive them closer to one another. A cut off distance is determined for each element so it will only interact with the

elements that are found within that distance. Physical integrity of cells is kept in the model by making that the attractive force between elements of the same cell is always larger than the one between elements from different cells (Newman, 2005). This model allows a wide range of cell behaviours to be implemented with ease. Cell growth and cell death are the result of respectively adding or removing elements from a cell. Cell division is the result of splitting the elements of a single cell into two new daughter cells. Given that cells can be composed of a large number of elements, asymmetric cell divisions can be implemented by distributing a different number of elements to each daughter cell (see Chapter 4). Active cell deformation (such as cell contraction) can be accomplished by modifying the equilibrium radius of the some of the cell's elements, and passive cell deformations will take place by a spatial rearrangement of the elements within a cell. Differential cell adhesion can be modelled by assigning different values to the attractive forces between elements from different cells. A modelling framework based on the SEM has been recently developed (Delile, Doursat, and Peyriéras, 2013). This work includes a number of cell behaviours such as cell signalling, adhesion, division, active cell migration and simple boolean gene networks, but doesn't implement epithelial cells with cellular and mechanical properties different from mesenchymal cells.

Chapter 4 of this thesis introduces a new modelling framework that allows to model the development of a wide range of systems involving epithelia, mesenchyme and extracellular matrix (Marin-Riera et al., 2016). The modelling framework uses the concept of subcellular elements (Newman, 2005) for mesenchymal cells and extracellular matrix and a modified implementation of that same concept for epithelial cells. That allows an integrative modelling of epithelial and mesenchymal cell mechanics and cell adhesion. The framework also implements the entirety of cell behaviours known to happen during animal development (Salazar-Ciudad, Jernvall, and Newman, 2003) and customizable gene regulatory networks. All this has been implemented in an open-source software called EmbryoMaker (Marin-Riera et al., 2016).

1.5 Predicting phenotypic variation arising from development and the developmental basis of evolutionary transitions

Knowing the ensemble of phenotypic variation that a developmental process can produce is relevant in order to make predictions on the evolution of a lineage. The variational properties will tell us which phenotypic variants are present at each generation for natural selection to act on, so they tell us in which directions the phenotype can change in evolution. The degree of stringency of natural selection can also be inferred by assessing what proportion of the phenotypes present in the variational properties are absent in observed natural phenotypic variation. This can be done at different levels, from the whole variational properties of a species or a developmental mechanism, to the local variational properties accessible by single small mutations from a given individual or a set of individuals within a population.

In order to know the variational properties of development in an organism, one ought to study all the genetic variants or mutants that are known to affect the phenotype through that developmental process (either for a whole species or for a population in a given generation). For most species and organs this is likely to be difficult and quite time consuming. Mathematical models of development offer a cheaper and faster alternative for exploring those variational properties. In addition, since models are built on explicit hypotheses and assumptions it is often the case that we can understand why the model produces certain phenotypes (while this is never the case in purely statistical approaches). The computational exploration of the variational properties of development is often performed through a parameter screening of the model. In a parameter screening a large number of parameter combinations are generated by changing all the parameter values and fed to the model in order to get the set of possible morphologies (Salazar-Ciudad and Jernvall, 2004; Prusinkiewicz et al., 2007) (fig. 1.1).

Even though variational properties are informative about what phenotypes can be generated by development, they do not allow to make accurate predictions about evolutionary change in populations. The variational properties for a certain developmental process include all the phenotypes resulting from genetic variants that do not alter the structure of the network of molecular/cellular/tissue interactions. However, variational properties do not provide information about the likelihood of a certain phenotype to arise due to a genetic mutation in another phenotype. In natural populations it will most likely be the case that the phenotypic variation resulting from the standing genetic variation does not account for the whole variational properties. In that situation, new phenotypes may appear in the population through small mutations in existing phenotypes. In each generation only those phenotypes that are "one mutation away" from the existing genetic variants are likely to appear (fig. 1.6). Thus, in order to predict the direction of evolutionary change it is important not only to know the variational properties of development, but also how phenotypes are distributed with respect to all the genetic variants (that is, how different phenotypes are connected by genetic mutations, Salazar-Ciudad and Jernvall, 2005; Huynen, Stadler, and Fontana, 1996). The distribution of phenotypes with respect to the genetic variants is highly dependent on the GPM (Salazar-Ciudad and Jernvall, 2004). This is especially relevant regarding the emergence of novel traits or structures (Newman and Müller, 2000; Muller and Wagner, 1991). The fossil record shows proof of a large number of evolutionary novelties in the form of abrupt phenotypic changes in many lineages (Gould and Eldredge, 1977). It can be interesting to find out what changes in the genotype and in development have driven those transitions and which, if any, the intermediate phenotypes are. Computational models of development can infer the genetic and developmental bases of phenotypic transitions and their intermediate phenotypes. By finding the sets of parameter combinations (the genotypes) that give rise to both the ancestral and novel phenotypes in the model, one can determine how many and which changes in the parameters are needed to get from the former to the latter through intermediate phenotypes. This will give us information about which aspects of development need to change in order for that novel phenotype to arise. Using a model of tooth development several evolutionary transitions present in the fossil record were reproduced (Salazar-Ciudad

and Jernvall, 2002) and even though the phenotypic changes involved were quite large, they required relatively small changes in one or two parameters, suggesting that major evolutionary novelties do not require necessarily extensive changes in genetic or developmental regulation. Similarly, a model of limb development was used to reproduce the spatial bone patterns of limbs of several early tetrapod species from the fossil record by keeping the parameters of the core signaling network and only modifying the rates of tissue growth (Zhu et al., 2010).

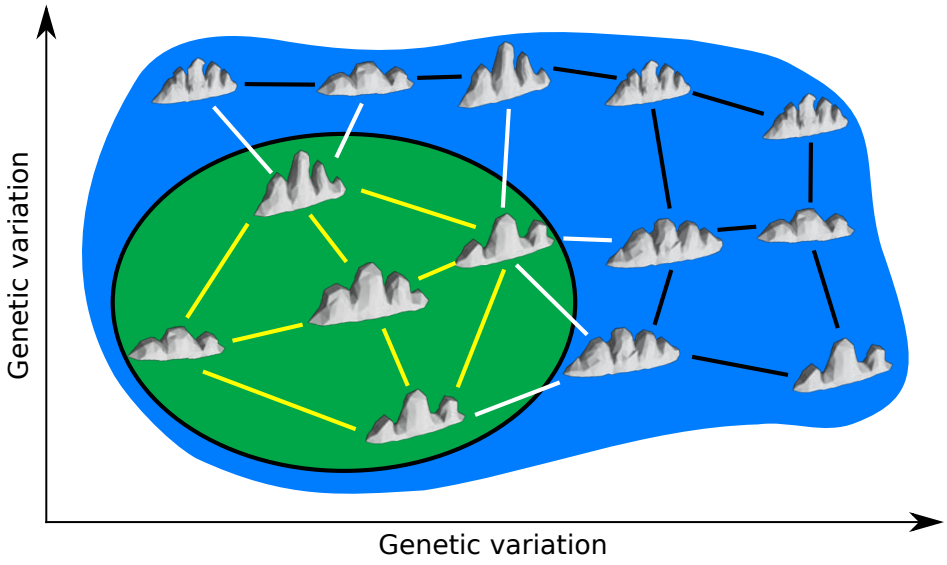

FIGURE 1.6: The variational properties of a certain structure are depicted as a distribution of phenotypes in the genotypic space. Each phenotype is connected to other phenotypes by a single genetic mutation. Given that populations use to have a limited amount of genetic variation, only a part of the total range of phenotypes can be present in the population (green circle). Any new mutation will originate from a genotype already existing in the population and thus the phenotype arising from the mutation will be either equal to the other existing phenotypes in the population (yellow lines) or a new phenotype one mutation away from the population's standing genetic variation (white lines), but never one of the phenotypes that are farther away (i.e. in the far right side of the variational properties). In order to reach those phenotypes, the pool of genetic variation will need to shift towards that direction, by means of natural selection or drift, until they are close enough so they can arise by single genetic mutations.

Models of development can also be used to understand the specific developmental origin of the standing phenotypic variation within a population. Gathering empirical data on how variable development is in a natural population is methodologically challenging. Computational models of development can be used to generate different sets of phenotypic variation by making screenings of one model parameter each time, creating different *in silico* populations. Then by assessing which *in silico* population resembles most the natural population we can infer which parameters of development may contribute to the natural phenotypic variation (Salazar-Ciudad and Jernvall, 2010). When comparing virtual populations generated by the tooth model with teeth from a natural population of seals, it was found that variation stemming from the model parameters involved in the diffusivity

of the inhibitor and the activation of the activator by itself were able to explain the largest part of the natural variation observed (Salazar-Ciudad and Jernvall, 2010).

Earlier mathematical models of tooth development (Salazar-Ciudad and Jernvall, 2002; Salazar-Ciudad and Jernvall, 2010) were able to predict the spatial arrangement of tooth cusps and their shape, but considered only the inner enamel epithelium (IEE) and the dental mesenchyme: the outer enamel epithelium (OEE), the follicular mesenchyme and suprabasal layer were not simulated. In these previous models the cervical loops were not represented in their entirety (only the part corresponding to the IEE) and the mechanics of their growth were implemented ad hoc: the cervical loops were forced to grow towards the mesenchyme with a lateral component that depended on the proliferation rate of the dental mesenchyme. Another model of tooth development (Takigawa-Imamura et al., 2015) implemented slightly more realistic cell mechanics, but only in two-dimensions and without considering neither the mesenchymal layer nor the presence of signalling centres regulating cell proliferation. Thus, current models of tooth development lack explanatory power regarding the aspects of morphogenesis affecting the overall shape of the tooth crown and tooth cusps and the role of cell mechanics on tooth morphogenesis. In order to achieve a better understanding of how tooth crown and cusp height and sharpness is determined a more realistic modelling of the mechanical interactions between the IEE, the OEE, the dental mesenchyme, the dental follicle and the suprabasal layer is required.

Chapter 5 of this thesis presents a new 3-dimensional model of tooth development that simulates the growth and mechanical interactions between all the tissue types that compose the tooth germ. This new tooth model is built as an application of a general model of embryonic development described in Chapter 4 of this thesis and implemented in the EmbryoMaker software (Marin-Riera et al., 2016).We use this model to explore the effect of differential cell growth and differential cell adhesion on tooth morphogenesis, and more specifically on cervical loop growth, through the mechanical forces exerted between cells during the period between bud stage and cap stage. The model aims to predict how the mechanical forces originating from growing tissues and cell adhesive contacts will affect the direction of growth of the cervical loops and the shape of the cusps, which will have a significant influence on the overall shape of the tooth crown. By analysing the variational properties of this new model we will be able to predict the role of cell adhesion and mechanics in the generation of morphological variation, something that was not accounted for in previous models.

The initial stage of the model is set after the induction of the PEK and simulates the emergence of the cervical loops from the tooth bud and later the formation of the first cusps within the tooth cap. The model predicts the separation of the epithelium into IEE and OEE and the mesenchyme into dental and follicular. Induction of the SEKs and Turing-like dynamics are not included in the model because they take place at later stages and are not required to address the questions raised in this study. Thus, this model does not intent to predict the relative position of tooth cusps within the tooth crown, as in earlier models (Salazar-Ciudad and Jernvall, 2002; Salazar-Ciudad and Jernvall, 2010), although these features will be included in the future. Accordingly, the model is aimed to predict how the

mechanical forces originated from growing tissues and cell adhesive contacts will affect the direction of growth of the cervical loops and the shape of the cusps, which will have a significant influence on the overall shape of the tooth crown.

Chapter 2

Objectives / Objectius

1. Provide theoretical predictions on how phenotypic variation and phenotypic complexity arise through the process of development, taking the morphogenesis of mammalian tooth as a case study and using a mathematical model of tooth development.
2. Provide theoretical predictions on how the complexity of the relationship between genotype and phenotype coming from development may affect the evolutionary dynamics of a population, taking tooth morphology as the phenotype under selection.
3. Expand the current model of tooth development to include new aspects of development that may be relevant for the origin of phenotypic variation.
4. Provide theoretical predictions on how the inclusion of these new features contribute to the generation of phenotypic variation in tooth development.

---

1. Generar prediccions teòriques sobre com el desenvolupament produeix la variació fenotípica i la complexitat morfològica, prenent la morfogènesi de la dent de mamífer com cas d'estudi i utilitzant models matemàtics del desenvolupament de la dent.
2. Generar prediccions teòriques sobre com la complexitat de la relació entre el genotip i el fenotip provinent del desenvolupament poden afectar les dinàmiques evolutives d'una població, prenent la morfologia de la dent com a fenotip sota selecció.
3. Expandir l'actual model de desenvolupament de la dent per tal d'incloure nous aspectes del desenvolupament que poden ser rellevants per entendre l'origen de la variació fenotípica.
4. Generar prediccions teòriques sobre com la inclusió d'aquests nous aspectes del desenvolupament contribueix a la generació de variació fenotípica en el desenvolupament de la dent.

Chapter 3

Predicting the effectiveness of natural selection on evolving populations *in silico* using models of development

The contents of this chapter were published in the form of a research article:

Salazar-Ciudad, Isaac and Miquel Marín-Riera (2013). "Adaptive dynamics under development-based genotype-phenotype maps." Nature 497.7449, pp. 361–4.

3.1 Abstract

Is natural selection able to find any arbitrary phenotype that can be produced by genetic variation? There is a long-lasting debate about processes limiting adaptation and, consequently, about how well adapted phenotypes are. Here we examine how development may affect adaptation by decomposing the genotype-fitness map into two mappings: one from genotype to morphology by means of a computational model of organ development, and one from morphology to fitness. In the latter map, the fitness of each individual is based on the similarity between realized morphology and optimal morphology. We use three different morphology-fitness maps based on how similarity is calculated: similarity is calculated for each trait (in terms of cell position individually), similarity is calculated for a large or a small number of morphological landmarks, and similarity is calculated at the level of the overall roughness of morphology. Evolution is simulated by applying, in every generation, the genotype-phenotype map and one phenotype-fitness map to each individual in the population, with mutation and drift. We show that the complexity of the genotype-phenotype map prevents substantial adaptation in some of these phenotype-fitness maps: sustained adaptation is only possible by "roughness" or "few-traits" phenotype-fitness maps. Our work adds developmental understanding to the long-standing question of which aspects of the phenotype can be effectively optimized by natural selection.

3.2 Introduction

The relationship between genetic and phenotypic variation, or the genotype-phenotype map (Wright, 1932; Waddington, 1957), has been argued to be

complex at all levels of organization in which it has been studied: RNA (Huynen, Stadler, and Fontana, 1996), proteins (Ferrada and Wagner, 2012) and development (Alberch, 1982; Salazar-Ciudad, 2006). If this relationship is complex, small genetic differences between a parent and a descendant do not necessarily lead to small phenotypic differences between them. In other words, genotypes producing adapted morphologies may not have any offspring with better-adapted morphologies. Then evolution by natural selection can often get trapped in local optima (Waddington, 1957; Kaufmann, 1993b).

There are a number of models of genotype-phenotype maps and adaptive dynamics (Wagner, 1994; Hansen and Wagner, 2001; Salazar-Ciudad and Jernvall, 2004). At the level of protein and RNA secondary structure (Huynen, Stadler, and Fontana, 1996) some models are able to provide realistic estimations of the map based on the rules of molecular interaction by which these phenotypes arise. For morphological phenotypes, understanding the genotype-phenotype map implies understanding the mechanisms of development (Alberch, 1982). Development has been repeatedly argued to constrain adaptation by affecting which morphological variation is possible and which is not, and by the complexity of the genotype-phenotype map it entails (Alberch, 1982; Salazar-Ciudad and Jernvall, 2004). It seems clear that development has an effect on adaptive dynamics but it is not evident how strong that effect is. Here we explore how a realistic approximation of this genotype-phenotype map provides a richer and more quantitative understanding of the limitations that development may impose on adaptation.

We employ a computational model of tooth development as an example of a genotype-phenotype map for a complex organ morphology. The tooth model mathematically summarizes the basic genetic and cellular interactions regulating tooth shape development (Salazar-Ciudad and Jernvall, 2002). The strength of those interactions is encoded by the values of the model parameters, and we take those values as a proxy for individual genotypes (see Appendix A). The developmental model produces a phenotype, a 3-dimensional morphology, from the parameters in each individual. As a result of model dynamics teeth with different parameter values can have different numbers of cells and different number of identifiable morphological landmarks such as tooth cusps. The model has previously been used to reproduce morphological variation at the level of natural populations (Salazar-Ciudad and Jernvall, 2002), and thus, represents current understanding in this systems of how development leads to adult morphology and its variation (a genotype-phenotype map). Thus, we take this model as informative about the actual complexity of the genotype-phenotype maps for a representative complex, functional organ.

We assume, for simplicity, that individual fitness is determined exclusively by the morphology arising from the developmental model. We use three kinds of phenotype-fitness maps, or natural selection criteria, to explore how and how much adaptation occurs, given our developmental genotype-phenotype map. These criteria assign fitnesses to individuals on the bases of three different ways of measuring the distance between each individual's morphology and a predetermined optimal morphology. Thus, the evolutionary process is simulated by applying these two mappings, genotype-morphology and morphology-fitness, to each individual. Each individual's

fitness determines its chances of contributing to the next generation. The simulations also include random mutation and drift. Mutation is implemented as changes in the developmental model parameters (see Appendix A).

The first selection criterion (fig. 3.1A), the Euclidean Morphological Distance (EMD), considers all cell-level traits in each teeth: this is the position of each cell in each three-dimensional tooth form. This is all the morphological resolution provided by the model: all other measurable traits are derived from these cell-level traits and thus EMD can be seen as taking into account all the morphological information or detail there is. In the EMD criterion the value of each cell-level trait in a form is compared against the value of the corresponding trait in an optimal morphology. These two morphologies can have different numbers of traits (see section 3.4). The fitness of each individual is then the sum of the fitness contribution of each cell-level trait. An individual is optimal when, for each cell-level trait, the distance to the optimum is close to zero (see section 3.4). Thus, all traits contribute to fitness.

The landmark-based selection criteria (fig. 3.1B) are like the EMD but based only on a small number of traits, the spatial coordinates of a small set of landmarks shared by all morphologies considered. Landmarks are chosen among the cusps present in teeth (see section 3.4). Different criteria use different number of traits: the few-traits criteria use 2 or 4 traits and the many-traits criteria use 8 or 13. These are chosen among the spatial coordinates of the landmarks either arbitrarily or from a Principal Component Analysis (PCA) of variation (see section 3.4).

In the Orientation Patch Count (OPC) selection criterion (fig. 3.1C) we measure the high-level (Kaufmann, 1993a) overall ruggedness of the tooth surface as the number of facets in the morphologies with different orientations in the x-y plane (fig. 3.1) (see Methods). Thus, fitness is not assigned by comparing single traits between a morphology and the optimum morphology but, instead, by comparing each trait with its neighbouring cell-level traits in the same individual and comparing the resulting OPC with that of the optimum (see section 3.4).

These three phenotype-fitness maps can be seen as lying on an idealized spectrum of possibilities from all traits being selected, to some traits being selected and to no traits being selected *per se*. They also represent three different views existing in the literature on phenotypic evolution. Optimality theory proposes that natural selection is the dominant force in evolution and that, consequently, one should expect most (Orzack and Sober, 1994), or at least many (Maynard Smith, 1978), phenotypes to be optimal. The EMD criterion is an extreme version of the idealized view that most phenotypes are optimal and that this can be studied by decomposing phenotypes into single traits that are themselves optimal (Gould and Lewontin, 1979; Orzack and Sober, 1994). A much less extreme version is to select for a large number of landmark-based traits to be optimal: the many-traits selection criterion. In fact, morphological adaptation has been traditionally understood by looking at a limited number of characters (Ji et al., 2002; Charles et al., 2011) and landmarks (Klingenberg, 2002). The Orientation Patch Count (OPC) (Evans et al., 2007) and other related measures (Bunn et al., 2011) are alternative ways to understand adaptation that are not directly related to specific landmarks but that yet consider all cell-level traits. More generally,

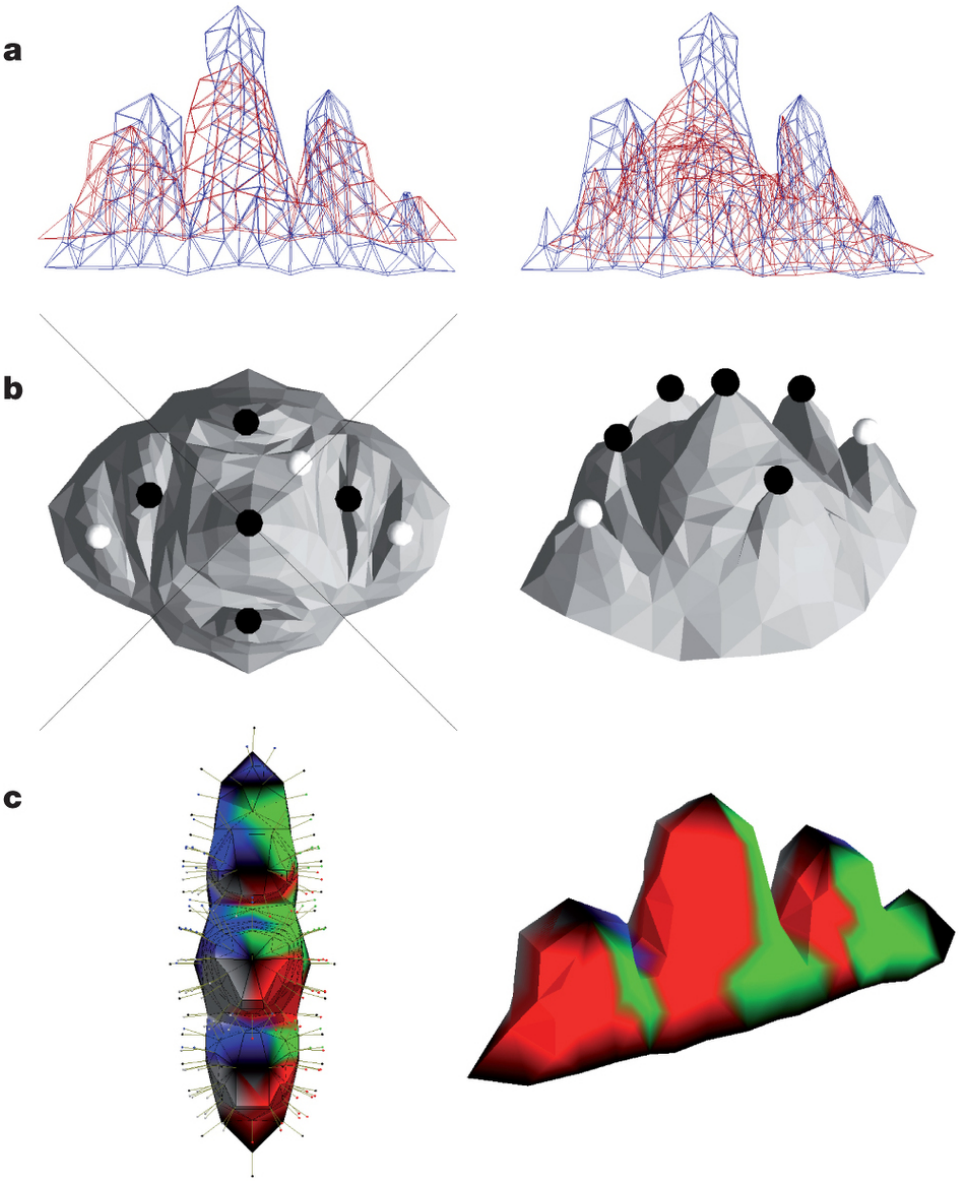

FIGURE 3.1: A, EMD is a measure of sheer disparity between two morphologies. It is defined as the sum of the distances of each cell in one morphology to the closest cell in the other morphology, divided by the sum of the surfaces of the two morphologies. The more a morphology resembles the optimal morphology, the smaller is the EMD between them. The figure depicts two pairs of morphologies, the optimal morphology in red, one in which EMD is small (left) and one in which EMD is larger (right). B, In the landmarks-based criterion we define the tips of five cusps on a morphology as landmarks and either take their coordinates (x,y and z), or their scores in a Principal Component Analysis (PCA) of variation as traits. Then the distances between these traits and the corresponding traits in the optimum are measured. C, OPC is a measure of morphological roughness. The orientation of the slope at each point is calculated and sets of adjacent points with the same orientation are grouped into patches. The total count of these patches is the OPC.

it has been suggested that phenotype-fitness maps should be degenerate, or many-to-one, in the sense that many different phenotypes should have

the same fitness (Alfaro, Bolnick, and Wainwright, 2005). Thus, we chose to implement OPC because it is a degenerate measure that has been used to study teeth previously (Evans et al., 2007; Santana, Strait, and Dumont, 2011; Godfrey et al., 2012).

3.3 Results and Discussion

Each evolutionary simulation started with all individuals having the same randomly chosen initial morphology. The optimal morphology in each simulation was chosen to be at a distance of 20%, 40%, 60% or 80% from the initial morphology (for each of the selection criterion; see Appendix A).

The simulations show that, initially, absolute fitness increases rapidly by relatively large steps but over generations these steps become smaller and less frequent (see Appendix A, fig. A.1). The EMD criterion does not reach the optimum often and, mostly, do not exhibit substantial adaptation (fitness does not increase much over evolutionary time) (fig. 3.2 and fig. A.2). This is the case even in simulations in which the initial and optimal teeth are only one mutation away. When the population is large (10000), the EMD criterion does in some cases reach the optimum but only when that optimum is at a phenotypic distance of 40% or less from the initial morphology in the evolutionary simulation. This population size is on the same order of magnitude than the estimated population sizes for mammals (Charlesworth, 2009). Most informative is, moreover, that morphologies at a 20% or 40% EMD from each other (fig. A.3, A.4) are barely distinguishable and yet natural selection can not reach one from the other (unless very large populations are considered). The many-traits criteria seldom reach the optimum (less than 40% of simulations for populations of 1000), but show substantial adaptation on average, specially for large populations. However, progressively higher levels of adaptation are attained for fewer trait selection criteria (see also fig. A.2).

Our results clearly indicate that adaptation for the EMD decreases rapidly with the phenotypic distance between initial and optimal morphologies (fig. 3.2C). The EMD only leads to substantial amounts of adaptation when the initial and optimal teeth are very similar or when both are simple unicuspid teeth (fig. 3.2D). In this case there is only a small amount of adaptive morphological change. In contrast, with the OPC criterion high degrees of adaptation are reached even for large distances between initial and optimal teeth (fig. 3.2). In fact, the OPC selection criterion often reaches the optimum, and leads to cumulative morphological changes that are larger, than the ones reached when the EMD and many-traits criteria selection criteria are employed (fig. A.6C).

We interpret our results to indicate that with complex genotype-phenotype maps adaptation can only occur if the overall genotype-fitness map is sufficiently simple. The OPC allows this by being degenerate, that is, the same OPC values are found in forms that differ in their morphological details (Evans et al., 2007). In figure 3.3 we have quantified the degeneracy in each phenotype-fitness map (see fig. A.7, A.8, A.9, A.10 for related measures). In degenerate maps, like OPC, genetically related individuals can have quite different phenotypes because of the complex genotype-phenotype map but

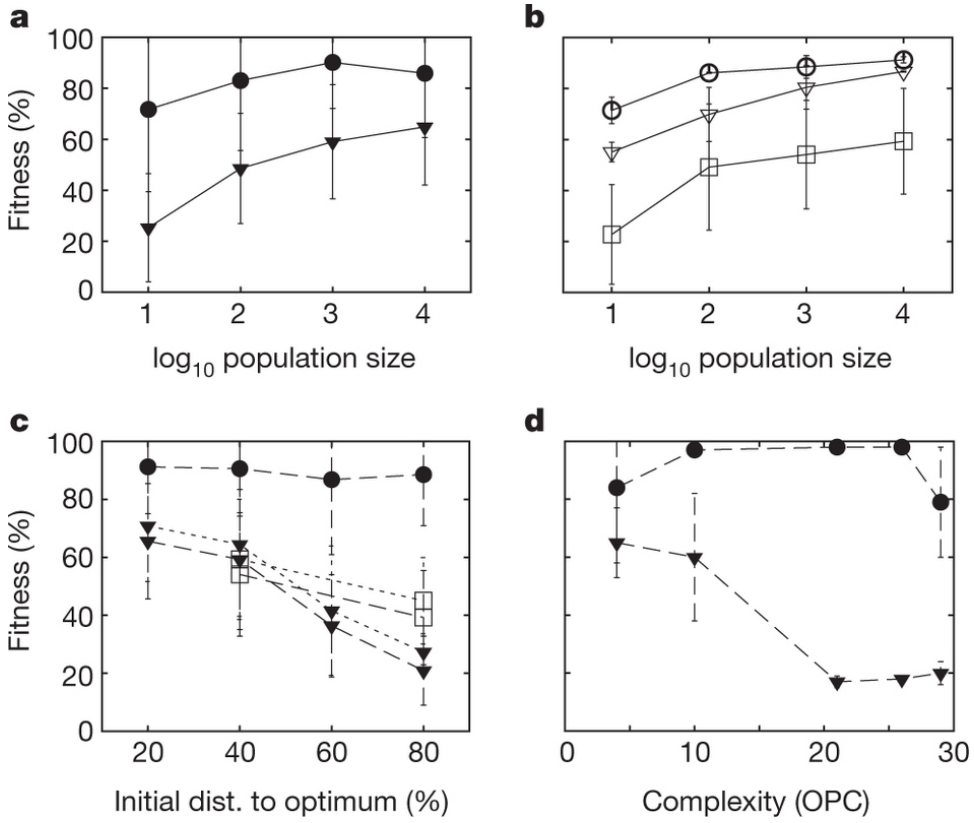

FIGURE 3.2: The EMD criterion rarely reaches the optimum and attains modest fitnesses. A, Average of the fitness reached in EMD and OPC simulations with different population sizes (n=345 for EMD, n=309 for increasing OPC and n=309 for decreasing OPC). The simulations start with morphologies at a distance of 40% from the optimum. For the landmark-based simulations progressively larger fitnesses are obtained as fewer traits are considered. B, the same as A using the PCA-based and non PCA-based landmark criteria for different number of traits (n=193 for raw traits, n=64 for PCA-based traits). Fitness increases with population size and decreases with the phenotypic distance between initial and optimal morphologies. C, Average final fitness in EMD and OPC simulations with optimal morphologies at different distances from the initial morphology with different population sizes (n=154 for EMD at population of 1000, n=60 for EMD at population 10000, n=47 for OPC at population 1000). D, average final fitness of EMD and OPC simulations with initial morphologies of different complexity (n=63). Complexity is calculated as the OPC.

can still have the same fitness. In this way, populations do not get easily trapped in local adaptive peaks.

The EMD and landmark phenotype-fitness maps are simple and smooth: there is a single peak phenotype and all other phenotypes have a fitness smoothly proportional to the distance to that phenotype (fig. 3.3 and related measures in fig. A.7, A.8, A.9, A.10). However, as these figures indicate, degeneracy is low and, consequently, there is no simplifying effect over the overall genotype-fitness map. Degeneracy is low because it is defined at the level of the traits that contribute to fitness. This pertains to all traits for OPC and EMD, many traits for the many-traits criterion and few for the few traits landmark-based criteria. The few-traits criteria work better for adaptive evolution simply because fewer traits need to be optimized.

However, this occurs only with a complex genotype-phenotype map: with a simple genotype-phenotype map the optimum is reached irrespectively of the number of traits it contains (Tenaillon et al., 2007). With a complex genotype-phenotype map, optimizing many traits implies moving farther in the phenotypic space and increasing the chances of getting trapped.

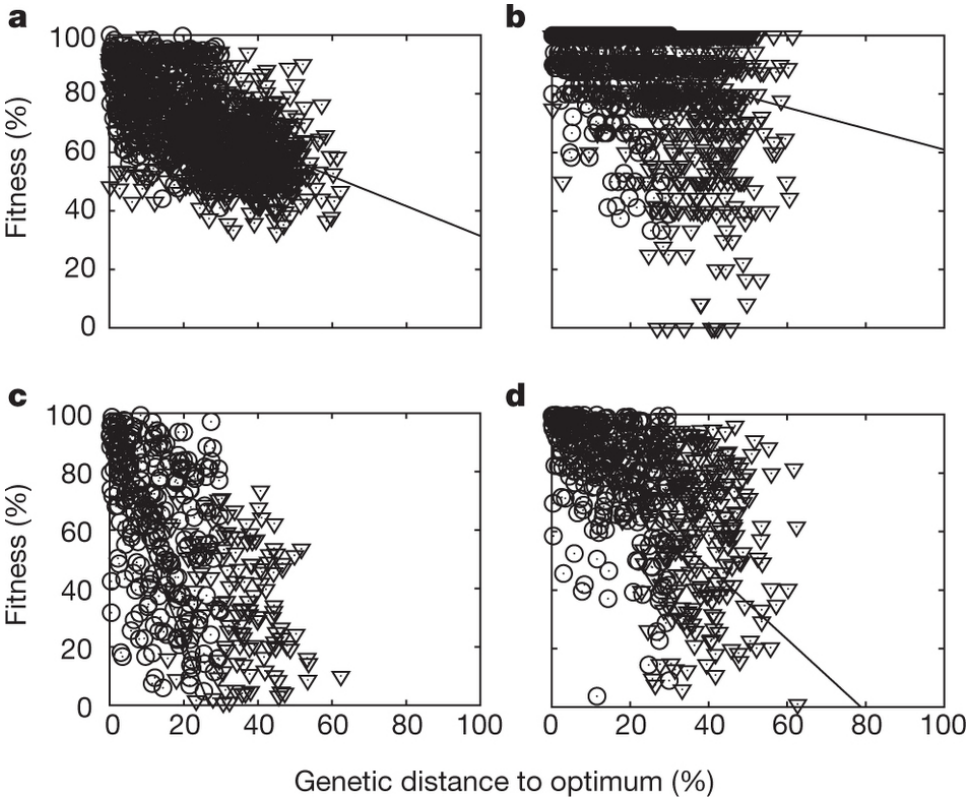

FIGURE 3.3: Comparison between the different genotype-fitness maps resulting from combining the genotype-phenotype map with the different phenotype-fitness maps. A large number of mutant offspring were obtained by mutation from the same initial individuals than in figure 3.2 (what we call the parents). Each mutant was different from its parent in 1 or 5 parameters. A, Fitness of each mutant phenotype (measured by taking its parent as the optimum) using the EMD criterion versus genetic distance between parent and mutant, measured as parameter distances between parent and mutant (see Appendix A). B, As in A but fitness measured by the OPC criterion. C, As in A but fitness measured by the landmark-based criterion with many traits (13) and, D, few traits (2). Linear regression equations: A, $y = -0,5871 + 87,70\ r^2 = 0,3511$ (n=699); B, $y = -0,3763x + 95,17\ r^2 = 0,06082$ (n=699); C, $y = 3,206x - 0,1429\ r^2 = 0,2632$ (n=242); D, $y = 1,199x + 95,68$ $r^2 = 0,1997$ (n=242). As it can be seen the genetic landscapes of EMD and landmark criteria are relatively well correlated ($r^2 = 0,3549$, $r^2 = 0,2632$ and 0,1897 respectively) in spite of the random mutation and the complexity of the phenotype. The genetic landscape induced by the OPC phenotype-fitness map on the model genotype-phenotype map is more uncorrelated ($r^2 = 0,06082$) and very degenerated: many mutants have the same OPC than their parent, even mutants that are genetically very distant from their parent. In fig. A.7 it can be seen that, in fact, most mutants have the same OPC than their parent.

In the case of teeth, the OPC has been shown to correlate with diet in rodents, carnivorans, primates and bats (Evans et al., 2007; Santana, Strait,

and Dumont, 2011; Godfrey et al., 2012); low and high OPC values correspond to animal and plant diets, respectively. It is still not known if this applies to other mammalian groups. However, the evolutionary adaptive transitions in tooth morphology have traditionally been understood on the bases of detailed analysis of specific characters (Ji et al., 2002; Charles et al., 2011) and landmarks (Klingenberg, 2002). Our suggestion is that the correlation between the OPC and diet may not necessarily arise from specific OPC values being the optimal morphological solution to specific diets. The OPC may just be the adaptive solution, among several possible ones, with the degenerate phenotype-fitness map that allows substantial degrees of adaptation despite the complexity of the genotype-phenotype map.

More in general, our results suggest that natural selection cannot find adaptive morphologies with most, or even many, of their traits being adaptive (as in EMD) because real genotype-phenotype maps are far too complex. Our analysis does not specifically take ecology into consideration: it is possible that in certain environments maximal fitness requires a very specific morphology characterized by a unique combination of many trait values. But our results suggest that this is only achievable if the initial phenotypes in the population are very similar to the optimal one (Fig. 3.2, fig. A.3, A.4). In contrast, substantial sustained morphological adaptation may have required degenerate phenotype-fitness maps, as in the OPC, or occurred only with respect to a small subset of the traits, as in the few-traits criteria. Our results should apply even when the selection criteria change over time or when no selection occurs for long periods of time. Adaptation should still mostly occur in the time periods when the OPC or the few-traits selection criteria are in place.

Our results do not preclude natural selection from having a crucial effect in most morphological traits. Specific selective responses in most traits in a morphology could occur because of correlations with few traits being directly adaptive (Alberch, 1982; Lande and Arnold, 1983) or because of selection on the overall roughness, or similarly degenerate properties, of morphology. This degeneracy may also facilitate the evolution (in the specific case of teeth) of upper and lower occlusion without affecting OPC (Polly, Le Comber, and Burland, 2005). We propose that the diversity of forms in evolution is in part result of the degenerate phenotype-fitness maps providing multiple solutions to the same problems.

Degenerate maps have been found in models of RNA folding and evolution (Huynen, Stadler, and Fontana, 1996). In these, the degeneracy occurs in the genotype-phenotype map and not in the phenotype-fitness map (some low level of degeneracy occurs also in our genotype-phenotype map). For a RNA sequence of a given length, only a limited number of phenotypes (that is secondary structure folds) occur and the number of these is much smaller than the number of possible sequences. In our model, in contrast, the phenotype has many more dimensions of variation than the genotype (see Appendix A).

There has been some controversy in the study of adaptive landscapes on how rugged or uncorrelated these landscapes are. Studies based on gene networks tend to view such landscapes as uncorrelated (Kaufmann, 1993b) while some studies coming from population genetics assume that those landscapes need to be correlated for adaptation to be possible (Orr, 2006). Our results present a potential point of connection between these

two views. The genotype-phenotype map would indeed be complex or rugged (as our current understanding of development indicates) but the phenotype-fitness maps would be simple or degenerate, as in the few-traits and OPC criteria (fig. 3.4). This effectively leads to an adaptive landscape that is rugged but not too rugged for adaptation to occur.

3.4 Methods

We use a computational model of evolution that is based on a developmental computational model that gives rise to three-dimensional morphologies, from genotypes represented by the values of the model parameters, and on a set of criteria of selection on those morphologies. Thus, the evolutionary model is implemented in the context of a population of individuals that perpetuates itself over generations and in which there is selection and mutation on the parameters of the developmental model. The developmental model is described in detail in the original publication where it was first used (Salazar-Ciudad and Jernvall, 2010). In the Appendix A, however, we describe how the evolutionary model deals with the aberrant morphologies that can arise from the developmental model when parameters get extreme values by mutation and some other details about how the evolutionary model uses the developmental model. The evolutionary model is simply in charge of applying mutation on the model parameters (genotypes) of individuals in the population, at a rateμper individual, calling the developmental model for each individual genotype, applying one of the three selection criteria in the phenotype (morphology) of each individual and determining which individuals contribute to the next generation based on the individual relative fitness calculated by the selection criteria. In brief, mutation is implemented by adding or subtracting a proportion of the value of a parameter, randomly chosen, in an individual. Selection is implemented by choosing each individual in a generation stochastically from individuals in the previous generation and making the probability of being chosen proportional to individual relative fitness. The details of how mutation and selection are implemented are described in Appendix A. The present article is specially focused in understanding the different evolutionary consequences of the selection criteria. The methods section describes these criteria in detail.

3.4.1 Euclidean Morphological Distance (EMD)

EMD is a measure of sheer disparity between two forms or morphologies (see figure 3.1A). This measure considers all cell-level traits (each cell position in 3D). It is defined as the mean distance of one cell in one morphology to the closest cell in the other morphology, corrected by size:

$$D = \frac{\sum_{k=1}^{n_1} d_{min12k} + \sum_{k=1}^{n_2} d_{min21k}}{n_1 + n_2} \tag{3.1}$$

Where D is the EMD between two morphologies (like an individual's morphology and the optimal morphology in a simulation), d_{min12k} is the

euclidean distance, in three-dimensional space, between cell k at morphology 1 and the cell in morphology 2 that is closest to that cell (again in three-dimensional space) and vice versa for d_{min21k}. n_1 and n_2 are the number of cells in morphology 1 and 2 respectively. Notice that this phenotypic distance can be applied to morphologies made of different number of cells. This is important because different genotypes lead, in the model, to morphologies with different number of cells and different number of morphological features (such as cusps). In each comparison morphologies are rescaled and rotated by a Procrustes algorithm, so that the final configuration has the minimal EMD possible between the two morphologies. To correct for size and make the distance dimensionless we divide the distance by the square root of the sum of the surface areas of the two morphologies,

$$d_{EMD} = \frac{D}{\sqrt{S_1 + S_2}} \tag{3.2}$$

where d_{EMD} is the final phenotypic distance, D is the distance resulting from the Procrustes algorithm and S_1 and S_2 are the surface areas of morphology 1 and 2 respectively.

The d_{EMD} resulting from comparing an individual's morphology to an optimal morphology is used to calculate the absolute fitness of that individual.

3.4.2 Landmark based distances

Landmarks are taken at tooth cusps. This criterion can thus only be applied to teeth with enough number of cusps (see fig. A.11, type 4). The first landmark is the height of the the highest cusp. We then divide the teeth into four quadrants (anterior, posterior, buccal and lingual) (see fig. 3.1B). The second, third, fourth and fifth landmarks are identified as the tallest cusps in the anterior, posterior, bucal and lingual quadrants respectively. Teeth with fewer cusps than required for measuring a given number of traits were given 0 absolute fitness. The traits could be chosen by two criteria:

3.4.3 Raw landmark based traits

We identify up to 13 quantitative traits as the components (x, y and z) of the position vectors of the 5 landmarks. Tooth developmental dynamics in the model and in real systems (Jernvall, 2000) ensure that the highest cusp forms in the center (it is also the first cusp to form). By taking this position as the point (0,0,0) we ensure that all model morphologies are in the same spatial reference system. For the central cusp, only the height is considered, because its X and Y position never change as a result of the tooth model dynamics. Simulations were run for a different number of traits, n_t, for 2,4,8 and 13. The phenotypic distance by the landmark criteria, d_{lan}, from a tooth to its optimum is calculated as the square root of the sum of differences between each trait and that trait optimal value:

$$d_{lan} = \sqrt{\sum_k = 1_t^n (c_k - o_k)^2} \tag{3.3}$$

Where n_t the number of traits under selection, c_k is the trait value of the k trait and o_k is the optimal value of the k trait. This measure is simply the distance between two points in a n_t-dimensional morphospace.

3.4.4 PCA based traits

We also defined the traits as the scores of the principal components (PC) resulting from a Principal Component Analysis (PCA) of variation on the landmarks. We calculate the PCA in a subset of the morphological space by taking the initial phenotypes and generating 300 random mutants (changing one to three parameter values per individual) and using them as a population sample to perform the PCA. Note that this PCA is only representative of a small hypervolume of the parameter space (that around the initial morphologies) not about the variation possible from the model in general (that is too vast and multidimensional to be efficiently characterized by a PCA). Note also that this method cannot be applied to the whole morphology since the PCA can only use datasets of the same dimensionality (that is the same number of landmarks), and different morphologies in the sample population often have different number of cells and different number of cusps. Thus, as above, we take 5 different landmarks (and thus 13 traits) and perform the PCA to get the principal components of variation. From the 13 possible PCs, we only take into account the ones representing a significant amount of variation (> 1%), that was 8 PCs. In the evolutionary simulations the PC loadings were used to calculate the scores of a given phenotype on the coordinate system of the PCs. The phenotypic distance for a given phenotype was calculated as in the raw landmark-based distances, but using the PC scores as trait values and comparing them with the optimal PC scores (as in equation 3.3). This way selection acts always, and independently, on all the PCs. The PCA was done with the PAST software (http://folk.uio.no/ohammer/past/).

3.4.5 Orientation Patch Count (OPC)

OPC is a multivariate measure of tooth surface complexity (Evans et al., 2007). It consists on classifying each cell-level trait on a surface by the orientation of the projection of its normal vector (vector normal to the tooth surface in that cell point) on the x-y plane, then grouping all adjacent cells with the same orientation into patches. The count of all patches on the morphology surface is the OPC value. We established 4 orientation classes corresponding to the 4 quadrants of the x-y plane (see figure 3.1C). A smoothing algorithm was applied onto the normal vectors in order to erase random noise from the morphology:

$$\vec{n_i}' = \vec{n_i} + \lambda \sum_{j=1}^{V} \vec{n_j} \tag{3.4}$$

where $\vec{n_i}'$ is the smoothed normal vector, $\vec{n_i}$ is the original normal vector, λ is a parameter of the smoothing algorithm, V is the total number of neighbour cells and $\vec{n_j}$ is the normal vector of a neighbour cell. We always ran the smoothing algorithm 5 times per phenotype with λ=0,1.

The phenotypic distance by the OPC criterion, d_{OPC}, is calculated as the absolute value of the relative difference between the OPC value of a morphology and the optimal OPC in each simulation,

$$d_{OPC} = |\frac{c - c_{opt}}{c_0}| \tag{3.5}$$

where c is the OPC value of a morphology, c_{opt} is the optimal OPC value and c_0 is the OPC value of the morphology at the beginning of the simulation.

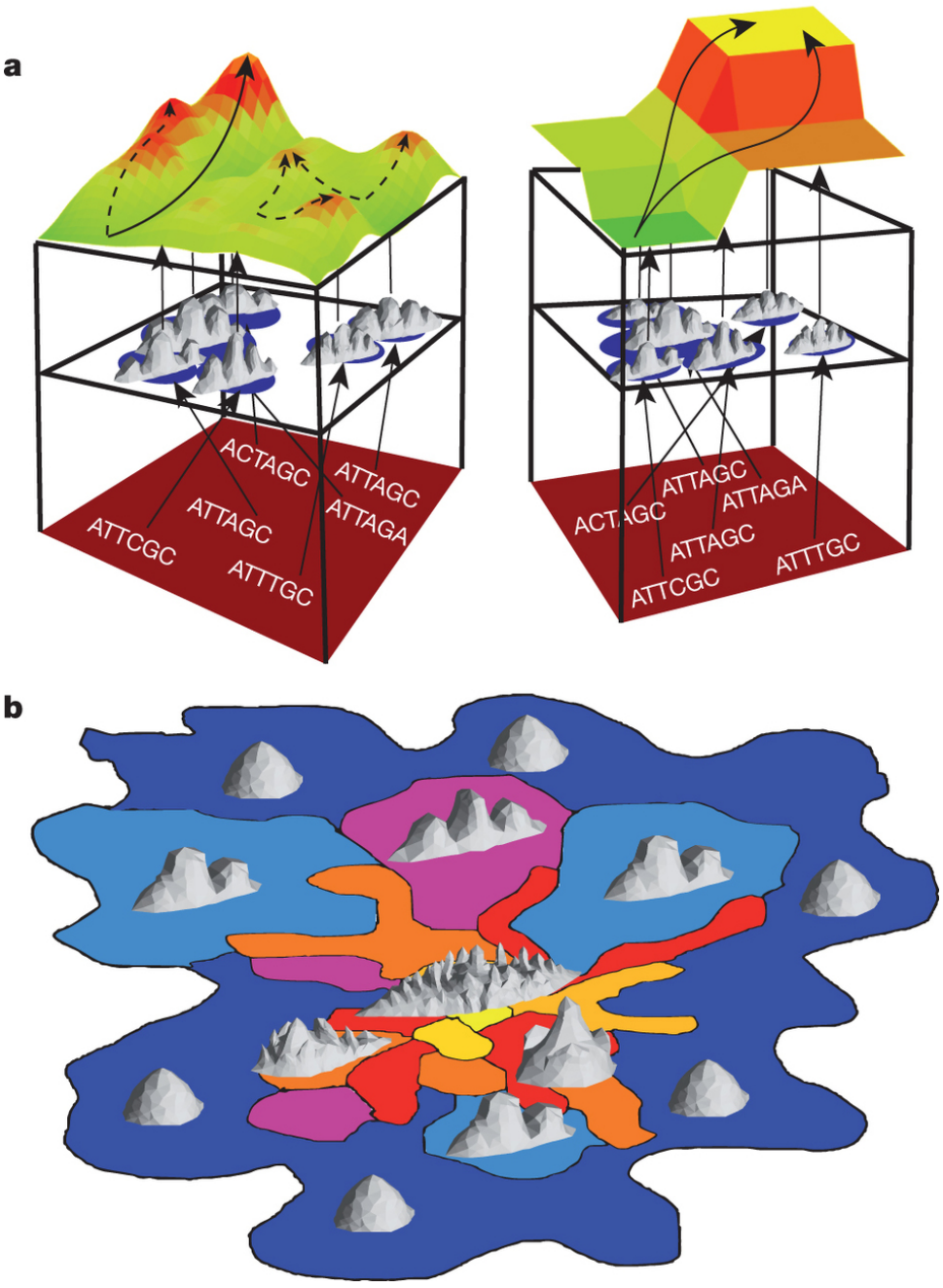

FIGURE 3.4: A, The three layers represent from below to above genetic variational space, morphological variational space (morphospace) and fitness space, respectively. An adaptive landscape is the direct mapping between the genetic space and fitness. The arrows connecting each layer represent the mappings between genotype and morphology (development) and between morphology and fitness (selection criterion). When a non-degenerate phenotype-fitness map is used (left), the complexity of the genotype-phenotype map gives rise to a highly rugged adaptive landscape and populations often get trapped in local peaks (dashed lines) and seldom reach the optimum (solid line). When a degenerate phenotype-fitness map is used (right) the landscape is smoothed and thus populations can attain the global optimum. B, Idealized representation in 2D of the tooth model parameter space. Regions in different shades represent hypervolumes of the parameter space where the resulting morphologies have similar complexity (for example the same number of cusps or the same OPC). The space occupied by complex morphologies is small, but is in contact with more regions of different complexity, so it is easier to change complexity by mutation when morphologies are complex (as indicated by figure A.12.

Chapter 4

Expanding models of animal morphogenesis in order to analyse the properties of development across systems

The contents of this chapter were published in the form of a research article:

Marin-Riera, Miquel et al. (2016). "Computational modeling of development by epithelia, mesenchyme and their interactions: a unified model." Bioinformatics (Oxford, England) 32.2, pp. 219–25.

4.1 Abstract

The transformation of the embryo during development requires complex gene networks, cell signalling and gene regulated cell behaviours (division, adhesion, polarization, apoptosis, contraction, extracellular matrix secretion, signal secretion and reception, etc.). There are several models of development implementing these phenomena, but none considers at the same time the very different bio-mechanical properties of epithelia, mesenchyme, extracellular matrix and their interactions.

Here we present a new computational model and accompanying open-source software, EmbryoMaker, that allows the user to simulate custom developmental processes by designing custom gene networks capable of regulating cell signalling and all animal basic cell behaviours. We also include an editor to implement different initial conditions, mutations and experimental manipulations. We show the applicability of the model by simulating several complex examples of animal development.

The source code can be downloaded from:
http://www.biocenter.helsinki.fi/salazar/software.html .

4.2 Introduction

Multi-scale computational models help in making the complexity of development more amenable to the human mind. They may be conceived as summaries of what is known about the development of an organ or embryo part. Such summaries, however, should be built upon explicit mechanistic hypotheses about how development works in these systems. Computational models can then provide explicit quantitative predictions about how

an embryo changes over time according to each specific mechanistic hypothesis. These quantitative predictions (e.g. quantitative morphology as the position of each cell in space and the levels of expression of different genes in space) can then be compared with experimental results to facilitate the rejection or provisional acceptance of a hypothesis (further experiments would be required for actual acceptance of an hypothesis).

There is a long list of models of pattern formation and morphogenesis involving cell-cell communication and biomechanics. Some of them simulate the development of different organs, such as teeth (Salazar-Ciudad, 2008; Salazar-Ciudad, 2010a), limb (Hentschel et al., 2004), turtle shell's scutes (Moustakas-Verho et al., 2014), to cite a few. In other cases, processes involving the whole embryo, such as gastrulation in the sea urchin, are reproduced (Cummings, 1990; Cummings, 1994; Cummings, 2005; Davidson et al., 1995). Others implement a framework of cell biomechanical interactions and gene regulation, including one or several cell behaviors, namely: cell division, apoptosis, growth, etc.

The Cellular Potts Model (CPM) (Graner and Glazier, 1992) defines a regular lattice in which cells occupy several contiguous lattice positions. Cells may extend to further lattice positions or retract from them by minimizing a Hamiltonian energy function based on cell adhesion, volume conservation and other mechanical properties. The CPM is especially suitable for developmental systems in which cells change their neighbors and shape in rather dynamic ways (such as in cell sorting and migration by chemotaxis and haptotaxis). However, tissues in which forces can be transmitted along a specific direction at long distances, such as in epithelia, may not be so easily grasped by the CPM. There are several models based on CPM that implement several cell behaviors such as cell division, cell migration (by chemotaxis), apoptosis, and cell-cell signalling (Hogeweg, 2000; Izaguirre et al., 2004; Starruß et al., 2014a).

In the vertex model (Honda, Tanemura, and Nagai, 2004) cells are defined by the contact surfaces with other cells or the media. Each cell is a polygon (or polyhedron in the 3D case) defined by a set of vertices and edges (and faces in the 3D case). Forces are calculated based on overdamped motion equations or a Monte Carlo algorithm. This model is suited to simulate dynamics of densely packed tissues like epithelia, but are not very well suited to simulate processes involving mesenchymal tissues or processes in which cells move freely in the extracellular space and seldom form tightly packed cell condensates. In some cases, cell signalling and gene regulatory networks have been implemented as well (Farhadifar et al., 2007; Smith et al., 2012; Pitt-Francis et al., 2009).

The viscoelastic model (or IBCell model) (Rejniak, 2007) combines an off-grid representation of cells as elastic bodies with a lattice Boltzmann Method to simulate intra and intercellular fluids. Thus, each cell is represented by a set of points forming an elastic body in a continuous 2D space but chemical substance concentration is calculated on a regular grid in discrete space. Then the position of each cell part on this grid has to be interpolated every time to correctly calculate how fluid fluxes in the grid affect cell shape. This implementation and level of detail makes the model computationally very costly and thus precludes, according to the authors (Tanaka, Sichau, and Iber, 2015), its extension to 3D problems. Several cell behaviors

have been implemented: cell proliferation (Rejniak, 2007), cell growth and cell-cell signalling (Merks et al., 2011).

The subcellular element model (SEM) (Newman and Müller, 2005; Sandersius and Newman, 2008) is an off-lattice model in which cells are made of groups of point elements that interact mechanically based on a potential equation. Motion is computed with Langevin overdamped dynamics. In some cases cells are modeled with a single subcellular element, thus approximating Drasdo's spheroid model (Drasdo, Hoehme, and Block, 2007; Delile, 2013). The original model does not implement any cell behaviors apart from cell-cell adhesion, but in a more recent implementation of the model (Delile, 2013), cell division, polarization, cell signalling, migration and some basic gene networks are implemented.

Here we present a new general multi-scale model of development, the EmbryoMaker, that differs from previous models in a fundamental way: the model recognizes that the bio-mechanical properties of epithelia, extracellular matrix (ECM) and mesenchyme are crucial to correctly understand many processes in early animal development and in organogenesis. Other models are adequate for either mesenchymal cells or flat epithelia, but not for both or for their interaction as it occurs in development (Biggs and Mikkola, 2014).

The EmbryoMaker is based on the SEM for mesenchymal cells but not for the epithelia and for the non-mechanical interactions. Epithelial cells are instead assumed to be made of cylindrical subcellular elements rather than spheres as cells in the SEM. In addition, epithelial-specific rules are used to capture how epithelia interact with mesenchyme and ECM and, in general, how they behave in development.

In addition, our model implements the most complete set of basic cell behaviours used by cells in animal development (cell growth and division, cell death, cell migration, polarization, ECM secretion, cell-cell adhesion, cell-cell signalling, epithelial to mesenchymal transition). This is done by a unified set of rules acting on the cells' subcellular elements. EmbryoMaker includes also a detailed implementation of cell's molecular mechanisms involved in regulation and cell-cell communication. This includes transcriptional and post-transcriptional gene networks and signalling by the diffusion of gene products in the extra-cellular space and the binding to their specific receptors. A fundamental aspect of the model is that the gene products present in each subcellular element can regulate its mechanical properties and activate or repress specific cell behaviours. This allows an explicit multi-scale coupling between microscale regulatory molecular processes and macroscale mechanical properties and cell behaviours.

4.3 Structure of the model

Our modelling framework is, thus, multi-scale; it includes cells, ECM, cell parts and regulatory molecules that interact in intracellular networks and molecules that diffuse between cells in the extracellular space and affect gene expression in other cells (see fig. B.1). The parameters of the model specify how much each molecule regulates those and also the intra and extracellular molecule regulatory network. The model allows to implement

any arbitrary gene network and initial condition. Thus, different implementations of the model will have different number of parameters according to how many genes and molecules it includes. The number of subcellular elements (also called nodes in here) in a cell and the number of cells can change due to cell growth and cell division.

4.3.1 Biomechanics

Subcellular elements or nodes represent a physical part of a cell. Mesenchymal nodes are spherical while epthelial nodes are cylindrical. Each cylinder consists of two elements that move independently, an apical and a basal one, which may have different mechanical properties. These two nodes are tied by an elastic spring (fig. 4.1F). The ECM consists of spherical nodes that do not belong to any cell.

Each node has a radius of interaction p^{ADD} that defines the maximum distance at which the node is in contact with other nodes and a radius of equilibrium p^{EQD} at which the mechanical interaction is at equilibrium (i.e. the interaction force is zero). Nodes i and j will experience a repulsive force when they are closer than their equilibrium distance ($d_{ij}^{EQD} = p_i^{EQD} + p_j^{EQD}$) and an attractive force when their distance is longer than d_{ij}^{EQD} but closer than d_{ij}^{ADD} (fig. 4.1B and Appendix B). The modulus f_{Aij} of the mechanical force vector is calculated as,

$$\begin{cases} f_{Aij} = k_{ij}^{REP}(d_{ij} - d_{ij}^{EQD}) & if\ d_{ij} < d_{ij}^{EQD} \\ f_{Aij} = k_{ij}^{YOU}(d_{ij} - d_{ij}^{EQD}) & if\ d_{ij} \geq d_{ij}^{EQD} \\ f_{Aij} = 0 & if\ d_{ij} \geq d_{ij}^{ADD} \end{cases} \tag{4.1}$$

Note that the elastic coefficients (k_{ij}^{REP} and k_{ij}^{YOU}) may be different in the repulsive and attractive case and depend on the mechanical properties of the nodes i and j (that in our model can be regulated by molecules). The direction of force vectors differ between the mesenchymal-mesenchymal, the epithelial-epithelial and the epithelial-mesenchymal interacion, since vectors need to be normal to the contact interface between the two elements (fig. 4.1C-E, see Appendix B). Our model, thus, makes the assumption, as in the SEM, that cell shape can be represented with some degree of accuracy by a set of movable points (the nodes) that adhere to each other (so that each cell is a cohesive entity by the cohesion between its nodes). This way adhesion between cells is represented also as adhesion between nodes. We also assume that no two cells can occupy the same physical space and that, thus, there would a repulsion force when two nodes get to close.

The model also assumes, in contrast to the SEM, that epithelial tissues always tend to minimize local curvature (Forgacs and Newman, 2005) and that epithelial cells always tend to orient their apical-basal axis normal to the local tissue surface. In order to evaluate the local curvature between apical nodes i and j we consider the vector $\vec{c_{ij}}$, connecting those two nodes, $\vec{s_{ik}}$ and $\vec{s_{jl}}$ as the vectors that define the elastic link to their basal counterparts and $\vec{m_{ijkl}}$ as the sum of $\vec{s_{ik}}$ and $\vec{s_{jl}}$ which defines the vector normal to the surface between i and j. The bending radial force tends to minimize the local curvature of the epithelial sheet, while the bending rotational force reorients the cylinders so that their longitudinal axis ($\vec{s}$) is always normal to the epithelial surface (fig. 4.1G, see also fig. B.4B,C). The bending radial

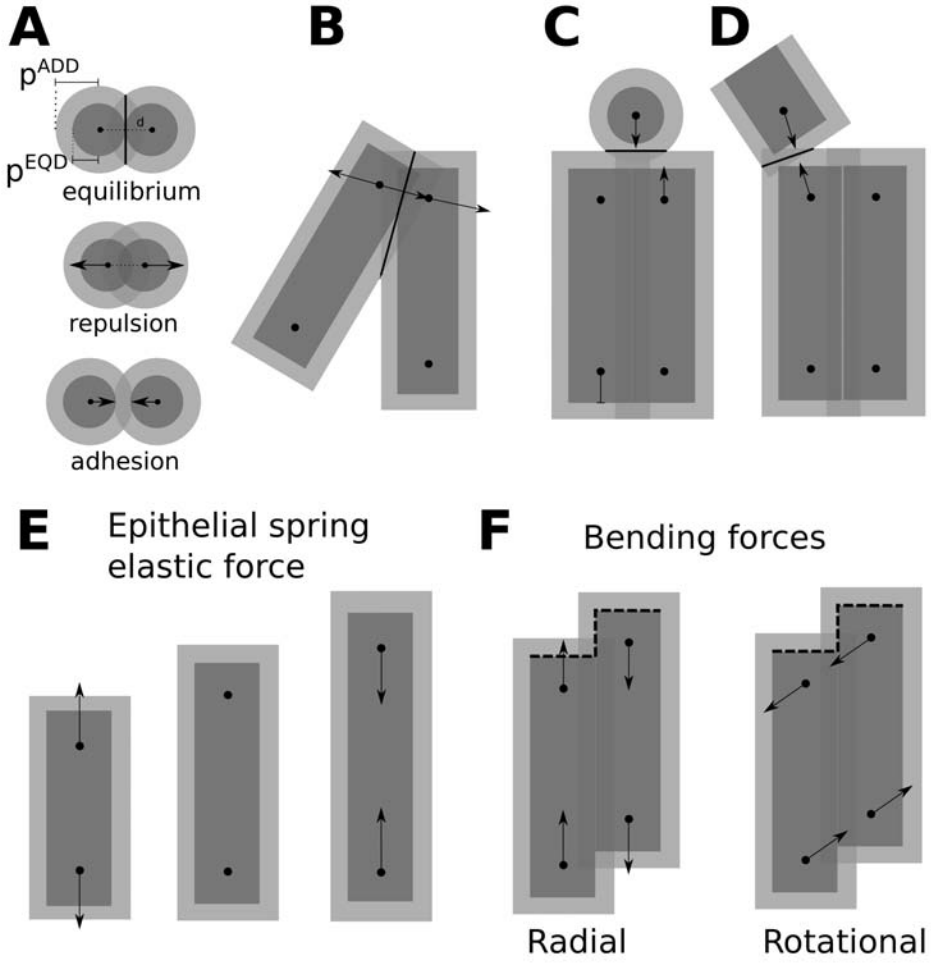

FIGURE 4.1: A) Mechanical interactions between spherical elements are determined by the distance between their centres and their distance of equilibrium. B, C, D) Mechanical interactions between two cylindrical elements or between a cylindrical and a spheric one act along a vector normal to the surface of interaction between the two elements. E) The two nodes composing a cylindrical element are tied by an unbreakable elastic spring. F) Epithelial bending is regulated by two different forces, a bending radial force and a bending rotational force. All arrows represent force vectors.

force is defined by the vector $\vec{f_{ESTij}}$ and the bending rotational force by $\vec{f_{ERTij}}$,

$$\vec{f_{ESTij}} = k_{ij}^{EST} \frac{\vec{m_{ijkl}} \cdot \vec{c_{ij}}}{|\vec{m_{ijkl}}|} \hat{m_{ijkl}} \tag{4.2}$$

$$\vec{f_{ERPij}} = k_{ij}^{ERP} \frac{\vec{s_{ik}} \cdot \vec{c_{ij}}}{|\vec{s_{ik}}|} \hat{c_{ij}} \tag{4.3}$$

Where k_{ij}^{EST} and k_{ij}^{ERP} derive from node's mechanical properties and the circumflex denotes unit vector. Note that the force vector $\vec{f_{ESTij}}$ will be normal to the epithelial surface and $\vec{f_{ERPij}}$ will be parallel to the vector

connecting i and j (see fig. 4.1G).

Motion is computed by an hybrid method, solving a system of PDEs assuming Langevin overdamped dynamics (see Appendix B) and then applying Monte Carlo random displacements to a certain proportion of nodes at each time step (see Appendix B). This way some stochasticity is introduced into the model.

4.3.2 Gene and molecular regulation

The model assumes that all molecules are contained within the system's nodes, whether they belong to a cell or to ECM. Diffusion is then the transport of molecules between nodes. Thus, both calculations of movement and specific molecule concentrations occur only on nodes and molecules carried by nodes respectively. We use the term regulatory molecules to refer to any molecule irrespectively of whether a molecule is a gene product or not.

Each regulatory molecule has a set of intrinsic characteristics that are assumed to be genetically encoded and thus do not change over model dynamics. We call those molecular parameters. For each regulatory molecule these are: diffusivity coefficient, intrinsic rate of degradation, regulatory interactions at the level of transcription and at the level of catalytic reactions, regulation of each node property and regulation of cell behaviours and cell properties (see Appendix B). Some regulatory molecules act as inter-cellular adhesion molecules, for which there is a matrix that specifies the binding affinity of each pair of adhesion molecules. Transcription occurs only in the nuclear node of each cell and we assume it follows a saturating function similar to Michaelis-Menten (as in many other previous models, Mjolsness, Sharp, and Reinitz, 1991):

$$Q_{ik} = \frac{\Phi\left(\sum_{l=1}^{n_g} t_{lk} g_{il}\right)}{1 + \Phi\left(\sum_{l=1}^{n_g} t_{lk} g_{il}\right)} \tag{4.4}$$

Where Q_{ik} is the rate of transcription of gene k in node i, g_{il} is the amount of transcriptional factor l in node i and t_{lk} is the strength by which each specific transcriptional factor k activates (positive t_{lk}) or inhibits (negative t_{lk}) the transcription of gene l. The sum is done through all the regulatory molecules and by definition only transcriptional factors have t_{lk} terms different from zero. The set of t_{lk} values in a model constitute the matrix T (whose size depends on the number of genes included in the model). Φ is a function that is equal to 0 for values of x smaller than 0 and equals to x when x is greater than 0 ($\Phi(x) = 0\ if\ x < 0$ and $\Phi(x) = x\ if\ x > 0$). This function is used to ensure that there is no such a thing as negative transcription (although t_{lk} can be negative and thus repress transcription).

The model represents two states of the same protein as two different regulatory molecules with different molecular properties. Non-transcriptional enzymatic reactions mediate the transformation of one regulatory molecule into another. Thus:

$$S_{ik} = \sum_{l=1}^{n_g}\sum_{j=1}^{n_g} r_{jlk} g_{ij} \frac{g_{il}}{1+g_{il}} - \sum_{l=1}^{n_g}\sum_{j=1}^{n_g} r_{jkl} g_{ij} \frac{g_{ik}}{1+g_{ik}} \quad (4.5)$$

The first term defines the rate of production of regulatory molecule k in node i due to the transformation of other molecule l into k catalysed by j. The second term defines the rate of loss of form k due to its transformation into molecule l mediated by catalysis by j. Each element r_{jlk} specifies the catalytic activity of regulatory molecule j on the transformation of regulatory molecule l into regulatory molecule k. As before we assume Michaelis-Menten kinetics similar to those known to occur in enzymatically catalysed reactions. For simplicity we assume K_M equals 1. As before the set of all r_{jlk} values constitutes matrix R.

4.3.3 Molecular diffusion and signalling

We assume that regulatory molecules can diffuse between nodes in a cell. Extracellular diffusible molecules can diffuse between nodes in different cells. For these molecules its concentration in a node represents concentration in the extracellular space close to the node. Diffusion can not occur in empty space but it can occur in aqueous media represented as ECM nodes. This is implemented by applying Fick's second law in the context of the irregular and changing mesh made by all the nodes that are not too far from each other (This distance is itself considered, see Appendix B for details). This method allows to implement diffusion in a way that is not computationally very costly. Since molecules are only defined within nodes, they cannot diffuse towards empty spaces or cavities. This limitation can be overcome, if necessary, by filling empty spaces with ECM nodes. Diffusible molecules may interact with membrane receptors. The model also considers membrane tethered ligand and receptor complexes, such as in the Notch-Delta signalling pathway (Meir, Munro, and Odell, 2002, see Appendix B).

4.3.4 Cell behaviors and regulation of node properties

The model is specifically designed to incorporate all animal cell behaviours as simple rules on cell nodes (fig. 4.2, see Appendix B and fig. B.5, B.6, B.7, B.8, B.9, B.10). This way the temporal and spatial scales of these rules are compatible with each other. Growth is increase in the sizes of the elements in a cell, until a maximum element size is reached (fig. 4.2A and fig. B.6). Only then a new small element is added to the cell, which can then undergo further growth. The model allows to choose the number of elements a cell is made of (though that number may fluctuate due to growth, division or death). Large number of elements per cell should be chosen when cell shape changes are a driving force in development. Otherwise cells can be chosen to comprise few, or even a single element. Cell division is the splitting of a cell's nodes between two daughter cells (fig. 4.2B and fig. B.7, see Appendix B). Apoptosis is implemented as a gradual decrease in the size of nodes in a cell (fig. 4.2C and fig. B.8, see Appendix B) until nodes, and eventually the cell, disappear. Adhesion is implemented as the strength of the attractive force between elements from different cells. Cell contraction, or expansion, is simply a limited decrease, or increase, in the

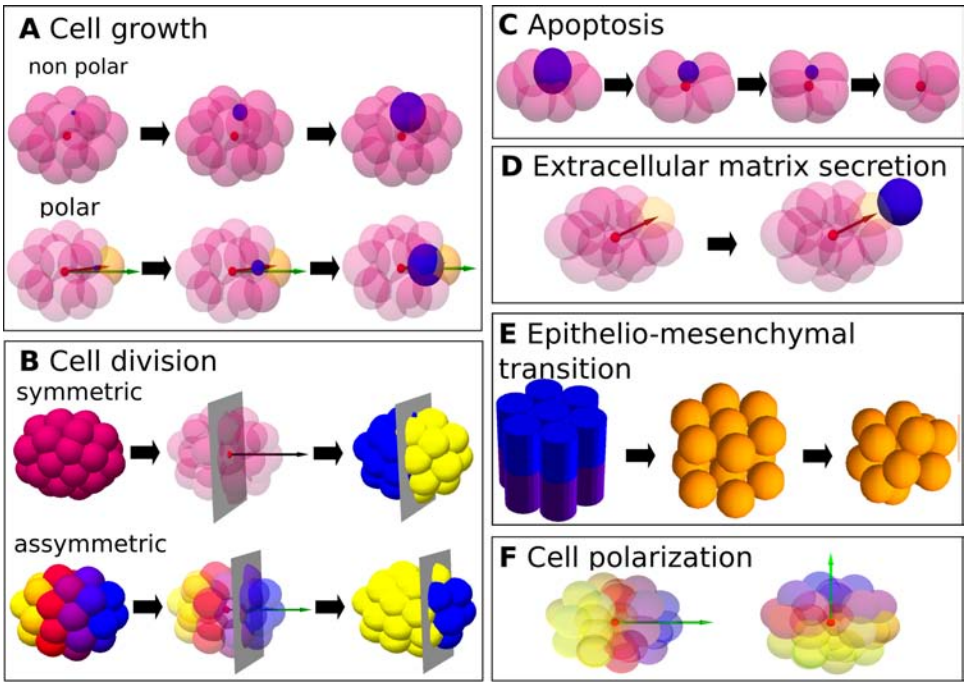

FIGURE 4.2: A) Cell growth. A new node (blue) is added in a random position within a cell (non polar growth), or in a direction determined by the polarization vector (polar growth). B) Cell division. The plane of division splits the cell into two daughter cells. In asymmetric division an intracellular gradient determines the relative size of daughter cells. C) Apoptosis. A node (in blue) decreases size until it disappears. D) Extracellular matrix secretion. A free (non-cellular) ECM node (blue) is secreted outside the cell in a random direction from a cellular node (yellow). E) Epithelial-mesenchymal transition. Epithelial cylinders transform into mesenchymal nodes. F) Cell polarization is defined as a 3D vector (green arrow) determined by an intracellular molecular gradient.

size of some elements in a cell. This can occur actively as a result of gene expression changes. Nodes can also have some degree of plasticity, which can be genetically regulated, and decrease or increase their size if put under strong compressive or tensile forces respectively by surrounding nodes. Migration is the result of random node movement (see Appendix B) biased by differential adhesion or chemoattractant gradients (although the mobility of elements can be increased by gene regulation to simulate cell body extensions like filopodia). Cell shape changes are the result of the relative movements of the elements in a cell. Extra-cellular matrix (ECM) secretion results from the production of ECM nodes by cells expressing a specific gene (fig. 4.2D and fig. B.9, see Appendix B). Cells can also undergo epithelial-mesenchymal transitions (fig. 4.2E and fig. B.10, see Appendix B).

An important aspect of our model is that all cell behaviours and elements' bio-mechanical properties (such as adhesion) can be affected by the regulatory molecules present in nodes. Specific equations apply for the molecular regulation of each cell behaviour, see Appendix B for details. Each column of matrix C specify how strongly each molecule positively or negatively regulates a certain cell behaviour.

Regulation of node mechanical properties at each time step follows:

$$p_i^l(t) = \Phi\left(p_i^l(0) + (1 - p_i^{DIF}) \sum_{k=1}^{n_g} e_{lk} g_{ik}\right) \tag{4.6}$$

Where $p_i^l(t)$ is the value of node property l in node i at time t and $p_i^l(0)$ is the value of that node property l in node i when the node was created (this is in the initial condition or when the node first arose through growth). p_i^{DIF} is the degree of differentiation in node i (differentiation slows down changes in nodes). The effect of molecule k on node property l is the model parameter e_{lk} of matrix E.

A slightly different rule applies to node size, p^{EQD}. At each time step this property is equal to:

$$p_i^{COD} = p_i^{GRD} + p_i^{PLD} + p_i^{VOD} \tag{4.7}$$

The first term is due to active contraction, the second to cell growth or apoptosis, the third to cell mechanical plasticity and the fourth to volume conservation. Cell contraction occurs when a molecule regulates negatively the node property p^{COD}. Since contraction is happening in the nodes, cells may have contraction in only part of its nodes, as it is necessary in a number of developmental processes such as in invagination by apical cell contraction. p^{COD} is calculated as in equation 4.7 above. The other terms are explained in the Appendix B.

The cell behaviours described above are in fact the most basic animal cells have. However, individual cells can do more complex behaviours by differentially activating these basic cell behaviours in different parts of the cell. This is possible in our model because different nodes in a cell can have different concentrations of different regulatory molecules (since transcription occurs in the nucleus node but other reactions and signalling

occurs in all nodes). This allows a rich set of spatially polarized cell behaviours. Thus, for instance, the orientation of cell division plane is by default normal to the longest axis of the cell following Hertwig's rule (Minc, Burgess, and Chang, 2011) but can be oriented in the model by the direction of intra-cellular gene product gradients through the nodes in a cell (that may arise from extra-cellular gradients through signalling, not affecting transcription). In a similar manner, intra-cellular gradients can regulate asymmetric division (Salazar-Ciudad, Jernvall, and Newman, 2003) quantitatively controlling how much larger is one daughter cell with respect to the other. Similarly, intracellular differences in regulatory molecule concentrations can lead to contraction of only specific parts of a cell, as in the apical contraction required for invagination of epithelia. Similarly cells can migrate in a polarized manner by preferentially extending more filopodia towards one side of the cell than to others. This can be reproduced in the model by differentially regulating the mobility of nodes along the cell.

4.3.5 Initial conditions, model parameters and model structure

For the application of the model to the study of pattern formation and morphogenesis in a specific experimental developmental system three sets of developmental parameters need to be specified by the user: 1) The network of interactions between regulatory molecules (the T and R matrices) 2) How these regulatory molecules affect each node property (E matrix) and cell behaviour (C matrix) 3) The initial conditions as the spatial location of each node and the quantity of each regulatory molecule in each of them.

4.4 Implementation

The EmbryoMaker software implements the model herein described and provides a user-friendly environment to design and manipulate custom developmental systems. An editor allows the user to design cells, nodes and gene expression in the initial conditions or at any time during a simulation. The NetworkMaker software allows to design and edit gene network topologies and set model parameters. A central panel displays each application network. We use the Cell Behaviour Ontology (Sluka et al., 2014) for the example model of sea urchin gastrulation (section 4.5). All software and documentation can be downloaded from:

"http://www.biocenter.helsinki.fi/salazar/software.html".

4.5 Application examples

In this section we provide a set of examples of how the model can be applied to animal development. The aim in these examples is not to understand their dynamics better, but to show that the model is realistic and general enough to consider many different animal systems. We have started by implementing a set of developmental mechanisms that have been proposed to be the most basic ones in animal development (Salazar-Ciudad, Jernvall, and Newman, 2003). This is the simpler developmental mechanisms able to produce pattern formation by using only one cell behaviour (see fig. 4.3 and fig. B.11, B.12, B.13, B.14, B.15, B.16 for the genotype-phenotype maps).

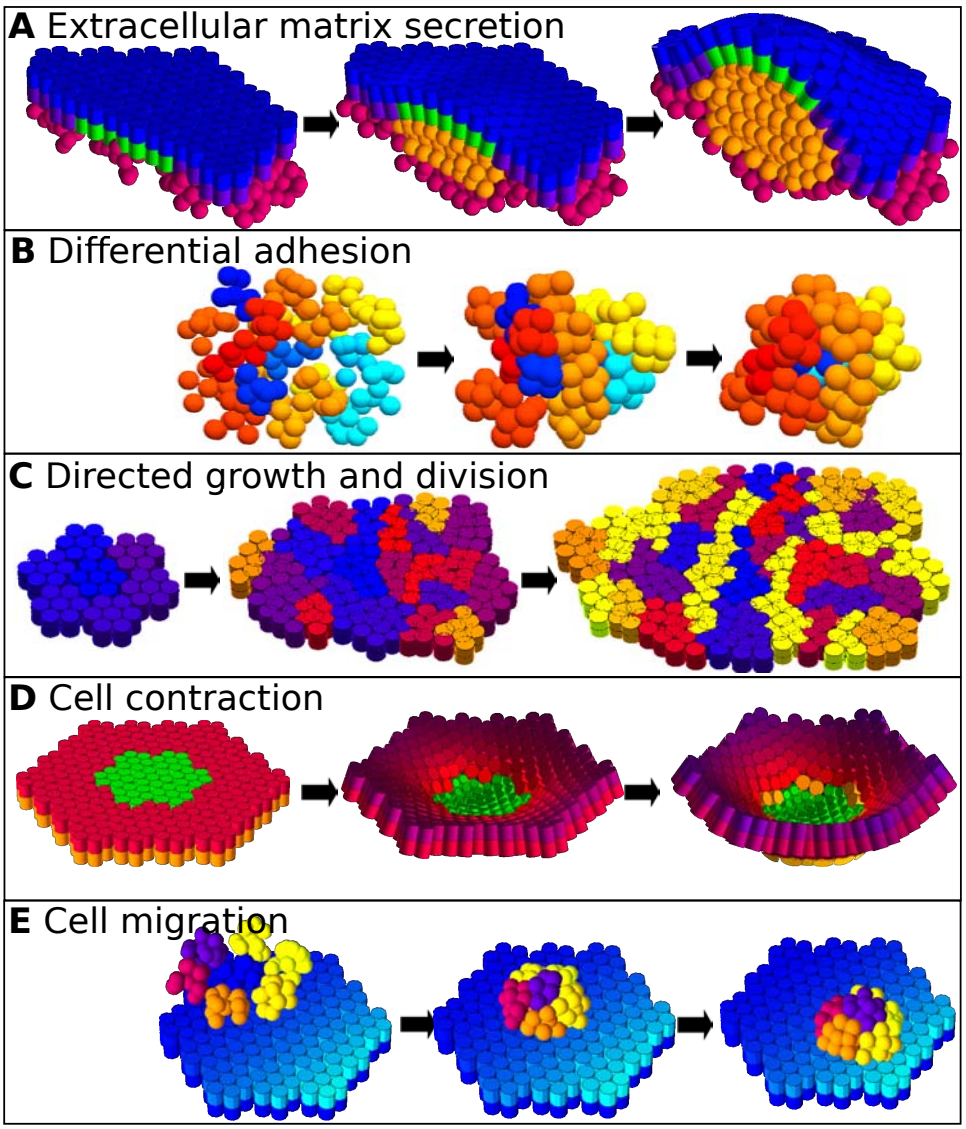

FIGURE 4.3: A) ECM (orange) secretion by epithelial cells (blue and green) in the space between an epithelium (blue and purple) and a mesenchyme (pink), driving the deformation of both tissues. B) Differential adhesion. Intercellular adhesion molecules (blue shades and red-yellow shades) drive the rearrangement of cells in space to maximize adhesivity contacts. C) Directed growth and division. Polarized cell growth and division of individual cells lead to a tissue elongation in the direction of a molecular gradient. Each cell is drawn in a different color or shade. D) Apical contraction of groups of epithelial cells (green) produces invagination of an epithelial sheet. The red shade represents the z axis. E) Cell migration. Mesenchymal cells (red, orange and yellow) show high adhesivity for the epithelial adhesion molecule. This molecule is expressed on a concentration gradient on the epithelium (light blue: high concentration, dark blue: low concentration). As a result mesenchymal cells progressively migrate, in a biased random walk, towards the high concentration regions.

We also apply the model to a paradigmatic example of animal morphogenesis: the sea urchin gastrulation (Lane et al., 1993). In this example we do not intend to provide any new fundamental understanding about the sea urchin development *per se*. Instead we use sea urchin gastrulation to explain the steps by which the model is applied to a specific system.

At the onset of gastrulation, the sea urchin embryo consists of a hollow sphere of epithelial cells with some mesenchymal cells scattered in the blastocoel cavity. At the vegetal pole of the embryo there is a group of cells called the vegetal plate. The early invagination of the vegetal plate arises from two combined processes: the apical constriction of the vegetal plate cells and the secretion of ECM on their apical side (Lane et al., 1993). Experiments by Lane *et al.* applied different drugs to sea urchin embryos to specifically inhibit each of those two processes. By applying either one drug or the other different degrees of invagination were achieved, with no invagination when both drugs were applied.

The initial conditions assumed in the application of our model consist of a hollow epithelial sphere surrounded by an external ECM layer acting as the semi-rigid egg cover of the sea urchin (hyaline layer). It is known that only cells in the vegetal plate are involved in apical constriction and ECM secretion so we assume a molecule (from now on called molecule A) to be present in the apical side of the most central vegetal cells in the embryo (in black in fig. 4.4) to regulate those two processes. The model assumes that actomyosin contraction in the apical cell cortex triggered by molecule A results in a decrease of apical cell surface, which is translated as a decrease in apical surface of the cell's cylinders. The initial conditions were set with the EmbryoMaker editor. We set the three different experimental conditions using NetworkMaker. In the first we set a e_{lk} value different from zero and negative so that A leads to apical contraction by decreasing the contraction radius, p^{COD}, as specified in equation 4.7 (with k being for contraction, and l being for molecule A). In the second experiment A promotes the secretion of ECM in the nodes where it is present according to the equation for ECM secretion:

$$\frac{\partial p_i^{ECM}}{\partial t} = \sum_{m=1}^{n_g} c_{ml} g_{im} \tag{4.8}$$

Where p_i^{ECM} is the rate at which ECM accumulates around node i. c_m^{ECM} is the C matrix element where the effect of molecule m (in this case A) on ECM secretion is specified and g_{im} is the concentration of m in node i. Once p_i^{ECM} reaches a value of one a ECM node is secreted near node i (see Appendix B for the equations for the molecular regulation of other cell behaviours). Thus, this equation simply assumes that the rate of secretion of ECM is a linear function of the concentration of the regulatory molecules that promotes this secretion.

In the third experiment, A promotes both processes at the same time. The first experiment is the one in which an applied drug inhibits ECM secretion, the second the one in which a drug inhibits apical contraction and the third the one in which no drugs are applied. The experiments thus only involve different values in the $e_{A\ COD}$ and $e_{A\ ECM}$ parameters. These were

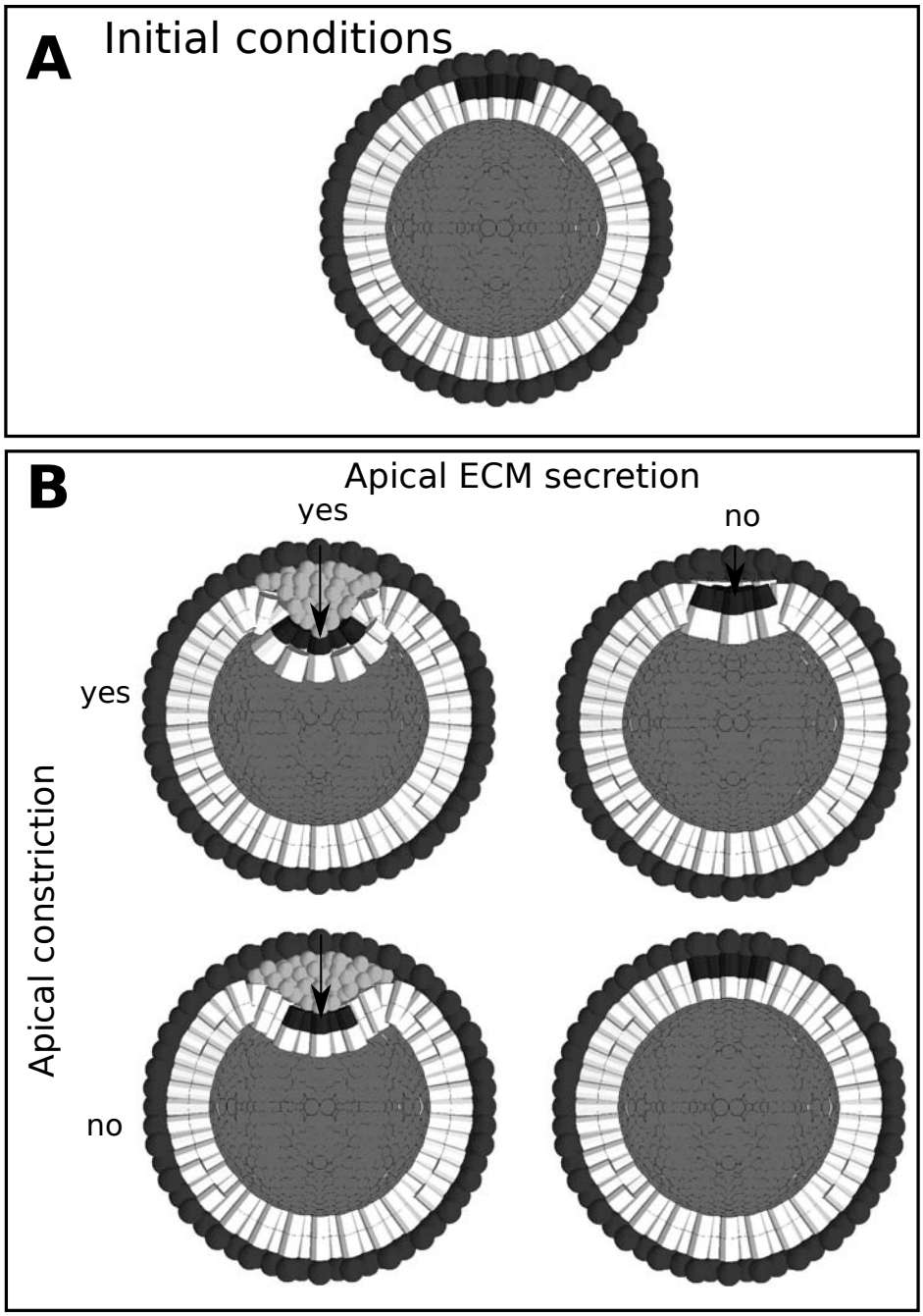

FIGURE 4.4: A) A hollow spherical epithelium surrounded by a rigid outer layer of ECM. A group of cells in the vegetal pole of the embryo express a specific gene product on their apical side (black). The "black gene product" may regulate cell constriction, ECM secretion or both. B) Different degrees of invagination, measured as the depth of the archenteron (black arrow), are achieved using each one of the mechanisms or both at the same time. All simulations were run 1500 time steps. Parameter values for: active contraction $e_{A\ COD}$ = -0.05 , active ECM secretion $e_{A\ ECM}$ = 0.9 , inactive contraction $e_{A\ COD}$ = 0.0 and inactive ECM secretion $e_{A\ ECM}$ = 0.9 . Other node properties were set homogeneous among all nodes for all simulations: p^{REC}=5.0 , p^{ADH}=5.0 , p^{EST}=10.0 , p^{ERP}=10.0 , p^{EQD}=0.15 , p^{ADD}=0.27 . The latter are standard values in which epithelia are stable.

manually set to match the morphologies observed in the two drug experiments. When these two values are used at the same time, in the third experiment, a deeper invagination resembling the one observed experimentally (Lane et al., 1993), when drugs are not added is found in the model simulations (fig. 4.4B). There are several earlier modelling approaches on sea urchin gastrulation (Cummings, 1990; Cummings, 1994; Cummings, 2005; Davidson et al., 1995) that undergo a more analytical and systematic study of the invagination process. However, as we mentioned above, our aim in this example is only to show that our model can be applied to real developmental systems.

The above examples consider only developmental mechanisms that involve only one or two cell behaviours at the same time. An additional example combines the regulation of several different cell behaviours to show how from simple initial conditions (fig. 4.5) a more complex gastrula-like embryo can be produced (see Appendix B section B.2 for a more detailed explanation of the observed dynamics).

4.6 Discussion

Our model differs from previous ones in implementing all the cell behaviours known in animal cells. In addition, it explicitly implements the molecular quantitative regulation of all cell mechanical properties and cell behaviours. Ours differs from previous models in explicitly considering the different mechanical properties of epithelial and mesenchymal cells and ECM and their interactions.

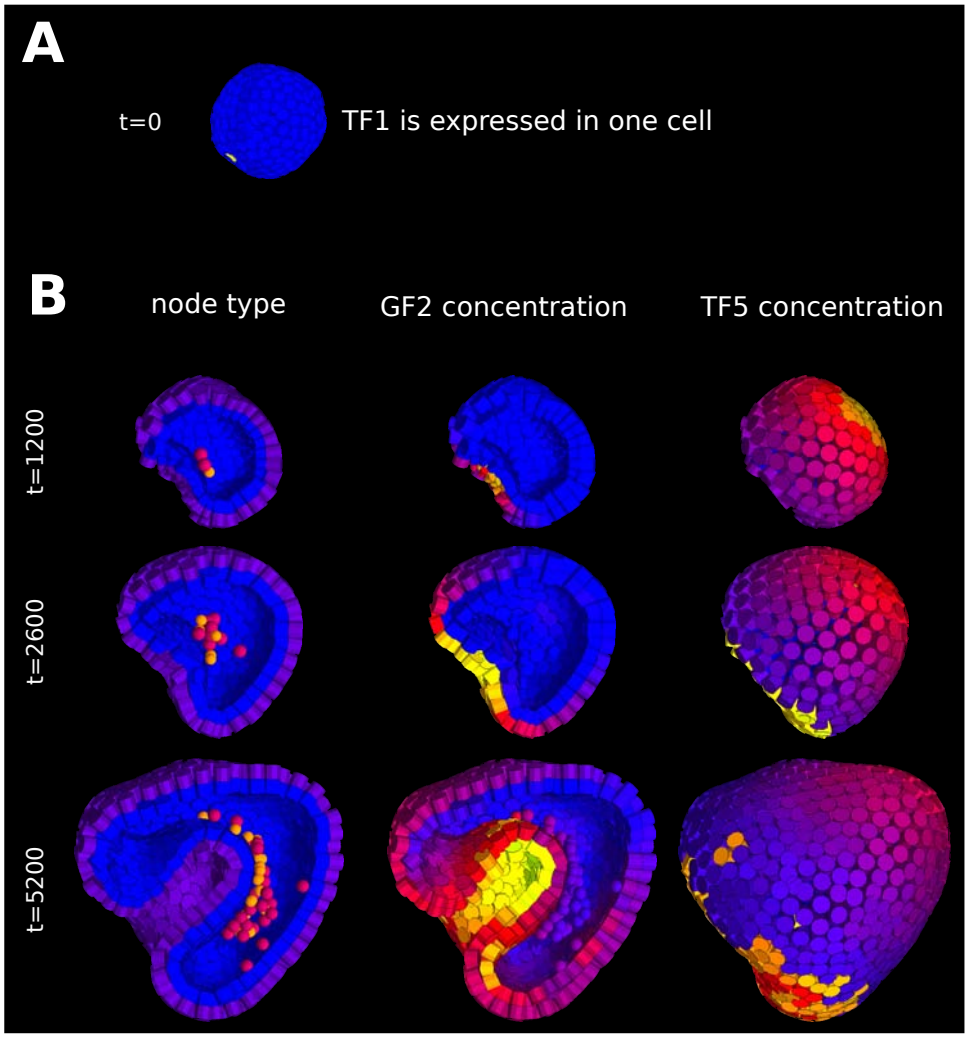

FIGURE 4.5: A) Initial conditions, hollow spheric epithelium with a single cell (yellow) expressing gene TF1. B) Outcome, after different number of iterations, of the complex developmental mechanism applied on the initial conditions in A. The left column shows, in section, the node types. Blue for basal side of cylinders, violet for the apical side of cylinders, red for mesenchymal cells and orange for extracellular matrix nodes. Middle and right column display concentrations of GF2 and TF5 respectively (yellow for high concentration, blue for low concentration). See Appendix B for details.

Chapter 5

A new model of early tooth development

5.1 Abstract

The mammalian tooth is a complex structure that shows great morphological variation across the phylogenetic tree and has a great influence on a species' diet. Morphological variation in teeth is mostly originated during development. Tooth morphogenesis involves mechanical forces mediated by growth and adhesive properties of epithelial and mesenchymal cells. In order to account for tooth morphology and its variation, we need to understand how moving cells and growing tissues exert mechanical forces on its surroundings during morphogenesis. Mathematical models of development integrate experimental knowledge and make quantitative predictions on the phenotype given a genetic or environmental perturbation. Previous mathematical models of tooth development did not implement the biomechanical aspects of tooth morphogenesis in detail. For that purpose we have built a new model of tooth development that implements realistic cell mechanics at all the cell layers involved in early tooth development. We have simulated tooth morphogenesis in a variety of scenarios assuming different rates of growth in the different tissues composing a tooth and different adhesive properties between them. The model predicts that the adhesive properties of cells within and between tissue types contribute significantly to the shaping of the tooth germ. More specifically, the model predicts that there are two main opposing forces that drive the direction of growth of the cervical loops (the two epithelial folds that will create the flanks of the tooth crown). Adhesive interactions between epithelium and mesenchyme drive the cervical loops to expand deeper into the jaw mesenchyme, whereas adhesive interactions between epithelium and the suprabasal layer (epithelial cells that compose the bulk of the tooth germ) drive the expansion of the cervical loops in the buco-lingual direction. The combination of these two processes determine the direction of growth of the cervical loops over time and the overall shape of the tooth crown and the sharpness of cusps. Our results provide new insights on how cell and tissue mechanics during tooth development can generate phenotypic variation and thus lead to evolutionary change.

5.2 Introduction

The mammalian tooth has been and it is being extensively studied both in the fields of development and evolution (Butler, 1983; Butler, 1995). From

the developmental point of view, it is an interesting system due to the complexity of its dynamics, which combine cell signalling and extensive cell movements (Salazar-Ciudad, 2008; Salazar-Ciudad, 2012). Teeth develop in relative isolation from the rest of the body. That allows experimental manipulation of the developing tooth germ with minimal interference on the morphogenesis of the whole embryo. From the ecological and evolutionary point of view, teeth are required for the breakdown of food in mammals and thus tooth morphology is closely related with the animal's diet (Evans et al., 2007). Moreover, teeth are often the only traces left by extinct mammalian species and so tooth morphology is very important to study the mammalian fossil record. For those reasons teeth are an especially relevant organ system at the place where the fields of developmental and evolutionary biology meet (evo-devo) (Salazar-Ciudad, 2012). From this latter perspective, the question is not so much how a wild type tooth is achieved, but rather how variation in its development will result in different tooth morphologies (Harjunmaa et al., 2014).

Tooth development consists of the growth and folding of an epithelial invagination, on the growth and condensation of an underlying mesenchyme (the dental papilla) and the differentiation of these tissues into the different cell types of the tooth crown. Such a process involves reiterative inductive interactions between epithelium and mesenchyme and the differentiation of non-proliferative signalling centres called the enamel knots (EK) that regulate the patterning and positioning of tooth cusps (Jernvall and Thesleff, 2000). The most studied case of tooth development is that of the mouse first molar, but most other molars seem to develop similarly.

At embryonic day 13 (E13), the prospective molar consists of a round epithelial invagination surrounded by a mesenchymal condensation called the tooth bud (fig. 5.1A left). A non-proliferative signalling centre called the primary enamel knot (PEK) appears at the distal tip of the bud (Jernvall and Thesleff, 2000). At this stage the epithelial bud consists of two cell populations, a basal layer of epithelial cells contacting the basement membrane and expressing P-cadherin, and a suprabasal cell population composing the bulk of the bud and expressing E-cadherin (Jussila et al., 2015).

By E15 two epithelial folds, called the cervical loops, have emerged at the bucal and lingual sides of the PEK, and the morphology is reminiscent of a cap (fig. 5.1B top). The cervical loops, which are composed of both basal and suprabasal cells, first grow radically (i.e. in the direction where the roots of the developed tooth will be, as opposed to occlusal) and later towards the mid line of the tooth germ, effectively enveloping the mesenchymal condensate. During the late cap stage, the PEK disappears and two secondary enamel knots (SEK) appear at its lingual and bucal sides (Jernvall and Thesleff, 2000). At the SEKs cell proliferation is downregulated while it is upregulated in the surrounding epithelia through the action of diffusible growth factors secreted by the SEKs (such as FGFs and SHH, Laurikkala et al., 2001; Jernvall et al., 1994). As a consequence the epithelium between SEKs grows and bends radically towards the dental mesenchyme forming a valley (Jernvall, 1995). Since there is no downward growth in the SEKs these are left behind and, consequently, form the cusps (with each SEK at the tip of each cusp).

By E16 the position and shape of the first two cusps starts to become evident. At that stage the cervical loops separate the mesenchyme in two

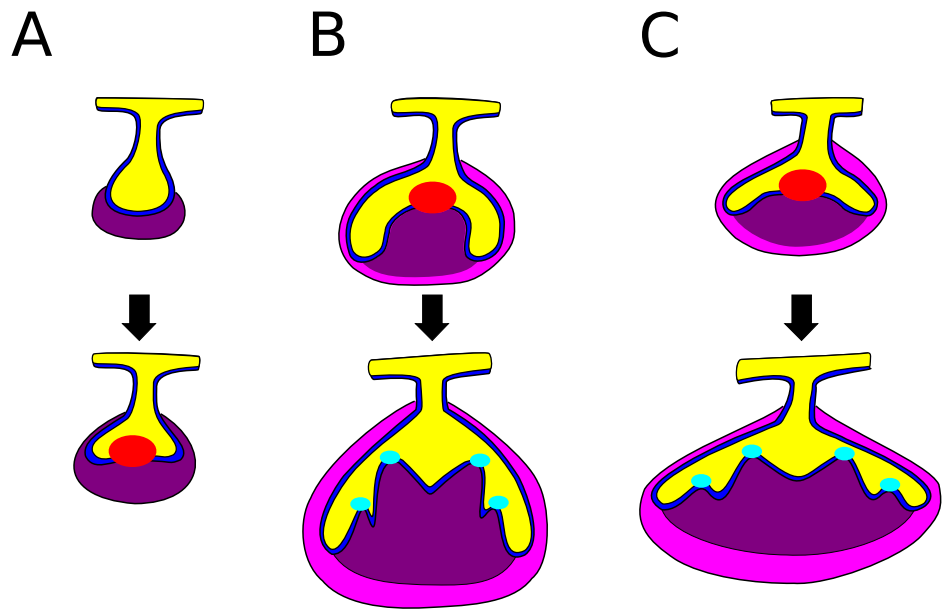

FIGURE 5.1: A, by E13, an epithelial bud (left, yellow) has invaginated from the oral epithelium into the jaw mesenchyme and a mesenchymal condensate has started to form (pink). Soon after that, a signalling centre called the primary enamel knot (PEK, red) appears at the distal tip of the bud and two epithelial folds start to grow in the bucal and lingual sides of the tooth germ. B, C, by E15 the growing cervical loops have eventually surrounded part of the mesenchyme (top). At this point the PEK has disappeared and new signalling centres called secondary enamel knots (SEK, cyan) are induced while the cervical loops continue to elongate (bottom). The direction of growth of the cervical loops during cap stage (B,C top) and later stages will determine the overall shape of the tooth crown and thus will influence the relative positions of cusps (B,C bottom). Cervical loops growing towards the jaw mesenchyme (B) will results in narrower and taller tooth crowns, whereas loops growing mainly in the buco-lingual direction (C) will result in wider and shorter tooth crowns.

populations: the dental mesenchyme enclosed by the loops (which will give rise both to the dentin forming odontoblasts and the tooth pulp) and the follicular mesenchyme surrounding the whole of the tooth germ (Rothová, Peterková, and Tucker, 2012). The former will give rise to the dentin forming odontoblasts and to the tooth pulp while the latter will not become part of the adult tooth. The suprabasal cells composing the bulk of the tooth germ start to vacuolate and secrete extracellular matrix, and at this stage they are called stellate reticulum. Two different epithelial populations can also be distinguished; the inner enamel epithelium (IEE) consisting of the epithelium enclosed between the two cervical loops, and the outer enamel epithelium (OEE) consisting of the epithelium facing outwards of the tooth germ. The final shape of the tooth crown and cusps corresponds to the shape of the IEE surface (fig. 5.1B,C), since this surface is where the enamel secreting ameloblasts will differentiate.

By E17 two additional SEKs appear anterior to the first two cusps and two days later, at postnatal day 0 (P0), the shape of the tooth crown is already determined and the ameloblast and odontoblast layers can be distinguished.

The overall morphology of the tooth is in most cases quite complex, but its general features can be described by: the size and shape of the tooth crown, the number and relative position of cusps and the specific shape of each cusp.

Cusps differ on how blunt or sharp they are. The way the epithelial and mesenchymal tissues grow and place themselves relative to one another will determine the final shape of the cusp. This means that not only cell signalling and gene expression changes, but also cell and tissue mechanics will contribute to the shaping of cusps. It has been proposed that the relative growth of the epithelium and the mesenchyme around the SEK contributes to cusp sharpness (Salazar-Ciudad, 2008). A high epithelial growth rate would result in sharp cusps whereas a high mesenchymal growth rate would result in blunt cusps (Salazar-Ciudad, 2008). In addition, the distal tip of the cervical loops will contribute too to the shape of the lingual-most and bucal-most cusps, thus the slope of the cervical loops will determine the sharpness of these specific cusps.

It has been proposed that the spatial positioning of the tooth cusps within the tooth crown is regulated through a combination of a Turing-like reaction-diffusion system and differential growth and adhesion (Salazar-Ciudad and Jernvall, 2002; Salazar-Ciudad and Jernvall, 2010). The height of the SEKs in the oclusal-radical axis will depend on their moment of appearance, the younger ones being lower and while the older ones being higher. This is the case because newer SEKs will appear at the edges of the IEE, and the IEE grows mainly in the radical direction due to the orientation of the cervical loops. Given that once they appear, SEKs do not seem to move in the oclusal-radical axis (Salazar-Ciudad, 2008), newer SEKs will result in shorter cusps and older ones in taller cusps (Jernvall, 2000). In that sense, the direction of growth of the cervical loop and consequently the slope of the IEE in respect to the occlusal-radical axis will determine the differences in height between cusps in a tooth. However, the mechanisms that determine the angle of growth of the cervical loops still remain to be elucidated.

The mechanical interactions between cervical loop epithelium, dental mesenchyme, follicular mesenchyme and suprabasal cells may have a role in determining the direction of growth of the cervical loop. On that account we hypothesize that four different processes could explain the direction of cervical loop and tooth morphogenesis between bud and cap stages.

1. Follicular mesenchyme or the jaw bone tissue might oppose a physical resistance to the growing tooth germ, thus preventing the cervical loops to grow in the buco-lingual axis.

2. Adhesive interactions between cells in the dental mesenchyme may regulate their level of compaction (i.e. the volume they occupy) and thus provide more or less space for the cervical loops to grow in the radical direction.

3. Adhesion between the dental mesenchyme and the enamel epithelium would result in the latter surrounding the former. This may create a pulling force initially directing the cervical loops radically and later medially towards the tooth germ mid line (Salazar-Ciudad, 2008). Cells in the suprabasal layer layer express different adhesion

molecules than the ones in the epithelium (Jussila et al., 2015; Zhao et al., 2015), thus

4. Differential adhesion between suprabasal cells and the apical side of the cervical loop epithelium may result in internal mechanical forces pulling the epithelium towards the centre of the tooth germ.

Mathematical models that integrate experimental knowledge and implement realistic dynamics of development can provide a preliminary framework in which to explore the consistency and morphological implications of mechanistic hypothesis and that can facilitate the understanding of such complex dynamics (Raspopovic et al., 2014; Salazar-Ciudad and Jernvall, 2005; Zhu et al., 2010). Earlier mathematical models of tooth development (Salazar-Ciudad and Jernvall, 2002; Salazar-Ciudad and Jernvall, 2010) were able to predict the spatial arrangement of tooth cusps and their shape, but considered only the IEE and the dental mesenchyme: the OEE, the follicular mesenchyme and suprabasal layer were not simulated. In these previous models the cervical loops were not represented in their entirety (only the IEE part was) and the mechanics of their growth were implemented ad hoc: the cervical loops were forced to grow in the radical direction with a lateral component that depended on the proliferation rate of the dental mesenchyme. Another model of tooth development (Takigawa-Imamura et al., 2015) implemented slightly more realistic cell mechanics, but only in two-dimensions and without considering neither the mesenchymal layer nor the presence of signalling centres regulating cell proliferation. In order to get a better understanding of how tooth crown and cusp height (and sharpness) is determined a more realistic model of the mechanical interactions between the IEE, the OEE, the dental mesenchyme, the dental follicle and the suprabasal layer are required.

In this study we present a new three dimensional model of tooth development that simulates the growth and mechanical interactions between all the tissue types that compose the tooth germ. This new tooth model is built as an application of a general model of embryonic development using the EmbryoMaker software (Marin-Riera et al., 2016) to address specific questions about cell mechanics during tooth morphogenesis. We use this model to explore the effect of differential cell growth and differential cell adhesion on tooth morphogenesis, and more specifically on cervical loop growth, through the mechanical forces exerted between cells during the period between bud stage and cap stage.

The model starts after the induction of the PEK and simulates the emergence of the cervical loops from the tooth bud and later the formation of the first cusps within the tooth cap. The model predicts the separation of the epithelium into IEE and OEE and the mesenchyme into dental and follicular. Induction of the SEKs and Turing-like dynamics are not included in the model because they take place at later stages and are not required to address the questions raised in this study. Thus, this model does not intent to predict the relative position of tooth cusps within the tooth crown, as in earlier models (Salazar-Ciudad and Jernvall, 2002; Salazar-Ciudad and Jernvall, 2010), although these features will be included in the future. Accordingly, the model is aimed to predict how the mechanical forces originated from growing tissues and cell adhesive contacts will affect the direction of growth of the cervical loops and the shape of the cusps, which will

have a significant influence on the overall shape of the tooth crown. For that purpose we have performed a parameter screening of the model focusing on the parameters related to differential growth and cell adhesion, as well as the effect of other mechanical parameters (such as cell incompressibility) on the overall shape of the tooth germ. After that, we have performed a set of *in silico* manipulations of tooth morphogenesis in order to test specific hypotheses on the role of mechanical interactions in determining the shape of the tooth crown and the tooth cusps.

5.3 Methods

5.3.1 Structure of the model

The tooth model was built using the EmbryoMaker modelling framework (Marin-Riera et al., 2016).

EmbryoMaker implements generic bio-mechanics of epithelial, mesenchymal cells and extracellular matrix (ECM), gene regulatory networks and a wide range of cell behaviours (cell division, cell adhesion, cell contraction, cell polarization, apoptosis, ECM secretion, etc.). In this model, each cell is made of a set of sub-cellular elements, elastic bodies that have a different shape depending on the cell type. Cells with no intrinsic polarization, such as mesenchymal and suprabasal cells are made of spherical elements, whereas epithelial cells, that have a marked apical-basal polarization due to their contact with the basement membrane, are made of cylindrical elements (being the long axis of the cylinder the apical-basal axis). Each cylinder is composed of two parts one representing the apical portion of the cell and the other the basal portion. By differentially applying forces onto each one of those elements epithelial cells are allowed to tilt. The ECM is made of spherical elements. Each element in mesenchymal cells or ECM and each of the two portions of a cylinder are called nodes. In this article, for simplicity, we chose each cell to be made of just one subcellular element, that is two nodes for epithelial cells and one for the rest. Since embryonic cells move in a highly viscous environment, we can assume overdamped motion kinetics (Purcell, 1977). Thus, movement of nodes is directly proportional to the mechanical forces exerted on them,

$$\frac{\partial \vec{r_i}}{\partial t} = \sum_{j=1}^{j=n_d} f_{Aij} \hat{u_{ij}} \tag{5.1}$$

where n_d is the number of nodes in the embryo, $\vec{r_i}$ is the position in three-dimensional space of node i, t is time (the model uses continuous time), f_{Aij} is the force modulus and $\hat{u_{ij}}$ is the unit vector between node i and node j for spherical nodes and an analogous property for cylinders. Nodes exert repulsive forces on surrounding nodes if the distance between their centres is smaller than their equilibrium distance (d_{EQD}) and exert attractive forces if the distance is larger than the equilibrium distance but shorter than the interaction distance (d^{ADD}) (fig. 5.2A). The modulus f_{Aij}

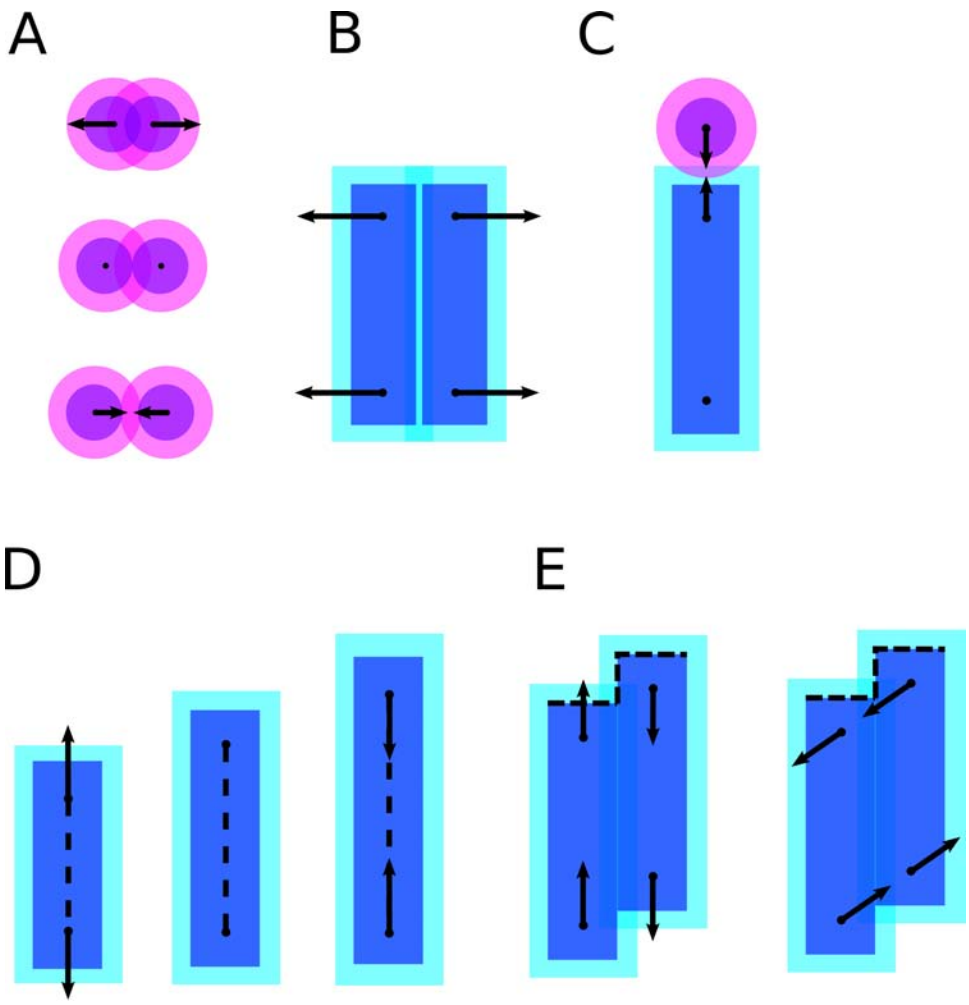

FIGURE 5.2: Epithelial (blue cylinders), mesenchymal and suprabasal layer (purple circles) cells are represented in the model as elastic bodies. A, when two cells are in contact with each other, elastic forces tend to position both cell centres at an equilibrium distance that is the sum of both cell radii (darker shade in the inner part of the cell). Thus, if they are closer than that distance, a repulsive force will apply, and if they are farther away than that an attractive force will apply. B, epithelial cells are composed of two elements that define the apical and basal portions of the cell and can move independently, but are connected through and elastic spring. Attractive and repulsive forces between epithelial cells are always applied in the direction normal to their lateral surfaces. C, attractive and repulsive forces between epithelial and mesenchymal cells are always applied in the direction normal to the apical or basal side of epithelial cells. D, the apical-basal length of epithelial cells is regulated by an elastic spring connecting the apical and basal elements, and tends to maintain the distance between both at an equilibrium value (specified as a property for each cell). E, epithelial bending forces contribute to keeping the epithelial cells as a sheet. Two types of bending force are implemented: radial (left) and rotational (right). Arrows indicate the direction of mechanical forces.

of the mechanical force vector is calculated as,

$$\begin{cases} f_{Aij} = k_{ij}^{REP}(d_{ij} - d_{ij}^{EQD}) & if\ d_{ij} < d_{ij}^{EQD} \\ f_{Aij} = k_{ij}^{YOU}(d_{ij} - d_{ij}^{EQD}) & if\ d_{ij} \geq d_{ij}^{EQD} \\ f_{Aij} = 0 & if\ d_{ij} \geq d_{ij}^{ADD} \end{cases} \tag{5.2}$$

where d_{ij} is the the distance between node i and j, d_{ij}^{EQD} is the equilibrium distance between node i and node j, which arises from the sum of the equilibrium radii, p_i^{EQD} and p_j^{EQD}, of nodes i and j, and is the interaction distance which is the sum of the node interaction radii p_i^{ADD} and p_j^{ADD}. k_{ij}^{REP} and k_{ij}^{YOU} are the elastic coefficients for repulsive and attractive forces respectively and they are determine by the mechanical properties of node i and j. The direction of force vectors differ between mesenchymal–mesenchymal, epithelial–epithelial and the epithelial–mesenchymal node interactions, since vectors need to be normal to the contact interface between nodes and nodes have different shapes in epithelial cells and mesenchymal cells (fig. 5.2A,B,C, see Chapter 4 and Appendix B for a detailed explanation). Attractive forces ($f_{Aij} > 0$) represent the tendency of contacting cells to stick together by means of adhesive contacts and the tension exerted by the actomyosin skeleton, whereas repulsive forces ($f_{Aij} < 0$) represent the tendency of cells to restore their equilibrium volume when they are compressed. The apical and basal nodes of epithelial cells are connected by an elastic spring that opposes any departure from an equilibrium distance between the apical and basal nodes of each cylinder (fig. 5.2D). The force generated by the spring connecting both epithelial nodes is calculated as follows,

$$f_{Sik} = k_{ik}^{HOO}(d_{ik} - p_i^{EQS})\hat{s_{ik}} \tag{5.3}$$

where $k_{ij}^{HOO} = p_i^{HOO} + p_j^{HOO}$ is the elastic coefficient of the spring (which is determined by the sum of the mechanical parameter p^{HOO} in both nodes), d_{ij} is the distance between node i and j, p_{ij}^{EQS} is the equilibrium length of the spring between node i and j and $\hat{s_{ij}}$ is the unit vector connecting the two epithelial nodes. Two additional force components are required in epithelial cells in order for them to organize as one layered sheets (fig. 5.2E). A radial force acts along the apical-basal axis of the cell and tends to restore displacements in that axis in respect to neighbouring cells in the epithelium, whereas a rotational force acts tangential to surface of the epithelium and tends to orient the apical-basal axis of cells normal to the epithelial plane. These forces are calculated as follows,

$$\vec{f_{ESTij}} = k_{ij}^{EST}\frac{\vec{m_{ijkl}} \cdot \vec{c_{ij}}}{|\vec{m_{ijkl}}|}\hat{m_{ijkl}} \tag{5.4}$$

$$\vec{f_{ERPij}} = k_{ij}^{ERP}\frac{\vec{s_{ik}} \cdot \vec{c_{ij}}}{|\vec{s_{ik}}|}\hat{c_{ij}} \tag{5.5}$$

where f_{ESTij} is the radial bending force and f_{ERPij} is the rotational bending force. We define c_{ij} as the vector connecting neighbouring node i and j, $\vec{s_{ik}}$ and $\vec{s_{jl}}$ as the vectors that connect each apical node to their basal counterparts and $\vec{m_{ijkl}}$ as the sum of s_{ik} and s_{jl} which defines the vector normal to the apical or basal surface between i and j. The radial bending force always acts on the direction of $\vec{m_{ijkl}}$ and is proportional to the deviation of the angle formed by $\vec{m_{ijkl}}$ and $\vec{c_{ij}}$ from 90º (the angle found when to cylinders are totally aligned). k_{ij}^{EST} is the sum of the mechanical parameter p^{EST} of nodes i and j. (see Chapter 4 and Appendix B for a detailed explanation). The rotational bending force is proportional to the deviation of the angle formed by $\vec{s_{ik}}$ and $\vec{c_{ij}}$ from 90º, but in this case the direction of

the force is parallel to $\vec{c_{ij}}$, thus promoting a tilting of the epithelial cylinder that reaches an equilibrium (that is the force modulus becomes 0) when the apical-basal axis of the epithelial cylinder is normal to the apical/basal cell surface. k_{ij}^{ERP} is the sum of the mechanical parameter p^{ERP} of nodes i and j. (see Chapter 4 and Appendix B for a detailed explanation).

In summary, the forces acting on an epithelial node are:

$$\frac{\partial \vec{r_i}}{\partial t} = \vec{f_{Sik}} \sum_{j=1}^{j=n_d} f_{Aij}\hat{u_{ij}} + \vec{f_{ESTij}} + \vec{f_{ERPij}} \tag{5.6}$$

where k is the node in the same cylinder than i and the sum is made over all the neighbouring nodes except for k.

Nodes are also containers of gene products. Gene products are produced by cells. Gene product transcription within a node is calculated as follows,

$$Q_{ik} = \frac{\Phi\left(\sum_{l=1}^{n_g} t_{lk} g_{il}\right)}{1 + \Phi\left(\sum_{l=1}^{n_g} t_{lk} g_{il}\right)} \tag{5.7}$$

where Q_{ik} is the rate of transcription of gene k in node i, g_{il} is the amount of transcriptional factor l in node i and each t_{lk} term is the strength by which each specific transcriptional factor k activates (positive t_{lk}) or inhibits (negative t_{lk}) the transcription of gene l. Φ is a function that is equal to 0 for values of x smaller than 0 and equals to x when x is greater than 0 ($\Phi(x) = 0$ if $x < 0$ and $\Phi(x) = x$ if $x > 0$). This function is used to ensure that there is not such a thing as negative transcription (although t_{lk} can be negative and thus repress transcription) (see Chapter 4 and Appendix B for a detailed explanation). Extracellular signal diffusion follows Fick's second law and takes place between nodes (see Chapter 4 and Appendix B for a detailed explanation). Because of that, we ensure that there are never empty spaces within the tooth in the model system. Extracellular signals can interact with their corresponding receptor to activate signal transduction. In order to model the kinetics of ligand receptor binding we consider three different molecular species, the free ligand, the inactive receptor and the receptor-ligand complex. The kinetics of receptor-ligand binding are as follows,

$$S_{ic} = a_1 g_{il} g_{ik} - a_{-1} g_{ic} \tag{5.8}$$

where l is the free ligand, k is the free receptor, c is the receptor-ligand complex, S_{ic} is the rate of production of c on node i, g_{ix} is the concentration of gene product x on node i, and a_1 and a_{-1} are the forward and backward constants for the ligand-receptor binding reaction respectively and are set as model parameters. (see Chapter 4 and Appendix B for a detailed explanation).

Gene products may affect the mechanical properties of cells and regulate cell behaviours. Given a mechanical property p, a gene product k will

affect its value as follows,

$$p_i^l(t) = \Phi\left(p_i^l(0) + (1 - p_i^{DIFF})\sum_{k=1}^{n_g} e_{lk}g_{ik}\right) \tag{5.9}$$

where $p_i^l(t)$ is the value of node property l in node i at time t and $p_i^l(0)$ is the value of that node property l in node i when the node was created (this is in the initial condition or when the node first arose through growth). Φ, as in equation 5.7, function ensures that node properties can become very small (or zero) but not negative.

Cell adhesion is mediated by a special type of gene product defined as adhesion molecule. Cells may produce adhesion molecules following equation equation 5.7. Given two neighbouring nodes, the strength of their mutual adhesion depends on their intrinsic adhesivity plus the product of the concentrations of each pair of adhesion molecule types expressed in each node and their binding,

$$k_{ij}^{YOU} = p_i^{ADH} + p_j^{ADH} + \sum_{l=1}^{n_{ADH}}\sum_{q=1}^{n_{ADH}}(g_{il}g_{jq}B_{lq}) \tag{5.10}$$

where k_{ij}^{YOU} is the elastic coefficient defined in equation 5.2, n_{ADH} is the number of different types of adhesion molecules, g_{il} is the amount of adhesion molecule l in node i, g_{jq} is the amount of adhesion molecule q in node j and B_{lq} is the affinity coefficient between adhesion molecules l and q (genetic parameter).

Cell proliferation is implemented as the splitting of one cell into two daughter cells of equal size. Each cell has a cell cycle variable P^{PHA} that can take values from 0 to 1. All cells divide when $P^{PHA} = 1$ and that variable resets to 0 in both daughter cells after they divide. Cell cycle progression depends on the presence of certain gene products that are set to regulate the cell cycle. Cell cycle progression is calculated as follows,

$$\frac{\partial P_h^{PHA}}{\partial t} = \frac{1}{n_h}\sum_{i=1}^{n_h}\sum_{m=1}^{n_g} c_{md}g_{im} \tag{5.11}$$

where h is the cell index, c_m specifies the ability of gene product m to regulate the cell cycle and g_{im} is the concentration of gene product m in node i (belonging to cell h). n_h is the number of nodes composing cell h (1 for mesenchymal and suprabasal cells and 2 for epithelial cells). Dividing the term by n_h ensures that cells with more nodes do not divide faster.

5.3.2 Model setup and initial conditions

The tooth model considers three tissues. 1) The enamel epithelium is composed of epithelial cells, each made of a cylinder, in contact with the mesenchyme; 2) the suprabasal layer is composed of cells consisting of single spherical nodes and they are located at the apical side of the epithelium and 3) mesenchymal cells, each made of a spherical node, located at the basal side of the epithelium (fig. 5.3A). Even though at later stages of tooth development the suprabasal cells become vacuolated cells and secrete a large amount of ECM, in the stages considered for this model the suprabasal layer consists of compact tissue (Sasaki et al., 1984).

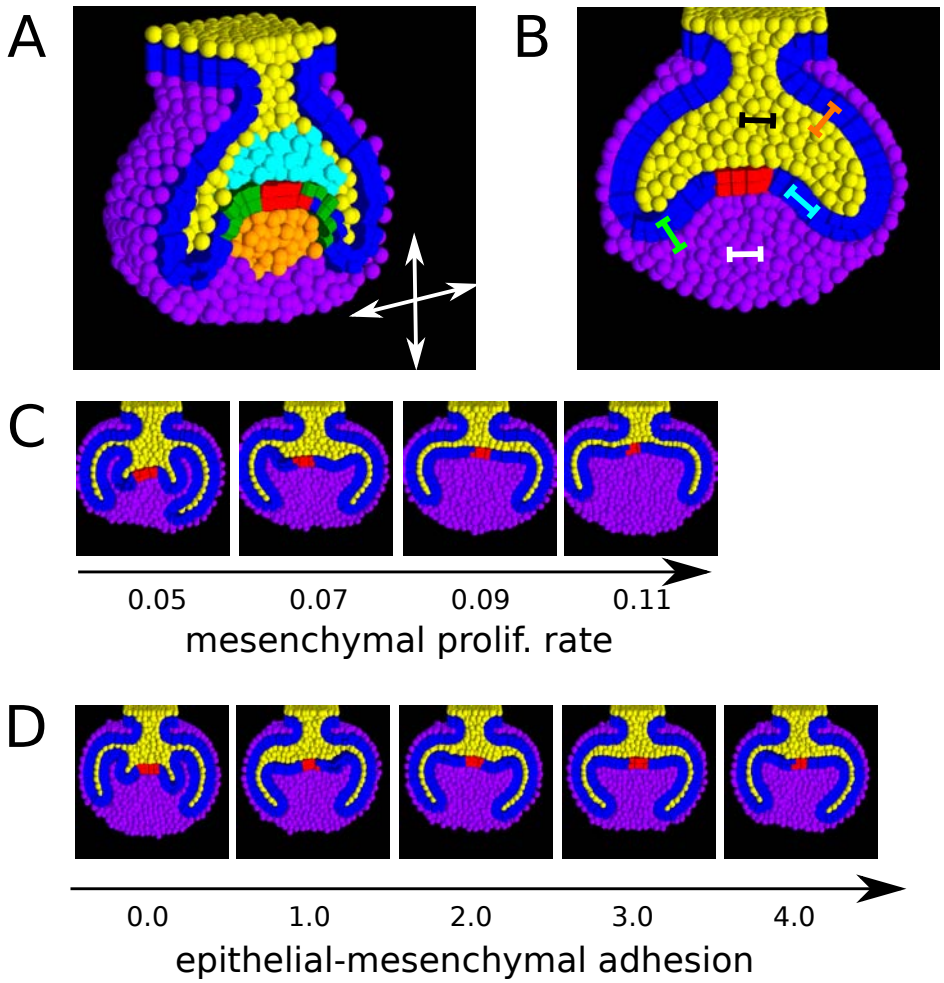

FIGURE 5.3: In the model 3 types of cells are specified: epithelial (blue cylinders), suprabasal (yellow spheres) and mesenchymal (purple spheres). A, differential cell proliferation in the model. An oblique view of a frontal section of the simulated tooth germ is shown. A group of epithelial cells is specified as a signalling centre (red cylinders), and secrete a growth factor that diffuses to the nearby tissues. Epithelial, suprabasal and mesenchymal cells that are within the diffusive range of the signal (green, cyan and orange) will divide at a certain rate. White arrows indicate the different spatial axes, occlusal (up), radical (down), bucal (left) and lingual (right). B, differential adhesion in the model. Each cell type expresses one specific type of adhesion molecule. Each couple of adhesion molecule types has a different binding affinity that is specified by model parameters. Color bars depict the different kind of adhesive interactions, epithelial homotypic (cyan), epithelial-suprabasal (orange), epithelial-mesenchymal (green), suprabasal homotypic (black) and mesenchyme homotypic (white). C, the resulting phenotype of different simulations is shown in which the mesenchymal proliferation rate was varied and the other parameters were kept constant. High values of mesenchymal proliferation lead to blunt, flat cusps while low values lead to thin, sharp ones. Epithelial and suprabasal proliferation rate are set to 0,25 and 0,03 respectively. D, the resulting phenotype of different simulations is shown in which the epithelial-mesenchymal adhesion was varied and the other parameters were kept constant. An effect on tooth germ shape similar to the one seen in C can be observed. All other adhesion parameters were set to 1,0.

All cells in a tissue express a specific gene product that acts as a transcription factor. All cells in a tissue express a tissue-specific adhesion molecule and a specific receptor for growth factors. In vivo, epithelial cells express different types of cadherins, a type of adhesion molecule that mediates cell-cell contacts. Cells in the enamel epithelium express a different type of cadherin than the ones in the suprabasal layer (Jussila et al., 2015; Zhao et al., 2015), but also integrins, a type of adhesion molecule that mediates cell-ECM contacts. Since mesenchymal cells are mostly surrounded by ECM they mostly express integrins, and thus strictly there is no such thing as cell-cell adhesion in the mesenchyme. However, for simplicity we assume that adhesion contacts between mesenchymal cells and between epithelium and mesenchyme are mediated by the ECM. In other words, the adhesion between two mesenchymal cells or between a mesenchymal and an epithelial cell considers the binding affinity of cell surface adhesion molecules (such as integrins) to the ECM and the adhesivity between the components of the ECM between both cells. By making each tissue express a different type of adhesion molecule in the model we can test hypotheses on how the differential adhesion between tissues affects tooth morphogenesis (fig. 5.3B).

In the model a group of epithelial cells is specified as a signalling centre reminiscent of the PEK by expressing a specific transcription factor. That transcription factor promotes the transcription and secretion of an extracellular signal (from now on the growth factor) that diffuses to the surrounding cells and interacts with each tissue-specific receptor. The binding between the growth factor an each type of receptor promotes the transcription of specific effector genes that regulate the progression of the cell cycle in each tissue. Thus, in the model only cells close to the signalling centre proliferate, while the rest remain non proliferative. This way we make sure cell proliferation takes places only in the inner enamel epithelium (IEE) and the dental mesenchyme, while the outer enamel epithelium (OEE) and the follicular mesenchyme remain non-proliferative, in accordance with experimental observations (Rothová, Peterková, and Tucker, 2012). We implement this simple signalling system in which each tissue expresses a specific effector molecule to systematically explore the effects of the proliferation rates in each different tissue on tooth morphology. The signalling centre transcription factor inhibits the transcription of growth factor receptors, thus those cells are insensitive to the growth factor they produce and never proliferate.

The period of tooth development considered in this study goes from the bud stage (E13) to the cap stage (E15). The initial conditions consist of an epithelial bud elongated in the anterior-posterior direction, covered by a layer of mesenchymal cells on the epithelial basal side. Cells at the borders of the epithelium are unable to move since they are mechanically attached to the oral epithelium. A horizontal unsurpassable barrier is put just above the edge of the tooth germ, emulating the boundary between the oral epithelium and the oral cavity. The temporal progression of the model has been solved using the Euler method (see Chapter 4 and Appendix B) and each simulation has been solved for 20000 time steps with a step size of 0,01.

Model parameter screening. Parameter screenings were carried out by simulating the transition from bud to cap stage with different parameter values. In each parameter screening some parameters were chosen to get different values and the rest were kept constant. A range of 3 or 5 different values were chosen for the selected parameters and all the combinations between them were generated and run in the model.

Tissue separation assays. Partial mesenchyme removal. At the bud stage initial conditions, all mesenchymal cells above a certain value of the z axis were removed, leaving only the ones closer to the signalling centre and the ones that presumably will form the dental papilla. After that the simulation was run normally.

Whole mesenchyme removal. A simulation was run from normal initial conditions until cap stage. Then all mesenchymal cells were removed, cell proliferation was repressed and then the simulation was further run until the tissue reached mechanical equilibrium.

5.4 Results

5.4.1 Mesenchymal cell proliferation and epithelial-mesenchymal adhesion contribute to the shape of the cusps

Figure 5.3A shows a frontal section of a simulated mouse molar at cap stage. Figure C.1A shows the developmental time sequence of the model molar until late cap stage (E15). A parameter screening on the rates of cell proliferation on the epithelium, suprabasal layer and mesenchyme was carried out in order to find the range of parameters for which wild-type molar phenotypes arises (fig. C.1B). Tooth morphogenesis was simulated starting from the bud stage and different values for the rates of cell proliferation in the basal epithelium, suprabasal layer and mesenchyme were set in each case. Wild type tooth phenotypes arose only for a restricted range of cell proliferation rates (fig. C.1B). In simulations where the rate of epithelial cell proliferation is high, the epithelium between the two cervical loops folds at one or several points, resulting in shapes reminiscent of tooth cusps (fig. C.1C). Figure 5.3C shows the effect of differential growth in the shape of cusps. Since mesenchymal proliferation only takes place under the signalling centre in the model, a high rate of mesenchymal cell proliferation will contribute positively to the final volume of the dental mesenchyme, which makes the tooth crown flatter and tooth cusps blunter (fig. 5.3D, first two pictures from the right). A low rate of mesenchymal proliferation, in contrast, will result in a high curvature at the tips of the cusps (i.e. sharp cusps, fig. 5.3C, first two from the left). A similar effect on tooth cusp morphology can be observed when the heterotypic adhesion strength between epithelium and mesenchyme is varied (fig. 5.3C). Tooth crowns with low value of epithelial-mesenchymal adhesion will tend to have sharper cusps while high values will result in blunter cusps. It has to be noted, however, that the heterotypic adhesion of epithelium and mesenchyme has an effect on how much the cervical loops envelope the dental mesenchyme. When the adhesion is lowest (fig. 5.3D first from left) the cervical loops don't entirely surround the mesenchyme under the signalling centre, leaving a

significant part of it outside the tooth crown. That means the amount of mesenchymal cells enclosed within the IEE and thus the size of the dental mesenchyme is dependent on the epithelial-mesenchymal heterotypic adhesive strength, which in turn may affect the bluntness of cusps the same way mesenchymal cell proliferation does.

5.4.2 Strong adhesion within the follicular mesenchyme and high epithelial-mesenchymal adhesion affects the angle of cervical loop growth

In order to assess the effect of mesenchymal cell adhesion on the direction of growth of the cervical loops in the model, we performed a parameter screening modifying the parameters related to the binding affinities between the adhesion molecules expressed in the epithelium and the mesenchyme. Figure 5.4A shows that simulations in which the homotypic adhesion strength between mesenchymal cells was high resulted in tooth germs that were narrow, with the cervical loops pointing downwards, whereas in simulations where mesenchymal homotypic adhesion was low, tooth germs were wide and cervical loops grew in the buco-lingual direction. No apparent differences were observed when the heterotypic adhesion between epithelium and mesenchyme was varied across simulations (fig. 5.4A).

In order to assess whether the effect of mesenchymal cell adhesion on the phenotype was mediated only by the follicular mesenchyme cells or also by the dental mesenchyme, we ran a set of simulations with the same parameter values as in figure 5.4A, but removing the prospective follicular mesenchyme at the initial conditions (fig. 5.4B). This allows to explore in isolation the effect of the dental mesenchyme itself on the direction of growth of the cervical loops. Removal of the follicular mesenchyme resulted in tooth germs that were wider in the buco-lingual axis compared to the ones that kept it (fig. 5.4A,B). In the simulations where follicular mesenchyme was missing, a high homotypic adhesion in the mesenchyme led the cervical loops to point radically, but that effect was more marked when the heterotypic adhesion between epithelium and mesenchyme was high (fig. 5.4B top right). Thus in the model, both a high homotypic adhesion of the dental mesenchyme plus a high adhesion between it and the inner enamel epithelium can drive the cervical loops to grow radically.

5.4.3 High homotypic adhesion strength of suprabasal cells results in buco-lingual orientation of the cervical loops

Figure 5.4 shows that in tooth simulations where there is neither homotypic mesenchymal nor heterotypic epithelial-mesenchymal adhesion (fig. 5.4 bottom left image both in A and B), the cervical loops tend to grow in the buco-lingual axis and barely in the occlusal-radical axis. That is, in absence of adhesion-mediated mechanical forces in the mesenchyme, the cervical loops mostly grow in the buco-lingual axis.

That led us to wonder what are the effects of epithelial and suprabasal cell adhesion in the shaping of the tooth germ. Figure 5.5A shows a set of simulations run with different values of homotypic adhesion of the suprabasal layer and heterotypic adhesion between epithelium and suprabasal cells. It

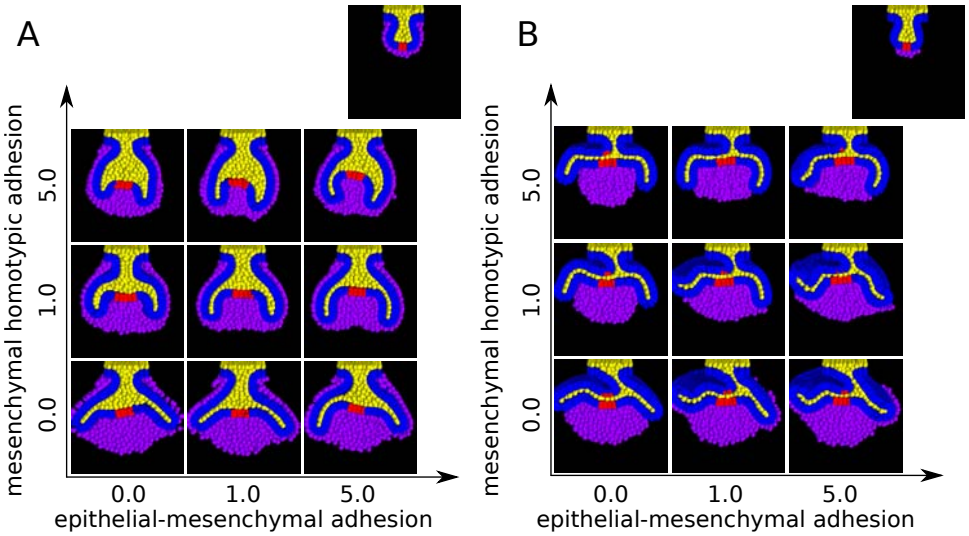

FIGURE 5.4: A, the resulting phenotype of different simulations is shown in which the mesenchymal homotypic adhesion and heterotypic epithelial-mesenchymal adhesion were varied and the other parameters were kept constant. Picture at the top indicates the initial conditions at bud stage. A strong adhesion between mesenchymal cells results in narrower tooth germs and cervical loops pointing radically. B, a set of simulations using the same parameter values as in A were run, but in this case the upper part of the mesenchyme was removed from the initial conditions (top picture), in order to simulate tooth morphogenesis in absence of follicular mesenchyme. Simulated tooth germs in absence of follicular mesenchyme are wider than the ones in A, especially when the adhesion strength in the mesenchyme is high. A high adhesion strength in the mesenchyme and between epithelium and mesenchyme is required for the cervical loops to point radically. Epithelial cells depicted in blue, suprabasal in yellow, mesenchymal cells in purple and epithelial signalling cells in red. All simulations were run for 20000 time steps. Epithelial homotypic adhesion, epithelial-suprabasal adhesion and suprabasal homotypic adhesion were set to 1.0 in all the simulations.

can be observed that, as long as there is a mechanical coupling between epithelium and suprabasal layer (i.e. heterotypic epithelial-suprabasal adhesion greater than 0), higher values of homotypic adhesion of the suprabasal layer will result in cervical loops growing in the buco-lingual direction rather than in the radical direction. This effect is more marked when the mesenchymal homotypic adhesion is equal to 0 (fig. 5.5B). It is worth noting that when there is no heterotypic adhesion between epithelium and suprabasal layer the cervical loops consistently grow radically (fig. 5.5A,B left columns). In other words, the cervical loop epithelia need to be mechanically attached to the suprabasal layer in order to grow in the buco-lingual direction. In addition, when simulations run with high and low values of epithelial homotypic adhesion were compared, no significant differences were observed (fig. C.2). We also observed the same effect of suprabasal layer homotypic adhesion on the orientation of the cervical loops for other tooth germ shapes (i.e. tooth germs with thicker cervical loops, fig. C.3). Thus, in the model the direction of growth of the cervical loops is greatly influenced by the adhesive properties of the suprabasal layer, but not by the ones of the epithelium.

In order to understand why does the increasing of the homotypic adhesion in the suprabasal layer results in cervical loops growing in the bucolingual direction, we analysed in more detail the mechanical forces at the cell level both in the epithelium and in the suprabasal layer in the model. In order to isolate the effect of forces acting on the epithelium and suprabasal layer only, we took simulated teeth at cap stage, removed the mesenchyme entirely and observed how the tooth germ would deform as a result of mechanical relaxation. In order to observe the effect of differential adhesion in epithelium and suprabasal layer during tissue relaxation, multiple separation assays were simulated by using the same cap stage tooth (simulation time = 8000, adhesion parameters used: epithelial-epithelial = 1.0 , epithelial-suprabasal = 1.0 , epithelial-mesenchymal = 1.0 , suprabasal-suprabasal = 1.0 , mesenchyma-mesenchymal = 1.0) and varying the parameters of epithelial-suprabasal heterotypic adhesion and suprabasal layer homotypic adhesion before separation of the mesenchyme and relaxation (fig. 5.5C,D). Using the same initial conditions allowed us to compare the deformation observed between the different separation assays. The way we change the adhesive properties of the tissue before separation of the mesenchyme and relaxation would be analogous to an experimental setting in which tooth germs are treated with drugs that affect the stability of adherens junctions and their coupling to the cytoskeleton for the different adhesion molecules expressed in the epithelium and the suprabasal layer. Cell proliferation was inhibited during these simulations in order to observe the sole effect of the mechanical relaxation in the tooth germ. In the model, both epithelium and mesenchyme are under a certain degree of compression exerted by the surrounding follicular mesenchyme during growth. Thus, when the mesenchyme is removed, the tooth epithelium tends to expand and the cervical loops elongate up to a certain degree (fig. 5.5C).

Our results show that when the suprabasal homotypic adhesion is high the cervical loops bend towards the buco-lingual axis. The same tendency can be observed for other combinations of mechanical parameters, such as epithelial homotypic adhesion and cell incompressibility (fig. C.4). During the mechanical relaxation, average cell mechanical stress (f_{Aij}, calculated as in equation 5.2 and 5.10, see section 5.3.1) in the epithelium and in the suprabasal layer were measured at each time step. Figure 5.5D shows the mechanical stress profiles during relaxation for the simulations depicted in fig. 5.5C. Both suprabasal layer and epithelium started with negative mechanical stresses (i.e. compression) that rapidly increased and reached a steady state, meaning that the system had reached a global mechanical equilibrium. Note that a global mechanical equilibrium means that the sum of all forces in the system equals 0, but at the same time the local sum of forces in a small region or over a specific tissue it's not necessarily 0. In the simulations without adhesion between the epithelium and the suprabasal layer (5.5D left column) both tissues reached mechanical equilibrium at 0 mechanical stress (i.e. neither compression nor tension). In contrast, when both tissues were mechanically coupled through heterotypic adhesion contacts (fig. 5.4D middle and right columns) epithelial cells only reached 0 mechanical stress when homotypic adhesion strength in the suprabasal layer was 0 (fig. 5.5D bottom row). When the suprabasal homotypic adhesion was high (fig. 5.5D top row), a steady state was reached in which

suprabasal cells had a positive mechanical stress (i.e. tension) and epithelial cells had a negative stress (i.e. compression). Thus, in the simulations where a high homotypic adhesion of the suprabasal layer was set and the cervical loops were more aligned with the buco-lingual axis, a global mechanical equilibrium was achieved in which epithelium and suprabasal layer were under compression and tension respectively along the length of the cervical loop, but they balanced each other through heterotypic adhesion contacts between them.

5.4.4 Cell incompressibility and other mechanical parameters do not qualitatively affect the shaping of the tooth germ in simulated teeth

Variation in other mechanical parameters within the model was explored in order to identify other possible factors that influenced the shaping of the developing tooth germ (fig. C.5). Tooth development was simulated varying cell incompressibility in epithelia, suprabasal layer and mesenchyme independently (fig. C.5A, B, C). It can be observed that high cell incompressibility in the suprabasal layer results in a larger volume of that tissue compared with the cases where incompressibility is low. High incompressibility in the epithelium results in a larger epithelial surface and longer cervical loops (compare images between fig. C.5A, B and C). No effect is observed when cell incompressibility is varied in the mesenchyme. That accounts for the fact that epithelium and suprabasal cells, but not mesenchymal, are under compression during morphogenesis in the model (fig. C.6). Variation in parameters related to epithelial bending forces didn't show any effect on tooth germ shape except when they were set to 0, in which case cells lost the ability to remain aligned in the epithelial plane and the tooth epithelium broke (fig. C.5D).

5.5 Discussion

We have built a model of tooth development that implements realistic cell mechanics of all the cell layers of the tooth germ, and it has provided predictions on how mechanical forces mediated by cell-cell adhesion may influence the growth of the cervical loops and the shaping of the tooth cusps in the model. Tooth morphogenesis is achieved in the model through the differential growth and adhesion of epithelia, mesenchyme and suprabasal layer, mediated by extracellular signalling from a non-proliferative epithelial signalling centre reminiscent of the primary enamel knot (PEK).

Earlier models of tooth development implemented the induction and patterning of the secondary enamel knots (SEK) through a Turing-like signalling network (Salazar-Ciudad and Jernvall, 2002; Salazar-Ciudad and Jernvall, 2010). Even though it is not possible to determine the final shape of the tooth crown without considering that process, it was out of the scope this study to implement the induction of the SEKs and the Turing-like dynamics, since they were not required for addressing the questions raised in this study, which focus on the orientation of the cervical loops during morphogenesis.

Recently, a two-dimensional model of tooth development has been presented that simulates the mechanical interactions between IEE, OEE and suprabasal layer during the transition from bud to cap stages (Takigawa-Imamura et al., 2015). The authors propose a mechanism by which two or more epithelial folds, reminiscent of cervical loops and tooth cusps, emerge from the tooth bud due to the mechanical instability created by the differential growth between epithelium and suprabasal layer. However, morphologies in the model seldom resembled real tooth morphologies. This was probably due to the fact that they did not include a signalling centre that would regulate localized growth and they did not include any kind of mesenchymal tissue, which did not allow to model any kind of mechanical interaction between epithelium and mesenchyme. Although their predictions regarding the emergence of epithelial folds are in accordance with our predictions on the same aspect (see section 5.4, fig. C.1), in our model the location of cell proliferation in specific regions has a large effect on how mechanical forces are distributed across tissues, which makes our model more realistic and more suited to address the questions related to tissue mechanics and cell adhesion raised in this study.

Differential growth between IEE and dental mesenchyme regulates the sharpness of tooth cusps. The model provides predictions about how differential cell growth contribute to the shape of the cusps. Our model supports the hypothesis that higher ratios of epithelial to mesenchymal growth rates will lead to sharper cusps and vice versa. Epithelial and mesenchymal growth rates determine the resulting surface area and volume of the tooth cusp respectively, thus high ratios of epithelial to mesenchymal growth will translate into high surface area to volume ratios of the cusp which means they will be sharp (fig. 5.3C). In contrast, model predictions on the effect of epithelial-mesenchymal cell adhesion on cusp shape are rather controversial. According to the differential adhesion hypothesis (Townes and Holtfreter, 1955) high heterotypic adhesion between epithelium and mesenchyme relative to mesenchymal homotypic adhesion should lead to high level of contact surface area between epithelium and mesenchyme relative to mesenchymal volume, which would result in sharp cusps. However, that is in contradiction with model predictions. In the model, higher heterotypic adhesion strengths result in the cervical loops surrounding a larger number of mesenchymal cells, which means the resulting dental mesenchyme will have a greater volume, thus mimicking the effect of increased mesenchymal proliferation on cusp shape (fig. 5.3D). It has been proposed that polarized cell division and directed cell rearrangements and migration in the mesenchyme play an important role in the morphogenesis of some developing systems such as the vertebrate limb (Boehm et al., 2010; Linde-Medina, Hallgrímsson, and Marcucio, 2015). Even though these cell behaviours were not included in this model for the sake of simplicity, we acknowledge they might have a role in determining the shape of tooth cusps.

Homotypic adhesion within the mesenchyme and heterotypic adhesion with the epithelium directs the growth of the cervical loops in the radical direction. The model predictions support our hypotheses regarding the role of cell adhesion in the follicular mesenchyme, dental mesenchyme

and suprabasal layer in determining the direction of growth of the cervical loops. Since follicular mesenchyme cells don't proliferate in the model (and neither they do in vivo, Peterkova et al., 2014), the overall growth of the tooth germ leads to a stretch of the follicular mesenchymal cell layer (fig. C.6A). That will result in the adhesive bonds between follicular mesenchyme cells generating tensile forces tangential to the surface of the OEE (fig. C.6) that oppose a mechanical resistance to the growth of the tooth germ in the bucal and lingual direction. The fact that in the model cervical loops grow in the bucal-lingual direction in the absence of adhesion in the mesenchyme (fig. 5.4A) and that simulated tooth germs are wider when the follicular mesenchyme is removed at the starting stage (fig. 5.4B) support that hypothesis. If that is case in vivo, we should expect cells within the tooth germ to be under compression and cells in the follicular mesenchyme to be under tension. It has been shown, although at earlier stages, that the cells conforming the mesenchymal condensate under the tooth bud are under mechanical compression (Mammoto et al., 2011), however there is not to our knowledge any evidence of the follicular mesenchyme to be under tension. Even though direct measurements of mechanical stresses during development have not been yet achieved on teeth, the presence of tensile forces could be inferred by assuming that a directional mechanical stress should result in cells being stretched in the same direction (Panousopoulou and Green, 2016). Additionally, at the stages of tooth development considered, the jaw mesenchyme is starting to differentiate into bone. Since mineralized bone tissue will present a rigid barrier to the growth of the tooth germ, we would expect it to have the same effect on the growth of the cervical loops as the compression exerted by the follicular mesenchyme.

The model predicts that the heterotypic adhesion between IEE and dental mesenchyme contributes to the growth of the cervical loops in the radical direction. By means of homotypic adhesive bonds, mesenchymal cells form a cohesive mass under the signalling centre, while cells at the edges of this mass will pull the IEE (and thus the cervical loop) radically through heterotypic adhesive interactions. As the cervical loops grow longer, they keep enveloping the dental mesenchyme and start growing towards the mid line of the tooth germ (fig. 5.4B). Thus, the model predicts that both the adhesion of the dental mesenchyme to the IEE and the compression that the follicular mesenchyme exerts on the cervical loops drive their direction of growth in the same manner. In order to experimentally test these two hypotheses, it would be necessary to isolate the effects of follicular and dental mesenchyme by removing the follicular mesenchyme but not the dental mesenchyme, as we did in the model (fig. 5.4B).

Homotypic adhesion of the suprabasal cells directs the growth of the cervical loops in the bucal-lingual axis. According to the model, mechanical stresses in the suprabasal layer mediated by homotypic and heterotypìc bonds also have an effect on the direction of growth of the cervical loops (fig. 5.5). During the elongation of the loops, suprabasal cells attached to the apical side of the cervical loop epithelium are stretched (fig. 5.5D), so as long as there is homotypic adhesive bonds between suprabasal cells they will pull the cervical loop towards the body of the tooth germ. This is supported by the fact that in the simulations where adhesion between epithelium and suprabasal layer is 0, but homotypic suprabasal adhesion is

greater than 0, the suprabasal cells don't follow the tip of the cervical loops, but instead remain in the centre of the tooth germ conforming a more or less round mass (fig. 5.5A,B left columns). Thus, stronger homotypic adhesive bonds in the suprabasal cells will result in larger mechanical tensions along the cervical loops as a result of the tendency of those cells to form a round body. That tension will act like an elastic cable running from tip to tip of the cervical loops through the main body of the tooth germ, forcing them to become straight and reducing the overall curvature of the IEE (fig. 5.5C,D, C.6B). Thus, the model predicts that as long as cervical loops form in the tooth bud and the suprabasal homotypic adhesion and epithelium-suprabasal heterotypic adhesion are greater than 0, the suprabasal cells will exert a tensile force along the length of the cervical loops. In order to exert this kind of mechanical stresses in vivo, suprabasal cells should be in close contact and have strong adhesive bonds with each other, thus we expect these mechanical forces to take place before suprabasal cells start to vacuolate and secrete extracellular matrix. As mentioned above, the presence of mechanical tension in the suprabasal layer could be inferred by assessing whether cells are elongated in the direction of growth of the cervical loop.

The direction of growth of the cervical loops is regulated by a balance of forces exerted between mesenchyme, epithelium and suprabasal cells. Our model suggests that the direction of growth of the cervical loops is greatly influenced by mechanical forces exerted by cells in the epithelium, mesenchyme and suprabasal layer by means of intercellular adhesive interactions. The range of orientations of the cervical loops spans from being close to parallel to the buco-lingual axis to being aligned with the occlusal-radical axis and even pointing towards the tooth germ mid line. In the model we have identified three factors that play antagonistic effects in determining that orientation. On the one hand, buco-lingual compression of the follicular mesenchyme and pulling of the cervical loops by the dental mesenchyme contribute to the radical and towards the mid-line growth of the cervical loops (fig. 5.1B). In the other hand, internal mechanical tensions from the suprabasal layer contribute to the growth of the cervical loops parallel to the buco-lingual axis (fig. 5.1C). Thus, the model predicts that the balance between these three factors will determine the direction of growth of the cervical loops and the overall shape of the tooth crown. We have checked the effect of other model parameters not related to cell adhesion on tooth germ shape, but variation in none of them contributes significantly this aspect of tooth germ shape (fig. C.5).

It has been shown in some invaginated epithelial organs, such as salivary and mammary glands (Gjorevski and Nelson, 2012; Larsen, Wei, and Yamada, 2006; Shih et al., 2015), that active cell migration at the tips of epithelial buds plays an important role in driving organ morphogenesis. It is not known yet whether cells at the tip of the cervical loops actively migrate towards the mesenchyme or are passively driven by tissue growth. Even if cells at the tips of the loops were actively migrating towards the mesenchyme by following some kind of spatial cues, we argue that they would still be greatly influenced by the mechanical forces exerted by the mesenchyme and the suprabasal layer and thus the predictions of this model would still be valid.

Taken together, this new model sheds some light into the less known, biomechanical aspects of tooth morphogenesis and their role in the emergence of some morphological features not usually considered in studies of tooth development. Even though the cervical loops are a transient structure in tooth development, their specific morphology will determine the overall shape of the tooth crown and influence the relative positions of the tooth cusps (fig. 5.1B,C). This model and the predictions provided by it thus emphasize the importance of the biomechanical aspects of development in order to account for phenotypic variation and evolutionary change.

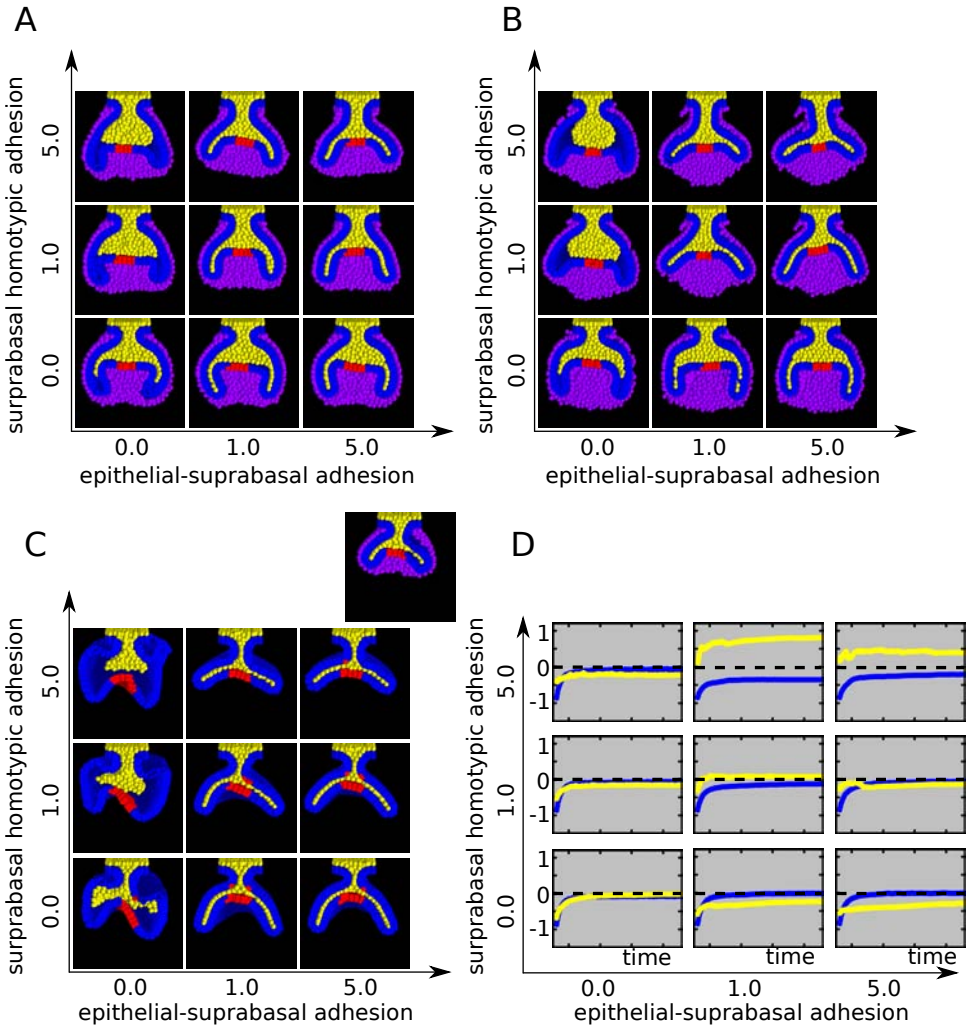

FIGURE 5.5: A, the resulting phenotype of different simulations is shown in which the mesenchymal homotypic adhesion and heterotypic epithelial-mesenchymal adhesion were varied and the other parameters were kept constant. Mesenchymal homotypic adhesion is set at 1.0 . A strong adhesion between suprabasal cells results in cervical loops growing in the buco-lingual direction. B, a set of simulations was run using the same parameter values, except for the mesenchymal homotypic adhesion that was set to 0. The effect on the direction of growth of the cervical loops is more marked than in A. All simulations were run for 20000 time steps in A and B. Epithelial homotypic adhesion was set to 1.0 and cell incompressibility to 0.5 for simulations in A and B. C, a simulated tooth at cap stage (top picture, 16000 time steps) was taken and all the mesenchyme was removed. Then the remaining epithelium was allowed to mechanically relax. The shape of the tooth germ is shown after relaxation for different values of suprabasal homotypic adhesion and epithelium-suprabasal heterotypic adhesion. The angle formed by the cervical loops respect the occlusal-radical axis increases after relaxation when the homotypic adhesion of the suprabasal cells is high. Epithelial homotypic adhesion is set to 1.0 and cell incompressibility is set to 3.0 for these simulations. D, for each simulation depicted in C, a temporal profile of the average cell mechanical stress in the epithelium (red) and suprabasal layer (green). Negative values mean compression and positive values mean tension. Higher homotypic adhesion in the suprabasal layer results in suprabasal cells being in a state of mechanical tension while the epithelium is in compression. Tissues were allowed to relax for 30000 time steps.

Chapter 6

Discussion

In this thesis we have studied the role of the genotype-phenotype map (GPM) in the distribution of phenotypic variation arising from development and its effect on the evolutionary dynamics of populations. By using the mammalian tooth as a model system, which is characterized by high complexity at both the level of phenotype and at the level of the GPM, we have simulated its development and evolution by means of mathematical models. This has provided us with predictions on how the complexity of the GPM affects the distribution of phenotypic complexity (see Chapter 3, Salazar-Ciudad and Marín-Riera, 2013) and morphological variation (see Chapter 5) with respect to genetic variation and and also predictions on how the GPM affects the effectiveness of natural selection in driving populations to an adaptive optimum (see Chapter 3, Salazar-Ciudad and Marín-Riera, 2013).

This has been achieved first by using a tooth development model previously published (Salazar-Ciudad and Jernvall, 2010) that accounted for the inductive (changes in gene expression patterning due to cell signalling) and morphogenetic (cell movements) events taking place in the inner enamel epithelium (IEE) of the developing tooth. Even though this model is able to reproduce realistic morphological variation regarding the positioning and shape of the tooth cusps, there are some aspects of tooth development that were not explicitly considered in this model, and so were implemented *ad hoc*. That imposes a number of limitations on the explanatory power of that model regarding the effect of mechanical interactions between the different cell layers of the tooth germ on the final shape of the tooth crown and tooth cusps.

Thus, in order to increase the predictive power and the level of realism not only of the tooth model, but also of other models of development, we designed a general modelling framework accounting for the dynamics of epithelia, mesenchyme and extracellular matrix (ECM) (Marin-Riera et al., 2016). This new modelling framework implements basic cell mechanics of epithelial, mesenchymal cells and ECM, the majority of cell behaviours known to happen during development (Salazar-Ciudad, Jernvall, and Newman, 2003) and custom gene regulatory networks. The modelling framework was shown to be able to predict the GPM and variational properties of the most basic developmental mechanisms (Salazar-Ciudad, Jernvall, and Newman, 2003) and also for other complex mechanisms arising from a combination of the basic ones (fig. 4.5, see Chapter 5).

Taking advantage of the new modelling framework we remade the tooth development model, now including all the cell layers involved and accounting for differential cell growth and cell adhesion between different

cell types and also for cell mechanics. By exploring the variational properties and GPM of the new tooth model we were able to make predictions on how the tooth crown is originates through the emergence of the cervical loops (see Chapter 1) and also on the role of cell adhesion and mechanics in the shaping of the tooth crown.

6.1 Complex GPMs preclude adaptation when there is selection on every morphological detail of the phenotype

The results presented in Chapter 3 (Salazar-Ciudad and Marín-Riera, 2013) predict that, contrary to some classic views in population genetics (Orzack and Sober, 1994; Maynard Smith, 1978), natural selection is generally unable to fine-tune a complex phenotype based on an optimal morphology. In other words, given enough time, natural selection can sculpt all aspects of the phenotype (Charlesworth, Lande, and Slatkin, 1982). The selection criterion of EMD (the one that takes into account all the details of the phenotype, see Chapter 3) represents exactly the optimality point of view, the search for a specific morphology characterized precisely on the bases of the values of all its morphological traits. This does not mean that there is no selection nor that selection is not a crucial force determining the direction of evolutionary change. This simply indicates that natural selection cannot "look" at morphology in a trait-per-trait basis and that, in any case, it should be "looking" at morphology in a different way in order for adaptive change to occur. Our model predicts that natural selection sees morphology on the basis of a complex holistic measure of the overall shape (such as the OPC, see Chapter 3). Notice that in tooth a change in the OPC value can be reached alternatively through changes in different parts of the phenotype and so it is not dependent on any specific morphological features, such as landmarks.

The conclusions of this study regarding the role of a complex GPM on adaptive evolution are summarized in figure 3.4A. Our model predicts that as long as the GPM is complex, the mapping between morphology and fitness needs to be degenerate, or many-to-one (Alfaro, Bolnick, and Wainwright, 2005) for adaptive change to occur. Our suggestion that natural selection is not looking at individual traits is not totally new. Classic work by Gould and Lewontin, already point out that natural selection does not act on individual traits (nor on alleles) but on the whole organism (Gould and Lewontin, 1979; Alfaro, Bolnick, and Wainwright, 2005). Our work reaches similar conclusions but from different lines of evidence; development and GPMs, with focus on functional organs and not on the whole organism as such. In addition, our model provides quantitative predictions on how much adaptive change a population can sustain under different selection regimes and at different levels of phenotypic complexity.

Even though the properties of both rugged and smooth adaptive landscapes have been extensively studied, there is little knowledge about whether the former or the latter are more frequent in nature. Theoretical predictions coming from the study of development (Alberch, 1982; Salazar-Ciudad and Jernvall, 2004; Salazar-Ciudad and Jernvall, 2005) point in the direction that

rugged landscapes are more prevalent due to the complexity of the underlying GPM. Our results however suggest a middle ground between the prevalence of either rugged or smooth landscapes in nature. GPMs tend to be complex and lead to rather rugged landscapes (as it is suggested from our present understanding of development), but this would force the mapping between phenotype and fitness to be relatively smooth so as to allow evolution by natural selection in globally not-so-rugged adaptive landscapes (such as in OPC).

The results on evolution by EMD from simple teeth (or by selection in a small number of landmarks) indicate that the fine-tuning of morphology may actually be possible in the case of very simple morphologies, which generally have simple GPMs. This suggests that in those lineages in which an increase in morphological complexity has occurred, there has probably been a change of how the phenotype is being selected (from fine-tuning selection to more OPC-like selection for more complex morphologies).

Our analysis of the variational properties also point out that simple unicuspid teeth are more likely to arise than complex multicuspid teeth when random parameters are given to the model (fig. A.12A) or when an existing tooth is mutated by changes in their model parameters (fig. A.12B). In the latter case it is also noticeable that while mutants originated from unicuspid teeth tend to have roughly the same complexity as the original phenotype, mutants originated from complex teeth show a large range of complexities. This is consistent with the larger disparity of mutant phenotypes arising from complex tooth morphologies in the model (Salazar-Ciudad and Jernvall, 2005). Experimentally it has recently been shown (Harjunmaa et al., 2012) that whereas mutations on single signalling pathways lead to a simplification of tooth morphology (in terms of OPC) one has to experimentally perturb multiple pathways at the same time in order to obtain mutants with higher morphological complexity. Our predictions are consistent with that. The question is then why are simple morphologies more likely to arise through mutations?. Is this a specific property of this model or something more general?.

In our model and in developing teeth, once a cusp starts forming it inhibits the formation of new cusps in its surroundings, which means it is relatively unlikely to develop a high number of cusps in a small tooth crown. In development this is due to lateral inhibition between cusps mediated by specific extracellularly diffusible molecular signals. OPC is not a direct measure of cusp number but it largely correlates with it. Since only some of the parameters of the model lead to a direct and overall increase in size and only large teeth can have many cusps it is to be expected that most combinations of parameter values produce relatively simple teeth. In fact, to fit a large number of cusps in a single tooth, the radius of lateral inhibition has to be decreased (for example by changes in the diffusivity or activity of the diffusing inhibitory signal) while at the same time the patterns of growth have to be changed so that cusps forming close to one another do not fuse. Both in the model and in real teeth, this involves tinkering the values of several parameters at the same time (Harjunmaa et al., 2012).

We expect that this anisotropy in the distribution of simple and complex phenotypes is also general, as it has been argued before (Salazar-Ciudad and Jernvall, 2005). This is expected because, in brief, morphology in development arises because cells interchange molecular signals and undergo

biomechanical changes that transform some early spatial distributions of cell types into different ones. The more interchanges of signals there are, the more the respective spatial patterns can differ from prior ones. Since the egg cell and the early embryo are spatially simple it is natural that to increase phenotypic complexity a large number of gene interactions are required. Since genetic mutations most often consist in one gene or gene interaction being defective it is to be expected that most mutations lead to a simplification of the genetic networks and of the phenotype (Harjunmaa et al., 2012).

In summary, the modelling of tooth development and evolution has provided interesting predictions on how morphological complexity can arise through changes in development, and how that same process can hinder evolutionary change unless a specific type of natural selection is at play.

6.2 A general modelling framework for epithelia, mesenchyme and ECM allows to reproduce a wide range of basic developmental mechanisms and, by extension, a wide range of complex developmental systems

The general modelling framework presented in Chapter 3 and implemented in the EmbryoMaker software (Marin-Riera et al., 2016) has been shown to reproduce the most basic developmental mechanisms (as in any gene network that regulates at least one cell behaviour and is involved in pattern transformation, fig. 4.3, Salazar-Ciudad, Jernvall, and Newman, 2003), that is a mechanism involving at least one cell behaviour and that is capable of pattern transformation (i.e. a change in the spatial distribution of cell types). We were able to use the modelling framework to explore the variational properties and GPM of each of these basic mechanisms in isolation. We have shown that it is possible to model complex developmental mechanisms as a combination of some of the basic ones (see fig. 4.4, 4.5 and Chapter 5). It is also important to note the fact that the framework considers epithelial and mesenchymal cells as different cell types with different mechanical properties, since most developmental systems that have been studied (especially in vertebrates) require the interaction between epithelia and mesenchyme (Biggs and Mikkola, 2014).

We argue that the ability of our model to reproduce the GPM of a wide range of both simple and complex developmental mechanisms, including epithelia, mesenchyme and extracellular matrix, comes from the way we implemented cell types, cell mechanics and cell behaviours, and now we will discuss why other types of model are not as suitable as ours in doing all this.

The Cellular Potts Model (Graner and Glazier, 1992; Izaguirre et al., 2004; Starruß et al., 2014a) can effectively model a wide variety of cell behaviours and very detailed changes in cell shape, as well as being capable of modelling both epithelial and mesenchymal cells. However, the method used to numerically solve these models is restricted to the Monte Carlo method only. In this method, potential energies are calculated for each part of the cell and extracellular space based on the local stability of the system.

Changes in cell shape happen asynchronously, one random lattice position at a time. Changes are stochastic but biased towards minimizing the energy locally (Graner and Glazier, 1992). The use of this method guarantees that the system reaches the equilibrium state (i.e. the end of development), but it cannot provide predictions on the real time dynamics that lead to the equilibrium state. This means that this model cannot predict the mechanical forces acting on cells and the extracellular matrix during development. In our model this is not a problem, since the cell movement and mechanical forces are defined as a system of partial differential equations. By numerically solving the system of equations we do not only obtain the equilibrium state of the system, but also realistic predictions on the temporal dynamics and the mechanical forces that act along the whole process.

The Vertex Model (Honda, Tanemura, and Nagai, 2004; Farhadifar et al., 2007; Smith et al., 2012; Pitt-Francis et al., 2009) is very effective at modelling the dynamics of cell shape changes, cell rearrangements and both cellular and supracellular scale biomechanics of densely packed tissues (see Chapter 1, Section 1.4). This means that vertex models are exceedingly good at modelling epithelial cells and tissues, but they are not so good at modelling mesenchymal tissues and extracellular matrix, since the vertex implementation relies on the fact that all cells are in close contact, with no extracellular spaces between them. In our model, cells are not defined by their contact interfaces (i.e. vertices and edges) but by their centres and cell-cell contacts are represented as interactions that are totally dependent on the relative distance and position of cells. This allows to implement highly dynamic cell rearrangements in 3D with minimal mathematical complexity, as well as it allows cells to lose all their contacts and separate from the tissue. Moreover, the vertex model is good at modelling some cell behaviours such as small cell rearrangements, cell shape changes and polarized cell divisions, but not others like epithelial-mesenchymal transitions and individual cell migration.

The viscoelastic or Immersed Boundary Cell Model (IBCell) (Rejniak, 2007; Tanaka, Sichau, and Iber, 2015) can effectively model cell shapes with high detail and thus a wide variety of cell behaviours related to both epithelial and mesenchymal cells can be potentially implemented (see Chapter 1, Section 1.4). Moreover, the immersed boundary method allows to model the extracellular environment in a quite realistic way, making it possible to simulate the dynamics of mesenchymal tissues as well as realistic diffusive and advective molecular dynamics. This model has the advantage over the vertex model that the cell-cell contacts can be dynamic, since single cell membranes are modelled separately and then dynamic adhesion contacts between different cell membranes can be implemented (Rejniak, 2007). However, that requires for the IBCell model to define cell membranes with a high level of detail, which means a high number of vertices for each cell. The high level of detail of cell membranes added to the fact that the flow dynamics are solved by means of a lattice means that a lot of computational power will be required to model large systems. This is not a problem in our model, since the level of detail of cells can be adjusted by setting the number of elements cells are composed of, up to the point that a whole cell may be represented by a single element. Thus, the large computational requirements of the IBCell model constitute a limitation on the size of the system

it can model. The high level of detail of the IBCell model also makes its implementation in 3D computationally heavy, thus most instantiations of this model are restricted to 2D (Rejniak, 2007; Tanaka, Sichau, and Iber, 2015).

The subcellular elements model (SEM, Newman, 2005) can model cell shapes and cell deformation with high detail, as well as a variety of cell behaviours and realistic cell mechanics. However, in the original version of the model cells are intrinsically amorphous, which means that they cannot assume a consistent apico-basal polarisation and therefore cannot have different cellular properties in different sides of the cell. In other words, there is no robust way to specify apical, basal and lateral sides of the cell with specific cellular properties (thus, they cannot organise in monolayered sheets characteristic of epithelia). Therefore, the original SEM is suitable for modelling mesenchymal-like tissues but not epithelia (Newman, 2005; Delile, Doursat, and Peyriéras, 2013). In our model we use the original implementation of the SEM to model mesenchymal cells and a different implementation to model epithelial cells. Epithelial cells are also made up of subcellular elements, but instead of being spherical as in the original SEM, they are cylindrical. Each cylinder conforms a portion of the cell stretching from the apical to the basal side, so an epithelial cell is composed of a group of cylinders placed next to one another. The bottom and top faces of the cylinder represent a portion of the apical and basal side of the epithelial cell respectively, and the lateral face accounts for the inner part of the cell, or part of the lateral cell membrane if the cylinder is at the edge of the cell. The lateral faces of cylinders can only interact with the lateral faces of other cylinders, whereas the apical and basal faces can interact with apical and basal sides of other cylinders or with mesenchymal or ECM elements. The assumed shape of the epithelial elements ensures that epithelial cells keep a consistent apical-basal polarisation and they organise as monolayered sheets. Our framework is the first one to use an extended implementation of the SEM that is able to model epithelial cells. In addition to that, our modelling framework implements a wide range of cell behaviours that were not present in the original SEM.

The SEM and, by extension, our model present some limitations compared to the models described above. Individual cells are defined by the spatial distribution of elements conforming the cell body. It is assumed that the elements located at the outermost part of the cell represent, at least partially, the cytoplasmic membrane. Since elements within a cell can be rearranged, the cell membrane may not always be represented by the same elements. This means the SEM is not very good at modelling phenomena that involve a detailed representation of the cell membrane and quantification of its properties, such as cell deformation by means of membrane surface tension. It is in theory possible in the model to infer the cell membrane by identifying which elements are located in the surface of the cell, although it would involve additional calculations and it actually presents several implementation challenges. Nonetheless, these membrane related phenomena tend to become irrelevant for the behaviour of the system at tissue level scales. Another limitation of the SEM is that, being an off-lattice model, the regions where there are no cell elements are basically void. This means diffusion of molecules cannot be calculated in those regions. This can be overcome in our model by filling the space between cells with extracellular matrix, which is represented in the model by a special type of

subcellular element. This way, molecules can diffuse both within cells between cellular elements and between cells through the extracellular matrix elements.

The EmbryoMaker framework aims to be a widespread platform for modelling morphogenesis of a variety of organs and systems. So far it has proved to be able to successfully model systems as different as tooth development (see Chapter 5) and early spiralian cleavage (Brun-Usan, Marin-Riera and Salazar-Ciudad, in preparation). This shows the potential of such a general framework to replace the classic idea of designing organ-specific models by the idea of designing models that can simulate multiple types of organs with different parameter settings. That will allow precise and quantitative comparisons between different systems within the same model, and will also allow to make predictions on the origin of evolutionary transitions between different types of organs (e.g. the transition from reptilian scales to avian feathers, Chang et al., 2009; Chuong et al., 2000).

6.3 A new and more detailed model of tooth development sheds light on the effect of differential cell adhesion and cell mechanics on the variational properties of teeth and its GPM

By applying the general modelling framework described on Chapter 4 to the case of tooth development, we have created a new model that includes a number of features that were lacking in previous models of tooth development (Salazar-Ciudad and Jernvall, 2002; Salazar-Ciudad and Jernvall, 2010), such as multiple cell layers of the tooth germ (fig. 5.1) and realistic cell mechanics.

Earlier models of tooth development implemented the induction and patterning of the secondary enamel knots (SEK) through a Turing-like, reaction-diffusion signalling network (Salazar-Ciudad and Jernvall, 2002; Salazar-Ciudad and Jernvall, 2010). Even though it is not possible to determine the final shape of the tooth crown without considering the number and relative position of the tooth cusps, they were not required for addressing the questions raised in this study, which focused on explaining morphological variation on the overall shape of the tooth crown. The reaction-diffusion signalling network is indispensable in order for the model to account for the patterning of the SEKs and the positioning of the tooth cusps in the tooth crown. Thus, it is planned that future versions of the model will include the reaction diffusion network and the patterning of tooth cusps.

So far, our model contributes with new predictions that the earlier models could not provide, regarding the emergence of the cervical loops from the tooth bud (see Chapter 1 Section 1.2.1) and the effect of differential cell adhesion on the direction of growth of the cervical loops and its role on determining the overall shape of the tooth crown.

Our model predicts that the transition between bud and cap stage (see section 1.2.1), in which the cervical loops emerge and envelop the dental mesenchyme, occurs due to a combination of differential cell growth and adhesion between the different cell types. A parameter screening focused

on the rates of cell proliferation in the different tissues indicates that the cervical loops only originate for a range of parameter values, characterized by a much higher rate of proliferation in the basal epithelium (cell layer contacting the mesenchyme) relative to the suprabasal layer (cells composing the bulk of the tooth germ) (see Section 5.4, fig. C.1). A two-dimensional model of tooth development, which simulated the mechanical interactions between IEE, OEE and suprabasal layer during the transition from bud to cap stages, was recently published (Takigawa-Imamura et al., 2015). The authors propose a mechanism by which two or more epithelial folds, reminiscent of cervical loops and tooth cusps, emerge from the tooth bud due to the mechanical instability created by the differential growth between epithelium and suprabasal layer. However, morphologies in the model seldom resembled real tooth morphologies. This was probably due to the fact that they did not include a signalling centre that would regulate localized growth and they did not include any kind of mesenchymal tissue, which did not allow to model any kind of mechanical interaction between epithelium and mesenchyme. Although their predictions regarding the emergence of epithelial folds are in accordance with our predictions on the same aspect, in our model the location of cell proliferation in specific regions has a large effect on how mechanical forces are distributed across tissues, which makes our model more realistic and more suited to address the questions related to tissue mechanics and cell adhesion raised in this study.

Our model also predicts that the direction of growth of the cervical loops is greatly influenced by mechanical forces exerted by cells in the epithelium, mesenchyme and suprabasal layer by means of intercellular adhesive interactions. The range of orientations of the cervical loops spans from being close to parallel to the buco-lingual axis to being aligned with the occlusal-radical axis and even pointing towards the tooth germ mid line. In the model we have identified three factors that play antagonistic effects in determining that orientation. On the one hand, buco-lingual compression of the follicular mesenchyme and pulling of the cervical loops by the dental mesenchyme contribute to the radical and towards the mid-line growth of the cervical loops (fig. 5.1B). In the other hand, internal mechanical tensions from the suprabasal layer contribute to orienting the cervical loops parallel to the buco-lingual axis (fig. 5.1C). Thus, the model predicts that the balance between these three factors will determine the direction of growth of the cervical loops and the overall shape of the tooth crown.

Surprisingly, the exploration of the variational properties of the new tooth model showed mostly gradual phenotypic variation when parameters were gradually varied (fig. 5.4, 5.5). This would suggest that the GPM for this model is quite simple, which might seem contradictory to the fact that the earlier tooth models, despite being less detailed than the current one, show a much more complex GPM characteristic of a morphodynamic mechanism (Salazar-Ciudad and Jernvall, 2004; Salazar-Ciudad and Jernvall, 2005; Salazar-Ciudad and Marín-Riera, 2013). This can be explained by the fact that one of the aspects of tooth development that contributes most to the complexity of the GPM is the reaction-diffusion signalling network, which is actually lacking in the current model. This could lead us to think that the aspects of tooth development related to differential cell adhesion and mechanics alone have a simple GPM and thus do not contribute to the complexity of the whole GPM of tooth. Nonetheless, this might not

be necessarily the case, since we have shown that differential cell adhesion controls the direction of growth of the cervical loops, which greatly determines the overall shape of the tooth crown, making it either tall and narrow or short and broad. Since the patterning of tooth cusps is dynamically determined by a reaction-diffusion mechanism, it is very sensitive on the shape of the patterning field (in this case the inner enamel epithelium). This, in turn, will affect the effective reach of the diffusing molecules that determine the induction of new SEKs and the positioning of new cusps in the developing tooth crown. Taking this into account, we propose that differential cell adhesion does not have a simple contribution to the GPM, but feeds into and increases the complexity of the GPM coming from the reaction-diffusion dynamics.

Taken together, this new model sheds some light into the less known biomechanical aspects of tooth morphogenesis and their role in the emergence of some morphological features not usually considered in studies of tooth development. This model and the predictions provided by it thus emphasize the importance of the biomechanical aspects of development in order to account for phenotypic variation and evolutionary change.

Chapter 7

Conclusions

1. Phenotypic complexity tends to be anisotropically distributed with respect to genetic variation in complex GPMs. In other words, it is more likely to get a simple phenotype rather than a complex one only by chance, and it is also more likely that complexity decreases rather than increase by a random mutation once the phenotype is already complex.

2. Complex GPMs stemming from development lead to rugged adaptive landscapes and consequently hinder adaptive evolution when the phenotype-fitness map (i.e. fitness function) considers all the details of the phenotype. In contrast, a many-to-one phenotype-fitness mapping results in a smoothening of the adaptive landscape, thus facilitating adaptive change.

3. It is possible to design a general 3-dimensional model of development including epithelia, mesenchyme and extracellular matrix and the majority of cell behaviours, and which is capable of reproducing the most basic developmental mechanisms that lead to pattern transformation, as well as their variational properties.

4. Mathematical modelling of tooth development predicts that differential tissue growth between the dental epithelium and the suprabasal layer is responsible for the emergence of the cervical loops, which are indispensable for the formation of the tooth crown.

5. Mathematical modelling of tooth development predicts that differential mechanical forces mediated by differential cell adhesion control the direction of growth of the cervical loops, which has a direct effect on the overall shape of the tooth crown and by extent on the way tooth cusps are patterned.

---

1. La complexitat fenotípica tendeix a estar distribuida de forma anisotròpica respecte la variació genètica en MGF complexes. És a dir, és més probable que el desenvolupament produeixi un fenotip simple que un de complex per pur atzar, i també és més probable que una mutació genètica aleatòria en un fenotip complex comporti una reducció de complexitat més que un increment.

2. MGFs complexes provinents del desenvolupament causen l'aparició de paisatges adaptatius rugosos i en conseqüència obstaculitzen l'evolució adaptativa quan el mapa fenotip-*fitness* (és a dir, la funció de *fitness*) té en compte tots els detalls del fenotip. Per altra banda, la presència d'un mapa fenotip-*fitness* "*many-to-one*" causa el suavitzament de la superfície del paisatge adaptatiu, cosa que facilita el canvi adaptatiu.

3. És possible dissenyar un model general 3D del desenvolupament que inclogui epiteli, mesènquima i matriu extracel·lular i la majoria de comportaments cel·lulars, i que sigui capaç de reproduir els mecanismes de desenvolupament més bàsics capaços d'una transformació de patró, així com les seves propietats variacionals.

4. El modelatge matemàtic del desenvolupament de la dent prediu que el creixement diferencial entre l'epiteli dental i la capa suprabasal és el responsable de l'aparició dels bucles cervicals, els quals són indispensables per a la formació de la corona dental.

5. El modelatge matemàtic del desenvolupament de la dent prediu que les diferències en les forces mecàniques mediades per l'adhesió diferencial entre les cèl·lules controlen la direcció de creixement dels bubles cervicals, la qual cosa té un efecte directe en el la forma global de la corona dental i per extensió en el patró espacial de les cúspides.

Appendix A

Supplementary Information Chapter 3

A.1 Genotype-phenotype map: Developmental model

The genotype-phenotype map is implemented through a model of tooth morphogenesis. The model mathematically summarizes the basic genetic and cellular interactions experimentally known to be important when the model was build (Salazar-Ciudad and Jernvall, 2010) and the overall logic by which these are coordinated to lead to final adult morphologies. The model has previously been used to produce the final morphology of seal teeth and its natural populational variation in three-dimensions. Thus the model can be taken as a state-of-the-art representation of what is currently understood about the developmental origin of complex morphologies and their variation.

The model involves a set of parameters (n=21) related to physical properties of the cells and tissues, the extracellular signal diffusivity and the strength by which these signals regulate each others expression. The genetic and physical (mechanical) interactions in the model are known to exist experimentally. Any mutation on a gene affecting one of these parameters will lead to a change on the parameter value, so we then take the parameters of the tooth model as a proxy for the genotype.

The model can be run under different parameter values. Its output, or morphology, is the position, in three dimensions (x,y and z) of each cell in the enamel epithelium, that is tooth surface, and the list of which epithelial cells are in direct contact.

All simulations start with the same simple spatial initial conditions: a small flat enamel epithelium ubiquitously expressing BMP4 at low levels (as in Salazar-Ciudad and Jernvall, 2010). It is important to note that phenotypic features like cell number, cusp number or morphology are not predetermined (this is not a geometric model), they just arise from developmental dynamics. For a detailed account of the experimental evidence on which the model is based, for a more detailed description of the model itself and for the actual algorithm and freely available code we refer the reader to the original publication presenting the model(Salazar-Ciudad and Jernvall, 2010).

The tooth model is specially suited for the questions addressed in this work. Most computational models of development focus only on single wild-type morphologies, or their lab mutants, and do not consider microevolutionary level variation (Shvartsman, Muratov, and Lauffenburger, 2002; Jaeger et al., 2004; Honda et al., 2008; Zhu et al., 2010; Hester et al.,

2011; Tamulonis et al., 2011). Most of these models consider only morphological transformations occurring in early or mid development and, thus, not on the morphologies more readily seen by natural selection. There are also several genotype-phenotype map models that are not based on development biology but are used to address valuable evolutionary questions (Kaufmann, 1993b; Wagner, 1994; Hansen and Wagner, 2001; Alvarez-Castro and Carlborg, 2007; Jarvis and Cheverud, 2009; Wagner, 2011; Barton and Turelli, 2004). Some consider only univariate morphologies while others consider gene network-like models in which the level of expression of each gene is treated as a trait. Most of those models do not consider extracellular cell signalling, extracellular signal diffusion in space, space itself or mechanical interactions between cells. All these epigenetic interactions and factors have repeatedly been argued (Newman and Comper, 1990; Newman and Müller, 2000) to confer a very specific set of variational properties and genotype-phenotype maps to actual embryonic development that are not present in mere gene network models (Wagner, 1994; Salazar-Ciudad, Garcia-Fernández, and Solé, 2000). Many of these interactions are, however, included in the tooth model. These other models do not actually propose how the genotype-phenotype map is. They represent, instead, a convenient way to implement a whole class of presumably possible genotype-phenotype mappings without having to precisely specify which of them are actually encountered in real systems and which ones are not. That is very convenient for a number of evolutionary questions (Fontana, 2002; Newman and Müller, 2000; Carter, Hermisson, and Hansen, 2005; Borenstein and Krakauer, 2008; Gjuvsland et al., 2007; Fierst and Hansen, 2010; Kaneko, 2012; Draghi et al., 2010) but for the questions addressed here, however, it is crucial to consider only as-realistic-as-possible genotype-phenotype maps. This is because we are not concerned on whether the map is complex but on how complex it is and on how that level of complexity precludes, or favours, some phenotype-fitness maps in having an important role in morphological evolution.

Another advantage of our model for the questions addressed in here is that the morphological dimensionality of the morphology is not prespecified but arises from model's dynamics. Different combinations of the model parameters lead to morphologies characterized by different number of cells, landmarks and cusps. This is a consequence of having cell proliferation and movement in the model. This does not occur in gene network models that identify each gene's expression as a phenotypic trait. It does not occur either in genetic landscape models that take each loci as being responsible for a distinct phenotypic trait (thus in these models genotypic and phenotypic space have the same dimensionality). Growth in our model produces that the parameter space, our proxy for the genetic space or genotype, can often have less dimensions than the morphological space since each cell accounts for 3 traits, its position in the x,y and z coordinates, and thus 3 dimensions. Not all variation in this space is possible, that is determined by the model dynamics. In the tooth model and in real teeth (Salazar-Ciudad and Jernvall, 2010), for example, the highest cusps are always the first to form and thus are roughly around the center of the tooth. In general, as it can be seen in figure A.12, simple teeth with one or few cusps arise much more easily in the model. That is probably a general property of models with growth and signalling leading to complex morphologies.

Many dimensions of change are possible but most variants tend to cluster in the part of the morphospace with the simplest morphologies. Since there is lateral inhibition in the model cusps can not form very close to each other. Thus, there are relatively few parameter combinations that lead to many cusps in a tooth: the inhibitor activity, diffusivity and its effect on the activator have to be low but not to low for cusps to fuse and at the same time the growth parameters have to be such that cusps are separated by enough growth in between them. In general it is likely that producing progressively more complex teeth requires a progressively more accurate fine-tuning of the model parameter values.

A.2 Evolutionary model

The evolutionary model (see figure A.13) starts with a constant sized, genetically homogeneous population. Then random mutations are generated in each generation at a rate μ (probability of mutation per individual; double mutations occur with probability μ^2). For each different genotype in each generation, the morphology is calculated through the tooth developmental model. Next, in each generation an absolute fitness is calculated for each individual morphology accordingly to how closely that individual morphology resembles that of an arbitrary optimal morphology. A different optimal morphology is used in different simulations and that optimum does not change over generations (each simulation between each initial and optimal morphologies was repeated three times with different random number seeds). Three different types of criteria or phenotype-fitness maps are used: the EMD criterion, landmark-based criteria (raw for 2 to 5 landmarks corresponding to 2,4,8 and 13 traits respectively and PCA-based for 5 landmarks) and the OPC criterion (the same single criterion is used through each simulation). Finally each individual fitness is used to determine which individuals in a generation contribute to the next generation. Reproduction is implemented by choosing, for each individual of the next generation, which individual of the current generation will be his parent. The probability of being chosen is equal to the individual's relative fitness. This way one individual, especially if it has high relative fitness, can be the parent of more than one individual in the next generation. This way also, there is genetic drift, but individuals with a higher relative fitness have a higher probability of having descendants in each generation (see figure A.1).

A size limit of 600 cells was put on the tooth development model in order to prevent the development of too large teeth, since teeth have a limited space in the jaw to develop. A minimum size was also established (50 cells). Thus, teeth having more than 600 cells or fewer than 50 were considered non-viable and a 0 fitness was assigned to them. All simulations of evolution were run with the same mutation rate, that is $\mu = 10^{-3}\, mutations\ individual^{-1}\ generation^{-1}$.

Random mutations are generated by selecting one of the 22 parameters of the tooth model at random and modifying its value k_j in the following way:

$$k'_j = k_j + k_j M \tag{A.1}$$

where k'_j is the "mutated" value of the parameter j, k_j is the original value and M is a random variable with a uniform distribution between [-0.1 , 0.1].

We use a fitness function that is linear with the phenotypic distance of each individual i, d_i, to the optimal morphology. The way this distance is calculated is different for each selection criterion (d_{EMD}, d_{lan} or d_{OPC}), as explained above, see below. The absolute fitness W_i of an individual i is:

$$W_i = \Phi\left(-\frac{sd_i}{d_0}\right) + 1 \tag{A.2}$$

Where Φ is a Heaviside function ($\Phi = 0$ if $x < 0$ and $\Phi = x$ otherwise), s is the selection coefficient and d_0 is the phenotypical distance between the morphology at the beginning of the simulation and the optimum. s determines the slope of the fitness function. For most of the simulations we used an s of 0.5, thus the absolute fitness of the initial morphology is 0.5 and the maximum fitness (when the population reaches the optimum) is 1. However, in further discussions and in figures we scale absolute fitness as the percentage of fitness attained relative to the difference between initial and maximum absolute fitness (from 0 to 100%). Other simulations were run with different values of s (see section A.10 and figure A.14). Note that in this model we select for a specific optimal morphology, and not for increases or decreases in a set of traits. Thus, absolute fitness has a maximal value of 1.

The relative fitness for each individual (w_i) is:

$$w_i = \frac{W_i}{\sum_{j=1}^{P} W_j} \tag{A.3}$$

A.3 Simulation dynamics and stagnancy criterion

A simulation was considered to attain the optimum when the most frequent morphology on the population had a d of 0 for OPC, equal or less than 6% for EMD and less or equal than 5% for the landmarks based criterion.

In many simulations the optimal morphology is not reached. The population, instead, settles in a non-optimal morphology, what we call a stasis morphology. The stagnancy criterion, as described below was used to decide when to stop those simulations.

The evolution of populational average fitness in our simulation can be described as periods of change, what we call adaptive leaps, and periods of stasis (see figure A.1).

In each generation, during a stasis period, we record the total number of mutants, m_t, that have arisen since the last adaptive leap and the number of those that are fitter than the stasis morphology, m_f. Most of the fitter mutants represent such a small increase in fitness that they are lost by genetic drift. In each generation we calculate r_m :

$$r_m = \frac{\sum |d_j - d_s|}{m_t} \tag{A.4}$$

Where d_j is distance between a mutant morphology and the optimum and d_s is the distance between the stasis morphology and the optimum.

For that calculation only mutants that are fitter than the stasis morphology are considered. r_m is calculated at every generation as long as m_t is equal or greater than 10000. Trials were made to assess that a minimum of 10000 random mutations was enough to explore all the possible mutant morphologies that could arise from a stasis morphology before the stagnancy criterion was applied. If r_m is lower than $3 \cdot 10^{-6}$ the stagnancy criterion is satisfied and the simulation is stopped. With this r_m value we ensure that the simulation does not stop if there is at least one mutant (for every 1000 mutants tried) that has a distance to the optimum that is 0,3% larger than that of the stasis morphology to the optimum. If an adaptive leap event happens, m_t and m_f are reset to 0 and the counting starts again.

Due to computational time constraints, we ran only a set of simulations (n=24 for EMD, n=48 for OPC, n=40 for the landmark-based simulations) using the stagnancy criterion. The rest of the simulations (n=321 for EMD, n=570 for OPC, half for OPC increase and half for decrease, n=153 for the landmark-based simulations) were run for 20000 generations and the ones with small population sizes (10 and 100) were run for 50000 generations. We checked all simulations individually in order to be sure that there weren't significant increases of fitness since long before the simulation was stopped.

A.4 Optimal phenotypes

Twenty different teeth were chosen as initial morphologies for the evolution simulations. The initial teeth was chosen to belong to four idealized morphological types (see figure A.11) that represent simple and common cusp spatial arrangements arising in the model. In the case of the landmark based criterion we only used type 4 teeth (see figure A.11) as initial morphologies because they are the only ones that have enough cusps to allow to define 5 landmarks.

We wanted to ensure that the optima are not developmentally impossible morphologies: that they are within the variational properties of the developmental model. This is because the purpose of this study is to assess the effect of a realistic genotype-phenotype map on adaptive dynamics, not the variational properties of the model itself. For this purpose, random mutants were created from each initial morphology. Then EMD, OPC and landmark-based phenotypic distances (for 2, 4, 8 and 13 traits) were calculated between each initial morphology and their mutants. As optimal morphologies for the OPC and EMD criterion we chose one parameter mutants that were at a distance of 20% and 40%. For 60% distances the optima had to be chosen among mutants in 5 parameters (since no phenotypes at these distances could be found by less mutations), and for distances of 80% among mutants in 10 parameters. For the PCA landmark-based criterion the optima were set by taking the scores of the initial phenotype and increasing or decreasing scores of all the PCs equally so the final phenotypic distance between the initial and the optimal phenotype was 40% or 80% (we also checked that these morphologies were possible from the developmental model).

A.5 Mutational screening

We used the following method to generate random phenotypic variation and measure some properties of the genotype-phenotype map and phenotype-fitness maps. Each initial morphology were taken as a 'parent', and a certain number of random mutants were generated from them. Each mutant was generated as in the evolutionary simulations (See section A.2, equation A.1), changing one or several of the model parameters of the parent at once. Then we could take measures on each mutant phenotype and compare them to the parent phenotype, or to an arbitrary optimum. This way we could also compare the model parameter differences between parents and mutants.

A.6 Distribution of complex phenotypes in the tooth model parameter space

Two approaches were used to assess how complexity was distributed on the parameter space of the tooth development model.

The first is a global sampling of the parameter space. This provides a general estimation of how likely are complex and simple morphologies from different zones of the parameter space. We took all the initial genotypes (n=25) used in all previous simulations and recorded the maximum and minimum value of each parameter. Then the range formed by the minimum and maximum value for each parameter was doubled and these ranges were used to delimit the parameter ranges for the global sampling of the parameter space. We created 300 random combinations of the 22 developmental parameters. For each combination of parameters the tooth model was run and the complexity was measured. We used OPC also as a measure of morphological complexity (as in Evans et al., 2007).

The second approach is a local sampling of the parameter space, done by means of a one parameter mutational screening (See section A.5), using all the initial morphologies that were used in our evolutionary simulations (25), which have OPC values ranging from 4 to 30. 300 mutants per parent phenotype were generated. The complexity of the mutant in relation to the complexity of the parent morphology was recorded (see figure A.12).

A.7 Measurement of the phenotype-fitness map degeneracy

In order to assess the properties of the different phenotype-fitness maps used in this study we performed one parameter and five parameters mutational screenings (see Section A.5) on the 25 initial morphologies used in the selection experiments. For each mutant, we calculated the genetic distance to the parent, d_{gen} as,

$$d_gen = \sqrt{\sum_i = 1^n_{par} \left(\frac{k_i - K_i}{K_i} \right)} \tag{A.5}$$

where n_{par} is the number of parameters of the tooth model, k_i is the value of parameter i in the mutant morphology and K_i is the value of the

same parameter in the parent. We also calculated the phenotypic distances on the bases of the phenotypic similarity between mutant and parent,with the EMD (d_{EMD}), the raw trait landmark-based (d_{lan}) and the OPC criteria (d_{OPC}). We then calculated the absolute fitness for each mutant taking its parent as the optimal morphology. Fitness is 0 when the phenotypic distance is 100%. Figure 3.3A, B show the distribution of fitness for EMD and OPC, respectively, as a function of the genetic distance to the parent. The correlation coefficient ($r^2 = 0,3549$ for EMD, $r^2 = 0,06082$ for OPC) gives an idea of the degree of degeneracy of each phenotype-fitness mapping, the lower it is, the more morphologies at different genetic distances can be at the same or similar phenotypic distance to the parent. Figure 3.3C, D shows the same results for the many and few traits landmark-based criteria.

Figure A.8A shows the d_{EMD} plotted against d_{OPC} of each mutant morphology in relation to the parent morphology. The degeneracy of the OPC phenotype-fitness map can also be seen here, as morphologies at the same distance in terms of OPC can be at quite disparate EMD distances from the parent.

In order to see how the different phenotype-fitness maps affect the topography of the whole genotype-fitness landscape, we took the phenotypes from the mutational screening, chose one of them as an arbitrary optimum (at 20%, 40% and 60% distance from the parent) and calculated the fitness of all of them in relation to that optima (Figure A.10).

A.8 Analysis of neutral networks in different phenotype-fitness maps

In order to further assess the degree of degeneracy of each phenotype-fitness map we performed a set of neutral-walk experiments similar to those in (Schuster et al., 1994). In each experiment we start with a parent morphology and submit it to successive rounds of mutation. At each round we mutate one of its parameters at random (see equation A.1) and evaluate the phenotypic distance to the initial morphology with one of the phenotype-fitness maps used in this study (d_{EMD} ,d_{OPC},d_{lan13} or d_{lan2}). If the distance of the mutant to the parent is lower than a certain threshold value we keep it and mutate it on the next round of mutation, if not we discard the mutant, keep the previous one and mutate that in the next round. We performed each walk for 100 rounds of mutation and then the genetic distance between the final and the initial morphologies was taken as a measure of how far away in the genetic space the walk proceeded. The threshold values for the EMD walks was of 6% and 1%. For the landmark-based walks the thresholds were of 5% and 1%. The threshold for the OPC walks was 0, that is the mutant had to have the same OPC value than the parent phenotype. For the EMD and OPC criteria 5 initial morphologies ranging from simple unicuspid teeth to complex multicuspid ones were used. 10 independent walks of 100 steps each were performed per initial morphology (figure A.9A). One initial morphology was used for the landmark-based criterion looking at different numbers of traits (2, 4, 8 and 13). 10 independent walks were performed for each number of traits.

A.9 Evaluation of final fitness of the simulations in terms of all selection criteria

In order to understand the relationships between the different phenotype-fitness maps used in this study we took each selection experiment with each phenotype-fitness map and calculated the fitness achieved at the end of the simulation in terms of the other fitness criteria. Thus, we evaluated the OPC-fitness in EMD simulations and the EMD-fitness of the OPC simulations (figure A.6A). We also evaluated the OPC-fitness and the EMD-fitness of the many trait and few trait landmark-based simulations (figure A.6B). We could not evaluate the landmark-based fitness of the EMD and OPC simulations because the landmark-based criterion can only be applied to a particular morphological category (see figure A.11, category 4) with enough number of cusps. To evaluate the OPC-based fitness of the EMD simulations we simply calculated the phenotypic distance, d_{OPC}, as in equation 3.1 (see section 3.4.5), taking the OPC of the initial phenotype as the initial OPC value and the OPC of the optimal morphology as the optimal OPC value. To evaluate the EMD-based fitness of the OPC simulations, as no specific morphology was chosen as optimum, we chose a random mutant (1 parameter changed) from the initial phenotype that had the optimal OPC value for that simulation and took it as an arbitrary optimum for the EMD criterion to be calculated. Then the phenotypic distance, dEMD, was calculated as in equation 1 and 2 (see section 3.4.1) taking this arbitrary morphology as the optimum for the EMD criterion. To evaluate the EMD-based fitness of the landmark-based simulations, as the optimum in these ones is a specific morphology, we took it as the optimum for calculating the phenotypic distance, d_{EMD}, as in equation 1 and 2 (see section 3.4.1). To evaluate the OPC-based fitness in the landmark-based simulations we did the same as described above for the EMD simulations.

A.10 Variation in the coefficient of selection

We ran a small set of simulations using the EMD and the raw landmark-based criteria with different values of the selection coefficient, s (see section A.2, equation A.2). Some were run with a smaller coefficient (s=1/4) and others with a greater one (s=4/3 and s=1). Selection coefficients of 1/4 and 1/8 made drift so important as to drive change and lead to the fixation of greatly maladaptive phenotypes, so we didn't try smaller selection coefficients. Population size was set to 100 and 20 different initial phenotypes were used. Optima at 40% phenotypic distance were used. Figure A.14 shows the mean absolute fitness achieved with different values of s. Note that in this figure, as in all others, absolute fitness is scaled as described in section A.2, that is all fitness values range between 0 and 100, being 0 the fitness of the initial condition and 100 the fitness at the optimum. There are no significant differences between the fitness attained with coefficients of selection ranging from 1/4 to 1 (Kruskal-Wallis test for EMD, n=160, K=0,7824, p=0,8537; many traits, n=75 K=0,4963, p=0,9196; few traits, n=75, K=0,3507, p=0,9502).

These results are not surprising since s has an effect on how fast a population can climb a fitness peak but not on the relative height of peaks (that

latter thing being what determines the final fitness). That relative height is a property of the shape of the adaptive landscape itself and thus it comes from the shape of the genotype-phenotype and phenotype-fitness maps. Note that a population would tend to climb towards the steepest of the fitness peaks near its location in the landscape but that in complex landscape, like in figure 3.4A, the steepest peak is not necessarily the tallest (and thus s has no effect on adaptation in the EMD and many landmark simulations).

A.11 PCA-based criterion. Response of single PCs to selection

The average response to selection for each PC in the PCA-based simulations is shown in figure A.5A. The response to selection is calculated as in equation A.2 for each PC, using the distance between the score at the end of the simulation and the optimal score and s=1. Note that a negative response means that the distance at the end of the simulation is greater than the initial distance. The response on higher rank order PCs (1 and 2) is significantly higher than in lower rank order ones (3 to 8) (Mann-Whitney test pairwise comparisons between PC1 and PC3 to PC8 gives a p<0,05, the same for comparisons between PC2 and PC3 to PC8; Mann-Whitney test between PC1 and PC2 gives a p=0,9037). Figure A.5B shows average population fitness and average response to selection for the eight PCs over simulation time (in generations) for a single simulation. Sudden changes in fitness correspond to the fixation of an adaptive variant in the population, which correlates with changes in several PCs. These changes normally decrease the distance to the optimum in some PC and increases it in others. This produces substantial variation in the level of adaptation in each PC in the different simulations (as seen in figure A.5A). This means that, while a PCA defines dimensions of variation that do not statistically covary between them in the sample population, that doesn't imply that there is a simple and modular genotype to phenotype relationship beneath them. That is, most mutations will affect the scores of several PCs at once (simply the PCA can not grasp all the complexity of the genotype-phenotype map).

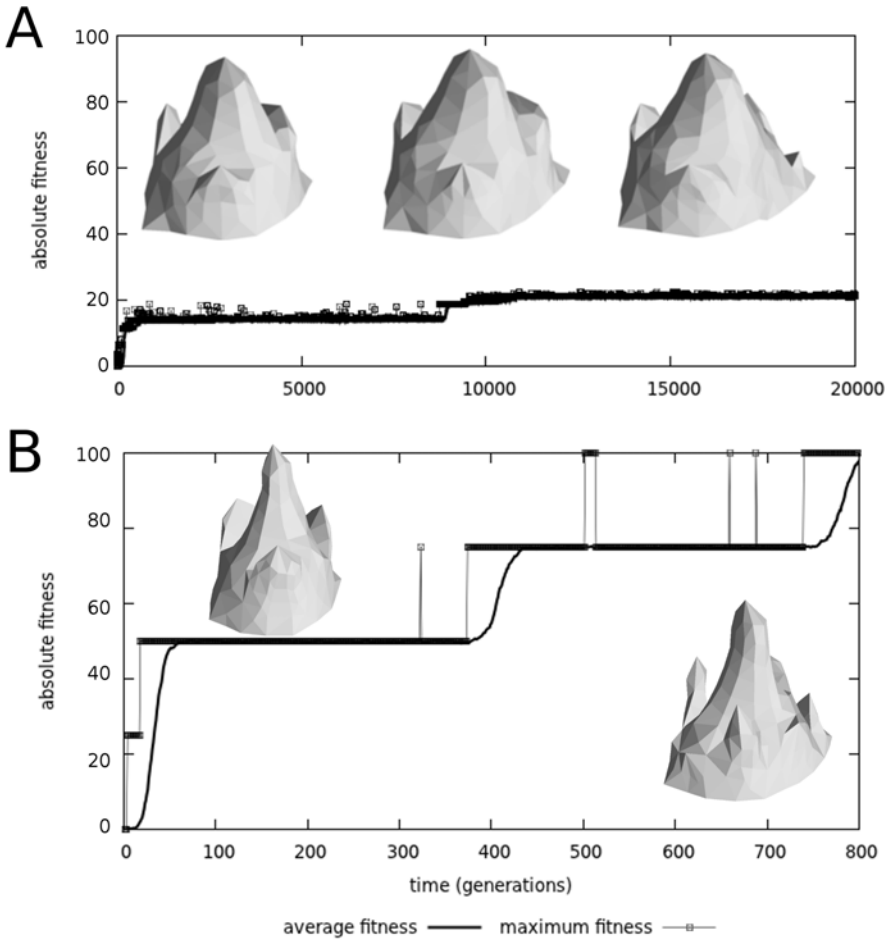

FIGURE A.1: Long periods of stasis and short periods of adaptive leap are observed. A, simulation using the EMD criterion selecting for an optimum at a distance of 40%. From left to right, the initial, final and optimal morphology are depicted respectively. B, simulation using the OPC criterion selecting for a higher optimum at a distance of 40%. From left to right, the initial morphology and the final morphology are depicted respectively. Note that the optimum is characterized by a single OPC value and that there are many morphologies that can have the same OPC.

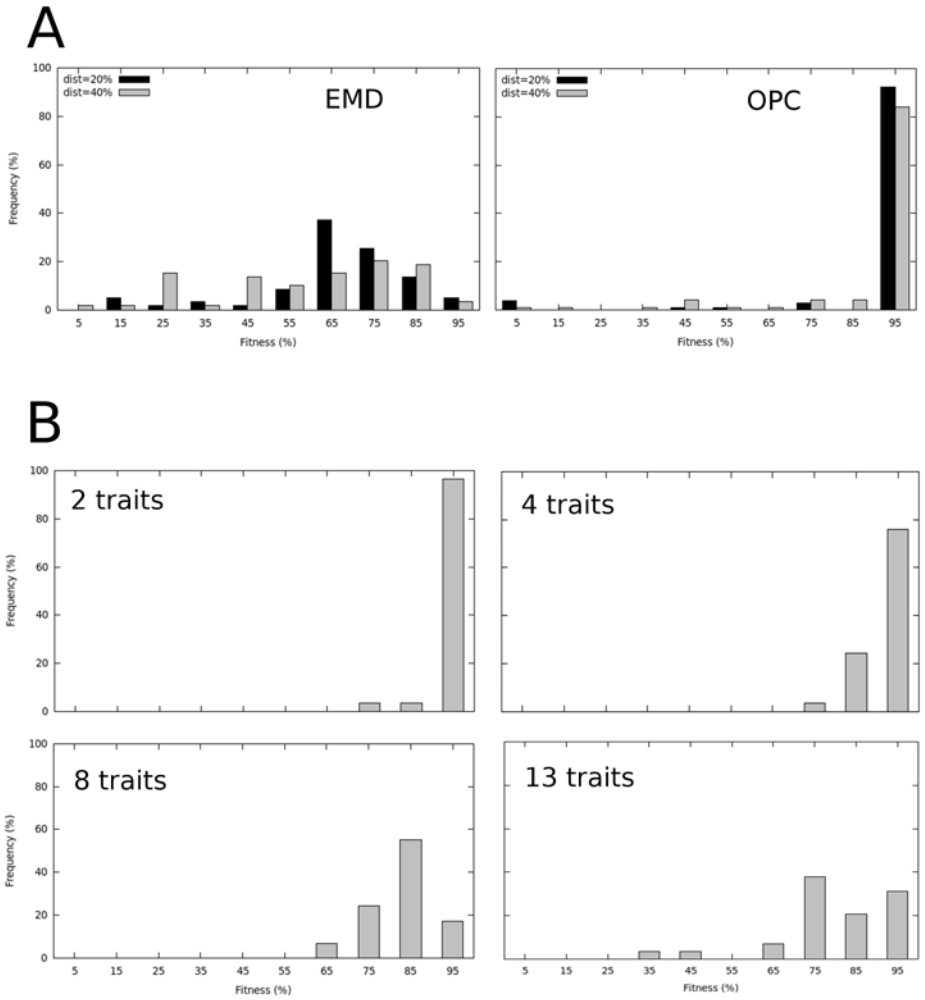

FIGURE A.2: A, the proportion of OPC simulations that arrive to the optimum is significantly higher than the proportion of EMD simulations, being the former close to 100%. It is also easier to arrive to the optimum when the distance from the initial morphology to the optimum is 20% than when it is 40%. B, The same for the raw landmark criteria simulations. When selection acts on few traits (i.e. 2 and 4) almost all simulations reach the optimum. When selection acts on a larger number of traits (8 and 13) most simulations remain trapped at a distance from it.

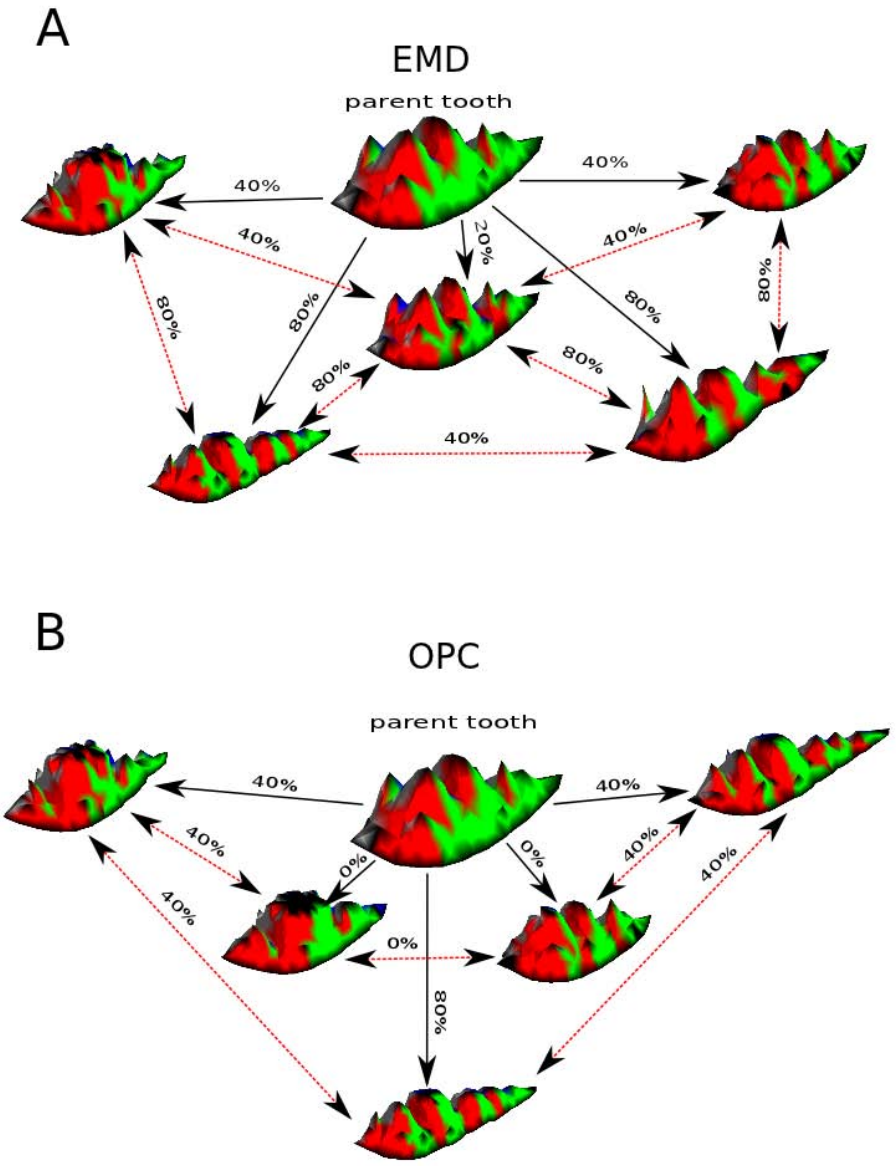

FIGURE A.3: A, distance relationships using the EMD criterion between some mutants and their parent morphology (in the center). These morphologies can be very different so the EMD between them may be large. This means that the distance between a morphology at a distance of 80% from the optimum and a morphology at a distance of 40% from the optimum need not be 40% and it rarely is. B, distance relationships using the OPC criterion (relative increases of the OPC value). In this case, morphologies at the same distance to the optimum are at a distance of 0 between them and morphologies at a distance of 80% and 40% from the optimum respectively are exactly at a distance of 40% between them.

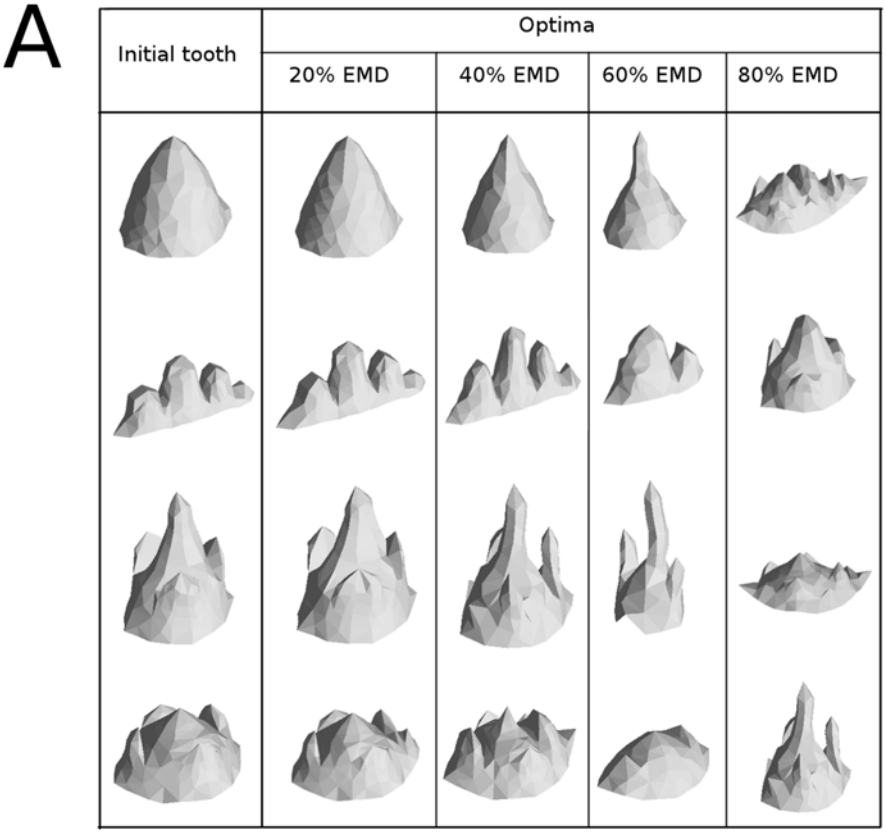

FIGURE A.4: Note that the distance is between each morphology in the left most column and morphologies in each column and that it is not implied, and it would rarely be the case, that morphologies in a column are at a proportional distance to morphologies in a column other that the left most. For example morphologies in the 20% EMD column are at 20% EMD to the left most column but not to the column to their right, the 40% EMD column (as it should be clear from figure A.3). Thus, that optimal adaptation can occur for optima at 20% and 40% EMD from an initial morphology does not imply that adaptation to more distant optima (lets say at morphology B at 80% EMD from that initial morphology) can occur by selecting for successive optima at 20% EMD from each other (as the artificial linear arrangement in this figure may seem to indicate).

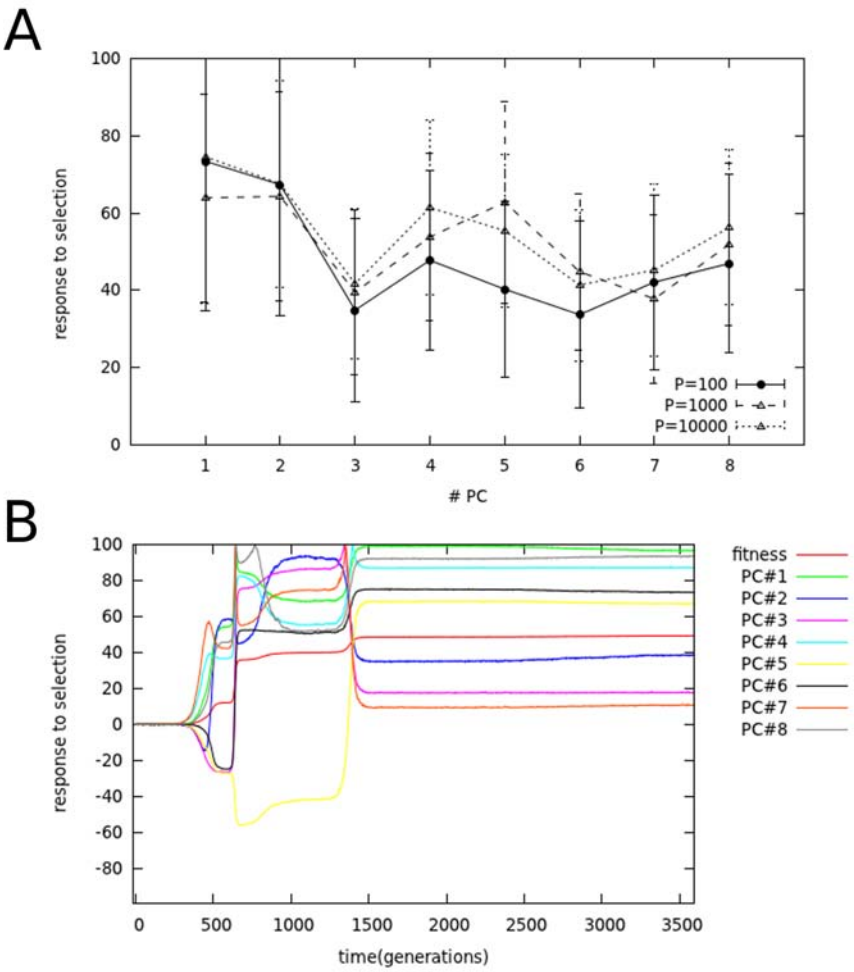

FIGURE A.5: Response to selection is calculated as the proportion of the distance between the initial score value and the optimal value attained at the end of the simulation (see section 11 Appendix A), which is 0 if the final value is the same as the initial value and 100 if is equal to the optimal value. A, average response to selection for each PC for a population size of 100, 1000 and 10000. B, average response to selection in an example simulation (population size = 10000) over simulation time (generations), average absolute fitness is also plotted. In this figure it can be seen that there are a number of adaptive leaps consisting in changes in the distance from a PC score and its optimal value. An increase in one PC tends to lead to decreases in the other PCs (although the overall fitness increases the adaptation at each PC does not increase linearly with time).

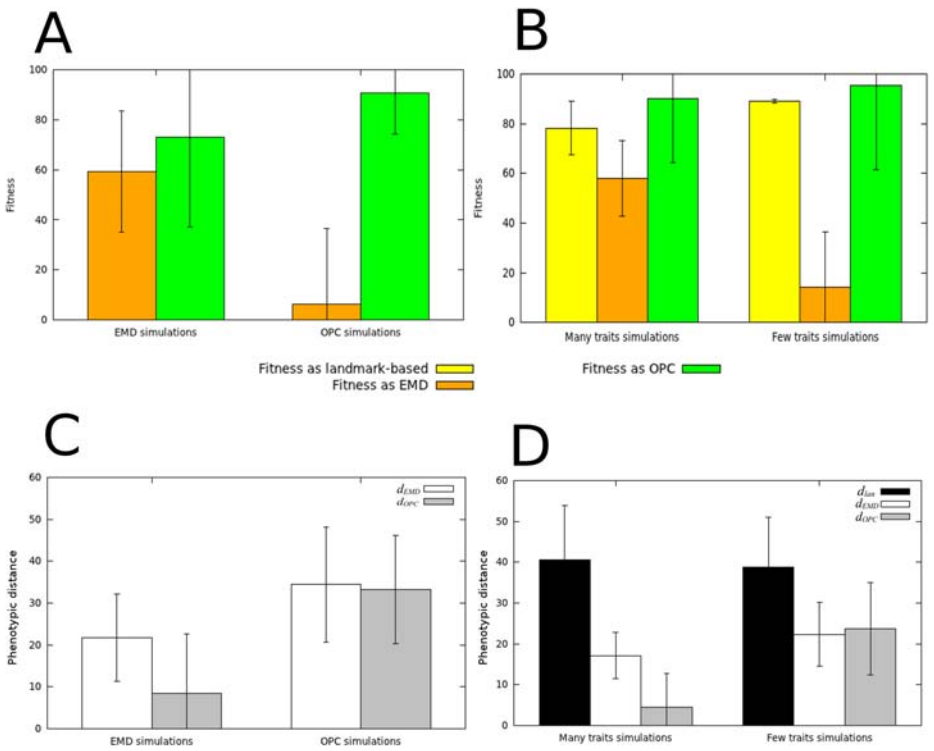

FIGURE A.6: Each cluster of columns corresponds to the set of simulations run under a certain phenotype-fitness map. A, B, evaluation in terms of fitness (see section 9 of Appendix A). Each column within a cluster indicates the average fitness attained when evaluated for a certain phenotype-fitness map (not necessarily the one that was actually used during the simulation). A, simulations selecting for EMD and OPC, evaluated only for EMD and OPC. B, simulations selecting for many and few landmark-based traits, evaluated for landmark-based distances, EMD and OPC. C, D, The same but evaluated in terms of relative amount of phenotypic change at the end of the simulation. This is measured as the phenotypic distance (d_{EMD}, d_{OPC} and d_{lan}) between the final phenotype and the initial phenotype. C, simulations selecting for EMD and OPC, evaluated only for d_{EMD} and d_{OPC}. D, simulations selecting for many and few landmark-based traits, evaluated for d_{lan}, d_{EMD} and d_{OPC}. The fitness values attained for the different criteria (A,B) in the EMD and many traits simulations are similar, because the number of traits under selection is high. In contrast, the EMD-fitness for the OPC and few traits simulations is low, because there are few traits or no specific traits at all (OPC) under selection and thus there are lots of morphologies equally adaptive other than an specific morphological optimum. The amount of change in terms of OPC is low in the EMD and many traits simulations because teeth at a given EMD distance may, or may not, be different in the number and orientation of cusps and thus, may , or may not, be different in the OPC. In fact, of the many morphological differences that give a given EMD only some of them change the OPC.

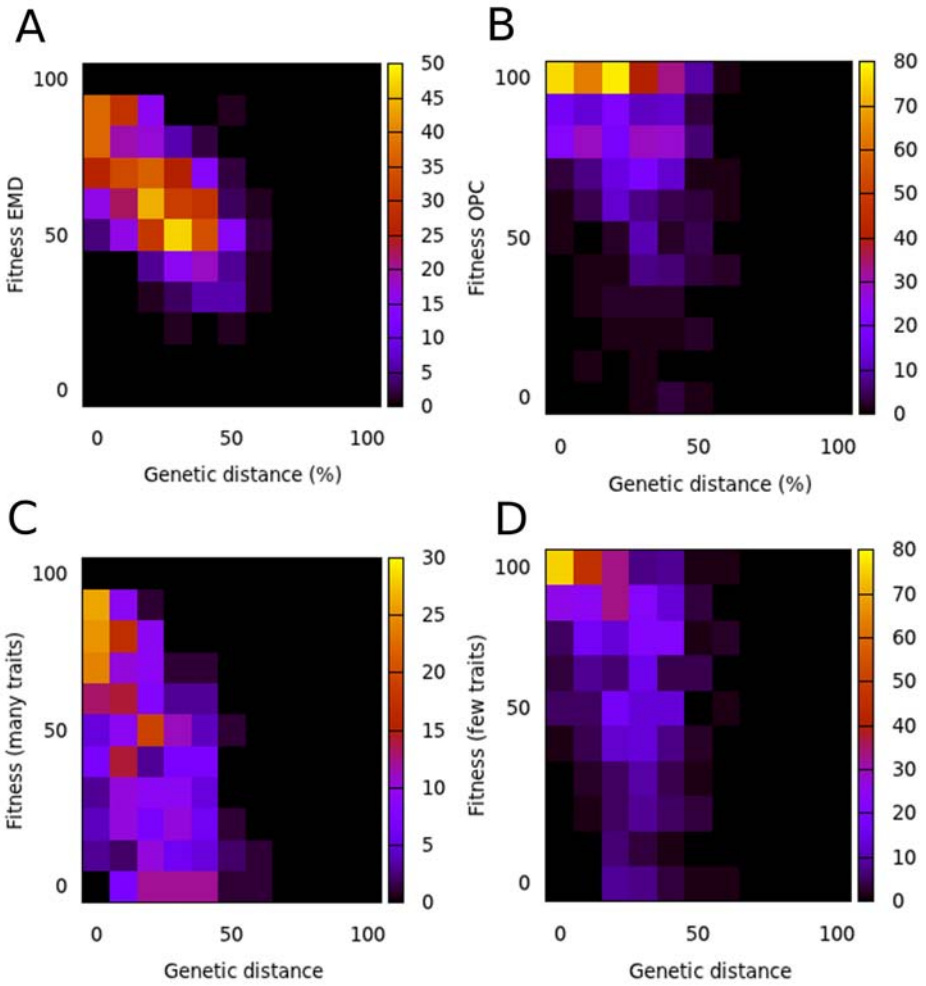

FIGURE A.7: The cartesian space is divided into a regular square grid and the color of each square cell corresponds to the number of data points falling within it. This way it is easier to see that a great portion of mutants have the same OPC value than its parent morphology (that is 100% fitness in B) even if there's a long genetic distance between them, showing a high degeneracy of the OPC phenotype-fitness map.

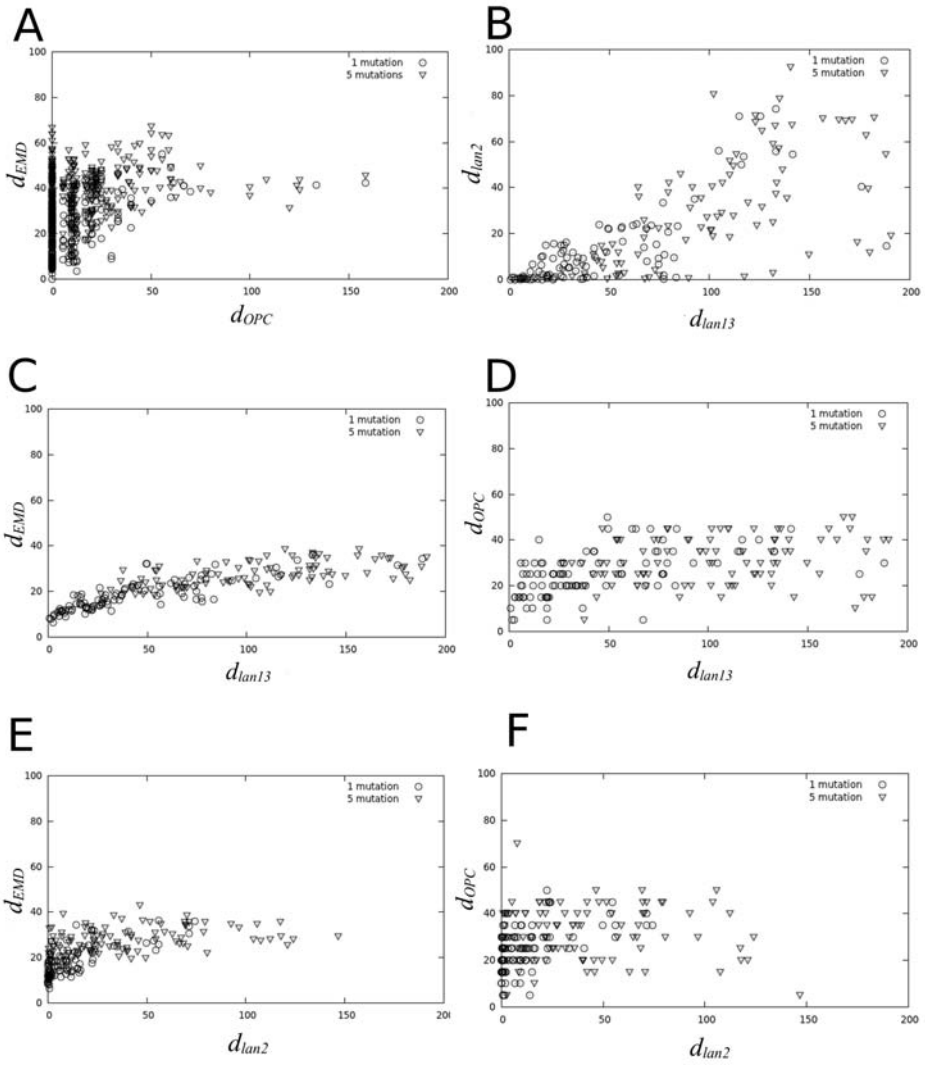

FIGURE A.8: A, EMD phenotypic distance (d_{EMD}) against OPC phenotypic distance (d_{OPC}) (n=670). B, many-traits (d_{lan13}) phenotypic distance against few-traits (d_{lan2}) phenotypic distance (n=100). C, d_{lan2} against d_{EMD} (n=100). D, d_{lan2} against d_{OPC} (n=100). E, d_{lan13} against d_{EMD} (n=100). F, d_{lan13} against d_{OPC} (n=100). The degeneracy of the OPC phenotype-fitness map can also be seen here, as morphologies at the same distance in terms of OPC can be at quite disparate EMD distances from the parent. Note that several morphologies, even if they suffered 5 mutations, are at an OPC distance of 0, that is, have the same OPC as the parent . This means the OPC phenotype-fitness map creates a neutral network of genotypes that may lead to rather disparate morphologies in terms of EMD. A similar thing occurs for the relationship between OPC and the landmark distances. This does not occur for the relationship between EMD and the landmark distances. In this case there is a relatively high linear correlation between these measures (specially for the many-landmarks distances).

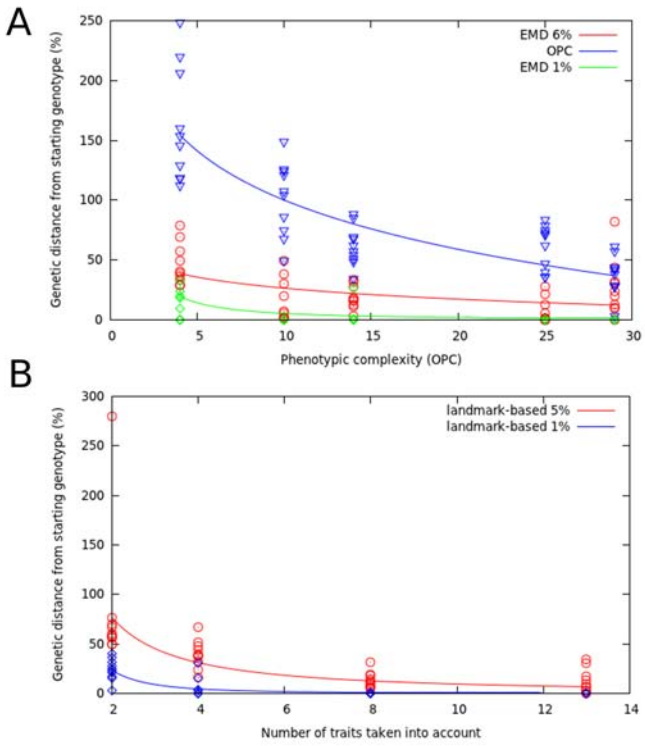

FIGURE A.9: A, using the EMD and OPC criterion, B using the raw-traits landmark-based criterion (see section 3.4.3). Final genetic distance from starting genotype is shown relative to the OPC (used as a proxy of tooth complexity) (A), or number of traits taken into account (B). Each point represents a neutral walk simulation (n=50 for EMD, n=50 for OPC and n=40 for landmarks-based), solid lines are logarithmic regressions fit to the sample points. For the OPC criterion, the walk was done along phenotypes of the same OPC value. For the EMD and landmark-based criteria, the walk was done along phenotypes at a distance equal or less than 5% or 1% from the initial phenotypes. It can be readily seen that the neutral walks by the OPC criterion go much further than the rest. This indicates that the this phenotype-fitness map is degenerate while the others, where only small distances are attained, are not.

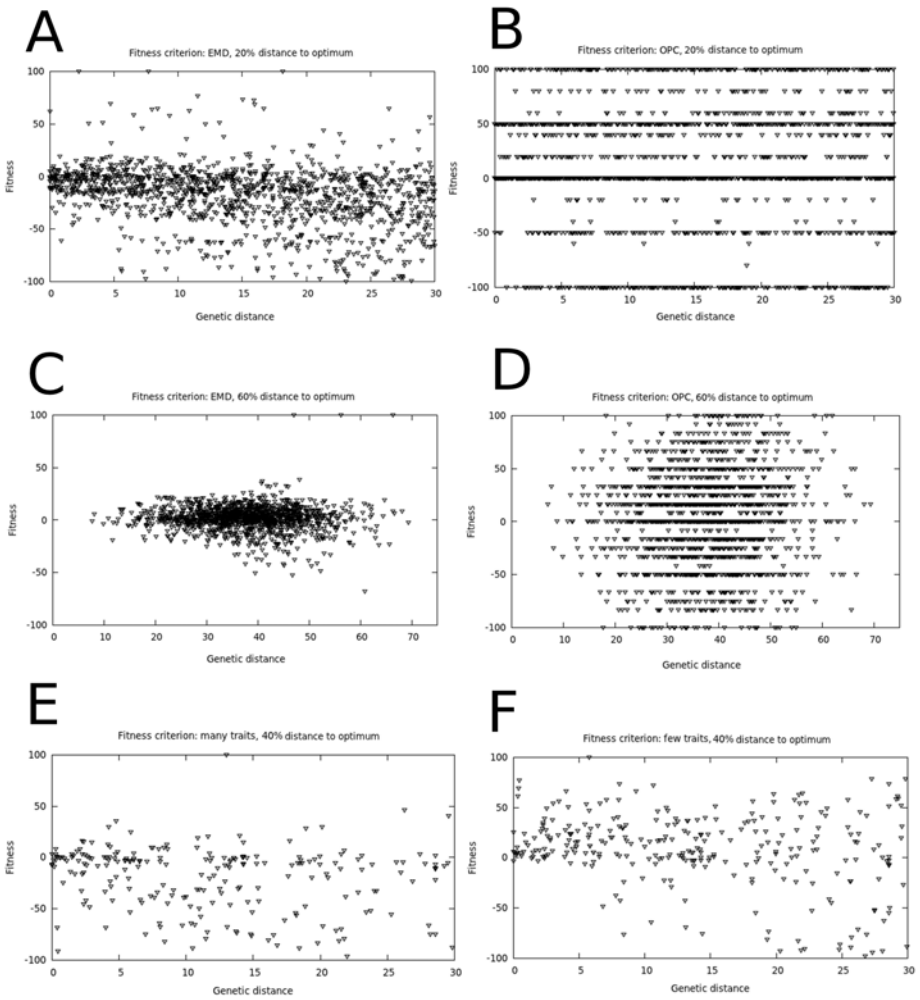

FIGURE A.10: Fitness is measured using the EMD criterion (A, C), the OPC criterion (B, D), many traits (E) or few traits (F). Optima at different distances were chosen: at 20% (A,B) and 60% (C,D) for OPC and EMD and 40% (E,F) for landmark-based. For optima at 20 and 40% mutants with one parameter changed (n=500 for OPC and EMD, n=300 for landmark-based). A fitness value of 0 is assigned to the phenotypes that are at the same distance to the optimum than the initial phenotype, and negative fitness if they are further from the optimum. It can be readily seen that for the OPC criterion there are some mutants that are in the optimum already for short genetic distances from the parent. This is also the case, to some extent, for the few landmarks criteria. It can be readily seen that this is not the case for the EMD and many landmark criteria.

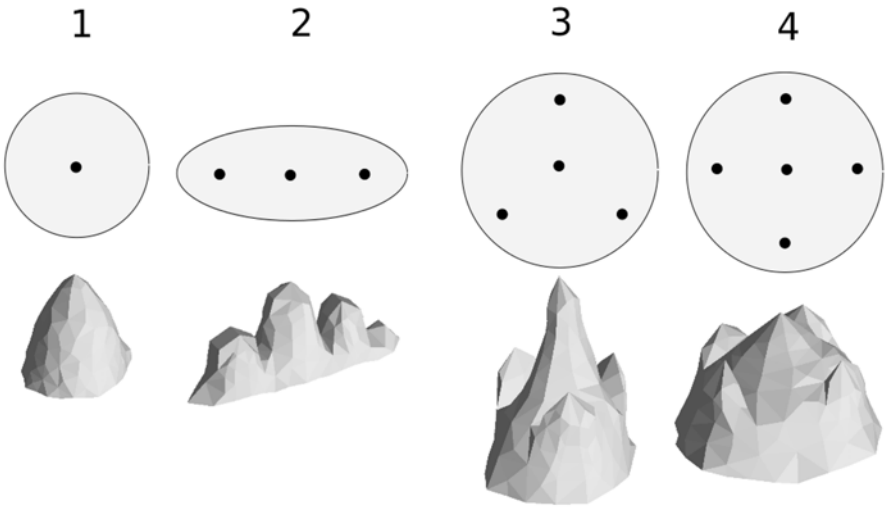

FIGURE A.11: Category 1 includes unicuspid teeth, category 2 three-cusped teeth in which cusps are roughly arranged in a straight line when projected in the x-y plane, category 3 includes four-cusped teeth roughly arranged in a equilateral triangle when projected in the x-y plane (the tallest cusp being in the middle), category 4 includes five-cusped teeth roughly arranged in a cross when projected in the x-y plane (the tallest cusp being in the middle) These differ in the number of tooth cusps and the way cusps are positioned around the central cusp. The upper schemes represent an upper view of the , the black dots represent the tips of the cusps. These categories are the most frequent morphologies produced by the model and thus a good representation of a big region of the tooth model available morphospace.

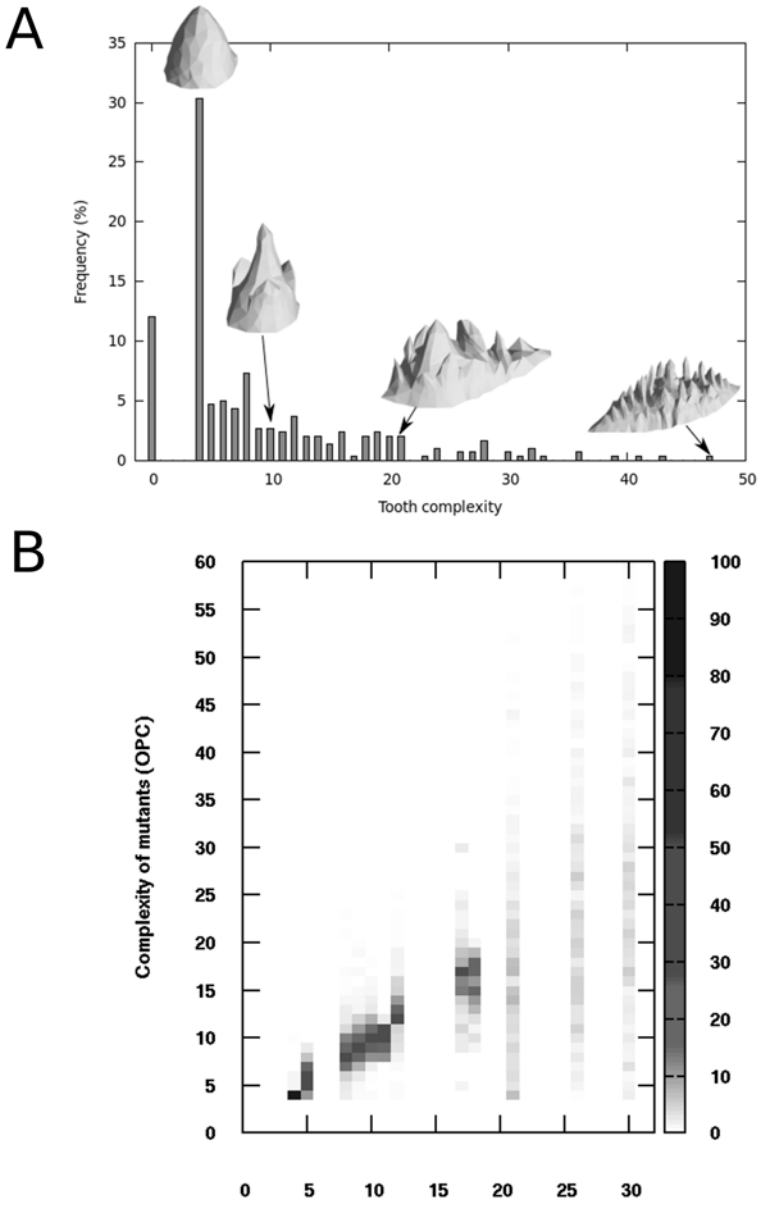

FIGURE A.12: A, global sampling of the tooth model parameter space: the plot shows the frequency of morphologies of different OPC values when giving random values to the tooth model parameters. Most of the morphologies arising from the model are simple, 30% of the sample are one cusp morphologies. In other words, there is a great number of combinations of parameters that give rise to low OPC morphologies and only a small number of combinations that result in high OPC morphologies. B, Local exploration of the tooth model parameter space. Morphologies of different OPC were taken as parents (X axis) and mutant offspring were generated from them, recording their OPC (Y axis). The shade in each point represents the frequency of mutant offspring of a certain OPC originated from a parent of a certain OPC (in boxes in the x-y plane). The diagonal line indicates the points where the OPC of mutants and parent morphologies are the same. For low OPC morphologies we see that most mutants have an OPC equal or similar to that of their parents. In contrast, high OPC morphologies give rise to offspring with a large range of OPC values (both higher and lower).

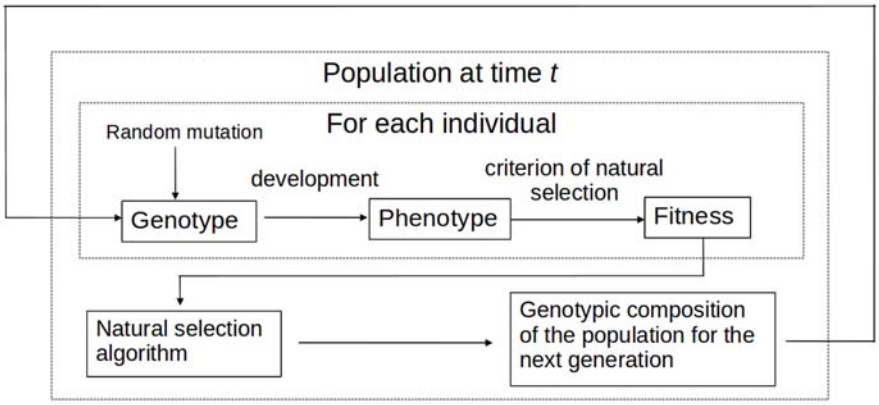

FIGURE A.13: Then the morphology is calculated from the genotype by the tooth development model. A criterion of natural selection evaluates the morphology and assigns an absolute fitness value to the individual. When this is done for all individuals, a stochastic selection algorithm that takes into account the relative fitness values of all the individuals determines, for each individual of the next generation which individual of the current generation will be its parent. The probability of being chosen as parent depends on the individual's relative fitness.

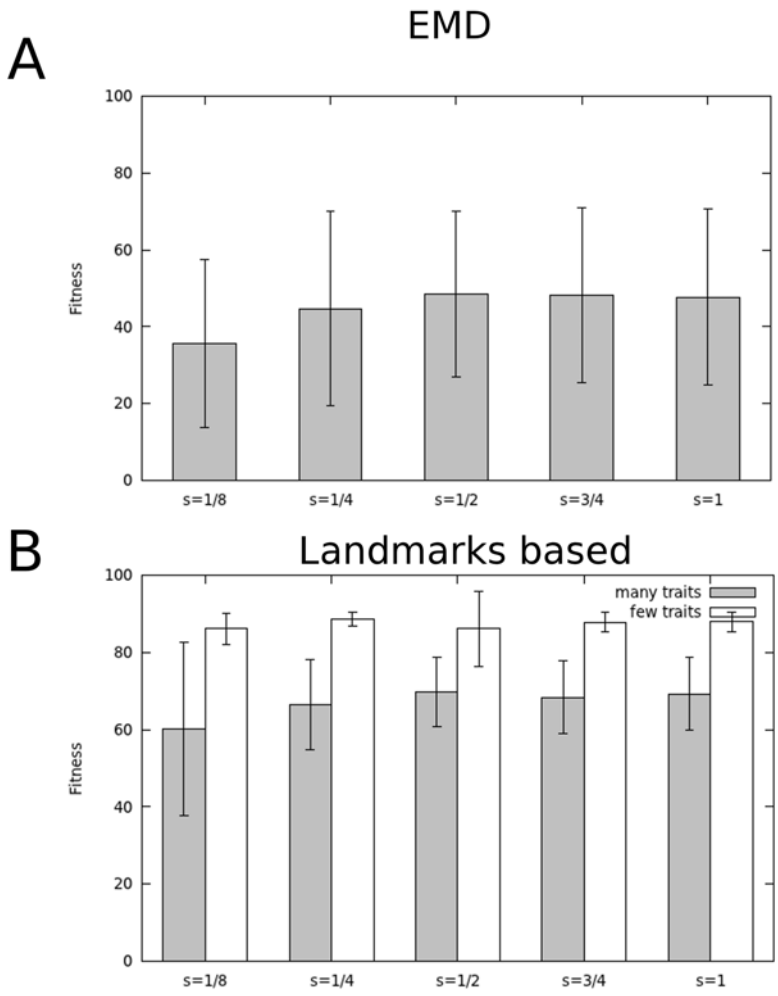

FIGURE A.14: A, simulations using the EMD criterion (n=61). B, simulations using the many traits and few traits landmark-based criteria (n=84). Simulations were run with a population size of 100, initial distance to the optimum of 20% and 40% for EMD simulations and of 40% for landmark-based simulations. The different s values were s=1/4, s=1/2, s=4/3 and s=1.

Appendix B

Supplementary Information Chapter 4

B.1 Introduction

This model considers cells, cell parts, extra-cellular matrix (ECM), gene products and other molecules involved in gene regulation. The model includes in a unified framework bio-mechanical and gene product interactions in development. The model also includes all basic cell behaviours known in animal cells (Salazar-Ciudad, Jernvall, and Newman, 2003). These are: cell growth (polar and non-polar), cell division (directed and non-directed), apoptosis, secretion of ECM and signalling molecules, reception of extracellular signals, cell contraction, cell adhesion and movement and shape change as a consequence of those.

The model includes mesenchymal cells, epithelial cells and ECM. Each cell can further differentiate to adopt a specific morphology and patterns of gene expression due to signalling and mechanical forces. Mesenchymal cells and ECM are made of spherical nodes while epithelial cells are made of cylinders. Each cylinder is made of two nodes, an apical and a basal one. The number, size and position (in a continuous 3D space) of nodes changes via model dynamics. As a result cells move and change their size and form. Changes in the spatial location and shape of cells configure the overall changes in an embryo's morphology.

All calculations, including molecule concentrations and diffusion, are made exclusively on the nodes but allowing diffusion of some molecules between nodes within and between cells (see below). The spatial distribution of those nodes represents the embryo's morphology and within each cell it represents cell shape. Each node has a set of mechanical properties and can accumulate different types of molecules. These properties are numerical values that affect the forces acting between nodes and are affected by the molecules present in nodes. Cells also have properties whose values are affected by the molecules present in each of its nodes.

In addition, the model includes some global parameters and a gene network that can be different depending on the organ or embryo part being studied. The gene network is specified by a n_g x n_g matrix (matrix T), a n_g x n_g x n_g matrix (matrix R), a n_g x n_{ga} matrix (matrix E) and a n_g x n_{gb} matrix (matrix C). n_g is the number of different molecules the model considers (a value specified by the user), n_{ga} is the number of node properties plus the number of cell properties and n_{gb} is the number of cell behaviours considered in the model. Each element, t_{ij}, in the T matrix specifies how strongly gene j activates (positive value) or inhibits (negative value) the

transcription of gene i. Each element, r_{ijk}, in the R matrix specifies how strongly molecule i catalyses the production of molecule k from molecule j. Thus, our model considers both transcription of genes and the modification of molecules that may be gene products (for example post-translational modifications of gene products) or not (such as many secondary messengers). We refer to both the gene products and other molecules that affect gene expression as regulatory molecules. Each element, e_{ij}, in the E matrix specifies how strongly molecule i increases or decreases node property j in the nodes where this molecule is present. Each element, c_{ij}, in the C matrix specifies how strongly gene i promotes or inhibits cell behaviour j in the nodes where it is expressed. In addition each molecule has its own intrinsic rate of degradation and diffusion. Regulatory molecules can *(i)* diffuse between any of the nodes of a cell, *(ii)* only between the apical or basal nodes of an epithelial cell, or, *(iii)* in the case of growth factors, in the space between cells.

Cells also have properties whose values are affected by the regulatory molecules present in each of its nodes. Cell behaviours are implemented in the model as specific rules of manipulation of these nodes and their distribution between cells. Which behaviour a cell performs in a given instant of time is affected by the regulatory molecules present in it (in a way specified by the user).

Our model does not impose any specific gene network but is a computational framework in which any arbitrary network can be implemented and in which these can regulate cell behaviours with realistic biomechanics. Thus, it is the user that decides which gene network and initial conditions the model uses according to the specific developmental system the user wants to study. The embryo editor and gene network software provide an easy way to do that without needing to program or understand the code (only the biological bases of the model need to be understood for that).

B.2 List of model elements

B.2.1 Subcellular elements and nodes

Subcellular elements are the smallest functional entities implemented in the model. They represent a physical portion of a cell. Mesenchymal subcellular elements are implemented as spherical elastic volumes, as has been done in the Subcellular Elements Model or SEM (Newman, 2005), and from now on they will be called mesenchymal nodes. In contrast, epithelial cells are made of sub-cellular elements with cylindrical shape. Each cylinder represents a portion of the cell extending from the apical surface to the basal surface. Thus, there are 3 different surfaces in each cylinder: apical and basal surfaces, which are in contact with the extracellular space or with mesenchymal nodes, and a lateral surface, which contacts the lateral surface of other cylinders from the same cell or the ones from other cells (fig. B.3). Each epithelial cylinder consists of two nodes, an apical and a basal one, which may have different properties. These two nodes are tied by an elastic spring (see section B.3.2). In order to represent the ECM a third type of subcellular element is implemented as spherical nodes that do not belong to any cell (see section B.3.3).

Node properties are named by a lower case p and a three letter superindex specific of each property and a second subindex for the specific node (e.g: p_i^{EQD} is the property "equilibirum distance" for node i; notice the superindex is not p_i multiplied EQD times).

B.2.2 Cells

Cells are functional entities in the model that can change their shape and perform a wide range of cell behaviours (see section B.6). Cell properties are named by a upper case P and a three letter superindex specific of each property and a second subindex for the specific cell. Thus, for example, P_i^{PHA} is property "PHA" of cell i. Each cell is composed of one or several subcellular elements and the shape of the cell, thus, is given by the relative positions of those elements. The number of nodes in a cell in a given moment depends on how many nodes it had initially, on how much has it grown, on whether it has divided or not and on two cell properties. The first, P^{MIN}, specifies the minimal number of nodes in a cell, this precludes cell divisions leading to any daughter cell with less than that number. The other one, P^{MAX}, specifies the maximal number of nodes a cell can have (beyond that number the cell may divide, if it is in right cell cycle stage, or stop growing). In those developmental systems where cell shape is not important the user can choose to have mesenchymal cells made of a maximum of only one node and epithelial cells of two nodes, a cylinder (this requires, in addition, setting the logical model parameter, L_1, to 1). These cell properties, like all other properties, can be genetically regulated. Each cell has one node specified as the nucleus, which only differs from the other nodes in that it's the only node in which transcription takes place. The fact that cells are made of nodes permits cells to have internal spatial asymmetries, that is, different nodes in a cell can have different properties (either different mechanical properties or different amounts of different regulatory molecules).

B.2.3 Gene products, regulatory molecules and genetic parameters

The amount of a regulatory molecule at a node can change during simulation time due to gene product transcription, diffusion between nodes, biochemical reactions and degradation (see section B.4). Each regulatory molecule has a set of properties, which we call generically gene parameters even if not all the regulatory molecules are necessarily gene products. These properties are actual model parameters in the sense that they are assumed to be genetically encoded and, thus, do not change during the simulation. Chemical transformation of regulatory molecules is implemented in the model as the transition of one type of molecule to another one by regulated catalysis. Thus, transformed molecules are not considered to have changed properties but abundance (this abundance of each molecular species being a property of each node). The only difference the model considers between gene products and other regulatory molecules is that gene products can be transcribed and translated while other regulatory molecules can only be transformed (we include different equations for transcription and catalysis of each chemical transformation). Transcription and

translation can be considered together by the transcription equations or can be considered separately by using the non-transcription catalysis equations for translation. Gene products can also be modified post-translationally by reaction catalysed by other regulatory molecules. In this model we treat each one of these modifications as a different gene product that can have its own genetic parameters. A specific gene product modification may be present or not in a node depending on the model dynamics and on the initial conditions.

These molecular parameters include diffusivity, rate of degradation, the ability to interact with other molecules (chosen by the user according to the system being studied) as specified in the T and R matrices and their regulation of node properties (E matrix) and cell behaviours (C matrix). In addition, the model includes a B matrix that specifies the affinity of membrane molecules that mediate cell-cell adhesion. By choosing the values of those parameters, for example by using the NetworkMaker software, the user can implement the number of genes and the the gene network he/she wants to study and how that affects node properties and cell properties and behaviours.

B.2.4 Global model paramenters

These are numerical values which, like the molecular parameters, do not change during a model simulation but that can be set to different values in different simulations of the model and do not have a direct correspondence with anything genetic. These include things such as the temperature and logic model parameters that specify some details about how the model is actually numerically implemented. These logic model parameters are identified by an L with a subindex specific to each parameter. Other model parameters are represented by an M with a subindex specific to each parameter (see section B.9).

B.2.5 Initial conditions

Those are the numerical values of all the nodes, and all the node and cell properties at time zero of a simulation (thus includes its location in 3D space and the amount of each regulatory molecule present in each node). These can be changed from simulation to simulation (with the embryo editor software or by manually editing the files that contain them). They are simply the stage in development from which we want to start to simulate development. This can be from a single cell with some internal spatial asymmetries (e.g., a zygote), or any arbitrary later stage in development (for example the second instar wing disc of Drosophila). Each initial condition is thus what we call a developmental pattern (Salazar-Ciudad, Jernvall, and Newman, 2003) and the model dynamics transform the initial pattern into a different later pattern according to the genetic parameters and model parameters. These parameters (e.g., the genes and gene networks that regulate which cell behaviours) and the initial conditions are subject to change between simulations depending on the question and the developmental system being studied by the user.

B.3 Mechanical forces

B.3.1 Node neighbouring and basic biomechanical interactions

Nodes have a size, specified by a radius in both spheres and cylinders. Two nodes adhere to each other if they come into contact (fig. B.3A), that is, if their distance is smaller than the sum of their radii. The radius is a node property, p^{ADD}. Adhesion brings these nodes closer, further decreasing their distance until the equilibrium distance between nodes is reached. This represents the generic property of cell adhesion. If the nodes are from different cells this adhesion can be increased or decreased by adhesion molecules expressed at the nodes, as we will later explain.

Cell parts such as nodes represent physical objects and thus two nodes cannot occupy the same spatial location. Nodes, thus, have a second radius, node property p^{EQD}, and two nodes start repelling each other if the distance between them is shorter that the sum of their p^{EQD} (fig. B.3A). If the distance between two nodes is exactly equal to the sum of their p^{EQD} radii then an equilibrium is reached in which these nodes neither attract nor repel each other.

From these forces the movement of node i due to its interaction with other nodes is,

$$\frac{\partial \vec{r_i}}{\partial t} = \sum_{j=1}^{j=n_d} f_{Aij} \hat{u_{ij}} \tag{B.1}$$

Where n_d is the number of nodes in the embryo, $\vec{r_i}$ is the position in three-dimensional space of node i, t is time (the model uses continuous time), f_{Aij} is the force modulus and $\hat{u_{ij}}$ is the unit vector between node i and node j for spherical nodes and an analogous property for cylinders. We assume that most developmental processes happen within highly viscous media (Purcell, 1977), thus we calculate movement through an overdamped equation of motion.

The force modulus between node i and j is,

$$\begin{cases} f_{Aij} = k_{ij}^{REP}(d_{ij} - d_{ij}^{EQD}) & if\ d_{ij} < d_{ij}^{EQD} \\ f_{Aij} = k_{ij}^{YOU}(d_{ij} - d_{ij}^{EQD}) & if\ d_{ij} \geq d_{ij}^{EQD} \\ f_{Aij} = 0 & if\ d_{ij} \geq d_{ij}^{ADD} \end{cases} \tag{B.2}$$

where d_{ij} is the distance at between node i and j, d_{ij}^{EQD} is the equilibrium distance between node i and node j (simply the sum of the equilibrium radii, EQD, node properties of nodes i and j, fig. B.3A) and d_{ij}^{ADD} is the sum of the node property ADD of node i and j,

$$\begin{cases} k_{ij}^{REP} = p_i^{REP} + p_j^{REP} \\ k_{ij}^{YOU} = p_i^{YOU} + p_j^{YOU} \\ d_{ij}^{EQD} = p_i^{EQD} + p_j^{EQD} \\ d_{ij}^{ADD} = p_i^{ADD} + p_j^{ADD} \end{cases} \tag{B.3}$$

p^{REP} and p^{YOU} are biomechanical properties of the nodes that specify how strong per unit distance are the repulsion and elasticity forces respectively between a pair of nodes. If i and j belong to different cells, k^{ADH}

is used instead of k^{YOU} and k^{REC} is used instead of k^{REP}. That is, we implement intercellular adhesion as an elastic force between cells. p^{REC} is different from p^{REP} because naturally cells may more strongly resist incoming matter from other cells than from the same cell. Between cells, in addition to the generic adhesion between nodes, there is an adhesion term that depends on which adhesion molecules are expressed in each of the nodes.

$$\begin{cases} k_{ij}^{REC} = p_i^{REC} + p_j^{REC} \\ k_{ij}^{ADH} = p_i^{ADH} + p_j^{ADH} + \sum_{l=1}^{n_{ADH}} \sum_{q=1}^{n_{ADH}} (g_{il} g_{jq} B_{lq}) \end{cases} \tag{B.4}$$

Where, n_{ADH} is the number of different types of adhesion molecules (a model parameter), g_{il} is the amount of adhesion molecule l in node i, g_{iq} is the amount of adhesion molecule q in node j and B_{lq} is the affinity coefficient between adhesion molecules l and q (genetic parameter). Notice that homotypic adhesion is allowed between different nodes. B_{pq} can be negative causing cells which express the p and q adhesion molecules to repel each other (as happens with some semaphorins (Bagnard et al., 1998), and ephrins (Wang and Anderson, 1997). Thus, the adhesion force is simply the product of how many adhesion molecules there are of each binding pair of a given type and the affinity of each pair of types of adhesion molecules.

If node density in space and the p_i^{ADD} of nodes are both large it becomes possible that two nodes could interact even if there are nodes between them. To avoid that unrealistic situation the model allows for three alternative algorithms to determine which nodes can effectively interact. Each simulation should be run with only one of these alternative methods:

1. Simple method. As described above, any node j that is at a distance (in 3D space) smaller than p_i^{ADD} + p_j^{ADD} from a given node (node i) is interacting with it.

2. Delaunay method. A tesselation of the 3D space is performed by the Delaunay triangulation algorithm, taking each node as a vertex. Then only the nodes that are connected by an edge in this tesselation and are at a distance smaller than the sum of their p^{ADD}, as above, interact.

3. Gabriel graph method. Similar to method 1, but for every two nodes at the right distance an additional criterion needs to be fulfilled. A sphere of a diameter equal to the distance between the two nodes is put in the mid-line between them. If there is any node within this sphere then those two nodes do not interact.

Method 1 is computationally faster and is the one used in the SEM (Newman, 2005). Method 3 is used in some similar models (Delile, Doursat, and Peyriéras, 2013). In most situations these three methods provide very similar results. Only when nodes extensively overlap in space (because overall large values in the p^{ADD} property in respect to node density in space) it is more realistic to use method 2 or 3 (method 3 being computationally more advantageous for only a slightly lower degree of realism). These two last methods simply preclude the interaction between two nodes that are close enough if there are other nodes in between. In that sense they implement a screening between nodes. These alternatives are specified, respectively by the logic model parameters L_3 and L_{11}.

B.3.2 Forces in epithelial nodes

In epithelia, the above forces work in a different way (although the rules to define neighbourhood, which nodes interact with which, is the same). Equations B.1 and B.2 are the ones used to calculate the forces in cylinders but the $\hat{u_{ij}}$, the direction of the force, and d_{ij} the distance between elements is calculated differently to take into account that epithelial elements have a cylindrical shape.

Direction of the mechanical interactions between epithelial cylinders

Epithelial cylinders have two types of surface: apical/basal and lateral. Those surfaces are defined by the orientation of the cylinder vector ($\vec{s_{ik}}$) connecting the two epithelial nodes (i and k) composing a cylinder.

$$\vec{s_{ik}} = \vec{r_i} - \vec{r_k} \tag{B.5}$$

Where r_i and r_k are the position vectors of nodes i and k respectively. Apical/basal surfaces can interact with mesenchymal or ECM nodes or with the apical/basal surface of other epithelial cylinders. Lateral surfaces can only interact with lateral surfaces of other cylinders. (fig. B.3B, C, D).

Case 1. Apical/basal cylinder surface against spheric node (fig. B.3C).

$$\vec{u_{ik}} = -\vec{s_{ik}} \tag{B.6}$$

In this case equation B.1 applies but $\vec{u_{ij}}$ is calculated with equation B.6. This simply reflects the fact that the contact area between the apical/basal side of the cylinder and the spheric node is always parallel to the apical/basal surface of the cylinder (fig. 4.1C).

Case 2. Apical/basal surface against apical/basal surface of different cylinders (fig. B.3D). In this case the interaction is calculated in a more complicated way. Let $\vec{m_{ijkl}}$ be a vector contained within the plane defined by the contact surface between the two cylinders (fig. B.3D). $\vec{m_{ijkl}}$ is calculated as the sum of the spring vectors from the two cylinders ($\vec{s_{ik}}$ and $\vec{s_{jl}}$),

$$\vec{m_{ijkl}} = \vec{s_{ik}} + \vec{s_{jl}} \tag{B.7}$$

where i and k are the nodes composing the first cylinder and j and l the nodes composing the second cylinder. $\vec{u_{ij}}$ has to be normal to the contact surface plane and at the same time has to be contained within the plane defined by $\vec{m_{ijkl}}$ and $\vec{c_{ij}}$ (fig. B.3D),

$$\vec{u_{ij}} = \vec{c_{ij}} - \vec{m_{ijkl}} \frac{\vec{m_{ijkl}} \cdot \vec{c_{ij}}}{|\vec{m_{ijkl}}|^2} \tag{B.8}$$

where $\vec{c_{ij}}$ is the vector connecting the two interacting nodes i and j (that is between nodes in the two different cylinders) (fig. 4.1D). The $\vec{u_{ij}}$ calculated in equation B.8 is then divided by its modulus (to become a unit vector) and then fed into equation B.1 to calculate the force as before.

There is a special case, when s_{ik} and s_{jl} are parallel and opposite in orientation, then $\vec{m_{ijkl}}$ is equal to 0. In that case, since the contact surface is normal to both $\vec{s_{ik}}$ and $\vec{s_{jl}}$, equation B.6 is used.

Case 3. Cylinder lateral surface against cylinder lateral surface (fig. B.3B). This case is similar to case 2, since $\vec{m_{ijkl}}$ is contained within the contact surface plane and $\vec{c_{ij}}$ has to be orthogonal to $\vec{m_{ijkl}}$ and contained in the plane defined by $\vec{m_{ijkl}}$ and $\vec{c_{ij}}$. Thus in this case $\vec{u_{ij}}$ follows equation B.8.

Note that, following Newton's third Law, the force acting on j must have the same modulus but opposite sign. Thus, in all cases,

$$\vec{u_{ij}} = -\vec{u_{ji}} \tag{B.9}$$

Distance between cylinders and between cylinders and spheres

Since the vector $\vec{u_{ij}}$ used to calculate mechanical forces is not always equal to $\vec{c_{ij}}$, the distance measure, d_{ij}, used to calculate the force modulus cannot be the actual distance between nodes i and j, which is the length of $\vec{c_{ij}}$. That is why the effective distance d_{ij} in equation B.2 is calculated in the case of cylinder as the projection of $\vec{c_{ij}}$ over $\vec{u_{ij}}$ (fig. B.3).

$$d_{ij} = \frac{\vec{c_{ij}} \cdot \vec{u_{ij}}}{|\vec{u_{ij}}|} \tag{B.10}$$

This simply reflects that in a cylinder the distances between interacting nodes are different (due to the cylindrical shape, fig. B.3).

Additional forces in epithelial cylinders

In addition epithelial nodes have two additional forces that are not present in other types of nodes.

Spring (fig. B.4A). The spring exerts an elastic force on the two nodes depending on the distance between them,

$$f_{Sik} = k_{ik}^{HOO}(d_{ik} - p_i^{EQS})\hat{s_{ik}} \tag{B.11}$$

Where $k_{ij}^{HOO} = p_i^{HOO} + p_j^{HOO}$ is the elastic coefficient of the spring (note p^{HOO} is a node property), p_i^{EQS} is the equilibrium length of the spring between node i and j (note $p_i^{EQS} = p_j^{EQS}$) and $\hat{s_{ij}}$ is the cylinder vector, the unit vector in the direction between both nodes in a cylinder. d_{ij} is simply the distance between node i and j.

Epithelial bending forces (fig. B.4B, C). We define two different forces that deal with bending of epithelial sheets. Due to external forces (e.g. a punctual force pushing one cylinder from the basal side) the apical or basal side of a given cylinders may slide in the apical-basal direction respect neighbouring cylinders (fig. B.4B, C). If we consider the relative position of the apical/basal nodes as a discrete representation of the continuous apical/basal epithelial surface, any displacement in the apical-basal direction will imply a local curvature of the epithelium. Thus, a radial force has to be defined that reduce local curvature. We define $\vec{c_{ij}}$ as the vector connecting

neighbouring node i and j, s_{ik} and s_{jl} as the vectors that define the elastic link to their basal counterparts and $m_{ijkl}^{\vec{}}$ as the sum of $s_{ik}^{\vec{}}$ and $s_{jl}^{\vec{}}$ which defines the vector normal to the apical or basal surface between i and j. The force vector f_{ESTij} is calculated as,

$$\begin{cases} \vec{f_{EST}}_{ij} = k_{ij}^{EST} \frac{m_{ijkl}^{\vec{}} \cdot c_{ij}^{\vec{}}}{|m_{ijkl}^{\vec{}}|} \hat{m_{ijkl}} & if \ \ \frac{m_{ijkl}^{\vec{}} \cdot c_{ij}^{\vec{}}}{|m_{ijkl}^{\vec{}}|} \geq M_{AMX} |c_{ij}^{\vec{}}| \\ \vec{f_{EST}}_{ij} = 0 & if \ \ \frac{m_{ijkl}^{\vec{}} \cdot c_{ij}^{\vec{}}}{|m_{ijkl}^{\vec{}}|} < M_{AMX} |c_{ij}^{\vec{}}| \end{cases} \tag{B.12}$$

This force always acts on the direction of $m_{ijkl}^{\vec{}}$ and is proportional to the deviation of the angle formed by $m_{ijkl}^{\vec{}}$ and $c_{ij}^{\vec{}}$ from 90º (the angle found when to cylinders are totally aligned). k_{ij}^{EST} is derived from a specific node property ($k_{ij}^{EST} = p_i^{EST} + p_j^{EST}$).

In order to minimize the sensitivity to small perturbations, we set a minimal value of the projection of $c_{ij}^{\vec{}}$ over $m_{ijkl}^{\vec{}}$ to apply $\vec{f_{EST}}_{ij}$, that is when the projection of $c_{ij}^{\vec{}}$ over $m_{ijkl}^{\vec{}}$ is less than the product of the distance between the two nodes and M_{AMX}, a model parameter, no surface tension force is applied.

If the force generating the local curvature is persistent, the epithelial surface will remain curved, and thus the epithelial cylinders will have to reorient in order to make their apical-basal vector s normal to the local epithelial surface, which is approximated by the vectors connecting the apical/basal node with their apical/basal neighbours. This is accomplished by applying a rotational force that will rotate the cylinders until their apical-basal vector $\vec{s}$ is normal to the surface plane at that position. The force vector f_{ERPij} is calculated as,

$$\begin{cases} \vec{f_{ERP}}_{ij} = k_{ij}^{ERP} \frac{s_{ik}^{\vec{}} \ \Delta c_{ij}^{\vec{}}}{|s_{ik}^{\vec{}}|} \hat{c_{ij}} & if \ \ k_{ij}^{ERP} \frac{s_{ik}^{\vec{}} \ \Delta c_{ij}^{\vec{}}}{|s_{ik}^{\vec{}}|} \geq M_{AMX} |c_{ij}^{\vec{}}| \\ \vec{f_{ERP}}_{ij} = 0 & if \ \ k_{ij}^{ERP} \frac{s_{ik}^{\vec{}} \ \Delta c_{ij}^{\vec{}}}{|s_{ik}^{\vec{}}|} < M_{AMX} |c_{ij}^{\vec{}}| \end{cases} \tag{B.13}$$

This force is proportional to the deviation of the angle formed by $s_{ik}^{\vec{}}$ and $c_{ij}^{\vec{}}$ from 90º, but in this case the direction of the force is parallel to $c_{ij}^{\vec{}}$, thus promoting a tilting of the epithelial cylinder that reaches an equilibrium (that is the force modulus becomes 0) when the apical-basal axis of the epithelial cylinder is normal to the apical/basal cell surface. k_{ij}^{ERP} is derived from a specific node property ($k_{ij}^{ERP} = p_i^{ERP} + p_j^{ERP}$).

In order to minimize the sensitivity to small perturbations, we set a minimal value of the projection of $c_{ij}^{\vec{}}$ over $s_{ik}^{\vec{}}$ to apply $\vec{f_{ERP}}_{ij}$, that is when the projection of $c_{ij}^{\vec{}}$ over $s_{ik}^{\vec{}}$ is less than the product of the distance between the two nodes and M_{AMX}, a model parameter, no surface tension force is applied.

Note that these forces act equally between any cylinders irrespective of whether these cylinders belong to the same cell or not. In that sense this force is a property of the epithelium as such.

In summary, thus, the forces acting on an epithelial node are,

$$\frac{\partial \vec{r_i}}{\partial t} = \vec{f_{Sik}} \sum_{j=1}^{j=n_d} f_{Aij} \hat{u_{ij}} + \vec{f_{EST}}_{ij} + \vec{f_{ERP}}_{ij} \tag{B.14}$$

where k is the node in the same cylinder than i and the sum is made over all the neighbouring nodes except for k.

This initial arrangement and the forces just described allow these cells to mechanically behave as epithelia for a broad range of realistic model parameters and initial conditions. Thus, epithelia behave as in two and a half dimensions, they can fold in complex ways in three-dimensional space but they retain a basically two dimensional structure with all cells in an epithelium binding to each other in their lateral sides (the apical and basal sides can bind to non-epithelial nodes or to nodes from different epithelium or even to a fold of the same epithelium).

B.3.3 Extracellular matrix

The ECM is represented by spherical nodes that do not belong to a cell. They interact with other nodes in the same fashion as mesenchymal nodes. ECM is usually composed of large proteoglycans and glycoproteins that swell the extracellular fluids and behave as a gel. Therefore by implementing the ECM as free nodes we capture their most relevant mechanical properties. ECM nodes can be secreted by cells (see below), they can also contain regulatory molecules and can harbour catalysis (but not transcription).

B.3.4 Node movement and noise

In addition to the movement equation defined at equation B.14 there is some noise in node movements. At each time step, a proportion M_{NOI} (a model parameter) of the nodes are chosen at random and are tentatively moved in a random direction for a random distance between 0 and p_i^{DMO} , a mechanical property of each node. For each node the potential mechanical energy is calculated, by integrating the same force equations shown in section B.3.1 and B.3.2, in the new position. If the potential energy in the new position is smaller than in the old position the movement is accepted. If not, the movement is accepted with a probability proportional to the difference in potential energy between the new and old positions and inversely proportionally to a temperature parameter, model parameter M_{TEM}, plus a node property defining the node's propensity to movement (p^{MOV}),

$$\begin{cases} P_{accept} = e^{-\frac{U_{after}-U_{before}}{M_{TEM}+p_i^{MOV}}} & if \quad U_{after} - U_{before} \geq 0 \\ P_{accept} = 1 & if \quad U_{after} - U_{before} < 0 \end{cases} \tag{B.15}$$

where P_{accept} is the probability of realization of the movement, U_{before} is the potential energy in the node position before movement and U_{after} is the potential energy after the movement. If the movement is not accepted the node is put back to its old position. This energy biased noise reflects the fact that noise can affect nodes' positions but it is unlikely to bring nodes into very energetically unfavorable positions (e.g. noise is very unlikely to bring a node from a cell inside another cell). This is a standard way to implement noise in many physical and biological systems (such as in SEM and in the Pott's model (Graner and Glazier, 1992).

At the level of cells and nodes this noise property, p^{MOV}, reflects in part the tendency of cells, especially mesenchymal cells, to temporarily

extent and retract cytoplasmatic projections (filopodia, pseudopodia and related structures) into the extracellular space. The likelihood of a pseudopodium retracting after being extended depends on whether it finds a suitable strong binding site (either in other cells or in the substrate). Also different types of cells tend to have pseudopodia of different lengths and tend to extend them with different frequencies. In individually migrating cells, in addition, the binding of those extensions is also relatively unstable so that cells can dynamically move over space. In our model this is captured by the p^{MOV} and p^{DMO} node properties. The movement of a node by noise can be represented as this node being the tip of a pseudopodium. p^{DMO} specifies how long pseudopodia can extend before being retracted and p^{MOV} specifies how labile this node binding is (the effect is simply to add to noise in equation B.15). Each node would then bind according to its p^{ADH} , plus the amount adhesion molecules expressed both in that node and in the one it is making contact with.

Noise allows the migration of individual cells over space (this is not possible in the model without noise). The two corresponding node properties are affected, as are all other node properties, by the regulatory molecules contained in nodes. This allows the implementation of individual cell chemotaxis. Basically, extracellular signal gradients induce intracellular regulatory gradients that differentially affect various properties within the cell. Similarly, haptotaxis is implemented given a spatial gradient adhesion molecules since motile cells will migrate towards regions where the concentration of adhesion molecules is higher (see fig. 4.5).

B.4 Gene expression, regulatory molecules and gene networks

Each regulatory molecule has a set of properties associated with it (which we call genetic parameters even if not all the molecules considered are gene products). These include the diffusivity of the molecule (D_i) and its intrinsic degradation rate (μ_i) and how they affect transcription, catalysis, node properties and cell behaviours. Each molecule, thus, has one row of its own in the T, R, E and C matrices.

The T, R, E and C matrices are the quantitative specification of the developmental mechanism used in a simulation (and thus would be different depending on the developmental system being simulated with the model).

In the model molecules can affect the concentration of each other in a node in two ways. Some regulatory molecules can affect the production *de novo* of gene products by affecting the transcription of genes and the translation of transcripts (these two processes can be considered together or separately with translation being a different kind of reaction). Some regulatory molecules can catalyse the reactions by which a certain type of molecule is transformed into another type of molecule. These latter reactions can be seen either as post-translational modifications of proteins, reactions involving non-protein molecules or simple binding of molecules leading to conformational changes in those (that then can change the properties of those molecules). Each different molecule can have different genetic parameters (e.g. they can affect other regulatory molecules in different ways or regulate different node properties or cell behaviours). Thus, for example, in

the model a certain protein and the same protein in a phosphorylated state are two distinct regulatory molecules that can have different genetic and regulatory properties (specified in different rows in the T, R, E and C). The same can happen when a regulatory molecule is bound to some other molecules. Binding can lead to conformational changes that may drive changes on the interactions of that regulatory molecule with other molecule and thus the bound form is considered to be a different molecule than the unbound form and then can have different genetic parameters (such as in T, R, E and C). This is often the case for receptors, the ligand unbound and bound forms of a receptor often can interact with different molecules.

B.4.1 Types of regulatory molecules

In our model the function of a regulatory molecule is specified by their genetic parameter matrices. Thus transcription factors are gene products that affect transcription of other genes. This implies that they have non-zero elements in their row in the T matrix specifying which gene products regulate other genes' transcription.

Enzymes are regulatory molecules that have at least one non-zero element in their row in the R matrix. These can represent regulatory molecules involved in signal transduction such as kinases, proteases, lipases, etc. or metabolic enzymes. The R matrix specifies which regulatory molecules catalyse the reaction between which regulatory molecules and with which activity per molecule. Regulatory molecules that catalyse the binding between molecules can also have non-zero elements in their row in the R matrix.

Regulatory molecules with a non-zero element in their row in the E matrix are regulating node properties. They can be for example myosins affecting contraction and thus p_i^{EQD} in specific nodes or molecules binding to the cytoskeleton to make nodes more stiff (p_i^{YOU}), etc.

In the same way regulatory molecules with non-zero elements in the C matrix are involved in regulating cell behaviours. These can be activators of mitosis, apoptosis, matrix secretion, etc.

Membrane adhesion molecules have a non-zero element in the first column of the E matrix. This element indicates the index of the adhesion molecule this regulatory molecule is. Then the B matrix indicates the affinity of binding between each adhesion molecule (each element in the B matrix is thus a model parameter as is the case for the other matrices).

There is nothing in the model restricting a regulatory molecule from being a transcription factor and an adhesion molecule at the same time or an enzyme and a transcription factor. All these functions are simply specified by the values in the T, R, E and C matrices. In addition, there is a genetic property called type. Each regulatory molecule can be of only one type. These are the following.

Extracellular diffusible signals or growth factors. These are regulatory molecules that can diffuse between nodes in different cells or between ECM nodes, as well as between adjacent nodes belonging to the same cell in contact with the cell surface. The amount of an extracellular signal in a node in the model does not represent how much of it there is the node but how much of it there is the extra-cellular space around the node. They cannot

diffuse across the apical basal axis of epithelial cells, that is between an apical node and a basal node of the same cylinder. These diffusible signals affect expression in a cell only be binding to specific membrane receptors. The bound forms may have different genetic parameters than the unbound form (for example it may have a different catalytic specificity or directly affect some node property). Note that mesenchymal cells tend to have rather irregular, fibroblast-like, shapes and thus most nodes should be expected to have some contact with the extra-cellular space. In the case of epithelial cells all nodes are in contact with the extra-cellular space, but extracellular signals cannot diffuse between nodes within the same cylinder.

Ligand binding receptors. They can only diffuse between adjacent nodes of the same cell in contact with the cell surface. Thus they cannot diffuse across the apical-basal axis of an epithelial cell. They can bind to extracellular diffusible signals located in the same node (if they are specified to do so) and form a receptor-ligand complex. The receptor-ligand complex may have different genetic parameters than the free receptor form.

Membrane tethered regulatory molecules. They can only diffuse between adjacent nodes of the same cell in contact with the cell surface. Thus they cannot diffuse across the apical-basal axis of an epithelial cell. They can mediate signalling by cell contact, thus they can interact and bind to membrane bound receptors located in adjacent nodes belonging to different cells.

Apically or basally localized molecules. These are intracellular regulatory molecules that are continuously being transported from one side of an epithelial cell (apical or basal) to the other. Molecules being transported to the apical side can freely diffuse between adjacent nodes from the same side, but are only transported along the the apical-basal axis of the cylinder when they are in the basal side. Molecules being transported to the basal side can diffuse freely between adjacent nodes from the same side, but are only transported along the apical-basal axis of the cylinder if they are in the apical side.

All other molecules. All other molecules can diffuse between all the nodes in a cell, but not between nodes of different cells.

B.4.2 Transcription

Transcription can only happen in the cell nucleus, which is located in a specific node within the cell. The rate of transcription of gene k in node i (provided that i is a nuclear node) is,

$$Q_{ik} = \frac{\Phi\left(\sum_{l=1}^{n_g} t_{lk} g_{il}\right)}{1 + \Phi\left(\sum_{l=1}^{n_g} t_{lk} g_{il}\right)} \tag{B.16}$$

Where Q_{ik} is the rate of transcription of gene k in node i, g_{il} is the amount of transcriptional factor l in node i and each t_{lk} term is the strength

by which each specific transcriptional factor k activates (positive t_{lk}) or inhibits (negative t_{lk}) the transcription of gene l (each of them is an element of matrix T). The sum is done through all the regulatory molecules and by definition only transcriptional factors have t_{lk} terms different from zero. Φ is a function that is equal to 0 for values of x smaller than 0 and equals to x when x is greater than 0 ($\Phi(x) = 0$ if $x < 0$ and $\Phi(x) = x$ if $x > 0$). This function is used to ensure that there is not such a thing as negative transcription (although t_{lk} can be negative and thus repress transcription).

Equation B.16 represents the binding of several transcriptional factors to the promoter of gene k. This is a saturating process that, for simplicity, is represented by a Hill equation of order 1. This means that when there are few activator factors the rate of transcription increases with the amount of these factors. But when there are many of these factors the rate of transcription does not increase as much with the amount of activator factors since the binding sites for each of them in the promoter are likely to be already occupied by them. The same (Salazar-Ciudad, Garcia-Fernández, and Solé, 2000; Salazar-Ciudad, Newman, and Solé, 2001) or similar (Reinitz and Sharp, 1995) equation has been used in previous models of gene networks in development.

B.4.3 Non-transcriptional catalysis

The rate of production of a regulatory molecule in a node depends on the product of the amount of each regulatory molecule that gives rise to it by the amount of each regulatory molecule that promotes this catalysis (all within the same node). The rate of production of a regulatory molecule k in node i is thus,

$$S_{ik} = \sum_{l=1}^{n_g}\sum_{j=1}^{n_g} r_{jlk} g_{ij} \frac{g_{il}}{1+g_{il}} - \sum_{l=1}^{n_g}\sum_{j=1}^{n_g} r_{jkl} g_{ij} \frac{g_{ik}}{1+g_{ik}} \quad \text{(B.17)}$$

The first term W_{ik} defines the rate of production regulatory molecule k in node i due to the transformation of other forms l into k catalysed by j. The second term U_{ik} defines the rate of loss of form k due to its transformation into other forms l mediated by catalyzation from j. The R matrix element r_{jlk} specifies the catalytic activity of regulatory molecule j on the transformation of regulatory molecule l into regulatory molecule k. Each term follows Michaelis-Menten kinetics in which $K_M = 1$.

B.4.4 Receptor-ligand binding

The kinetics of receptor-ligand binding need to be implemented differently, since in this case the two reactants (receptor and ligand) give rise to a single product (the receptor-ligand complex). Thus, the receptor-ligand complex is represented by a single regulatory molecule c which can be formed by

binding the ligand l and the receptor k and can also be dissociated giving rise to l and k. The rate of change of c on node i is,

$$\begin{cases} S_{ic} = a_1 g_{il} g_{ik} - a_{-1} g_{ic} \\ a_1 = r_{clc} = r_{ckc} \\ a_{-1} = r_{ccl} = r_{cck} \end{cases} \tag{B.18}$$

Where a_1 and a_{-1} are the forward and backward reaction constants respectively. Note that this equation will be applied instead of B.17 only for molecules that have been specified as receptors, which is set as a gene property.

The kinetics of binding between receptors and membrane tethered ligands are implemented differently, since the binding happens in the interface between different cells and thus receptor and ligand are located in different nodes. For that reason the receptor-ligand complex is not represented by a single regulatory molecule, but by two. If ligand l is expressed in node i and receptor k is expressed in node j, then the receptor-ligand complex will be represented by the ligand bound form o in node i and receptor bound form p in node j. The rate of change of o and l in nodes i and j respectively is,

$$\begin{cases} S_{io} = S_{jp} = a_1 g_{il} g_{jk} - a_{-1} g_{io} g_{jp} \\ a_1 = r_{klo} = r_{lkp} \\ a_{-1} = r_{kol} = r_{lpk} \end{cases} \tag{B.19}$$

B.4.5 Molecule degradation

In addition all molecules have a basal degradation rate. That is calculated as,

$$M_{ik} = \mu_k g_{ik} \tag{B.20}$$

where μ_k is the intrinsic rate of degradation of molecule k (a model genetic parameter).

Enzyme dependent degradation (in the sense of a molecule promoting the degradation of another molecule, i.e. protease) can be implemented by defining a regulatory molecule k corresponding to a molecule labelled for degradation and having a second regulatory molecule (the enzyme catalysing degradation) capable of transforming other molecules into this molecule k. This molecule might then have a large μ_k.

B.4.6 Diffusion

Diffusion is implemented as transfers of molecules between nodes (including ECM nodes). This transport follows Fick's second law of diffusion,

$$\frac{\partial q}{\partial t} = -D\nabla^2 q \tag{B.21}$$

Where q is concentration of a molecule, D is the diffusion coefficient of that molecule and $\nabla^2 q$ is the second derivative of the concentration in 3D space. We calculate transfers of matter between pairs of nodes. Since

we only make calculations in the nodes diffusion is essentially discrete (although non-uniformly) and this equation is roughly approximated by,

$$O_{ik} = D_k \sum_{j=1}^{n_v} \left(\frac{g_{ik} - g_{jk}}{d_{ij}} \right) \tag{B.22}$$

Where g_{ik} is the amount of molecule k in node i, t is time, D_k is the diffusivity coefficient of molecule k, n_v is the number of nodes within the maximum radius of diffusion from node i and d_{ij} is the distance between node i and j. Both this distance and n_v depend on how nodes are arranged in space. The maximum radius of diffusion is two times the maximal p^{ADD} in the embryo in a given iteration multiplied by M_{DIF}, a model parameter. This ensures an optimal accuracy even if there are changes in the sizes of the nodes in the embryo over time.

We only consider diffusion between nodes that are closer than the maximum radius of diffusion. The amount of molecules interchanged between two nodes in each instant of time decreases with the distance between these two nodes. Thus, after some distance this diffusion becomes negligible. Increasing this distance exponentially increases the number of node pairs to be considered, thus making the calculations very expensive computationally for a small gain in accuracy. M_{DIF} is a model parameter that will usually take values between one and few node radii.

Since diffusion is only calculated between existing nodes, it cannot happen within empty cavities (without cells but filled with fluid) in the embryo, such as blastocoels. To ensure effective diffusion of molecules across those cavities, ECM nodes need to be added in order to fill them. In growing embryos those cavities will most likely grow, or decrease, so ECM should be actively secreted and/or degraded by cells (see section B.6.8).

The same node neighbourhood used for mechanical interactions is used to determine which nodes will transfer molecules between them (see section B.3.1).

In some cases we might want to simulate a developing system with open boundary conditions, where diffusive molecules would be lost through the borders. In those cases, the nodes making the boundary of the system can be set as boundary nodes, p^{BOR} node property, and then they will tend lose intercellular diffusing molecules at a rate,

$$O_{ik(boundary)} = D_k \left(\sum_{j=1}^{n_v} \left(\frac{g_{ik} - g_{jk}}{d_{ij}} \right) - g_{ik} \right) \tag{B.23}$$

B.4.7 Total amount of a molecule in a node

In summary, the rate of change of the amount of primary gene product k (produced by transcription) in node i is,

$$\frac{\partial g_{ik}}{\partial t} = Q_{ik} - M_{ik} + O_{ik} \tag{B.24}$$

For other molecules this is,

$$\frac{\partial g_{ik}}{\partial t} = S_{ik} - M_{ik} + O_{ik} \tag{B.25}$$

B.5 Regulation of node properties

B.5.1 Node properties

Most node properties have already been described when describing mechanical forces. See section B.5.7 for a list and summary of those. Each of these values can be modified by the amounts of specific regulatory molecules in a node. Each element e_{lm} in the E matrix describes the effect of regulatory molecule m on node property l. The value of node property l at time t in node i is then,

$$p_i^l(t) = \Phi\left(p_i^l(0) + (1 - p_i^{DIFF})\sum_{k=1}^{n_g} e_{lk}g_{ik}\right) \tag{B.26}$$

where $p_i^l(t)$ is the value of node property l in node i at time t and $p_i^l(0)$ is the value of that node property l in node i when the node was created (this is in the initial condition or when the node first arose through growth). Φ, as in equation B.16, function ensures that node properties can become very small (or zero) but not negative. p_i^{DIF} is the degree of differentiation in node i (differentiation slows down changes in nodes). The amount of change in node properties is then related to how much of the molecules regulating these properties there is in a node and how strongly they regulate them, as specified in each element e in the E matrix. For simplicity, this regulation is supposed to be instantaneous compared with the rate at which nodes move or with the rate at which regulatory molecules are catalysed. Equation B.25 applies to all node properties except for p_i^{EQD} and p_i^{DIF} that are explained later.

B.5.2 Node and cell differentiation

In this model cell differentiation is defined as the process leading a cell to stop any developmental cell behaviour or cell movement. Thus, the moment when all cells in the embryo are differentiated marks the end of the developmental process in our model. The level of differentiation in a node depends on the expression of certain regulatory molecules, its rate of change is,

$$\frac{\partial p_i^{DIF}}{\partial t} = \Psi\left(\sum_{m=1}^{n_g} e_{ma}g_{im}\right) \tag{B.27}$$

where a is the index of the column in matrix E specifying how regulatory molecules regulate differentiation and $\Psi(x)$ is a function that is 1 if $x > 1$ and it's equal to x if $x < 1$. As stated in equation B.26, the effect of regulatory molecules on node properties is diminished by the level of differentiation of the node, being 0 when the differentiation level reaches 1. At the moment all nodes reach a differentiation level of 1, the developmental process has ended and thus the simulation stops.

B.5.3 Regulation of node radii

The p^{EQD} of a node is the sum of four other node properties that correspond to four different cell processes.

$$p_i^{EQD} = p_i^{COD} + p_i^{GRD} + p_i^{PLD} + p_i^{VOD} \tag{B.28}$$

The value of p^{EQD} in each node is updated in each iteration according to the values of these other four node properties. The first term is coming from node active contraction (due to myosin and related molecules), the second is coming from cell growth and apoptosis, the third from cell mechanical plasticity and the fourth from cylinder volume conservation. Having p^{EQD} determined by four independent terms allows contraction, growth, plasticity and volume conservation to occur at the same time in a cell. For example it is important that parts of the cell can contract while the cell is growing (and that would not be possible if growth and contraction would act directly on p^{EQD} since one growth would increase p^{EQD} and contraction would decrease it: so there may not be much change overall). Cell contraction is realized when a regulatory molecule regulates negatively the node property p^{COD}. Since contraction is happening in the nodes, cells may have contraction in only part of its nodes, as it is necessary in a number of developmental processes such as in invagination by apical cell contraction. p^{COD} is calculated as in equation B.26 above. The other terms are explained in the following sections and when explaining cell growth.

In addition, the model includes a minimum and a maximum for any p^{EQD}, these are model parameters M_{EMI} and M_{EMA}, but those can be set arbitrarily small or large (so that they have no effect on the model dynamics).

Special rules also apply to p^{ADD}. Any decrease or increase in p^{EQD} during a time step is also applied to p^{ADD} so that the difference between p^{EQD} and p^{ADD} does not change because of changes in p^{EQD}. This way the two node properties defining their effective size change together.

B.5.4 Node and cell plasticity

Cells are not totally elastic. They are also viscous, as explained in section above, and can also accommodate incoming pressures by actively or passively changing the cytoskeleton. This is specially relevant in the case of the highly packed cells encountered in epithelia. With plasticity (that can be totally deactivated by setting logical model parameter L_9 to zero) epitehlial nodes accommodate to compression from other nodes by reducing their p^{EQD}. For each node i compression is calculated as the mean difference between the distance to its neighbours and the equilibrium distance to them,

$$Z_i = \frac{\sum_{j=1}^{n_v}(d_{ij} - d_{ij}^{EQD})}{n_v} \tag{B.29}$$

Where n_v is the number of neighbours in contact with i and d_{ij}^{EQD} is as in equation B.2. If Z_i is smaller than 0 then the node is under compression. Then the change in p_i^{PLD} per unit time is,

$$\frac{\partial p_i^{PLD}}{\partial t} = p_i^{PLA} Z_i \tag{B.30}$$

where p_i^{PLA} is a node property (that can be directly regulated by regulatory molecules) specifying how plastic node i is.

B.5.5 Volume conservation in cylinders

The two nodes that compose an epithelial cylinder are mostly independent from one another, nonetheless they together represent a single subcellular element and thus certain mechanical deformations in one node may affect the other node. For instance, if an epithelial cell contracts its apical surface (for example by means of actomyosin activity on the cell cortex) the volume of the apical half of the cell would decrease and, by volume conservation, the basal half should increase in volume. This kind of deformation can also occur if an epithelial sheet bends passively (from forces generated somewhere else than in the bending part; Lane et al., 1993). In this case, cells can accommodate by adopting slightly wedged shapes.

Epithelial cylinder volume conservation is implemented in the model as transfers of volume between the apical and basal nodes of the cylinder, depending on the level of compression and/or tension there is on each side. In order to make effective the transfer of volume between sides p^{VOD}, is defined as a component of p^{EQD} due to volume conservation of the cylinder, that will tend to correct any deviations from the equilibrium volume due to deformation. Its rate of change over time is,

$$\frac{\partial p_i^{VOD}}{\partial t} = p_i^{VOC}\left(\frac{p_i^{GRD} + p_j^{GRD} - p_i^{EQD} - p_j^{EQD}}{2}\right) \tag{B.31}$$

Where j is the other node belonging to the same cylinder as i. This is essentially the average difference between p^{EQD} and p^{GRD} in a cylinder multiplied by the node property p_i^{VOC} ($p_i^{VOC} = p_j^{VOC}$). Note that the sum of p^{GRD} (the contribution coming from growth) from the two nodes as an equilibrium volume of the cylinder.

B.5.6 Spatial fixation of nodes

The model allows to simulate an embryo part or organ and consider the rest of the embryo in a simpler way. This is done by simulation only the organ and having some special conditions in its boundary to represent the rest of the embryo. For example, if the system is expanding it should feel a restorative force in the boundaries, due to the resistance to compression of the tissues beyond the system.

This is accomplished by setting the node property p^{FIX} to either 1 (elastic fixation) or 2 (complete fixation) in the nodes in the borders. A node with p^{FIX} of 1 will be tied to an elastic spring anchored to the position of that same node at time 0 of the simulation and with an equilibrium length of 0. Thus, an additional force component will act on these nodes,

$$\vec{f_{FI}} = p_i^{KFI}\left(\vec{r_i}(t) - \vec{r_i}(0)\right) \tag{B.32}$$

Where p^{KFI} is a node property that determines the elastic constant of the spring in the border that multiples the vector going from the actual position of node i to the position it had at time 0. A node with a p^{FIX}

of 2 will be completely fixed on space, thus making the boundaries of the system totally rigid.

B.5.7 Summary of node properties

- Intercellular adhesion: p^{ADH}
- Intracellular elasticity: p^{YOU}
- Cell compressibility between nodes of the same cell: p^{REP}
- Cell compressibility between nodes of different cells: p^{REC}
- Filopodia extensibility: p^{DMO}
- Filopodia instability: p^{MOV}
- Node plasticity: p^{PLA}
- Degree of differentiation: p^{DIF}
- Equilibrium radius: p^{EQD}
- Contraction component of p^{EQD}: p^{COD}
- Contraction component of p^{EQD}: p^{GRD}
- Contraction component of p^{EQD}: p^{PLD}
- Amount of ECM stored: p^{ECM}
- Fixation of node in space: p^{ADD}
- Elastic constant of fixation: p^{KFI}
- Open boundary node: p^{BOR}
- Rotational bending force coefficient: p^{ERP}
- Radial bending force coefficient: p^{EST}
- Apical-basal elasticity: p^{HOO}
- Apical-basal equilibrium distance: p^{EQS}
- Volume conservation component of p^{EQD}: p^{VOD}

B.6 Cell behaviours

B.6.1 Cell shape change and contraction

Cell morphology is determined in this model by the size and relative position of the nodes composing a cell. Thus, cell morphology can change due to passive processes, such as deformation by mechanical stresses, or due to active processes, like genetically regulated contraction or expansion of nodes within the cell.

Cell contraction occurs as explained in section B.5.3. Contraction by changes in p^{EQD} is a way for cells to produce intrinsic forces than then

can spread and affect neighbouring cells. Note that changes in p^{EQD} can be both increases and decreases and that even decreases will induce forces if the shrinking nodes are bound to other nodes (as it would often be the case).

B.6.2 Cell polarisation and internal cell asymmetries (fig. B.5)

Epithelial cells are by definition polarized in the apical-basal axis but can also have a polarization in the plane of the epithelium (what is often call planar cell polarity (Simons and Mlodzik, 2008). Mesenchymal cells can also be polarized. In our model the polarization of a cell h is described by a 3D vector, $P_h^{\hat{POL}}$, a cell property. This vector arises from the asymmetrical distribution of regulatory molecules within the nodes in a cell. First, a polarization score s_{hi} is calculated for each node i in cell h,

$$s_{hi} = \sum_{m=1}^{n_g} c_{ma} g_{im} \tag{B.33}$$

where a is the index of the column in the C matrix corresponding to the effect of regulatory molecules on cell polarity (see fig. B.5). The polarization vector is then,

$$P_h^{\vec{POL}} = \sum_{i=1}^{n_h} (s_{hi} - s_{hc}) \vec{r_i} \tag{B.34}$$

Where s_{hc} is the score of the node closest to the centroid of cell h, n_h is the number of nodes in cell h and is the position vector of node i. This is simply an average of each node position weighted by its score, compared to that of the most central node. This vector is then divided by its module to find the polarization vector itself (the unit vector $P_h^{\hat{POL}}$). Polarization, thus, arises from asymmetries in the distribution of some molecules within the cell. These asymmetries can be present already in the initial conditions or arise during development. In development, and in our model, these asymmetries can arise by signalling between cells (or between cells and the environment) that do not lead to changes in gene expression. For example, a cell can secrete a growth factor that then, by diffusion, reaches some nodes in a neighbouring cell with high concentration and some, more distant nodes of the same cell, with a lower concentration. If this signal elicits a signal transduction that promotes the catalysis of some regulatory molecule (for example the phosporylation of some protein) then there would be a gradient in the amount of this regulatory molecule within the cell (as in fig. B.5). Signalling without changes in transcription are implemented in the model by extracellular signals activating the catalysis of specific regulatory molecules (without transcription) that then reproduce this gradient inside the cell. Note this is more difficult to produce if the signal leads to transcription since transcription occurs in the nucleus and from there diffuses in all directions within the cell (the gradient is thus from the nucleus irrespective of the direction of the gradient of the extracellular signal).

B.6.3 Cell growth (fig. B.6)

Cell growth is implemented as a progressive addition of nodes within a cell. This can happen only if the nodes within a cell are not too compressed, this is if,

$$\sum_{i=1}^{n_h} Z_{ih} < M_{MCO} \tag{B.35}$$

Where n_h is the number of nodes in cell h. Z_{ih} is the level of compression on node i of cell h. Calculated for equation B.28. New nodes are added one at a time and with a small size M_{MID}, a model parameter. Then, as long as the nodes have a p^{EQD} smaller than M_{MAE}, a model parameter, they will grow at a rate,

$$\frac{\partial p_i^{GRD}}{\partial t} = \sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{mb} g_{im} \tag{B.36}$$

Where n_h is the number of nodes belonging to cell h and c_{mb} is the effect of regulatory molecule m on cell growth (column b in the C matrix). Only when all nodes in the cell have p^{GRD} equal to M_{MAE} a new node is added to cell h. Even though growth is a process at a cell level, there is only one node increasing at a time, thus the rate of growth of that node depends on the amount of growth-inducing molecules located throughout all the cell (more than one node per iteration can be added if the logical model parameter L_{16} is set to 1). The node properties of the new node are set as the initial node properties of a random node from the same cell. Gene and molecule expression in the new node is set the same value than the node closest to it (this way the the smoothness of spacial molecular gradients is not perturbed by cell growth). Depending on where the new nodes are put in space we distinguish between non-polar and polar cell growth.

Non-polar cell growth In mesenchymal cells the position of the new node is chosen at random within the boundaries of the cell. In the case of epithelial cells, a new cylinder is added in a random position within the boundaries of the cell, but its orientation in the apical-basal axis is the same as that of the cylinders closest to it (to the three closest ones or to all the cylinders in a cell if there is less than three cylinders in the cell).

Polar cell growth Polar growth occurs with a probability,

$$P_{h(polar)} = \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{md} g_{im} \tag{B.37}$$

where d is the index of the column in the C matrix corresponding to the effect of regulatory molecule m on the probability of polar growth (see fig. B.5). The position of the new node along the polarization axis is determined by finding the node which is farthest from the cell centroid and at the same time closest to the direction of polarization of the cell. This is done by calculating the dot products between the polarization vector $\vec{P_h^{POL}}$ and the vector connecting the cell centroid with each node. The node giving the largest dot product will determine the direction in which the new node will

be added. The new node is added at a position in the line between the cell centroid and this node and at a distance that is 80% of the distance between the centroid and this node.

B.6.4 Cell division (fig. B.7)

Symmetric cell division. Cell division is implemented by splitting an existing cell into two new daughter cells. In symmetric division both daughter cells inherit roughly the same number of nodes.

The triggering of division in a cell depends on two factors: the progression of the cell cycle and the number of nodes in the cell. Progression of the cell cycle is specified by the cell property P^{PHA} and can take values from 0 to 1 (being 1 when division takes place). The rate of increase in P^{PHA} on any cell h is,

$$\frac{\partial P_h^{PHA}}{\partial t} = \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{md} g_{im} \tag{B.38}$$

Where d is the index of the column in the C matrix corresponding to the effect of regulatory molecules on cell phase progression. n_h is diving the sum to ensure that just having more nodes does not affect phase progression. This is thus the sum of the contributions of all the regulatory molecules affecting that cell behaviour in all the nodes of a cell. In addition, for a cell to divide it is also required that it has at least P_{MIN} nodes (a cell property). Also if a cell has more than P_{MAX} (another cell property) it divides irrespectively of its phase (P_{MAX} can, however, be set to be arbitrarily large). As we later explain the values of these cell properties, as all other ones, can be modified by regulatory molecules.

In nature, it is often the case that the plane of division is normal to the longest axis of the cell, what is commonly referred as Hertwig's rule (Minc, Burgess, and Chang, 2011), or normal to the polarization axis of the cell (as specified in section B.6.2). The longest cell axis (Hertwig vector: P_h^{HER}, fig. B.8) is calculated by means of a 3D linear regression of nodes' positions. The actual division vector (the vector normal to the plane of division) is calculated as a weighted average of the Hertwig and polarization vectors,

$$P_h^{DIV} = (1 - w_h) P_h^{\hat{POL}} + w_h P_h^{\hat{HER}} \tag{B.39}$$

The weighing factor for any cell h, w_h, is calculated as a function of the concentration of regulatory molecules in cell h that affect this weight. This is,

$$w_h = \frac{1}{\sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{me} g_{im}} \tag{B.40}$$

Where e is the index of the column in the C matrix corresponding to the effect of regulatory molecules on polarized cell division. Thus, if $W_h = 1$ the division vector is equal to the Hertwig vector and if $W_h = 0$ it is equal to the polarization vector.

The actual plane of division is normal to the division vector $P_h^{\vec{DIV}}$ and is passes through the centroid of the cell. This plane splits the cell in two and nodes in one side are assigned to one daughter cell and nodes in the other

to the other. The former nucleus of the cell loses its identity (it becomes a normal node), and after the division the most central node (the node closest to the new centroids) of each new cell is chosen as the nucleus.

Asymmetric cell division. In asymmetric division the size of the two daughter cells is different (one daughter cell has more nodes than the other). In this case, the position of the division plane along the division vector $P_h^{\vec{DIV}}$ does not pass by the physical center of the cell (the centroid), but depends on the spatial distribution of certain regulatory molecules within the cell. Each node i in cell h gets a score determined by summing all the molecules affecting the asymmetry of division,

$$s_{hi}^{ASY} = \sum_{m=1}^{n_g} c_{mf} g_{im} \tag{B.41}$$

Where f is the index of the column in the C matrix corresponding to the effect of regulatory molecules on asymmetric cell division. Then the division plane is placed at the point in the axis defined by $P_h^{\vec{DIV}}$ where the sums of the scores of nodes at each side of the plane are equal. Thus, the more skewed the distribution of those molecules, the more asymmetric is the cell division. If the gene product distribution is uniform then the plane of division appears on the centroid of the cell.

If the division is very asymmetric and the cell has not a very regular shape then daughter cells with isolated nodes can be produced (that is nodes in a cell not having physical contact with each other). Since this situation is biologically unrealistic outcome of cell division, the physical integrity of potential daughter cells is checked before cell division. If a daughter cell has unconnected nodes, the division plane is moved again to a position closer to the centroid, until the two new cells have all their nodes connected.

B.6.5 Cell death (fig. B.8)

Cell death or apoptosis is implemented in this model as inverted cell growth, that is when a cell is dying nodes start to decrease in size (p^{GRD}) until a minimum size is reached, then the node is deleted from the simulation. When all nodes belonging to a cell disappear then the cell also does so. The rate of decrease of p^{GRD} and p^{ADD} are equal to,

$$\frac{\partial p_i^{GRD}}{\partial t} = \frac{\partial p_i^{ADD}}{\partial t} = -\sum_{m=1}^{n_g} c_{mg} g_{im} \tag{B.42}$$

where g is the index of the column in the C matrix corresponding to the effect of regulatory molecules on apoptosis.

B.6.6 Cell adhesion

Cell adhesion is integrated in the mechanical part the model (see section B.3). Each node has a basal adhesivity plus the one given by the expression of adhesion molecules, which depends on the affinity of the adhesion

molecules expressed in each node. As discussed in section B.3, this includes also the possibility to implement repulsion between cells.

B.6.7 Epithelial-mesenchymal transition (EMT) (fig. B.9)

EMT is implemented as a discrete transformation of an epithelial cell to a mesenchymal one. Each epithelial cylinder is converted into two mesenchymal nodes by changing the identity of the two epithelial nodes and removing the elastic spring between them. Before that, in order to keep the spatial continuity of the future mesenchymal cell, the apical and basal sides of the epithelial cell are brought closer to each other by reducing the length of the elastic springs. The transition is regulated by a cell property P_h^{EMT} that is progressively increased due to gene expression,

$$\frac{\partial P_h^{EMT}}{\partial t} = \frac{1}{n_h} \sum_{m=1}^{n_g} c_{mk} g_{im} \tag{B.43}$$

where k is the index of the column in the C matrix corresponding to the effect of regulatory molecules on EMT. n_h is diving the sum to ensure that just having more nodes does not affect P_h^{EMT}. When P_h^{EMT} reaches a value of 1, the transition is realized.

B.6.8 Secretion

ECM secretion (fig. B.10) ECM is represented as free spheric nodes, which can be secreted by any type of cell provided that there is expression of molecules regulating its secretion. Given that ECM nodes in the model represent finite amounts of large fibrous extracellular molecules like proteoglycans or collagen, those have to first accumulate within the cell before being secreted as one ECM node. The rate at which those products accumulate within a node is,

$$\frac{\partial p_i^{ECM}}{\partial t} = \sum_{m=1}^{n_g} c_{ml} g_{im} \tag{B.44}$$

where l is the index of the column in the C matrix corresponding to the effect of regulatory molecules on ECM secretion and p_i^{ECM} is the amount of accumulated ECM products within node i. Once p_i^{ECM} reaches a value of the model parameter M_{ECM} a node is secreted near node i and p_i^{ECM} is set back to 0. The p^{GRD} of the ECM node will be equal to M_{ECM}, correlating in a way the amount of ECM components and the volume of the node. The other components of p^{EQD} will be set to zero. The p^{REC} and p^{ADD} of a new ECM node i are determined at the moment of secretion depending on gene expression in the node that secreted it,

$$p_i^{REC} = \sum_{m=1}^{n_g} c_{mr} g_{i'm} \tag{B.45}$$

$$p_i^{ADD} = p_i^{EQD} + \sum_{m=1}^{n_g} c_{ms} g_{i'm} \tag{B.46}$$

Where i' is the node that secreted node i, r is the index of the column in the C matrix corresponding to the effect of regulatory molecule m on the amount of ECM matter secreted, s is the index of the column in the C matrix corresponding to the effect of regulatory molecule m on the amount of adhesive ECM matter. Note that ECM nodes can not have p^{YOU} or p^{REP} since they do not belong to a cell.

Some regulatory molecules expressed in the cell may be secreted along the ECM, such as adhesion molecules or other compounds that are tightly bound to the ECM fibrous components. Those molecules will be transferred to the ECM node at the time it is secreted. The molecules secreted this way can not diffuse between nodes but can react with molecules that are diffusing between ECM nodes (e.g. bind to extracellular signals). The proportion of regulatory molecule m that would be secreted with the node i and the proportion that would remain in the original node i' are,

$$\begin{cases} g_{im} = c_{mt} g_{i'm} \\ g_{i'm} = (1 - c_{mt}) g_{i'm} \end{cases} \tag{B.47}$$

where t is the index of the column in the C matrix corresponding to the propensity of regulatory molecules to be secreted.

In epithelial cells, the new ECM node is placed at a small distance of the node that secreted it, in the direction of the apical-basal axis (either apical and basal epithelial nodes can secrete ECM). In mesenchymal cells the ECM node is placed in at a random position in the line going from the cell centroid to the node that is secreting the node (see fig. B.10). If the cell has only one node then the ECM node is secreted in a random direction (at a short distance).

ECM degradation can be specified as a gene property in the C matrix. The degradation of ECM nodes is implemented similarly as in the case of cell death. The presence of a gene product specified as an extracellular proteases in an ECM node will promote a decrease in the node's size at a certain rate. When the ECM node reaches a minimum size, it disappears. The rate at which an ECM node shrinks due to protease mediated degradation is,

$$\frac{\partial p_i^{GRD}}{\partial t} = \frac{\partial p_i^{ADD}}{\partial t} = -\sum_{m=1}^{n_g} c_{mg} g_{im} \tag{B.48}$$

Where g is the index of the column in the C matrix corresponding to the effect of regulatory molecules on ECM degradation.

Diffusible extracellular signals. The process of secretion of extracellular diffusible molecules happens when an intracellular gene form is transformed via catalytic activity into an extracellular form. The latter will immediately start to diffuse to other cells or to the ECM. Extracellular diffusible molecules can act as extra-cellular signals by binding to cell surface receptors or by affecting the mechanical properties of the ECM nodes and even mediate its degradation, acting as extracellular proteases (Shapiro, 1998).

B.7 Regulation of cell properties

In the following section we explain how the cell properties $P^{\vec{POL}}$ (cell polarity), P^{PHA} (cell cycle progression), P^{EMT} (EMT transition) are changed.

The cell centroid is also a cell property, it is simply the average of the positions of each of its nodes in 3D space (it is thus a 3D vector). This centroid is calculated in each iteration. P^{MIN} and P^{MAX} are regulated by regulatory molecules,

$$\frac{\partial P_h^{MIN}}{\partial t} = \frac{1}{n_h} = \sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{mq} g_{im} \tag{B.49}$$

Where q is the index of the column in the C matrix corresponding to the effect of regulatory molecules on the minimal number of nodes for cell division P^{MIN}. n_h is the number of nodes in cell h, and it is diving the sum to ensure that just having more nodes does not affect P^{MIN}.

$$\frac{\partial P_h^{MAX}}{\partial t} = \frac{1}{n_h} = \sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{mw} g_{im} \tag{B.50}$$

where w is the index of the column in the C matrix corresponding to the effect of regulatory molecules on the maximal number of nodes allowed before cell division P^{MAX}. n_h is the number of nodes in cell h, and it is diving the sum to ensure that just having more nodes does not affect P^{MAX}.

B.8 Numerical integration

Differential equations are numerically integrated by the explicit Euler method or, optionally, by the explicit fourth-order Runge-Kutta method. These methods can be used with a fixed time step, with an adaptive time step or by a dynamic time step. The logic model parameters L_7, L_{10} and L_{15} specify how the numerical integrations are performed.

The adaptive step-size integration is done by the standard step-doubling procedure. When the integration is not adaptive the value of δ, the integration time step, can be either set constant (when logical model parameter $L_7 = 1$) or dynamic (default, $L_7 = 0$) over time, depending on the maximum node movement length at each time step. The dynamic δ value is calculated as,

$$\delta_t = \frac{M_{DDA}}{|\vec{r_{max}}|} \tag{B.51}$$

where δ_t is the value of δ at time t, $|\vec{r_{max}}|$ is the length of the longest movement vector in the system at that time and M_{DDA} is a model parameter that specifies the value of δ_t when $|\vec{r_{max}}|$ is unity. This ensures that when changes in node positions occur very fast (large $|\vec{r_{max}}|$) the calculations are done with higher accuracy (smaller δ_t). In that sense M_{DDA} also specifies how accurate the calculations are (higher accuracy when low M_{DDA} values).

We further control the value of δ by setting its maximum value, model parameter M_{DMA}. If δ is larger than that value, δ is set equal to that value. At the same time if δ is smaller than the model parameter M_{DMI} (a model

parameter) then δ is set equal to that value. Real time increases by δ_t per time iteration.

The numerical integrations can also be done with a constant δ by setting the logical parameter L_{10} to 1. In that case δ is equal to M_{DMI}.

B.9 Model parameters

B.9.1 Numerical model parameters

- M_{TEM}. Temperature analogue, this is how much noisy movements that are energetically unfavourable are likely to happen (see section B.3.4).
- M_{NOI}. Proportion of nodes to which noise is applied in each iteration. If δ is dynamic this proportion is weighted by δ_t (see section B.3.4)
- M_{MCO}. Maximal compression allowed in a cell to allow growth in it (see section B.6.3).
- M_{EMI}. Minimal p^{EQD} allowed.
- M_{EMA}. Maximal p^{EQD} allowed.
- M_{MAE}. p^{EQD} all nodes in a cell should have before adding a new node.

  Implementation model parameters. These are parameters controlling the numerical implementation of the model, this is the accuracy of the model. They have no biological meaning as such.
- M_{DIF}. Maximum radius of diffusion. By default its value is equal to 2. Values larger than that have a negligible effect on accuracy and largely decrease the speed of the model.
- M_{MID}. p^{EQD} given to the new added nodes by growth. Any value that is small compared to the average p^{ADD} of nodes (or compared to M_{MAE}) would produce the same model dynamics.
- M_{ECM}. Amount of extra-cellular matrix that has to accumulate in a node before an ECM node is secreted. This essentially controls how much ECM there needs to be for the model to consider that ECM as a node.
- M_{DMI}. Minimum δ. In any case, the numerical integration step may not be below this value. The lower this value is, the more accurate are model calculations but the slower the model would be. This parameter is only meaningful if delta is dynamic ($L_7 = 0$, that is the default).
- M_{DDA}. Accuracy of the numerical integration. This parameter is only meaningful if δ is dynamic ($L_7 = 0$, that is the default).
- M_{DMA}. Maximal value of δ allowed. This parameter is only meaningful if δ is dynamic ($L_7 = 0$, that is the default).

- M_{MNN}. Maximal number of nodes any node can interact with. If there is more than that number the program crashes. These is no optimal way to avoid that effect, since these neighbours need to be stored in a temporary matrix and there are system restrictions in the size of those. In addition, there is no way to predict how many nodes a node will interact with since this is a result of model dynamics. If this value is large the program would run slower.

- M_{AMX}. In the mechanical interaction between two cylinders, the minimum curvature in order to apply the epithelial rotational force and the epithelial surface tension force (see section B.3.2).

- M_{RMA}. Maximum node length of movement when δ is dynamic ($L_7 = 0$).

- M_{MAN}. Maximum number of nodes allowed in a simulation. Only applicable when ($L_2 = 1$).

- M_{GAB}. Sets the size of the exclusion sphere when using the Gabriel method to build the node's neighbourhood (only when $L_3 = 1$).

- M_{DFE}. Maximum p^{EQD} allowed for an epithelial node due to deformation (see section B.5.3)

There also some model parameters that only apply if some logical model parameters are set to 1. This is they have no meaning in the default version of the model but are relevant for some alternative versions of the model activated by some logical model parameters. They are explained in the next section.

B.9.2 Logic model parameters

Here we describe their non-default values and how these change the functioning of the model. All logical model parameters are by default set to 0 and that defines the canonical version of the model. By setting any of those parameters equal to 1 different variations of the model can be activated (all of them are only small variations). These options can be manually altered in the input and output files or can be edited through the gene network viewer.

- L_1. If set to 1 the model considers that each mesenchymal cell is made of a single node and each epithelial cell is made of a single cylinder. The initial conditions have to be designed consistent with that (each mesenchymal cell should have a single node and each epithelial cell a single cylinder). When this is the case slightly different rules apply in some cell behaviours.

  In the case when cells are composed of one node/cylinder, internal asymmetries cannot arise (except in epithelial cells, but only in the apical-basal axis), thus polarization has to be determined by the molecules present in the surrounding cells. In this case, the cell will polarize in the direction of a tissue-level molecular gradient. Thus, the s scores (equation B.33) are not calculated for the nodes within the same cell

(since there is only one, or two) but for the nodes of the neighbouring cells. Then, $\vec{P^{POL}}$ is calculated as in equation B.34 for the non single-node case using the s scores of the neighbouring cells.

Cells are normally supposed to double their size before dividing. Thus, in the case when cells are composed of one node/cylinder they only need to add one node/cylinder before dividing. In this case, cell growth happens at the same time as cell division, by adding a new node/cylinder and then splitting the cell in two (see section B.6.4).

In the case when cells are composed of one node/cylinder, they doesn't need to reach a minimum size nor a maximum size in order to divide. When P_h^{PHA} reaches the value of one, a new node/cylinder is added in the direction of the division vector and the cell is split in two, each node/cylinder belonging to a different cell.

The division vector, as explained above, depends on the weighted sum of $\vec{P^{POL}}$ and $\vec{P^{HER}}$, and in this case $\vec{P^{HER}}$ has to be calculated differently, since the shape of the cell cannot be determined by the relative position of its nodes (since there is only one). In this case we assume the shape is determined by the distances between the cell and its neighbours. If the distance between the cell and some of its neighbours is longer in a certain direction we assume that the cell is elongated in that direction. Thus, for simplicity, P_h^{HER} will be equal to the vector connecting the two neighboring cells which are farthest from one another.

Chemotaxis cannot be implemented as explained above when cells are composed of one node/cylinder. In this case, cells may move in a random fashion, but biased by the direction of cell polarization. Expression of certain regulatory molecules may determine the strength of this bias,

$$\vec{r_i^{noise}} = \vec{X} + P_h^{POL} \sum_{m=1} n_g c_{mb} g_{im} \tag{B.52}$$

where $\vec{r_i^{noise}}$ is the movement vector of node i by noise, $\vec{X}$ is a random unit vector with a spherical distribution, P_h^{POL} is the cell's polarization vector, c_{mb} is the element of the C matrix specifying the effect of regulatory molecule j on chemotaxis and g_{im} is the expression of regulatory molecule m on node i.

- L_2. If set to 1 stops the simulation after a threshold number of nodes are reached. This number is an optional model parameter, M_{MAN}.
- L_3. If set to 1 the Gabriel algorithm to determine which nodes interact with each other is used. This allows nodes to screen the interaction between other nodes (as explained in above). This option is often used combined with L_{11} set to 1.
- L_4. If set to 1 it disables epithelial bending forces.
- L_5. If set to 1 stops the simulation when any node gets a p^{EQD} or p^{ADD} value three times larger than its original value.

- L_6. If set to 1 forces apoptosis of all cellular nodes that are not interacting with any other one after a number of iterations determined by M_{TAL}, a model parameter.
- L_7. If set to 0 it uses explicit Euler method for the numerical integration, if set to 1 it uses explicit fourth-order Runge-Kutta for node movement and explicit Euler for the rest of equations. If set to 2 it uses explicit Runge-Kutta for node movement and regulatory molecule equations.
- L_8. If set to 1 stop the simulation once all cells are fully differentiated.
- L_9. If set to 1 it allows node plasticity.
- L_{10}. If set to 1 fixed value δ is used instead of dynamic δ.
- L_{11} If set to 1 calculates the neighbourhood between nodes by using the Delaunay tesselation (see section B.3.1). Otherwise neighbourhood is determined simply by distance.
- L_{12}. If set to 0, the mechanical interaction between two cylinders from the apical/basal side is calculated as described in section B.3. If set to 1 it is calculated as if they were two mesenchymal nodes (see section B.3).
- L_{13}. If set to 1 volume conservation in cylinders is implemented.
- $L14$. If set to 1 the diffusion of p^{EQD} components is allowed. This is there is diffusion of p^{GRD}, p^{COD} and p^{PLD} between mesenchymal nodes from the same cell and between epithelial nodes from the same side. This simply reflects that since the cytosol and membrane of different parts of a cell communicate there would be a natural redistribution of matter between cell parts, this is between nodes. This redistribution process is analogous to a diffusion in the sense that the flux of matter would be from nodes with higher values on those properties to nodes with lower values. The diffusivity of that process is model parameter, M_{DID}.
- L_{15}. If set to 1 uses adaptive step-size and fourth-order explicit Runge-Kutta numerical integration for node movement. If set to 2 it uses it also for regulatory molecules equations.
- L_{16}. If set to 1 the simulations are run by allowing more than one node to be added per cell per iteration due to growth.
- L_{17}. If set to 1 and dynamic δ is used returns the control to the user in the embryo display after the number of real time units that have been run (by default returns the control after the specified number of iterations and not of real time units). This option is only valid if the model is run with the user graphical interface (see EmbryoMaker user manual).
- L_{18}. If set to 1 noise is implemented without considering energies.

- L_{19}. If set to 1 runs the model only by energy biased noise (Monte Carlo Method). This in general does not change model outcomes but makes the simulations much more slower.

- L_{20}. If set to 1, epithelial nodes from one side may consider as neighbours epithelial nodes from the opposing side. By default those neighbour connections are not considered. This may be relevant when two different epithelial surfaces are close to each other, but it is not necessary when the same epithelium folds over itself, since the nodes in contact belong to the same surface.

- L_{21}. By default ($L_{21} = 0$) single element cells divide by adding a new, full sized cell next to it. When $L_{21} = 1$ the new cell has a smaller size and then gradually grows until it has full size.

B.10 Implementation of the model in the EmbryoMaker software

B.10.1 Structure of the code

The source code is written in fortran90 and is organized in different functional fortran modules. The most relevant modules are listed below.

- *general.mod.f90*. Declarations of the main variables that are used in common by the rest of the modules. These are global model parameters and node and cell properties. The set of all node properties are declared in a derived type fortran 90 variable. The same occurs for cell properties. The main variables used by the other modules are a matrix of nodes and cells properties (one element per cell and node). Essentially the rest of the code is mostly operations on those matrices (including re-dimensioning them).

- *model.mod.f90*. Manages the temporal progression of the developmental simulations and calls the subroutines in the neighbouring, biomechanical, genetic and nexus modules (once per iteration with Euler and several times with Runge-Kutta).

- *neighboring.mod.f90*. Contains the subroutines to calculate the neighbour relations between nodes.

- *biomechanic.mod.f90*. Contains the subroutines that calculate the mechanic interactions and displacement of nodes.

- *energy.mod.f90*. Contains the subroutines that calculate energy potentials for nodes that are used in energy-biased random movements.

- *genetic.mod.f90*. Declares the regulatory molecules and their parameters used in a specific instance of the model. It also contains the subroutines for transcription and non-transcriptional regulation.

- *nexus.mod.f03*. Contains the subroutines that implement the molecular regulation of node and cell properties and the calls to the cell behaviours. It also contains subroutines for some simple cell behaviours.

- *growth.mod.f90*. Contains the subroutines that implement cell growth.
- *death.mod.f90*. Contains the subroutines that implement cell death or apoptosis.
- *mitosis.mod.f90*. Contains the subroutines that implement cell division.
- *ecm.mod.f90*. Contains the subroutines that implement secretion of extracellular matrix.
- *single_node.mod.f90*. Contains certain subroutines that are used in the case cells are composed of one node.
- *pinta.mod.f90*. This file contains two modules: a view_modifier module to control how the embryo is seen (rotation, zooming, sectioning, etc.) and a function_plotter module that contains all the subroutines that draw nodes and controls the menu. This latter module is the one including the OpenGl and glut calls.
- *editor.mod.f90*. Contains the code required to manually edit the embryo.
- *ic.mod.f90*. Contains a set of subroutine for simple initial conditions.
- *initial.mod.f90*. Contains several initialization subroutines.
- *io.mod.f90*. Contains the hard-disc input/output subroutines.
- *OpenGl_gl.f90*, *OpenGl_glu.f90* and *OpenGl_glut.f90* define fortran interfaces for the OpenGl, GLU and glut functions, and have been taken from the f03gl project (`http://www-stone.ch.cam.ac.uk/pub/f03gl/index.xhtml`).

B.10.2 Input/Output format

EmbryoMaker and NetworkMaker use a custom I/O format. The same file written by the program as output can be read as input file as well. It is basically a text file listing all the model parameters and variables that are used by the software, including node positions and gene expression levels. The names of the parameters and variables are indicated in the file, so it is possible to edit the file manually. In that sense, both the editor tools of EmbryoMaker and NetworkMaker can be used to edit Input/Output files with a more intuitive graphic interface. By default, as explained in the manual, EmbryoMaker writes all the output files from a given run into a folder with a number (a different number for each run) within a folder called output.

B.11 Detailed explanation of the developmental mechanism shown in figure 4.5

Figure 4.5 shows an example of a combination of basic developmental mechanisms (contraction, polar growth, ECM secretion and hierarchic inductive mechanisms). The figure shows how from the network depicted in the left and the initial conditions in the center left (time zero) the patterns in the

right of the figure will arise over time. It is important to note that nothing else than this network and the initial conditions are specified to the model (there are no pre-patterns or changes in the rules of the model over time). The developmental patterns shown in the right of the figure simply arise from model dynamics. These pattern transformations can be explained qualitatively. The initial pattern consists of a hollow spherical epithelium (in the simulation shown in the figure each element is a cell) in which one cell expresses transcription factor 1 (TF1) (all other regulatory molecules are either not expresses or expressed homogeneously in all cells in the initial condition). TF1 promotes an epithelial-mesenchymal transition (EMT) and thus the single cell expressing that gene detaches from the epithelium and moves randomly in the interior of the blastula. TF1 activates the transcription of TF6, a gene that promotes cell motility, cell proliferation and extracellular matrix secretion. TF1 promotes the production and secretion of growth factor 1 (GF1). As a result, while this EMT is still taking place, GF1 reaches the nearby epithelial inner surface. All epithelial cells express a receptor for GF1 (R1) at the same level in the initial conditions. Thus, GF1 bind its receptor in the epithelial cells that are close enough to the proliferating mesenchymal cells (daughters of the cell originally expressing TF1 and that thus also express GF1). The activated receptor (RGF1) activates the transcription of TF2. TF2 activates cell contraction in the outer surface of the epithelium (this is a decrease in the p^{EQD} in these nodes), thus mediating a slight invagination in the epithelium next to where the mesenchymal cells are. TF2 also mediates the production of a second growth factor (GF2). In this case, GF2 is only secreted in the outer surface of the epithelial that end up expressing it. GF2 binds to receptor R2 which is expressed in all epithelial cells from the initial conditions. The activated receptor RGF2 activates the transcription of TF3, which mediates the increase in size of the outer side of the epithelial cells where it becomes expressed. Since TF2 strongly inhibits transcription of TF3, this one will be expressed only around the territory where TF2 is expressed, but without overlap. This means that while contraction in the TF2 territory promotes a concavity in the epithelium, TF3 mediated expansion in the surrounding cells will have the opposite effect, promoting a convex curvature surrounding the TF2 mediated concavity. Also, since TF3 expression relies on GF2 signalling, this means that the farther from the TF2 territory (GF2 source) the lower the concentration of TF3, thus creating a gradient of TF3 along the whole epithelium, being highest close to the TF2 territory and lowest at the opposite side. TF3 also promotes cell cycle progression where it is expressed, meaning that cells close to the TF2 territory will divide more rapidly than the ones farther from it, but not the ones within the concavity. TF3 also inhibits the transcription of another transcription factor (TF5) that is homogeneously expressed in all the epithelium. Thus, TF5 forms a gradient opposite to the one formed by TF3, meaning that is lowest near the TF2 territory and highest in the opposite side. TF5 also promotes polarization of cells, meaning that cells will become polarized along the gradient formed by TF5. This means that the cells located on the opposite side of the TF2 territory will divide in the direction of this gradient, thus promoting an elongation of the whole embryo in that direction. Close and within the invagination this oriented proliferation leads to a deepening of the invagination towards the inside of the embryo.

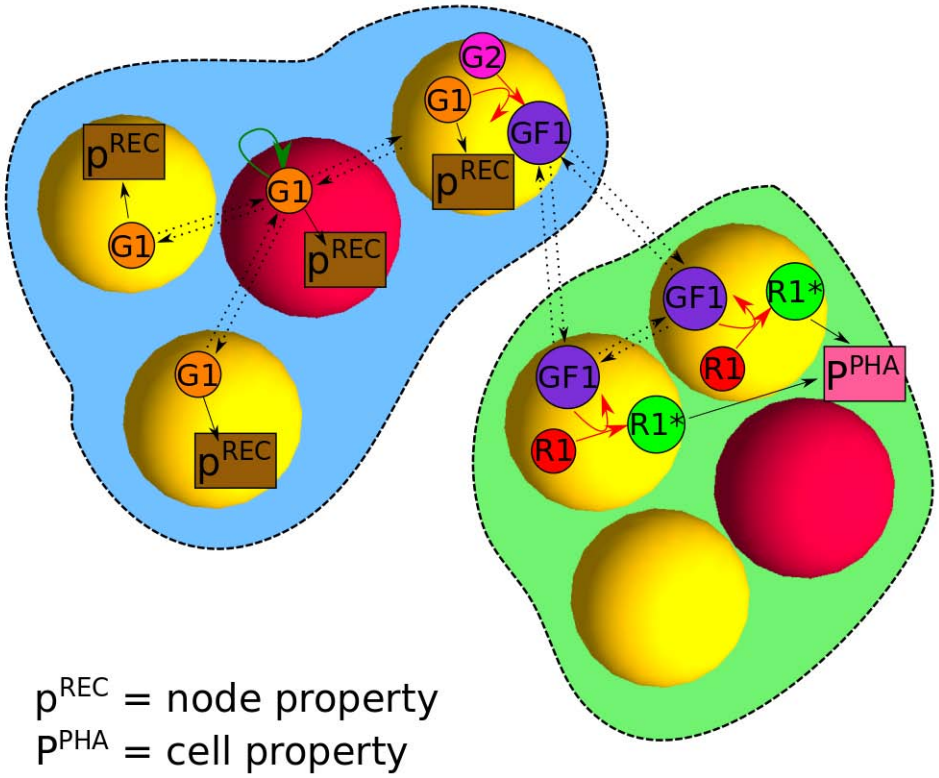

FIGURE B.1: On the macroscale, cells are represented by sets of integrated mechanical bodies (subcellular elements) that physically interact with each other by means of different mechanic equations. The mechanic coefficients used in these equations are affected quantitatively by the amount of certain regulatory molecules present on each node. The amount of different types of regulatory molecules within each node is determined by the microscale dynamics, which include gene transcription, enzymatic reactions, spatial diffusion and molecule degradation. Here, two different cells are shown. Subcellular elements, or nodes, are represented as spheres. All nodes belonging to the same cell are enveloped by a dashed line. The red node in each cell contains the cell nucleus, thus only in those nodes transcription will take place. Regulatory molecules are indicated as coloured circles with a letter and a number: G for a generic regulatory molecule, GF for a secreted growth factor, R for a receptor of secreted growth factors and R* for the activated receptor (the receptor bound to its growth factor ligand). Node properties are indicated with a lowercase p within a rectangle and cell properties are indicated with an uppercase P within a rectangle. In the figure we put the example that the P^{REC} node property and P^{PHA} cell properties are affected. Green arrows represent transcriptional regulatory interactions (see equation 4.4 in the main text), red arrows indicate enzymatic reactions (see equation 4.5 in Chapter 4), black solid arrows represent regulation of node or cell properties (see equation 4.6 in Chapter 4) and doted arrows indicate diffusion between nodes (equations B.21 and B.22).

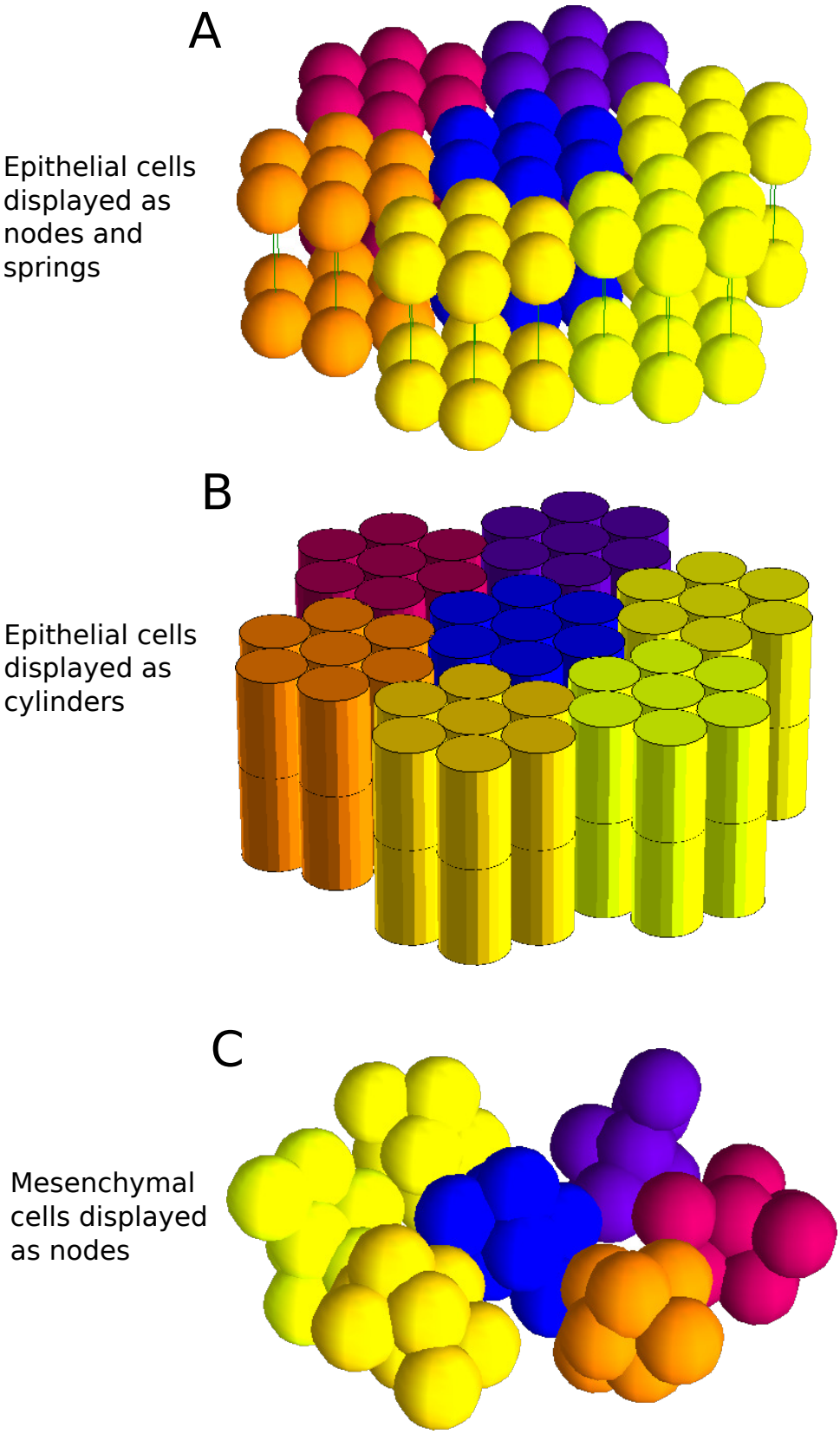

FIGURE B.2: A, a small epithelial sheet made up of 7 epithelial cells, each one made up of 14 nodes. Spheres represent nodes and green lines represent the springs connecting two nodes forming a cylinder. Nodes of the same color belong to a single cell. B, the same as A, but displaying the form of the cylinders. C, a small group of 7 mesenchymal cells made up of 10 nodes each cell.

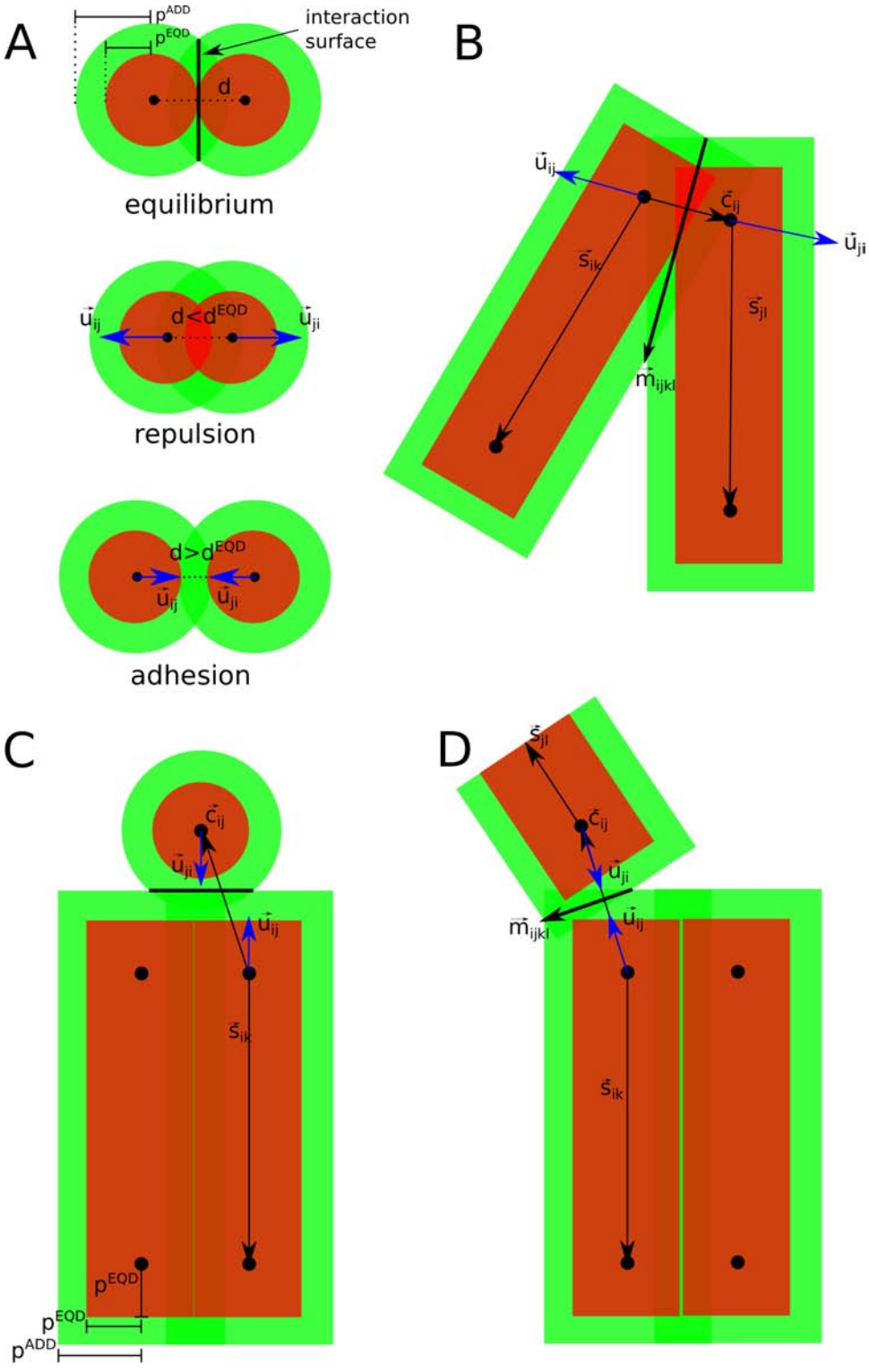

FIGURE B.3: A, when two spheric nodes (either mesenchymal or ECM) are at a distance closer than d^{ADD} they feel either an attractive force if they are closer than d^{EQD}, or a repulsive force if they are farther than d^{EQD}. The direction of the force goes from the center of one node to the center of the other. B, D, when two cylinders interact the same repulsive and attractive forces act, but the direction of the force is always normal to the contact surface of the cylinders. C, When a spheric node interacts with a cylinder apical or basal face, the direction of the force is always parallel to the apical-basal axis of the cylinder.

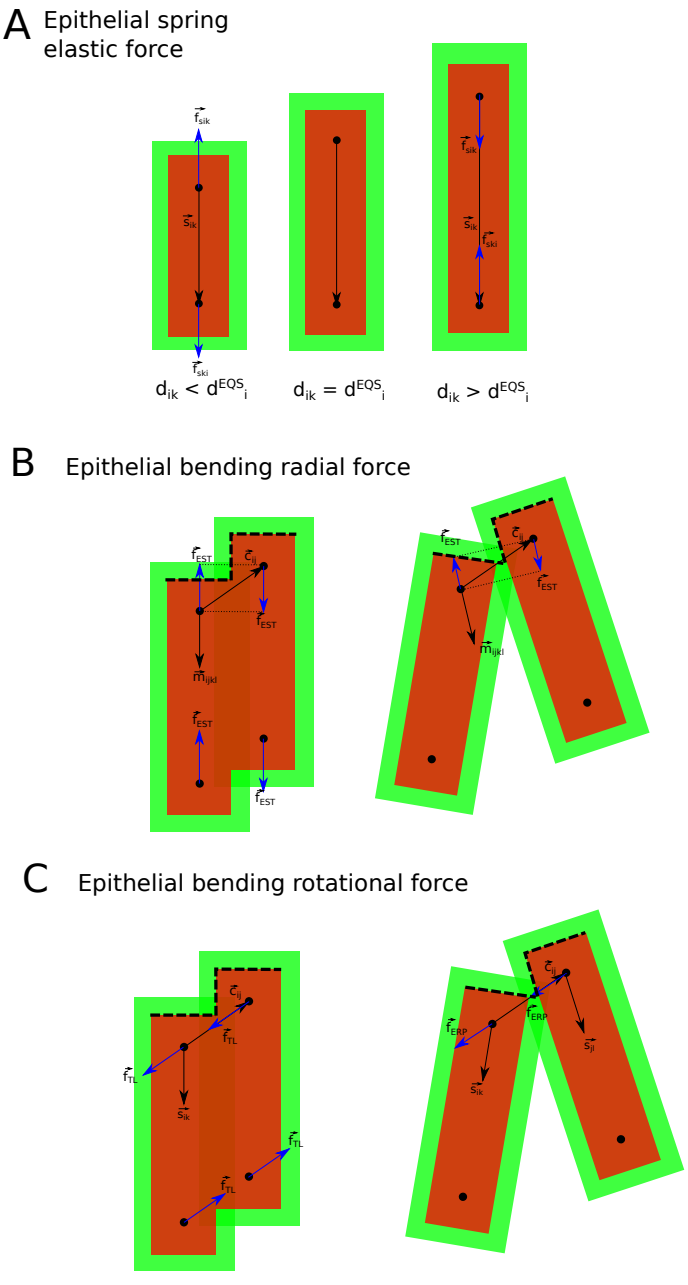

FIGURE B.4: A, the two nodes composing a cylinder are connected by an unbreakable spring. Elastic forces will always follow the direction of that spring. B, C, epithelial bending forces tend to put two cylinders in a position in which the angle between the vector connecting the two apical (or basal) nodes and the apical-basal axis is $\pi/2$. B, The bending radial force applies on a direction normal to the apical/basal surface. C, The bending rotational force applies in the direction connecting the two epithelial nodes from the same side.

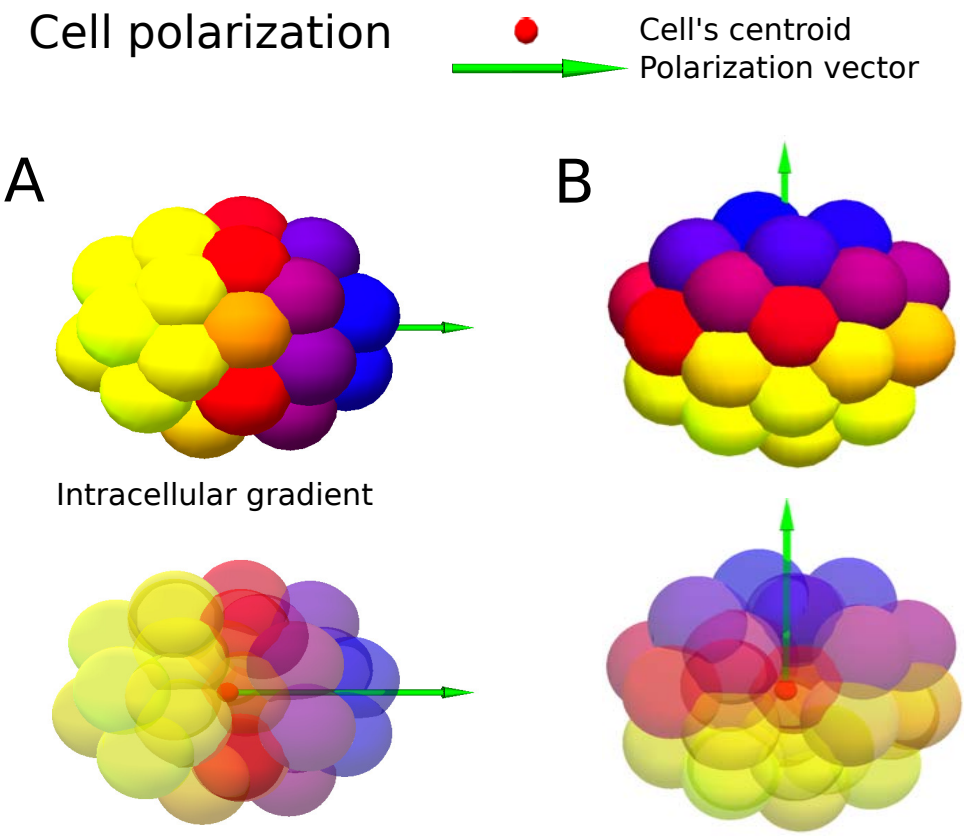

FIGURE B.5: Polarization is calculated as the sum of node-centroid vectors weighted by the difference in molecular concentration respect the actual geometrical centroid of the cell. In other words, the polarization vector (green arrow) will mostly point towards the part of the cell where the concentration of the molecule is higher (A and B, blue shaded nodes). As B shows, polarization vector is independent of cell shape.

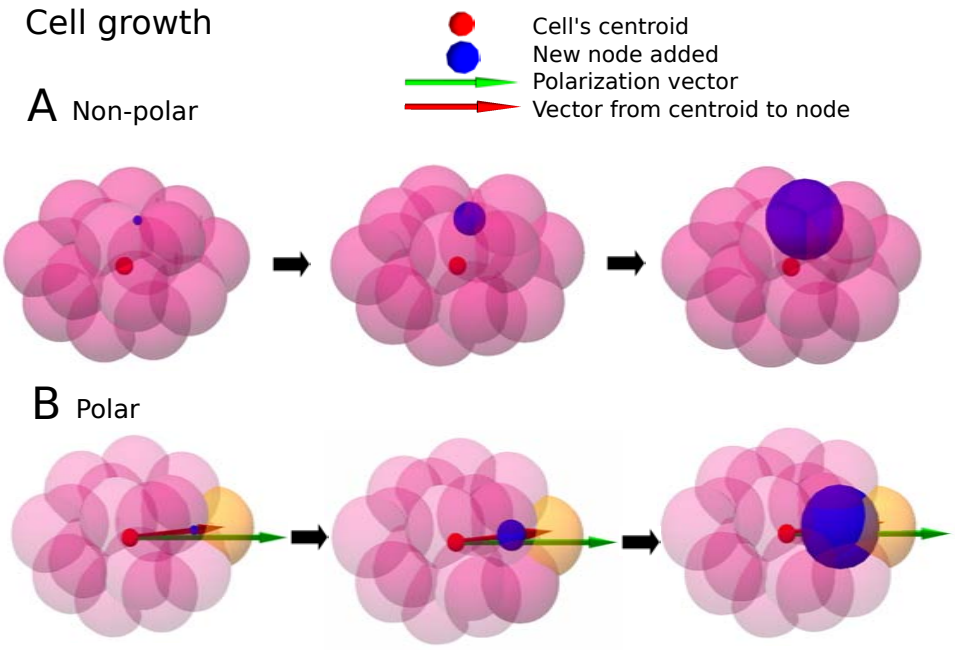

FIGURE B.6: A, non polar growth. A new node (blue) is added in a random place within the cell. B, polar growth. The most external node in the direction of the polarization vector is chosen. Then a new node is added at 80% of the distance from the cell's centroid to that node. In epithelial cells the same is done, but just using the basal layer of nodes, and then completing the cylinder adding one more node to the corresponding position in the apical layer.

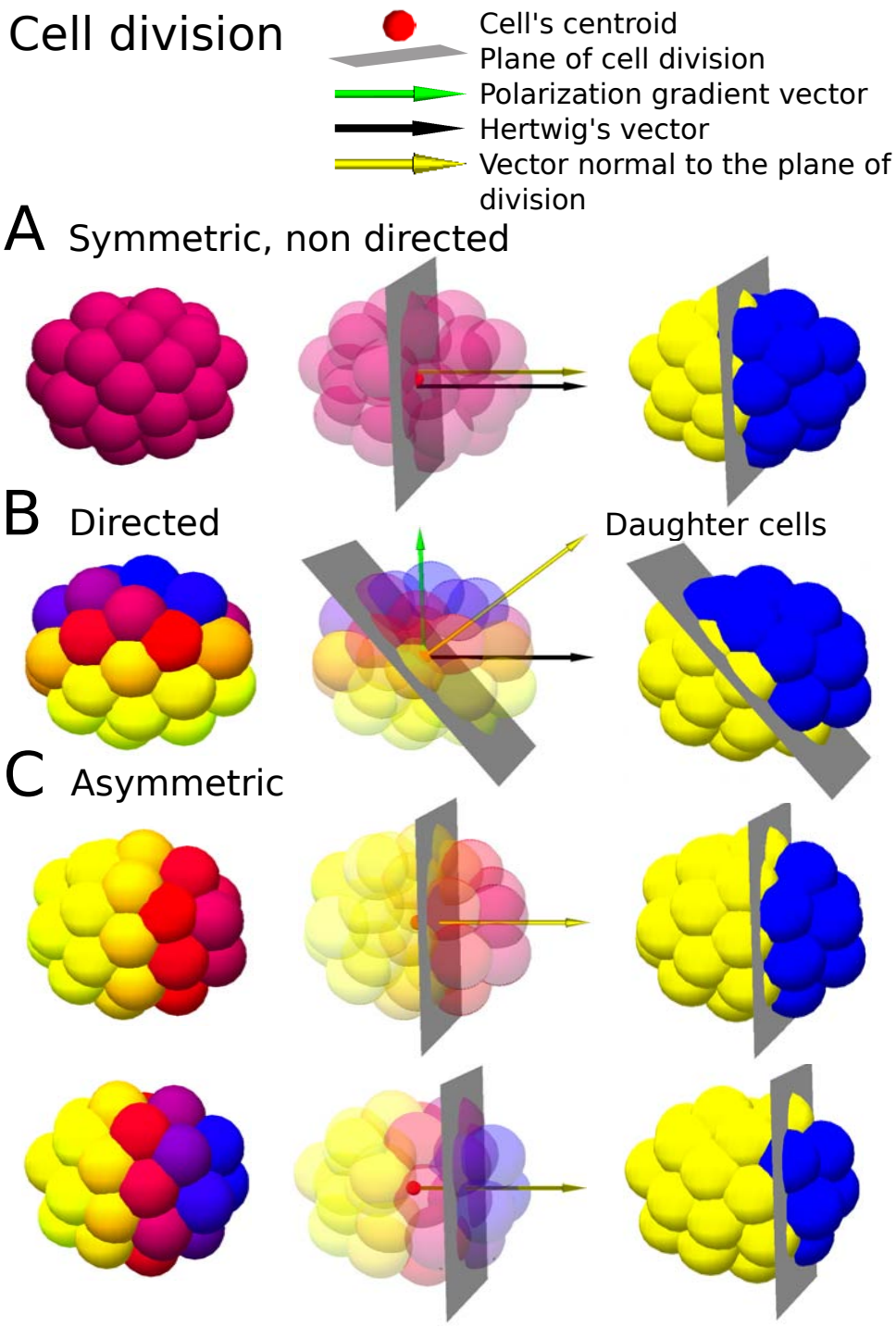

FIGURE B.7: A, symmetric, non-directed. The plane of division passes through the cell's center of mass, or centroid (small red solid ball) and the vector normal to that plane (yellow arrow) is the longest axis of the cell (Hertwig's or shape vector, black arrow). B. Directed mitosis. The plane of division passes through the cell's centroid, but the direction of the plane is the weighted sum of two vectors, a cell polarity dependent vector (green arrow), and the Hertwig vector. C, in assymetric cell division, once the vector normal to the division plane (yellow arrow) has been established, the plane of division is set away from the centroid of the cell so one daughter cell is larger than the other. The displacement of the plane depends on the direction and intensity of a particular spatial molecular gradient within the cell. B and C can be combined to get directed assymetric mitosis. In all cases, once the plane is set, the nodes at each side of it will be assigned to a single daughter cell.

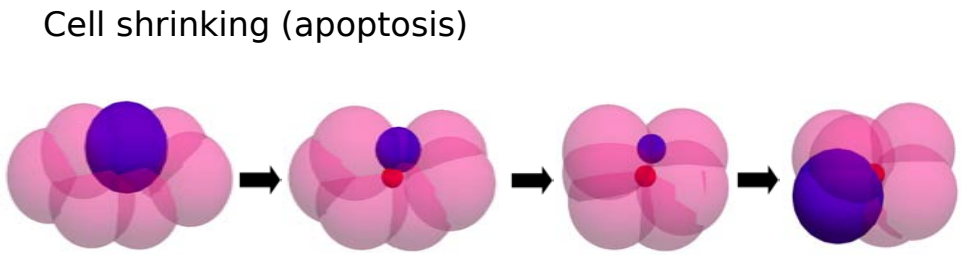

FIGURE B.8: When there are gene products promoting apoptosis within a cell there is a continued decrease in size by shrinking one node at a time (in blue). When the node reaches a minimum size (a model parameter) it disappears and another random node begins to shrink (rightmost picture). This process keeps going until the last node in the cell disappears and so does the cell.

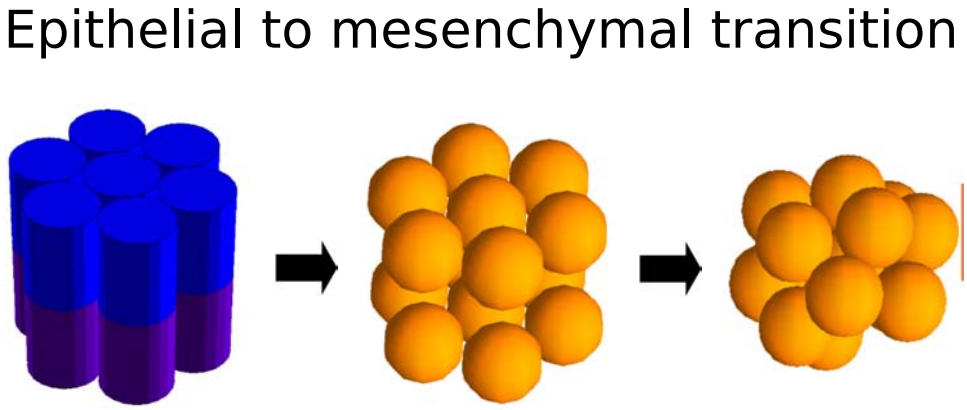

FIGURE B.9: An epithelial cell, represented as a set of cylinders is transformed into a mesenchymal cell by turning each cylinder into a pair of spheric nodes. Since any spring or surface tension forces are exerted on the nodes once they undergo the EMT, the nodes lack the typical arrangement of epithelial cells and become more irregular (right).

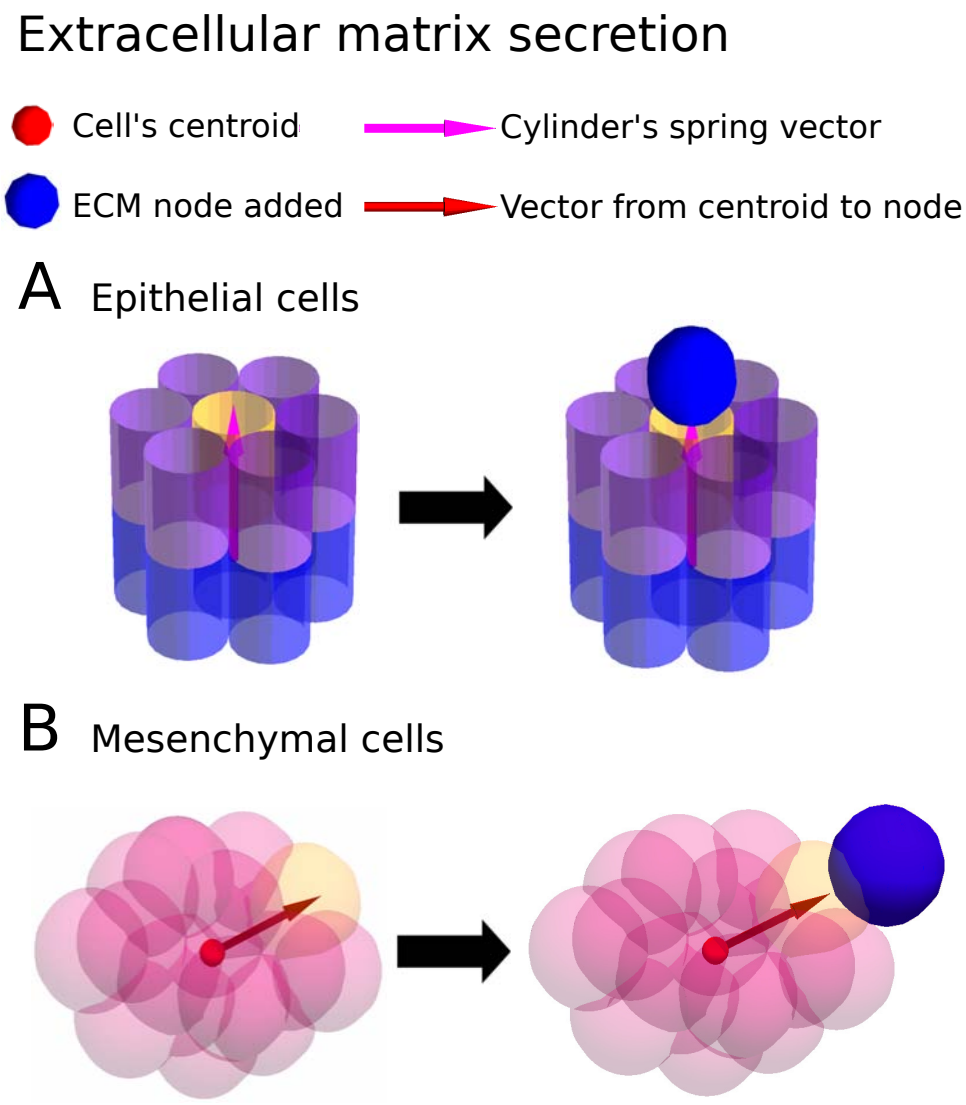

FIGURE B.10: ECM is represented as free spheric nodes (depicted in blue) that can be secreted by cells. A, when an epithelial cell secretes an ECM node, it appears close to the tip of the cylinder (in yellow) where the ECM products were accumulating. B, when a mesenchymal cell secretes an ECM node, a vector from the node to the centroid of the cell is chosen (red arrow), next the node closest to that vector and farthest from the centroid (in yellow) is selected. Then, an ECM node is added close to that node and in the opposite direction from the centroid of the cell. That way, intracellular deposition of ECM is prevented.

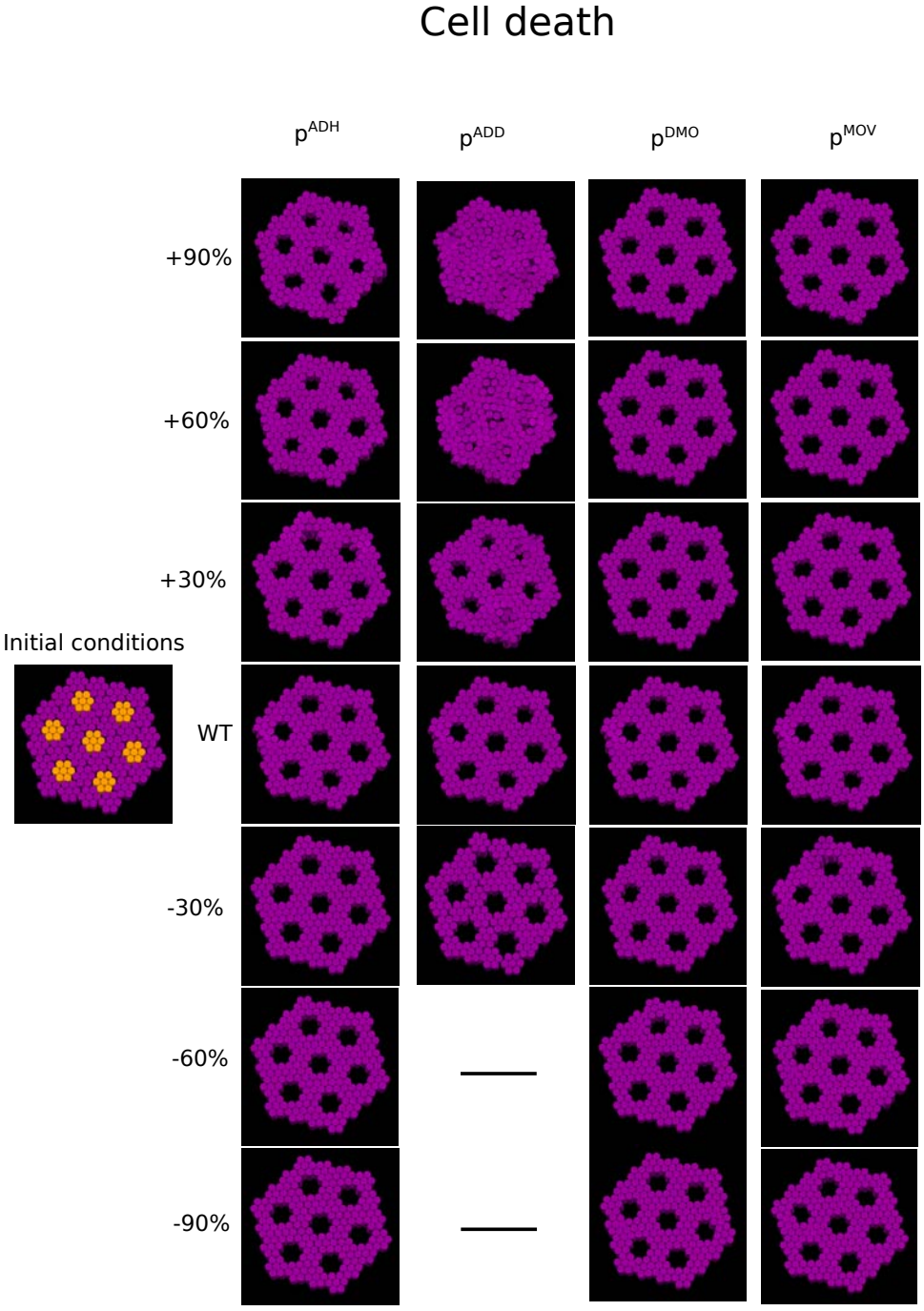

FIGURE B.11: Different simulations were run starting from the same initial conditions (left picture) except for one parameter that was modified. Each column shows the final phenotype when a single parameter is modified by a percentatge relative to the value of the wildtype (central row of pictures). Seven non-adjacent cells of an epithelial sheet, each one composed of seven cylinders, express an apoptotic factor (marked in orange in the initial conditions) promoting cell death. After the cell death, the existence and shape of the resultant gaps in the final patterns depend of many parameters. For example, we show in this simplified parameter space that the remaining cells are able to close the gaps if the node's radius (p^{ADD}) is large enough as to interact with the nodes in the other side of the gap. Variations in the remaining node properties (intracellular adhesion, radius and filopodia unstability, respectively p^{ADH}, p^{ADD} and p^{MOV}) do not seem to have a strong effect on final phenotypes. In this genotype map, as well as in many following ones, the lack of final phenotypes corresponding to -60% and -90% decreases of maximum radius of interaction between nodes stems from the fact that in those cases, the radius of interaction is lower than the radius of equilibrium, thus preventing interactions between nodes.

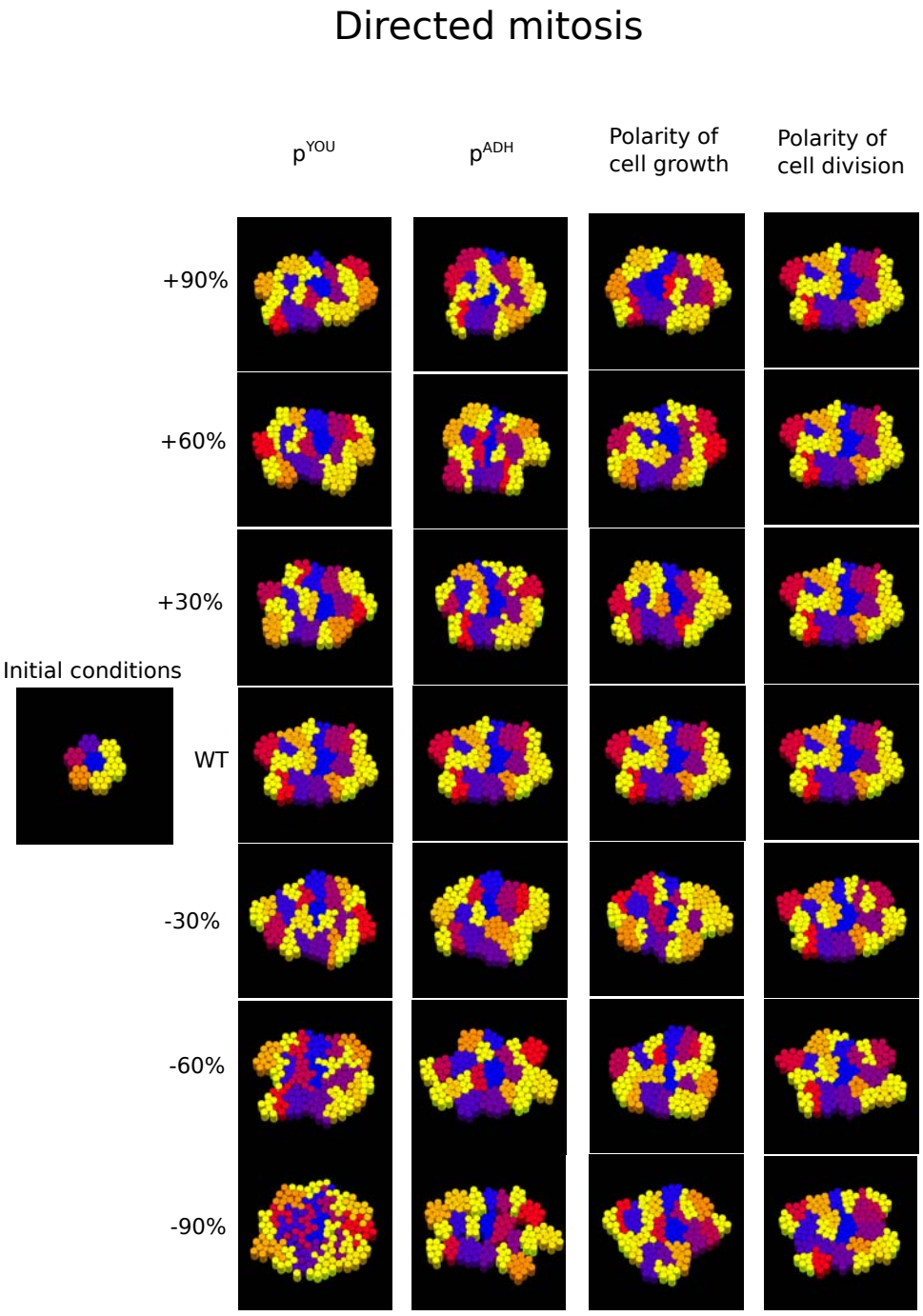

FIGURE B.12: Different simulations where run starting from the same initial conditions (left picture) except for one parameter that was modified. Each column shows the final phenotype when a single parameter is modified by a percentage relative to the value of the wild type (central row of pictures). A small epithelial sheet composed of seven cells (each one depicted in a different colour) is allowed to grow in a medium displaying a chemical gradient (from left to right, not shown). When cells grow, new nodes are preferentially added in the direction of the gradient and, when cells divide, the plane of cell division is also oriented according to that gradient. The combined effects of these mechanisms produces that epithelium gets elongated in the same direction of the gradient. This elongation is more pronounced under some parameter combinations, but fades out under others (e.g. reduced cell-cell adhesion: p^{ADH} < wild type p^{ADH}). Irregular cell shapes found in (p^{YOU}< 60%) stem from the fact that in these cases the intracellular adhesion between nodes practically equals extracellular one, so the cohesive forces keeping cell integrity vanish.

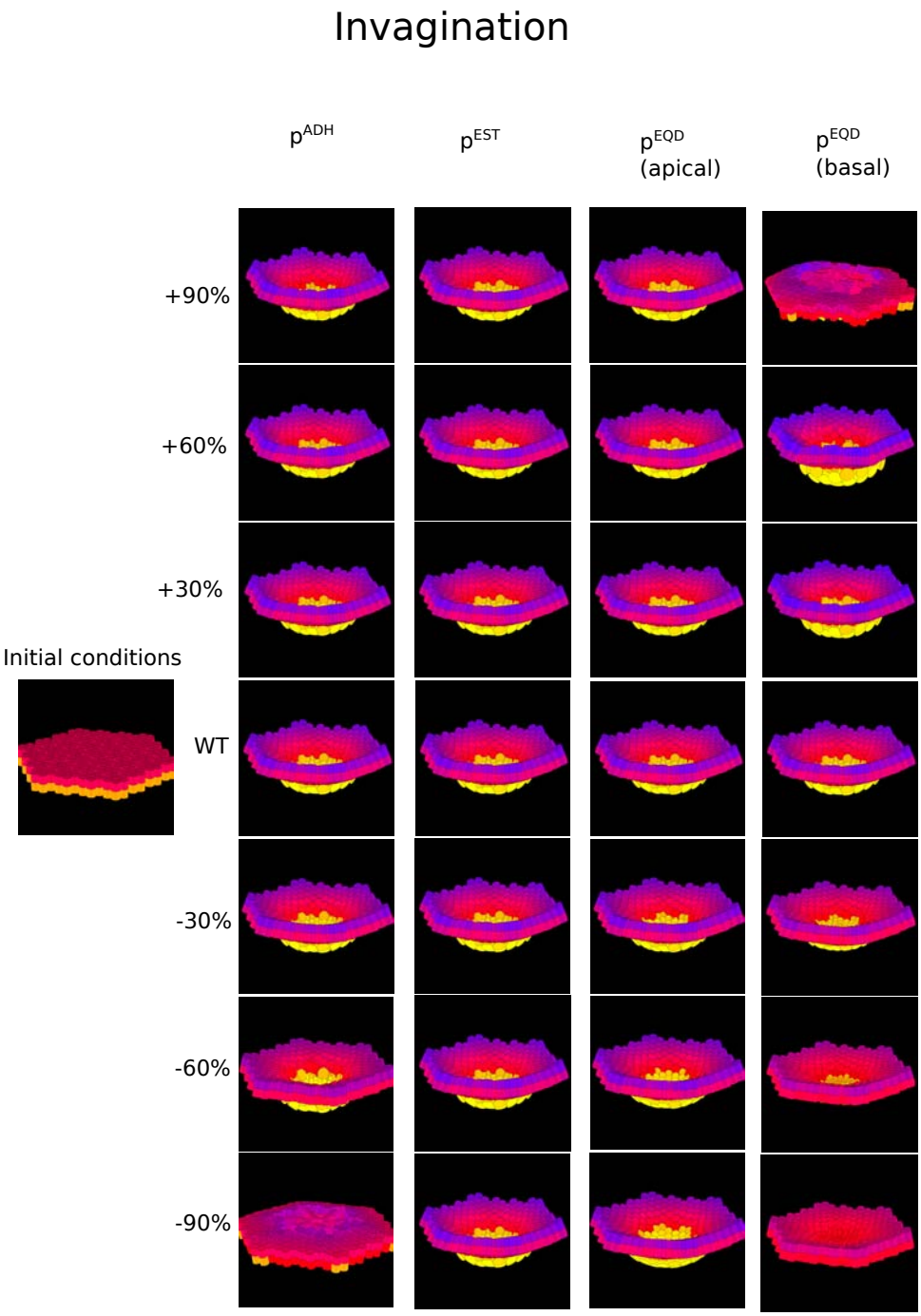

FIGURE B.13: Different simulations where run starting from the same initial conditions (left picture) except for one parameter that was modified. Each column shows the final phenotype when a single parameter is modified by a percentage relative to the value of the wild type (central row of pictures). An hexagonal epithelial sheet invaginates when cells located in its central area perform apical cell contraction. However, the depth of such invagination varies according to many parameters (in this plot, the deeper nodes are from their initial conditions position, the more yellow they are coloured). Under some parameter combinations, invagination does not occur. This happen when apical contraction is too low ($p^{EQD} <$ 90%), or when epithelial cells lack cohesivity between them (either by decreased cell-cell adhesion ($p^{ADH} <$ 90%), or by extreme apical contraction ($p^{EQD} >$ 90%)). Radial component of epithelial surface tension (p^{EST}) and equilibrium distance (p^{EQD}) seem to have only a mild effect on apical contraction, at least for the range of values presented in here.

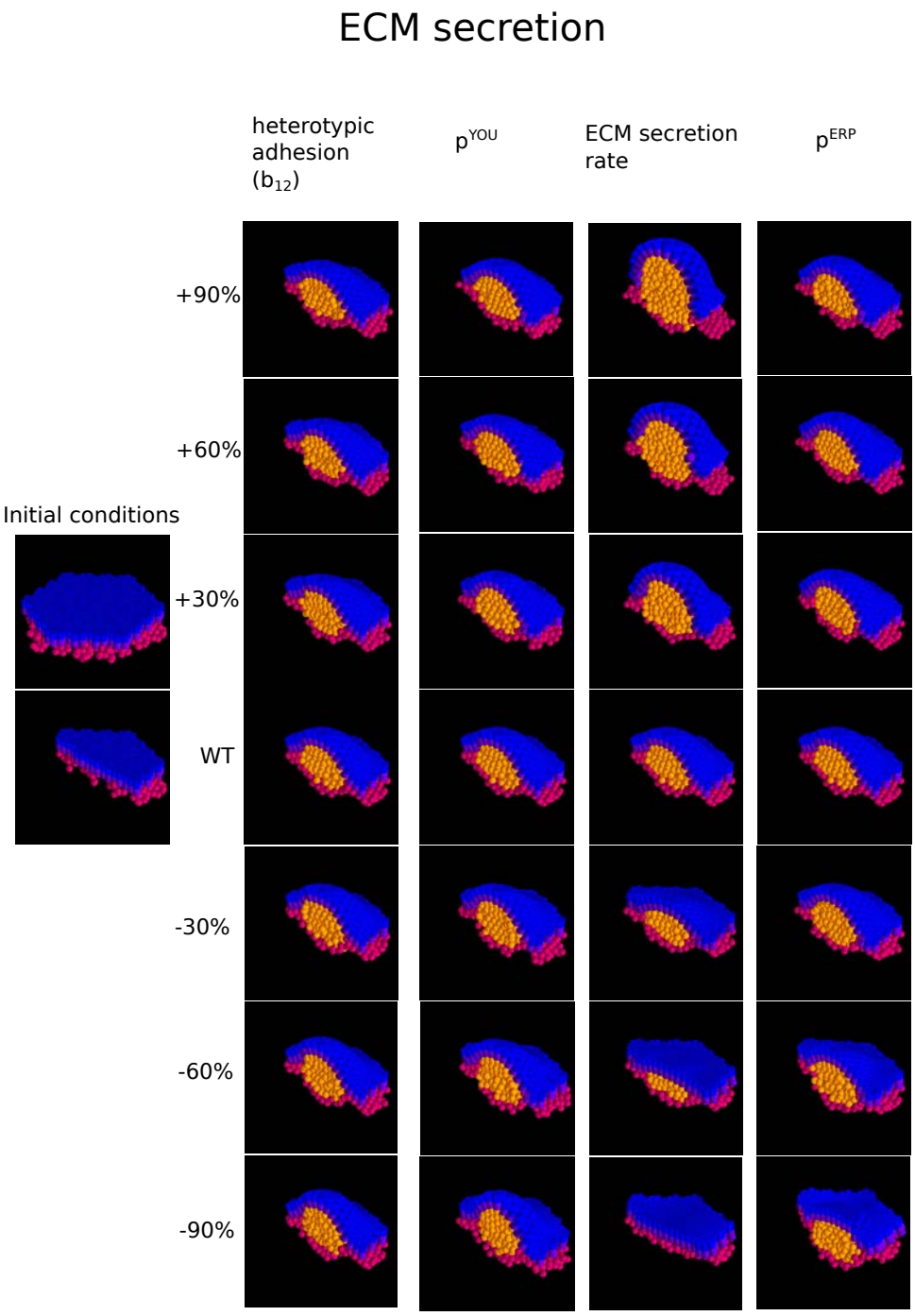

FIGURE B.14: Different simulations where run starting from the same initial conditions (left picture) except for one parameter that was modified. Each column shows the final phenotype when a single parameter is modified by a percentage relative to the value of the wild type (central row of pictures). In these simulations, extracellular matrix (orange nodes) is secreted by epithelial cells (in blue) in an extracellular space previously occupied by mesenchymal cells (pink). A section of the system is depicted for a better visualization. The mechanical forces resulting as a result of ECM secretion deform both the epithelial sheet and the underlying mesenchyme, but the relative magnitudes of such deformations depend on many parameters. When heterotypic adhesion is high, ECM nodes tend to spread between epithelial and mesenchymal cells, but they tend to form a rounded aggregate when this adhesion diminishes. Not surprisingly, deformation also strongly depends on the rate of ECM secretion, even deformation is not accomplished if the rate is too low. When the bending accommodation force (p^{ERP}) diminishes, epithelial cells become less adaptive to bending, so epithelia is only slightly deformed.

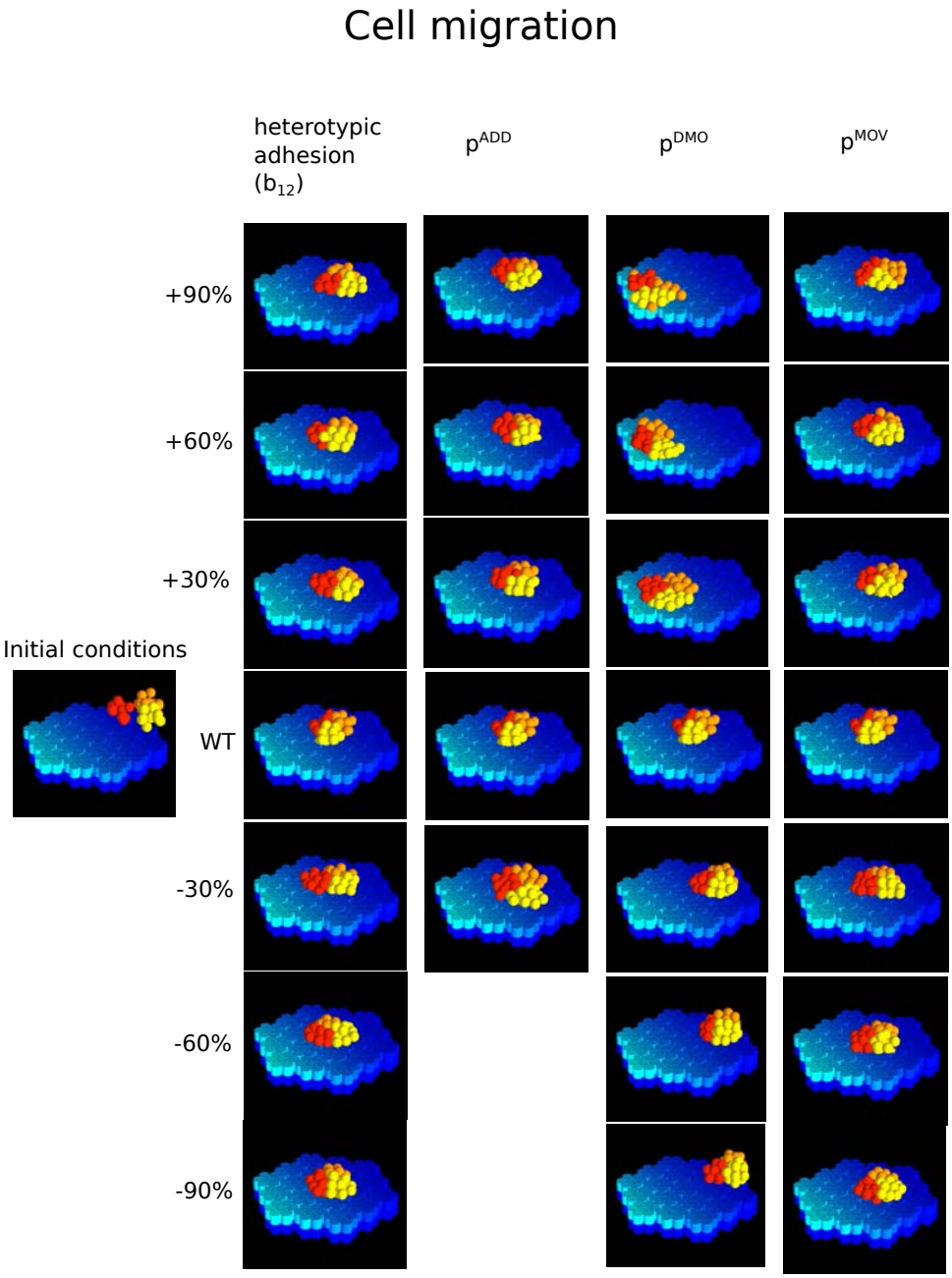

FIGURE B.15: Different simulations were run starting from the same initial conditions (left picture) except for one parameter that was modified. Each column shows the final phenotype when a single parameter is modified by a percentage relative to the value of the wildtype (central row of pictures). A mass of three mesenchymal cells (green) migrates over a hexagonal epithelial sheet that has a gradient in the expression of its adhesion molecles. Thus, by means of noise and differential adhesion, mesenchymal cells get displaced, from upper right part of the epithelium, towards areas with high concentration of adhesion molecules (yellowish nodes). Notice that effective cell migration is only promoted by some parameter combinations (e.g. very short pseudopodia (p^{DMO} < wild type p^{DMO}) prevents migration).

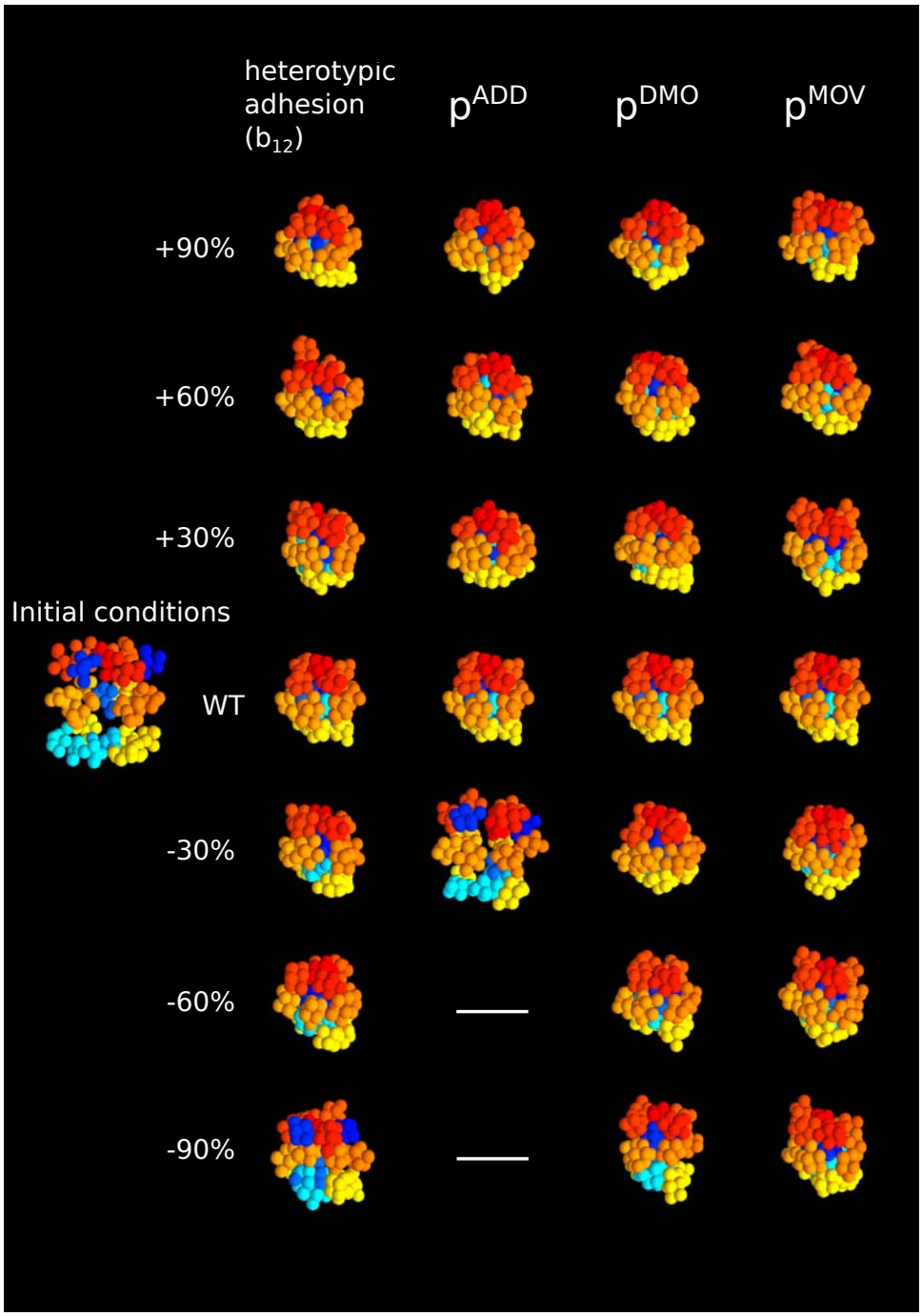

FIGURE B.16: Different simulations, changing one parameter at a time, where run starting from the same initial conditions (left picture). Each column shows the final pattern when a single parameter is modified by a percentage relative to the value of the wild-type (central row of pictures). The parameters explored are, from left to right: heterotypic affinity between the two types of adhesion molecules, the maximum radius of interaction between nodes, the maximum length of movement by noise per time step, and the likelihood of the node to make an unfavourable movement by noise. The lack of final phenotypes corresponding to -60% and -90% decreases of maximum radius of interaction between nodes stems from the fact that in those cases, the radius of interaction is lower than the radius of equilibrium, thus preventing interactions between nodes (cells fail to keep their internal cohesion).

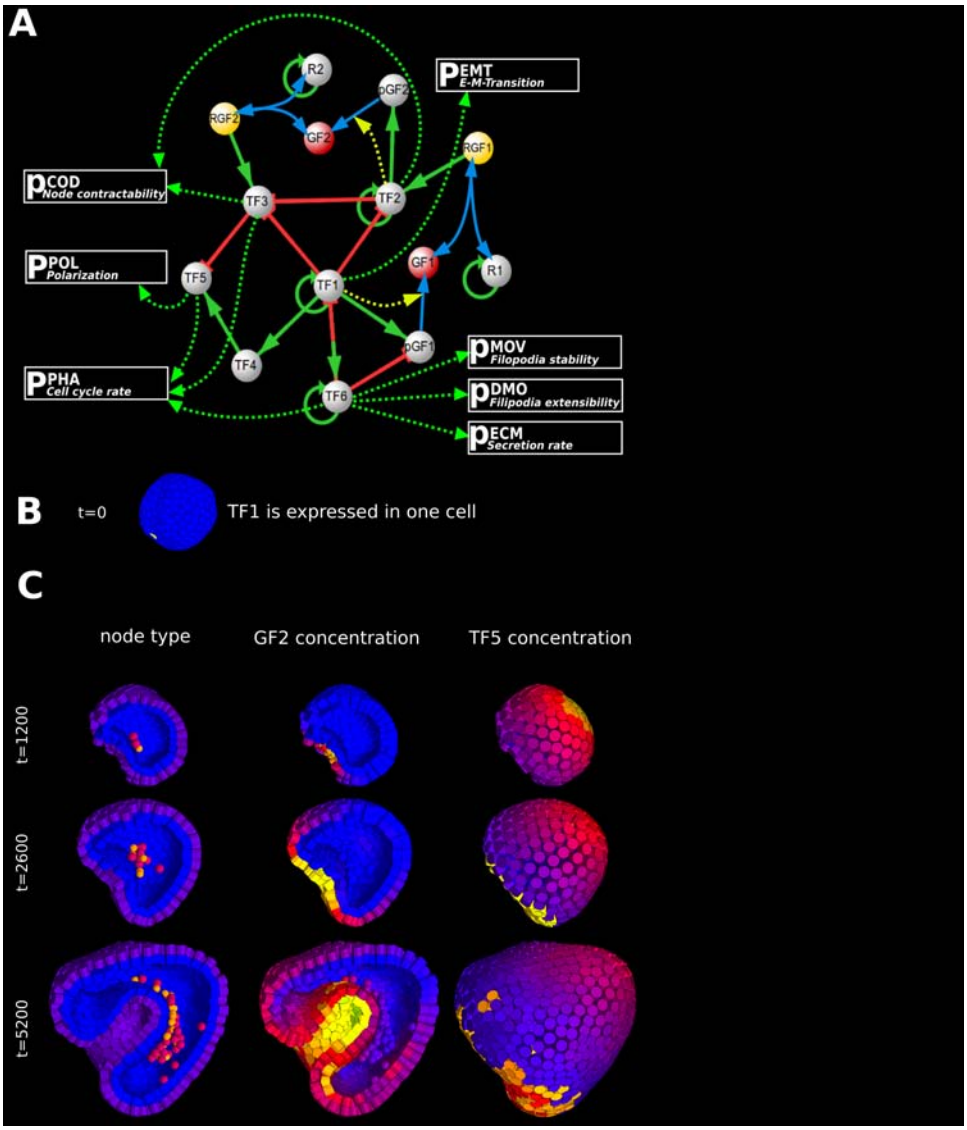

FIGURE B.17: A) Schema of the gene regulatory network. TF, transcription factor, pGF, Growth factor transcript, GF, secreted growth factor, R, receptor, RGF, receptor-ligand complex. Solid green and red arrows depict positive and negative transcriptional regulation respectively. Solid blue arrows depict chemical reactions. Yellow dashed arrows indicate catalysis. Green dashed lines indicate regulation of cell behaviours or node properties. B) Initial conditions, single cell (yellow) expresses gene TF1. C) Outcome, after different number of iterations, of the developmental mechanism in A on the initial conditions in B. The left column shows, in section, the node types. Blue for basal side of cylinders, violet for the apical side of cylinders, red for mesenchymal cells and orange for extracellular matrix nodes. Middle and lower row display concentrations of GF2 and TF5 respectively (yellow for high concentration, blue for low concentration). See section B.11 for details.

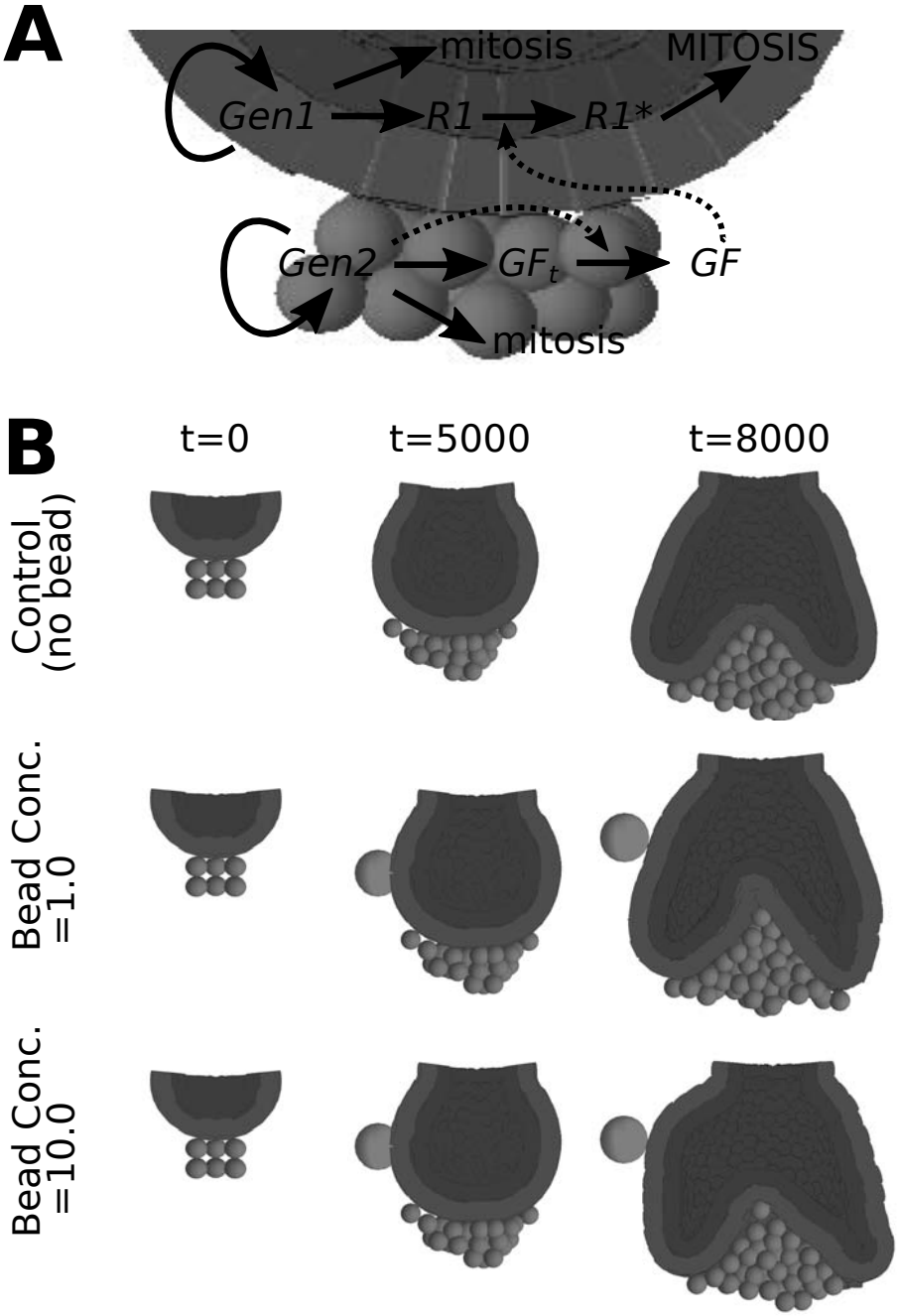

FIGURE B.18: A growing epithelial bud is in contact with a growing mesenchymal condensation. In this simulation each cell is made of a single element. Mesenchymal cells express a transcriptional factor that activates its own transcription and the transcription of a growth factor precursor. This precursor requires a chemical modification to become a growth factor that diffuses between cells in the extracellular space. This reaction is catalysed by the transcriptional factor. The growth factor diffuses in the extracellular space and binds to its receptor. This receptor is expressed only in the epithelium (where an auto-activatory transcriptional factor promotes its expression). The activated receptor enhances cell division in the epithelium. This leads to the bending of the epithelium close to the mesenchyme and finally enclosing the mesenchymal condensate (since there is an adhesion affinity between both tissues). Each row shows a time sequence. First row simulates development without bead addition, second to fifth show the same simulation but adding a bead at time step 10000 that releases the same growth factor the mesenchyme is secreting (beads with different concentrations of the growth factor). At the highest concentrations an ectopic epithelial bud can be seen beneath the bead. All pictures show a section of a three-dimensional system.

Appendix C

Supplementary Information Chapter 5

Default model parameters and node properties.

In all simulations, cells were represented with one single node (basal epithelial cells were composed on a single cylinder, that consists of two nodes).

The cell neighbouring method used was based on a Gabriel tessellation (Chapter 5).

Epithelial cell plasticity (i.e. permanent deformation mediated by mechanical forces, see Chapter 5) was deactivated in all simulations, thus all cells behaved like ideal elastic bodies.

Time progression was solved in all simulations by the Euler method, using a fixed time step size. Random cell movements via a Monte-Carlo method were not included in the simulations.

Node properties (see Chapter 5) for all cell types were set to the following values unless indicated otherwise.

- p^{EQD} = 0,25
- p^{ADD} = 0,50
- p^{REC} = 0,50
- p^{EST} = 1,00
- p^{ERP} = 1,00
- p^{EQS} = 0,125
- p^{HOO} = 5,00
- p^{MOV} = 0,00
- p^{DMO} = 0,00

Gene network and characterization of cell types.

A total of 31 gene products were specified in the system, with several gene-gene gene-cell behaviour interactions.

- Gene product 1: Epithelial marker. Gene expressed in all epithelial cells at a constant level. Regulates the expression of epithelial specific adhesion molecule. Promotes its own expression. Promotes expression of epithelial adhesion molecule.

- Gene product 2: Suprabasal cell marker. Gene expressed in all suprabasal cells at a constant level. Regulates the expression of suprabasal cell specific adhesion molecule. Promotes its own expression. Promotes expression of suprabasal adhesion molecule.
- Gene product 3: Mesenchymal cell marker. Gene expressed in all mesenchymal cells at a constant level. Regulates the expression of mesenchymal cell specific adhesion molecule. Promotes its own expression. Promotes expression of epithelial adhesion molecule.
- Gene product 4: Epithelial organ marker. Gene specifying the epithelial territory corresponding to the tooth germ. Promotes its own expression. Promotes expression of epithelial and suprabasal growth factor receptor.
- Gene product 5: Oral epithelium marker. Gene specifying the epithelial territory surrounding the tooth germ corresponding to the oral epithelium. Promotes its own expression.
- Gene product 6: Mesenchymal organ marker. Gene specifying the mesenchymal territory corresponding to the tooth germ. Promotes its own expression. Promotes expression of mesenchymal growth factor receptor.
- Gene product 7: Jaw mesenchyme marker. Gene specifying the mesenchymal territory outside the tooth germ. Promotes its own expression.
- Gene product 8: Epithelial adhesion molecule. Adhesion molecule expressed only in basal epithelial cells.
- Gene product 9: Suprabasal adhesion molecule. Adhesion molecule expressed only in suprabasal cells.
- Gene product 10: Mesenchymal adhesion molecule. Adhesion molecule expressed only in mesenchymal cells.
- Gene product 11: Signalling centre initial marker. Gene specifying a small group of epithelial cells at the tip of the tooth bud as signalling centre cells. These cells are specified at the initial conditions. The initial territory sends a signal that elicits the specification of surrounding cells as signalling centre, making it grow larger. Promotes its own expression.
- Gene product 12: Growth factor. Signalling molecule secreted by the signalling centre. Interacts with epithelial, suprabasal and mesenchymal receptors.
- Gene product 13: Epithelial receptor: inactive. Inactive form of the epithelial receptor for the growth factor.
- Gene product 14: Suprabasal receptor: inactive. Inactive form of the suprabasal receptor for the growth factor.
- Gene product 15: Mesenchymal receptor: inactive. Inactive form of the mesenchymal receptor for the growth factor.

- Gene product 16: Epithelial receptor: activated. Activated form of the epithelial receptor that results from the interaction with the growth factor.
- Gene product 17: Suprabasal receptor: activated. Activated form of the suprabasal receptor that results from the interaction with the growth factor.
- Gene product 18: Mesenchymal receptor: activated. Activated form of the mesenchymal receptor that results from the interaction with the growth factor.
- Gene product 19: Epithelial growth effector. Gene product that is activated as a result of the activation of the epithelial growth factor receptor. It promotes cell cycle progression at a certain rate.
- Gene product 20: Suprabasal growth effector. Gene product that is activated as a result of the activation of the suprabasal growth factor receptor. It promotes cell cycle progression at a certain rate.
- Gene product 21: Mesenchymal growth effector. Gene product that is activated as a result of the activation of the mesenchymal growth factor receptor. It promotes cell cycle progression at a certain rate.
- Gene product 22: Not used.
- Gene product 23: Not used.
- Gene product 24: Not used.
- Gene product 25: Signalling centre expanding signal. Signal secreted by the initial population of signalling centre cells. It promotes the specification of surrounding cells to signalling centre fate. This signal is only secreted by the initial population of cells, but not by the newly specified signalling centre cells, thus the signalling centre territory expands only as far as the reach of the diffusing signal.
- Gene product 26: Expanded signalling centre marker (epithelial). Gene product expressed by cells receiving the signalling centre expanding signal. Cells that express this gene will secrete the growth factor signal. The expression of this product will strongly inhibit expression of any growth factor receptor.
- Gene product 27: Expanded signalling centre marker (suprabasal). Gene product expressed by cells receiving the signalling centre expanding signal. Cells that express this gene will secrete the growth factor signal. The expression of this product will strongly inhibit expression of any growth factor receptor.
- Gene product 28: Signalling centre receptor (epithelial) inactive. Inactive form of the epithelial receptor signalling centre expanding signal.
- Gene product 29: Signalling centre receptor (suprabasal) inactive. Inactive form of the epithelial receptor signalling centre expanding signal.

- Gene product 30: Signalling centre receptor (epithelial): activated. Activated form of the signalling centre receptor that results from the interaction with the signalling centre expanding signal. The activation of this receptor activates the expression of the expanded signalling centre marker.

- Gene product 31: Signalling centre receptor (suprabasal): activated. Activated form of the signalling centre receptor that results from the interaction with the signalling centre expanding signal. The activation of this receptor activates the expression of the expanded signalling centre marker.

Degradation rate was set to 0.5 for all gene products and diffusion coefficient was set to 0.5 for all diffusive signals.

Varying the diffusion coefficient for the growth factor had the effect of varying the number of proliferative cells in the different tissues, thus it had the same effect as varying the rate of proliferation of cells which received the growth factor.

Varying the diffusion coefficient of the signalling centre expanding signal had the effect of varying the size of the signalling centre, which in turn meant varying the quantity and reach of the growth factor being secreted. Thus in only influenced the relative proliferation rates between cell types.

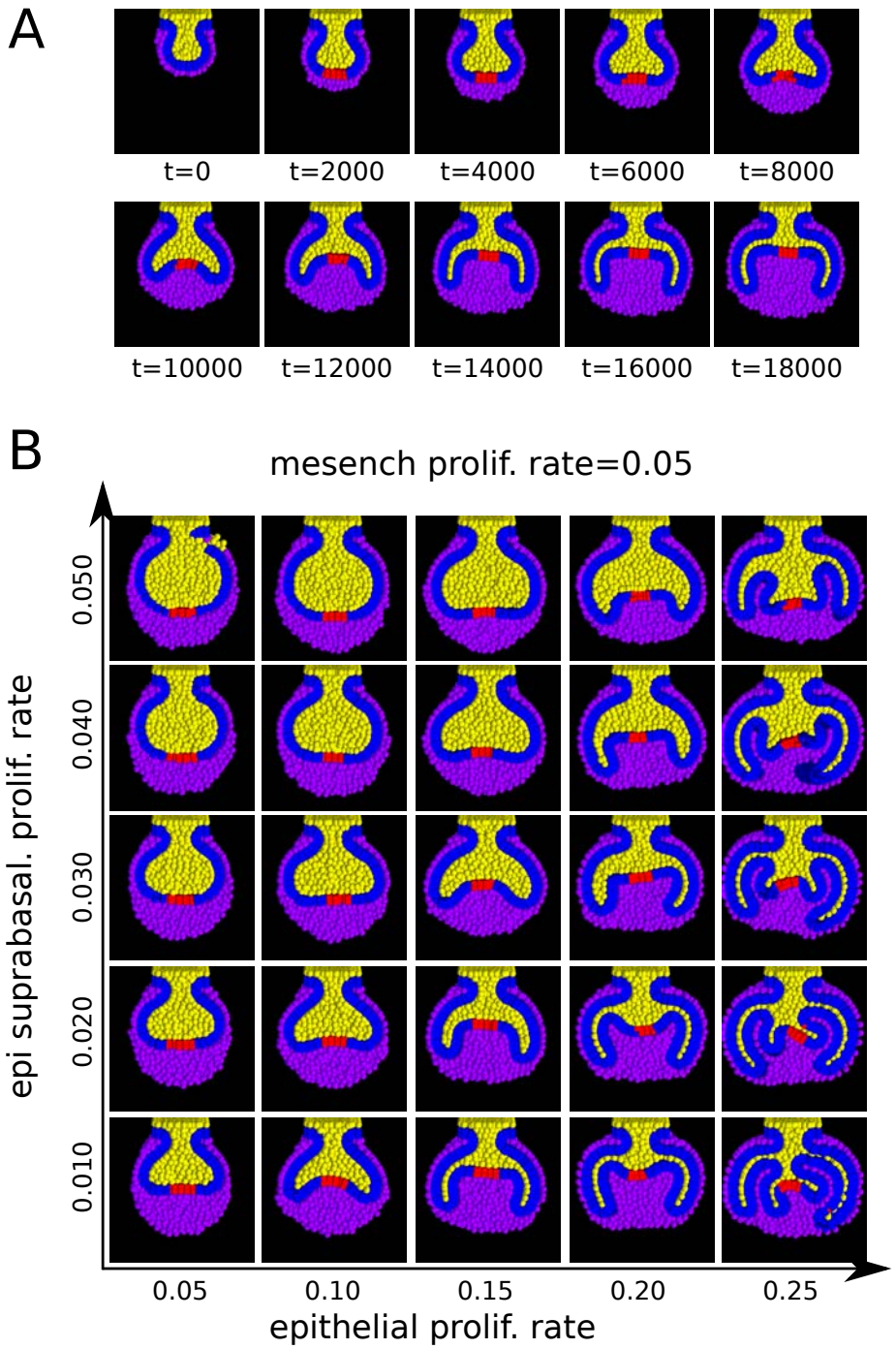

FIGURE C.1: A, time sequence of tooth morphogenesis in the model (frontal section). B, frontal sections of model tooth germs are shown as a result of running simulations with different proliferation rates in the epithelium (blue), suprabasal layer (yellow) while keeping the rest of the parameters constant. Epithelial cells composing the signalling centre are depicted in red. A high proliferation rate of epithelial cells in the developing tooth tends to promote the formation of epithelial folds, whereas high proliferation rates at the suprabasal layer preclude the formation of those, leading to tooth germs that look like enlarged bulbous buds. An intermediate range of values of the ratio of epithelial to suprabasal proliferation rates results in wild-type looking tooth germs, with two well defined cervical loops that separate the dental from the follicular mesenchyme. All simulations were run for 20000 time steps.

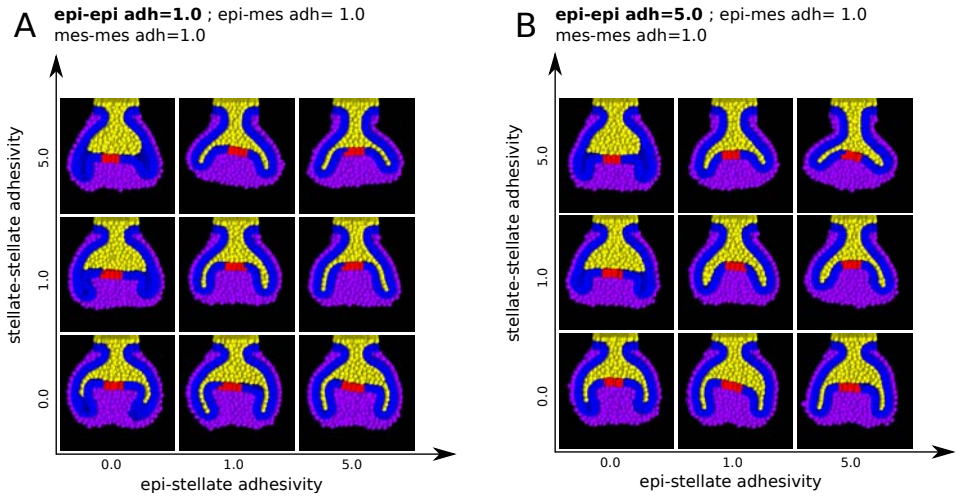

FIGURE C.2: The resulting phenotype of different simulations is shown in which the suprabasal homotypic adhesion, heterotypic epithelial-suprabasal adhesion and epithelial homotypic adhesion were varied while the other parameters were kept constant. Picture at the top indicates the initial conditions at bud stage. A, variation in suprabasal homotypic adhesion and epithelial-suprabasal heterotypic adhesion when epithelial homotypic adhesion is low. B, variation in suprabasal homotypic adhesion and epithelial-suprabasal heterotypic adhesion when epithelial homotypic adhesion is high. Slight differences in the shape of the cervical loops can be observed when epithelial homotypic adhesion is varied, but than accounts for the fact that higher epithelial homotypic adhesion will result in a higher tendency of epithelial sheets to reduce their curvature. Epithelial cells depicted in blue, suprabasal in yellow, mesenchymal cells in purple and epithelial signalling cells in red. All simulations were run for 20000 time steps. Epithelial-mesenchymal homotypic adhesion and mesenchymal homotypic adhesion were set to 1.0 in all simulations.

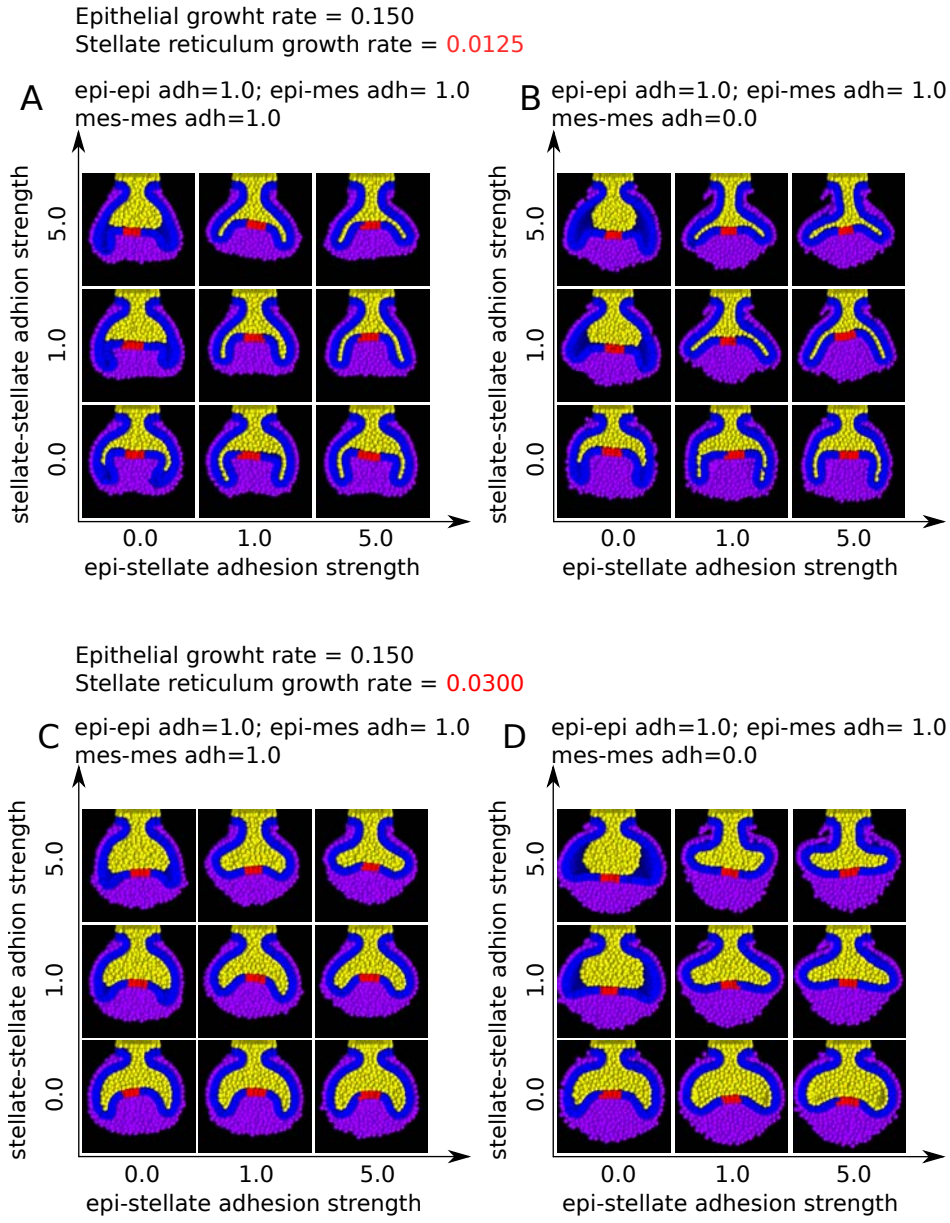

FIGURE C.3: A, B show the same parameter exploration as in figure 5.5A, B. C, D show the result of the same parameter exploration shown in figure A and B in tooth germs that developed thicker cervical loops. That was achieved by increasing the parameter that regulated the proliferation rate of the suprabasal layer in C and D compared to the tooth germs shown in A and B. It can be observed that when the cervical loops contain a larger number of suprabasal layer cells the effect of suprabasal layer homotypic adhesion on the orientation of the cervical loops is even more marked. Epithelial cells depicted in blue, suprabasal layer in yellow, mesenchymal cells in purple and epithelial signalling cells in red. All simulations were run for 20000 time steps.

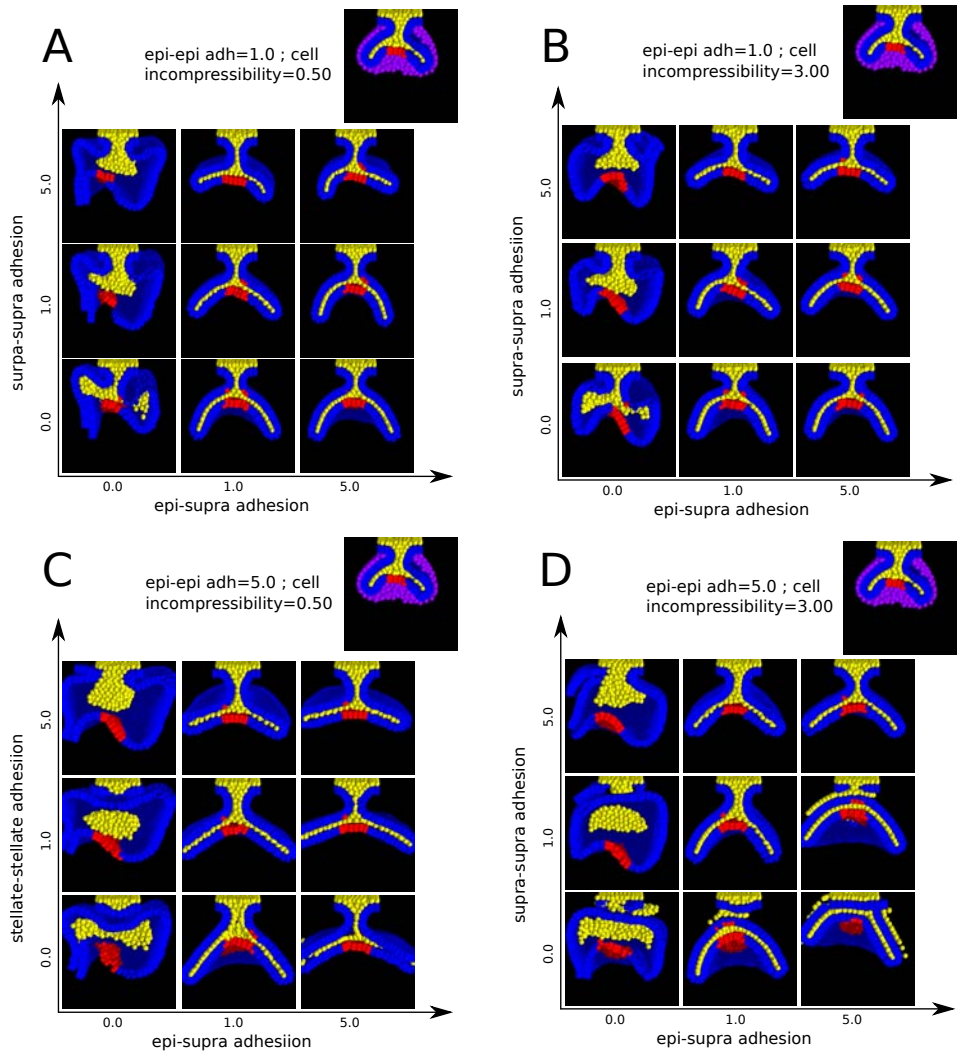

FIGURE C.4: Mesenchyme separation experiments as shown in figure 5.5C, were run for different combinations of parameter values for epithelial homotypic adhesion and cell incompressibility. A, epithelial homotypic adhesion set to 1.0, cell incompressibility set to 0.50. B, epithelial homotypic adhesion set to 1.0, cell incompressibility set to 3.00. C, epithelial homotypic adhesion set to 5.0, cell incompressibility set to 0.50. D, epithelial homotypic adhesion set to 5.0, cell incompressibility set to 3.00.

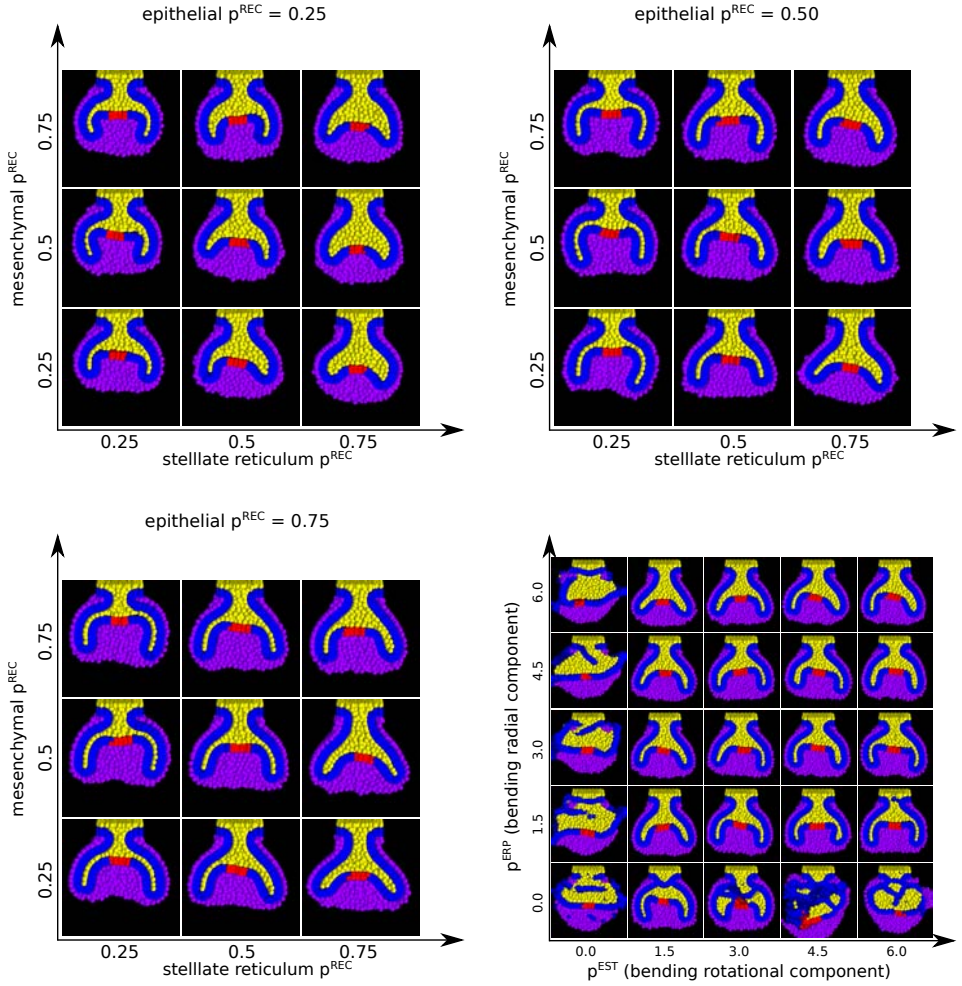

FIGURE C.5: The resulting phenotype of different simulations is shown in cell imcompressibility is varied independently in the epithelium, suprabasal layer and mesenchyme (A, B, C) while the other parameters were kept constant, as well as the cell mechanical parameters related to epithelial bending forces (D). A, B, C show the results of varying cell imcompressibility in the suprabasal layer and the mesenchyme for low (A), intermediate (B) and high (C) values of epithelial incompressibility. High incompressibility in the epithelium often results in longer cervical loops whereas high incompressibility in the suprabasal layer leads to a larger volume of the tooth germ. Incompressibility in the mesenchyme doesn't have an apparent effect on tooth germ shape. D, variation in the mechanical parameters related to epithelial bending forces is shown. Tooth germ shape is highly disturbed in the cases when one of the parameters is set to 0, since bending forces are required to keep the integrity of the epithelial tissue when it is bent. Variation in the parameters result in slight variations in tooth germ shape. Epithelial cells depicted in blue, suprabasal in yellow, mesenchymal cells in purple and epithelial signalling cells in red. All simulations were run for 20000 time steps.

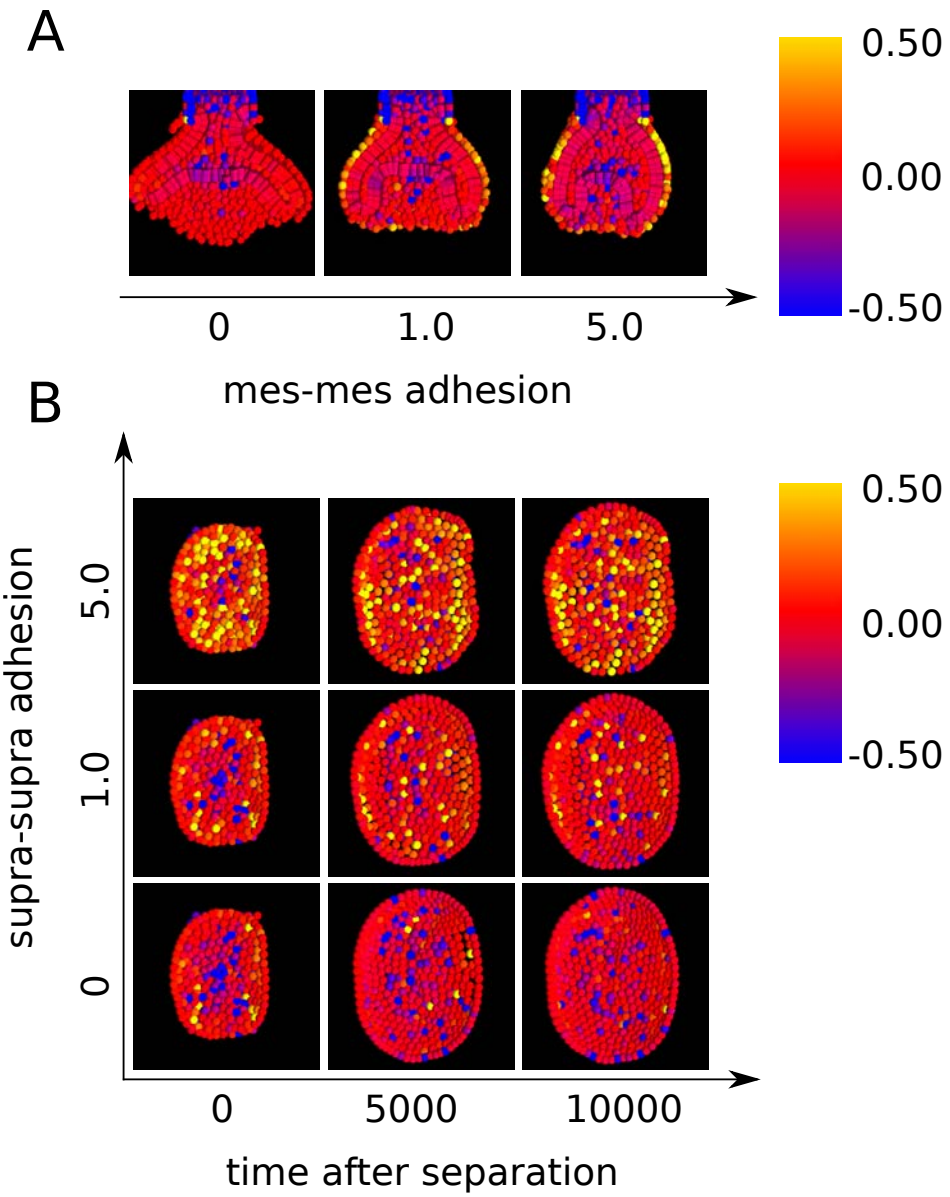

FIGURE C.6: Tooth germ sections are shown in which the colour of each cell is related to the intensity and sign of their mechanical stress. Negative values mean compression and positive values mean tension. A, Current cell mechanical stress in tooth germs at cap stage, run with different values of mesenchymal homotypic adhesion. High values of mesenchymal adhesion result in higher tensile stresses in the follicular mesenchyme and higher compression in the interior of the tooth germ. All other adhesion strengths were set to 1.0 . B, Cell mechanical tension at different time points during the relaxation after mesenchyme separation as in figure 5.5C. Three simulations are shown with different values of suprabasal homotypic adhesion (corresponding to the right most column in figure 5.5C). The whole tooth germ is show but only suprabasal cells are depicted. Higher suprabasal homotypic adhesion results in larger tensile stresses in suprabasal cells leading to cervical loops oriented in the buco-lingual direction. Tooth germs are oriented so the radical side is shown.

Bibliography

Alberch, P (1982). "Developmental constraints in evolutionary processes". In: *Evolution and {Development}*. Ed. by J T Bonner. Dahlem {Workshop} {Reports} 22. Springer Berlin Heidelberg, pp. 313–332. ISBN: 978-3-642-45534-6, 978-3-642-45532-2. DOI: 10.1007/978-3-642-45532-2_15.

Alfaro, Michael E, Daniel I Bolnick, and Peter C Wainwright (2005). "Evolutionary consequences of many-to-one mapping of jaw morphology to mechanics in labrid fishes." In: *The American naturalist* 165.6, E140–54. ISSN: 1537-5323. DOI: 10.1086/429564.

Alvarez-Castro, José M and Orjan Carlborg (2007). "A unified model for functional and statistical epistasis and its application in quantitative trait Loci analysis." In: *Genetics* 176.2, pp. 1151–1167. ISSN: 0016-6731. DOI: 10.1534/genetics.106.067348.

Bagnard, D et al. (1998). "Semaphorins act as attractive and repulsive guidance signals during the development of cortical projections." In: *Development (Cambridge, England)* 125.24, pp. 5043–53. ISSN: 0950-1991.

Barton, N H and M Turelli (2004). "Effects of genetic drift on variance components under a general model of epistasis". In: *Evolution* 58.1991, pp. 2111–2132. ISSN: 0014-3820. DOI: 10.1554/03-684.

Biggs, Leah C. and Marja L. Mikkola (2014). "Early inductive events in ectodermal appendage morphogenesis". In: *Seminars in Cell and Developmental Biology* 25-26, pp. 11–21. ISSN: 10963634. DOI: 10.1016/j.semcdb.2014.01.007.

Boehm, Bernd et al. (2010). "The role of spatially controlled cell proliferation in limb bud morphogenesis". In: *PLoS Biology* 8.7. ISSN: 15449173. DOI: 10.1371/journal.pbio.1000420.

Borenstein, Elhanan and David C. Krakauer (2008). "An end to endless forms: Epistasis, phenotype distribution bias, and nonuniform evolution". In: *PLoS Computational Biology* 4.10.

Bunn, Jonathan M. et al. (2011). "Comparing Dirichlet normal surface energy of tooth crowns, a new technique of molar shape quantification for dietary inference, with previous methods in isolation and in combination". In: *American Journal of Physical Anthropology* 145.2, pp. 247–261.

Butler, P. M. (1956). "The ontogeny of molar pattern". In: *Biological Reviews* 31.1, pp. 30–69. ISSN: 1464-7931. DOI: 10.1111/j.1469-185X.1956.tb01551.x.

Butler, P M (1983). "Evolution and mammalian dental morphology". In: *J Biol Buccale* 11.4, pp. 285–302.

Butler, P. M. (1995). *Ontogenetic aspects of dental evolution*.

Cai, Danfeng and Denise J Montell (2014). "Diverse and dynamic sources and sinks in gradient formation and directed migration." In: *Current opinion in cell biology* 30, pp. 91–8. ISSN: 1879-0410. DOI: 10.1016/j.ceb.2014.06.009.

Carter, Ashley J R, Joachim Hermisson, and Thomas F Hansen (2005). "The role of epistatic gene interactions in the response to selection and the

evolution of evolvability." In: *Theoretical population biology* 68.3, pp. 179–96. ISSN: 0040-5809. DOI: 10.1016/j.tpb.2005.05.002.

Chang, Cheng et al. (2009). "Reptile scale paradigm: Evo-Devo, pattern formation and regeneration". In: *International Journal of Developmental Biology* 53.5-6, pp. 813–826. ISSN: 02146282. DOI: 10.1387/ijdb.072556cc.

Charles, Cyril et al. (2011). "Regulation of tooth number by fine-tuning levels of receptor-tyrosine kinase signaling." In: *Development (Cambridge, England)* 138.18, pp. 4063–4073. ISSN: 0950-1991. DOI: 10.1242/dev.069195.

Charlesworth, Brian (2009). "Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation." In: *Nature reviews. Genetics* 10.3, pp. 195–205. ISSN: 1471-0064. DOI: 10.1038/nrg2526.

Charlesworth, Brian, Russell Lande, and Montgomery Slatkin (1982). "A Neo-Darwinian Commentary on Macroevolution". In: *Evolution* 36.3, p. 474. ISSN: 00143820. DOI: 10.2307/2408095. URL: http://www.jstor.org/stable/2408095?origin=crossref.

Chuong, Cheng Ming et al. (2000). "Evo-Devo of feathers and scales: Building complex epithelial appendages. Commentary". In: *Current Opinion in Genetics and Development* 10.4, pp. 449–456. ISSN: 0959437X. DOI: 10.1016/S0959-437X(00)00111-8.

Cickovski, Trevor M. et al. (2005). "A framework for three-dimensional simulation of morphogenesis". In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2.4, pp. 273–287.

Cummings, F.W. (1990). "A model of morphogenetic pattern formation". In: *Journal of Theoretical Biology* 144, pp. 547–566.

— (1994). "Aspects of growth and form". In: *Physica D* 79.2-4, pp. 146–163.

— (2005). "Interaction of morphogens with geometry". In: *Physica A* 355, pp. 427–438.

Darwin, Charles (1859). *On the Origin of the Species*. Vol. 5, p. 386. DOI: 10.1016/S0262-4079(09)60380-8.

Davidson, L.A. et al. (1995). "How do sea urchins invaginate?" In: *Development* 121, pp. 2005–2018.

Delile, Julien (2013). "PhD Thesis Julien Delile". In: Paris 7.

Delile, Julien, René Doursat, and Nadine Peyriéras (2013). *Computational Modeling and Simulation of Animal Early Embryogenesis with the MecaGen Platform*, pp. 359–405. ISBN: 9780124059269. DOI: 10.1016/B978-0-12-405926-9.00016-2.

Draghi, Jeremy a et al. (2010). "Mutational robustness can facilitate adaptation." In: *Nature* 463.7279, pp. 353–355. ISSN: 0028-0836. DOI: 10.1038/nature08694.

Drasdo, D., S. Hoehme, and Michael Block (2007). "On the role of physics in the growth and pattern formation of multi-cellular systems". In: *Journal of Statistical Physics* 128, pp. 287–345.

Evans, Alistair R et al. (2007). "High-level similarity of dentitions in carnivorans and rodents." In: *Nature* 445.7123, pp. 78–81. ISSN: 0028-0836. DOI: 10.1038/nature05433.

Farhadifar, R. et al. (2007). "The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing". In: *Current Biology* 24, pp. 2095–2104.

Ferrada, Evandro and Andreas Wagner (2012). "A comparison of genotype-phenotype maps for RNA and proteins". In: *Biophysical Journal* 102.8, pp. 1916–1925. ISSN: 00063495. DOI: 10.1016/j.bpj.2012.01.047.

Fierst, Janna L. and Thomas F. Hansen (2010). "Genetic architecture and postzygotic reproductive isolation: Evolution of Bateson-Dobzhansky-Muller incompatibilities in a polygenic model". In: *Evolution* 64.3, pp. 675–693.

Fontana, Walter (2002). "Modelling 'evo-devo' with RNA." In: *BioEssays : news and reviews in molecular, cellular and developmental biology* 24.12, pp. 1164–77. ISSN: 0265-9247. DOI: 10.1002/bies.10190.

Forgacs, G. and S.A. Newman (2005). *Biological physics of the developing embryo*. Cambridge University Press.

Gjorevski, Nikolce and Celeste M. Nelson (2012). "Mapping of Mechanical Strains and Stresses around Quiescent Engineered Three-Dimensional Epithelial Tissues". In: *Biophysical Journal* 103.1, pp. 152–162. ISSN: 00063495. DOI: 10.1016/j.bpj.2012.05.048.

Gjuvsland, Arne B. et al. (2007). "Statistical epistasis is a generic feature of gene regulatory networks". In: *Genetics* 175.1, pp. 411–420.

Godfrey, Laurie R et al. (2012). "Dental topography indicates ecological contraction of lemur communities." In: *American journal of physical anthropology* 148.2, pp. 215–27. ISSN: 1096-8644. DOI: 10.1002/ajpa.21615.

Gould, S J and R C Lewontin (1979). "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme." In: *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)* 205.1161, pp. 581–598. ISSN: 0962-8452. DOI: 10.1098/rspb.1979.0086.

Gould, S.J. and N. Eldredge (1977). "Puncuated equilibria: the tempo and mode of evolution reconsidered". In: *Paleobiology* 3.2, pp. 115–151.

Graner, François and James A. Glazier (1992). "Simulation of biological cell sorting using a two-dimensional extended Potts model". In: *Physical Review Letters* 69.13, pp. 2013–2016. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.69.2013.

Guo, Yina et al. (2014). "Branching patterns emerge in a mathematical model of the dynamics of lung development." In: *The Journal of physiology* 592.2, pp. 313–24. ISSN: 1469-7793. DOI: 10.1113/jphysiol.2013.261099.

Hansen, T F and G P Wagner (2001). "Modeling genetic architecture: a multilinear theory of gene interaction." In: *Theoretical population biology* 59.1, pp. 61–86. ISSN: 0040-5809. DOI: 10.1006/tpbi.2000.1508.

Harjunmaa, Enni et al. (2012). "On the difficulty of increasing dental complexity". In: *Nature* 483.7389, pp. 324–327. ISSN: 0028-0836. DOI: 10.1038/nature10876.

Harjunmaa, Enni et al. (2014). "Replaying evolutionary transitions from the dental fossil record". In: *Nature* 512.7512, pp. 44–48. ISSN: 0028-0836. DOI: 10.1038/nature13613.

Hentschel, H G E et al. (2004). "Dynamical mechanisms for skeletal pattern formation in the vertebrate limb." In: *Proceedings. Biological sciences / The Royal Society* 271.1549, pp. 1713–22. ISSN: 0962-8452. DOI: 10.1098/rspb.2004.2772.

Hester, Susan D. et al. (2011). "A multi-cell, multi-scale model of vertebrate segmentation and somite formation". In: *PLoS Computational Biology* 7.10. ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1002155.

Hogeweg, P (2000). "Evolving mechanisms of morphogenesis: on the interplay between differential adhesion and cell differentiation." In: *Journal of theoretical biology* 203.4, pp. 317–333. ISSN: 0022-5193. DOI: 10.1006/jtbi.2000.1087.

Honda, Hisao and Tatsuzo Nagai (2015). "Cell models lead to understanding of multi-cellular morphogenesis consisting of successive self-construction of cells." In: *Journal of biochemistry* 157.3, pp. 129–36. ISSN: 1756-2651. DOI: 10.1093/jb/mvu088.

Honda, Hisao, Masaharu Tanemura, and Tatsuzo Nagai (2004). "A three-dimensional vertex dynamics cell model of space-filling polyhedra simulating cell behavior in a cell aggregate". In: *Journal of Theoretical Biology* 226.4, pp. 439–453. ISSN: 00225193. DOI: 10.1016/j.jtbi.2003.10.001. arXiv: 0842283368.

Honda, Hisao et al. (2008). "Computer simulation of emerging asymmetry in the mouse blastocyst." In: *Development (Cambridge, England)* 135.8, pp. 1407–1414. ISSN: 0950-1991. DOI: 10.1242/dev.014555.

Huynen, M A, P F Stadler, and W. Fontana (1996). "Smoothness within ruggedness: the role of neutrality in adaptation". In: *Proceedings of the National Academy of Sciences* 93.1, pp. 397–401. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.93.1.397.

Izaguirre, J.A. et al. (2004). "CompuCell, a multi-model framework for simulation of morphogenesis". In: *Bioinformatics* 20, pp. 1129–1137.

Jaeger, Johannes et al. (2004). "Dynamic control of positional information in the early Drosophila embryo." In: *Nature* 430.6997, pp. 368–71.

Jarvis, J. P. and J. M. Cheverud (2009). "Epistasis and the evolutionary dynamics of measured genotypic values during simulated serial bottlenecks". In: *Journal of Evolutionary Biology* 22.8, pp. 1658–1668.

Jernvall, J (1995). *Mammalian Molar Cusp Patterns: Developmental Mechanisms of Diversity*. Acta zoologica Fennica. Finnish Zoological and Botanical Pub. Board. ISBN: 9789519481432.

— (2000). "Linking development with generation of novelty in mammalian teeth." In: *Proceedings of the National Academy of Sciences of the United States of America* 97.6, pp. 2641–2645. ISSN: 0027-8424. DOI: 10.1073/pnas.050586297.

Jernvall, J et al. (1994). "Evidence for the role of the enamel knot as a control center in mammalian tooth cusp formation: non-dividing cells express growth stimulating Fgf-4 gene." In: *The International journal of developmental biology* 38.3, pp. 463–9. ISSN: 0214-6282.

Jernvall, Jukka and Irma Thesleff (2000). "Reiterative signaling and patterning during mammalian tooth morphogenesis". In: *Mechanisms of Development* 92.1, pp. 19–29. ISSN: 09254773. DOI: 10.1016/S0925-4773(99)00322-6.

Ji, Qiang et al. (2002). "The earliest known eutherian mammal." In: *Nature* 416.6883, pp. 816–822.

Jussila, M. et al. (2015). "Suppression of epithelial differentiation by Foxi3 is essential for molar crown patterning". In: *Development*, pp. 3954–3963. ISSN: 0950-1991. DOI: 10.1242/dev.124172.

Kaneko, Kunihiko (2012). "Evolution of Robustness and Plasticity under Environmental Fluctuation: Formulation in Terms of Phenotypic Variances". In: *Journal of Statistical Physics* 148.4, pp. 686–704.

Kaufmann, Stuart (1993a). *The origins of order*. Vol. 209, p. 709. ISBN: 0195058119. DOI: 10.1002/bies.950170412.

— (1993b). *The origins of order*. Vol. 209, p. 709. ISBN: 0195058119. DOI: 10.1002/bies.950170412.

Keränen, S V et al. (1998). "Association of developmental regulatory genes with the development of different molar tooth shapes in two species of rodents." In: *Development genes and evolution* 208.9, pp. 477–86. ISSN: 0949-944X.

Klingenberg, Christian Peter (2002). "Morphometrics and the role of the phenotype in studies of the evolution of developmental mechanisms." In: *Gene* 287.1-2, pp. 3–10. ISSN: 0378-1119.

Lande, R. and S.J. J Arnold (1983). "The measurement of selection on correlated characters". In: *Evolution* 37.6, pp. 1210–1226. ISSN: 0014-3820. DOI: 10.2307/2408842.

Lane, M.C. et al. (1993). "A role for regulated secretion of apical extracellular matrix during epithelial invagination in the sea urchin". In: *Development* 117, pp. 1049–1060.

Larsen, Melinda, Cindy Wei, and Kenneth M Yamada (2006). "Cell and fibronectin dynamics during branching morphogenesis." In: *Journal of cell science* 119.Pt 16, pp. 3376–84. ISSN: 0021-9533. DOI: 10.1242/jcs.03079.

Laurikkala, J et al. (2001). "TNF signaling via the ligand-receptor pair ectodysplasin and edar controls the function of epithelial signaling centers and is regulated by Wnt and activin during tooth organogenesis." In: *Developmental biology* 229.2, pp. 443–55. ISSN: 0012-1606. DOI: 10.1006/dbio.2000.9955.

Linde-Medina, Marta, Benedikt Hallgrímsson, and Ralph Marcucio (2015). "Beyond cell proliferation in avian facial morphogenesis." In: *Developmental dynamics : an official publication of the American Association of Anatomists*. ISSN: 1097-0177. DOI: 10.1002/dvdy.24374.

Mammoto, Tadanori et al. (2011). "Mechanochemical Control of Mesenchymal Condensation and Embryonic Tooth Organ Formation". In: *Developmental Cell* 21.4, pp. 758–769. ISSN: 15345807. DOI: 10.1016/j.devcel.2011.07.006.

Marée, Athanasius F. M., Verônica A. Grieneisen, and Paulien Hogeweg (2007). "The Cellular Potts Model and Biophysical Properties of Cells, Tissues and Morphogenesis". In: *Single-Cell-Based Models in Biology and Medicine*. Basel: Birkhäuser Basel, pp. 107–136. DOI: 10.1007/978-3-7643-8123-3_5.

Marin-Riera, Miquel et al. (2016). "Computational modeling of development by epithelia, mesenchyme and their interactions: a unified model." In: *Bioinformatics (Oxford, England)* 32.2, pp. 219–25. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btv527.

Maynard Smith, John (1978). "Optimization theory in evolution". In: *Annual review of Ecology and systematics* 9, pp. 31–56. ISSN: 0066-4162. DOI: 10.1007/b130886.

Meir E. von Dassow, G., E. Munro, and G.M. Odell (2002). "Robustness, flexibility, and the role of lateral inhibition in the neurogenic network". In: *Current Biology* 12, pp. 778–786.

Menshykau, Denis, Conradin Kraemer, and Dagmar Iber (2012). "Branch mode selection during early lung development". In: *PLoS Computational Biology* 8.2. ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1002377.

Merks, R.M. et al. (2011). "VirtualLeaf: an open-source framework for cell-based modeling of plant tissue growth and development". In: *Plant Physiology* 155, pp. 656–666.

Merks, Roeland M H et al. (2008). "Contact-inhibited chemotaxis in de novo and sprouting blood-vessel growth." In: *PLoS computational biology* 4.9, e1000163. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000163.

Minc, N., D. Burgess, and F. Chang (2011). "Influence of cell geometry on division-plane positioning". In: *Cell* 144, pp. 414–426.

Mjolsness, E., D.H. Sharp, and J. Reinitz (1991). "A connectionist model of development". In: *Journal of Theoretical Biology* 152, pp. 429–456.

Moustakas-Verho, J. E. et al. (2014). "The origin and loss of periodic patterning in the turtle shell". In: *Development* 141.15, pp. 3033–3039. ISSN: 0950-1991. DOI: 10.1242/dev.109041.

Muller, G B and G P Wagner (1991). "Novelty in Evolution: Restructuring the Concept". In: *Annual Review of Ecology and Systematics* 22.1, pp. 229–256.

Newman, S a and W D Comper (1990). "'Generic' physical mechanisms of morphogenesis and pattern formation." In: *Development (Cambridge, England)* 110.1, pp. 1–18. ISSN: 0950-1991.

Newman, Stuart A. and Gerd B. Müller (2000). *Epigenetic mechanisms of character origination*.

Newman, Stuart a. and Gerd B. Müller (2005). "Origination and innovation in the vertebrate limb skeleton: An epigenetic perspective". In: *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 304.6, pp. 593–609. ISSN: 15525007. DOI: 10.1002/jez.b.21066.

Newman, T J (2005). "Modeling multicellular systems using subcellular elements." In: *Mathematical biosciences and engineering : MBE* 2.3, pp. 613–624. ISSN: 1551-0018. DOI: 10.3934/mbe.2005.2.613. arXiv: 0504028 [q-bio].

Orr, H. Allen (2006). "the Population Genetics of Adaptation on Correlated Fitness Landscapes: the Block Model". In: *Evolution* 60.6, p. 1113. ISSN: 0014-3820. DOI: 10.1554/05-701.1.

Orzack, Steven Hecht and Elliott Sober (1994). "Optimality Models and the Test of Adaptationism". In: *The American Naturalist* 143.3, p. 361.

Osterfield, Miriam et al. (2013). "Three-Dimensional Epithelial Morphogenesis in the Developing Drosophila Egg". In: *Developmental Cell* 24.4, pp. 400–410. ISSN: 15345807. DOI: 10.1016/j.devcel.2013.01.017. arXiv: 84875329495.

Paluch, Ewa and Carl-Philipp Heisenberg (2009). "Biology and physics of cell shape changes in development." In: *Current biology : CB* 19.17, R790–R799. ISSN: 1879-0445. DOI: 10.1016/j.cub.2009.07.029.

Panousopoulou, Eleni and Jeremy B. A. Green (2016). "Invagination of Ectodermal Placodes Is Driven by Cell Intercalation-Mediated Contraction of the Suprabasal Tissue Canopy". In: *PLOS Biology* 14.3, e1002405. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1002405.

Peskin, Charles S. (1972). "Flow patterns around heart valves: A numerical method". In: *Journal of Computational Physics* 10.2, pp. 252–271.

Peterkova, R. et al. (2014). "Three-dimensional analysis of the early development of the dentition". In: *Australian Dental Journal* 59.SUPPL. 1, pp. 55–80. ISSN: 18347819. DOI: 10.1111/adj.12130.

Pitt-Francis, J. et al. (2009). "Chaste: a test-driven approach to software development for biological modelling". In: *Computer Physics Communications* 180, pp. 2452–2471.

Polly, P. David, Steven C. Le Comber, and Tamsin M. Burland (2005). "On the Occlusal Fit of Tribosphenic Molars: Are We Underestimating Species Diversity in the Mesozoic?" In: *Journal of Mammalian Evolution* 12.1-2, pp. 283–299. ISSN: 1064-7554. DOI: 10.1007/s10914-005-4866-8.

Prusinkiewicz, Przemyslaw et al. (2007). "Evolution and development of inflorescence architectures." In: *Science (New York, N.Y.)* 316.5830, pp. 1452–1456. ISSN: 0036-8075. DOI: 10.1126/science.1140429.

Purcell, E M (1977). "Life at Low Reynolds Number". In: *American Journal of Physics* 45, pp. 3–11.

Raspopovic, J et al. (2014). "Modeling digits. Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients." en. In: *Science (New York, N.Y.)* 345.6196, pp. 566–70. ISSN: 1095-9203. DOI: 10.1126/science.1252960.

Reinitz, J and D H Sharp (1995). "Mechanism of eve stripe formation." In: *Mechanisms of development* 49.1-2, pp. 133–58. ISSN: 0925-4773.

Rejniak, K, Harvey J. Kliman, and Lisa J. Fauci (2004). "A computational model of the mechanics of growth of the villous trophoblast bilayer". In: *Bulletin of Mathematical Biology* 66.2, pp. 199–232. ISSN: 00928240. DOI: 10.1016/j.bulm.2003.06.001.

Rejniak, K.A. (2007). "Modelling the development of complex tissues using individual viscoelastic cells". In: *In: Anderson,A., Rejniak,K, editors. Single-Cell-Based Models of Biology and Medicine*. Birkhäuser Basel.

Rothová, Michaela, Renata Peterková, and Abigail S. Tucker (2012). "Fate map of the dental mesenchyme: Dynamic development of the dental papilla and follicle". In: *Developmental Biology* 366.2, pp. 244–254. ISSN: 00121606. DOI: 10.1016/j.ydbio.2012.03.018.

Salazar-Ciudad, I, J Garcia-Fernández, and R V Solé (2000). "Gene networks capable of pattern formation: from induction to reaction-diffusion." In: *Journal of theoretical biology* 205.4, pp. 587–603. ISSN: 0022-5193. DOI: 10.1006/jtbi.2000.2092.

Salazar-Ciudad, I and J Jernvall (2004). "How different types of pattern formation mechanisms affect the\revolution of form and development". In: *Evolution and Development* 6.1, pp. 6–16.

Salazar-Ciudad, I., S. a. Newman, and R. V. Solé (2001). "Phenotypic and dynamical transitions in model genetic networks I. Emergence of patterns and genotype-phenotype relationships". In: *Evolution and Development* 3.2, pp. 84–94. ISSN: 1520541X. DOI: 10.1046/j.1525-142X.2001.003002084.x.

Salazar-Ciudad, Isaac (2006). "Developmental constraints vs. variational properties: How pattern formation can help to understand evolution and development." In: *Journal of experimental zoology. Part B, Molecular and developmental evolution* 306.2, pp. 107–25. ISSN: 1552-5007. DOI: 10.1002/jez.b.21078.

Salazar-Ciudad, Isaac (2008). "Tooth morphogenesis in vivo, in vitro, and in silico." In: *Current topics in developmental biology* 81, pp. 341–71. ISSN: 0070-2153. DOI: 10.1016/S0070-2153(07)81012-X.

— (2010a). "A two level mutation-selection model of cultural evolution and diversity". In: *Journal of Theoretical Biology* 267.2, pp. 171–185. ISSN: 00225193. DOI: 10.1016/j.jtbi.2010.08.021.

— (2010b). "Morphological evolution and embryonic developmental diversity in metazoa." In: *Development (Cambridge, England)* 137.4, pp. 531–539. ISSN: 0950-1991. DOI: 10.1242/dev.045229.

— (2012). "Evolution in Biological and Non-biological Systems: The Origins of Life". In: *Biological Theory* 7.1, pp. 26–37. ISSN: 1555-5542. DOI: 10.1007/s13752-012-0066-y.

Salazar-Ciudad, Isaac and Jukka Jernvall (2002). "A gene network model accounting for development and evolution of mammalian teeth". In: *Proceedings of the National Academy of ...* 99.12, p. 12. ISSN: 0027-8424. DOI: 10.1073/pnas.132069499.

— (2005). "Graduality and innovation in the evolution of complex phenotypes: Insights from development". In: *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 304.6, pp. 619–631. ISSN: 15525007. DOI: 10.1002/jez.b.21058.

— (2010). "A computational model of teeth and the developmental origins of morphological variation." In: *Nature* 464.7288, pp. 583–586. ISSN: 0028-0836. DOI: 10.1038/nature08838. arXiv: 77950095599.

Salazar-Ciudad, Isaac, Jukka Jernvall, and Stuart A. Newman (2003). "Mechanisms of pattern formation in development and evolution". In: *Development* 130.10, pp. 2027–2037. ISSN: 09501991. DOI: 10.1242/dev.00425.

Salazar-Ciudad, Isaac and Miquel Marín-Riera (2013). "Adaptive dynamics under development-based genotype-phenotype maps." In: *Nature* 497.7449, pp. 361–4.

Sandersius, Sebastian a and Timothy J Newman (2008). "Modeling cell rheology with the Subcellular Element Model." In: *Physical biology* 5.1, p. 015002.

Santana, Sharlene E., Suzanne Strait, and Elizabeth R. Dumont (2011). "The better to eat you with: Functional correlates of tooth structure in bats". In: *Functional Ecology* 25.4, pp. 839–847. ISSN: 02698463. DOI: 10.1111/j.1365-2435.2011.01832.x.

Sasaki, T et al. (1984). "Intercellular junctions in the cells of the human enamel organ as revealed by freeze-fracture." In: *Archives of oral biology* 29.4, pp. 275–86. ISSN: 0003-9969.

Schuster, P et al. (1994). *From sequences to shapes and back: a case study in RNA secondary structures.* DOI: 10.1098/rspb.1994.0040.

Setty, Yaki et al. (2008). "Four-dimensional realistic modeling of pancreatic organogenesis." In: *Proceedings of the National Academy of Sciences of the United States of America* 105.51, pp. 20374–20379. ISSN: 0027-8424. DOI: 10.1073/pnas.0808725105.

Shapiro, Bruce E, Elliot M Meyerowitz, and Eric Mjolsness (2013). "Using cellzilla for plant growth simulations at the cellular level." In: *Frontiers in plant science* 4, p. 408.

Shapiro, Steven D (1998). "Matrix metalloproteinase degradation of extracellular matrix: biological consequences". In: *Current Opinion in Cell Biology* 10.5, pp. 602–608. ISSN: 09550674. DOI: 10.1016/S0955-0674(98)80035-5.

Shih, Hung Ping et al. (2015). "ECM Signaling Regulates Collective Cellular Dynamics to Control Pancreas Branching Morphogenesis". In: *Cell Reports* 14.2, pp. 169–179. ISSN: 22111247. DOI: 10.1016/j.celrep.2015.12.027.

Shvartsman, Stanislav Y, Cyrill B Muratov, and Douglas a Lauffenburger (2002). "Modeling and computational analysis of EGF receptor-mediated cell communication in Drosophila oogenesis." In: *Development (Cambridge, England)* 129, pp. 2577–2589.

Simons, Matias and Marek Mlodzik (2008). "Planar cell polarity signaling: from fly development to human disease." In: *Annual review of genetics* 42, pp. 517–40. ISSN: 0066-4197. DOI: 10.1146/annurev.genet.42.110807.091432.

Sluka, James P et al. (2014). "The cell behavior ontology: describing the intrinsic biological behaviors of real and model cells seen as active agents." In: *Bioinformatics (Oxford, England)* 30.16, pp. 2367–2374. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btu210.

Smith, A.M. et al. (2012). "Incorporating chemical signalling factors into cell-based models of growing epithelial tissues". In: *Journal of Mathematical Biology* 65, pp. 441–463.

Stadler, Peter F. (2006). "Genotype–Phenotype Maps". In: *Biological Theory* 1.3, pp. 268–279. ISSN: 1555-5542. DOI: 10.1162/biot.2006.1.3.268.

Starruß, J. et al. (2014a). "Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology". In: *Bioinformatics* 30, pp. 1331–1332.

Starruß, Jörn et al. (2014b). "Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology." In: *Bioinformatics (Oxford, England)* 30.9, pp. 1331–2. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btt772.

Steinberg, Malcolm S. (1996). "Adhesion in Development: An Historical Overview". In: *Developmental Biology* 180.2, pp. 377–388. ISSN: 00121606. DOI: 10.1006/dbio.1996.0312.

Swat, Maciej H et al. (2012). "Multi-scale modeling of tissues using CompuCell3D." In: *Methods in cell biology* 110, pp. 325–66. ISSN: 0091-679X. DOI: 10.1016/B978-0-12-388403-9.00013-8.

Tada, M. and C.-P. Heisenberg (2012). "Convergent extension: using collective cell migration and cell intercalation to shape embryos". In: *Development* 139.21, pp. 3897–3904. ISSN: 0950-1991. DOI: 10.1242/dev.073007.

Takigawa-Imamura, Hisako et al. (2015). "Tooth germ invagination from cell–cell interaction: Working hypothesis on mechanical instability". In: *Journal of Theoretical Biology* 382, pp. 284–291. ISSN: 00225193. DOI: 10.1016/j.jtbi.2015.07.006.

Tamulonis, Carlos et al. (2011). "A cell-based model of Nematostella vectensis gastrulation including bottle cell formation, invagination and zippering". In: *Developmental Biology* 351.1, pp. 217–228.

Tanaka, S., D. Sichau, and D. Iber (2015). "LBIBCell: a cell-based simulation environment for morphogenetic problems". In: *Bioinformatics* 31, pp. 2340–7.

Tenaillon, Olivier et al. (2007). "Quantifying organismal complexity using a population genetic approach." In: *PloS one* 2.2, e217. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0000217.

Townes, Philip L and Johannes Holtfreter (1955). "Directed movements and selective adhesion of embryonic amphibian cells". In: *The Journal of Experimental Zoology* 128.1, pp. 53–120. ISSN: 0022-104X. DOI: 10.1002/jez.1401280105.

Waddington, C H (1957). "The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser." In: *The strategy of the genes A discussion of some ...* Pp. ix +–262 . DOI: 10.1007/3-540-32786-X_7.

— (2012). "The epigenotype. 1942." In: *International journal of epidemiology* 41.1, pp. 10–3. ISSN: 1464-3685. DOI: 10.1093/ije/dyr184.

Wagner, Andreas (1994). "Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization." In: *Proceedings of the National Academy of Sciences of the United States of America* 91.10, pp. 4387–91. ISSN: 0027-8424. DOI: 10.1073/pnas.91.10.4387.

— (2011). "Genotype networks shed light on evolutionary constraints". In: *Trends in Ecology and Evolution* 26.11, pp. 577–584. ISSN: 01695347. DOI: 10.1016/j.tree.2011.07.001.

Wang, Hai U. and David J. Anderson (1997). "Eph Family Transmembrane Ligands Can Mediate Repulsive Guidance of Trunk Neural Crest Migration and Motor Axon Outgrowth". In: *Neuron* 18.3, pp. 383–396. ISSN: 08966273. DOI: 10.1016/S0896-6273(00)81240-4.

Wright, Sewall (1932). *The roles of mutation, inbreeding, crossbreeding and selection in evolution*. DOI: citeulike-article-id:1586133.

Zhao, Shao-yun et al. (2015). "Spatiotemporal expression pattern of E-cadherin and P-cadherin during mouse tooth development." In: *Beijing da xue xue bao. Yi xue ban = Journal of Peking University. Health sciences* 47.1, pp. 42–6. ISSN: 1671-167X.

Zhu, Jianfeng et al. (2010). "Bare bones pattern formation: A core regulatory network in varying geometries reproduces major features of vertebrate limb development and evolution". In: *PLoS ONE* 5.5. ISSN: 19326203. DOI: 10.1371/journal.pone.0010892.