

Coupling of genotype-phenotype maps to noise-driven adaptation, showcased in yeast polarity

Daalman, Werner

DOI

[10.4233/uuid:0dec7916-fa0c-44ef-85ef-0694d3eb25b3](https://doi.org/10.4233/uuid:0dec7916-fa0c-44ef-85ef-0694d3eb25b3)

Publication date

2020

Document Version

Final published version

Citation (APA)

Daalman, W. (2020). *Coupling of genotype-phenotype maps to noise-driven adaptation, showcased in yeast polarity*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:0dec7916-fa0c-44ef-85ef-0694d3eb25b3>

Important note

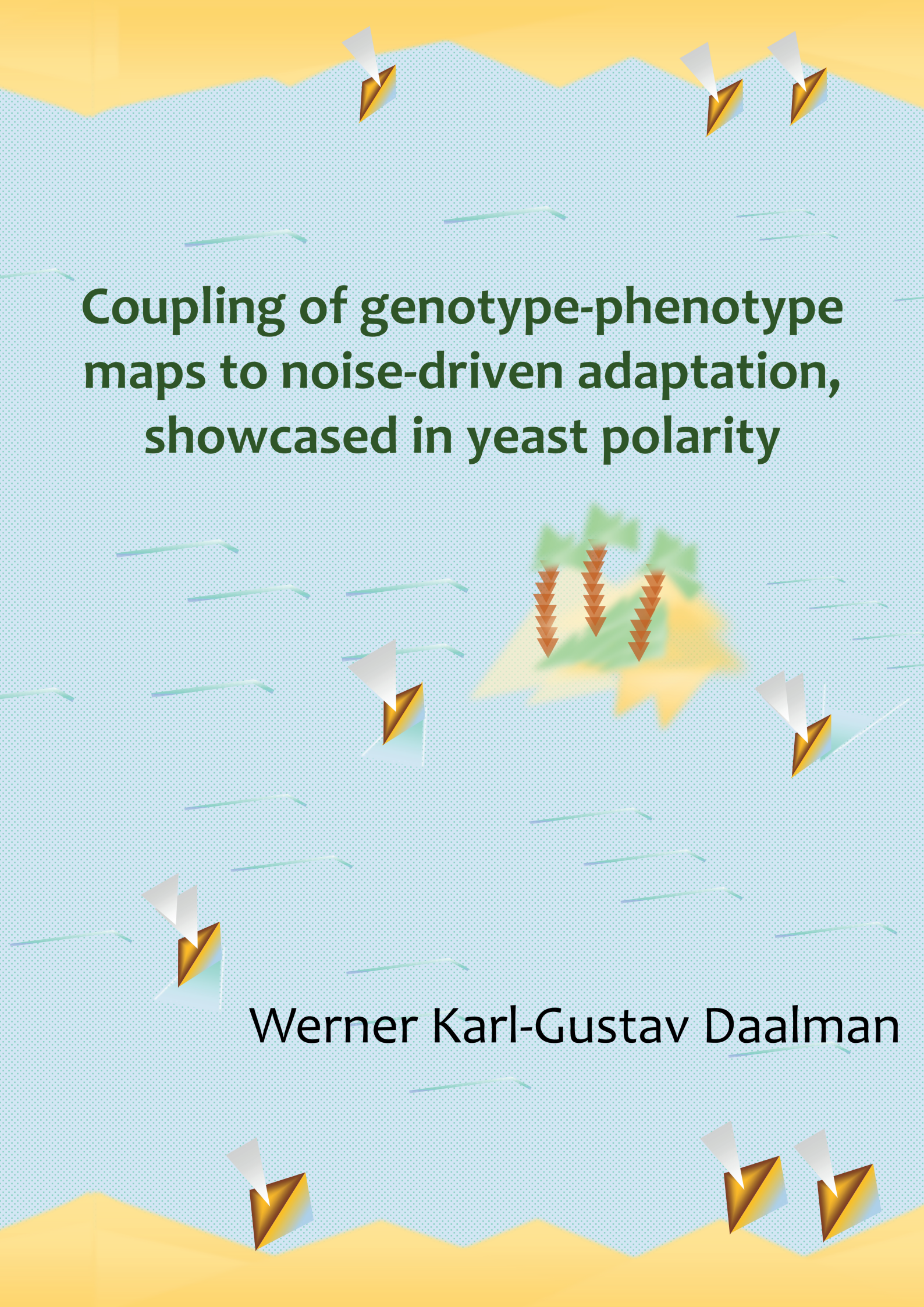
To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



Coupling of genotype-phenotype maps to noise-driven adaptation, showcased in yeast polarity

Werner Karl-Gustav Daalman

Coupling of genotype-phenotype maps to noise-driven adaptation,
showcased in yeast polarity

Dissertation
for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T. H. J. J. van der Hagen,
chair of the Board of Doctorates
to be defended publicly on
Wednesday 13 May 2020 at 10:00 o'clock
by

Werner Karl-Gustav DAALMAN
Master of Science in Applied Physics, Delft University of Technology,
the Netherlands
Born in Rotterdam, the Netherlands

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. A. M. Dogterom	Delft University of Technology, promotor
Dr. ir. L. Laan	Delft University of Technology, copromotor

Independent members:

Prof. dr. ir. S. J. Tans	Delft University of Technology
Prof. dr. P. Hogeweg	Utrecht University
Prof. dr. C. L. Wyman	Delft University of Technology / Erasmus University Rotterdam
Dr. R. Hermsen	Utrecht University

Other member:

Prof. dr. E. Frey	Ludwig Maximilian University of Munich, Germany
-------------------	---

This work was supported by the Netherlands Organization for Scientific Research (NOW/OCW), as part of the Frontiers of Nanoscience (NanoFront) program.



Keywords: self-organization, evolution, epigenetics, yeast polarity, genotype-phenotype map

Printed by: Ipskamp printing

Cover by: Werner Karl-Gustav Daalman

Copyright © 2020 by Werner Karl-Gustav Daalman

Casimir PhD series 2020-10

ISBN 978-90-8593-437-0

An electronic copy of this dissertation is available at repository.tudelft.nl

Table of contents

Samenvatting	9
Abstract	11
1 Introduction.....	13
1.1 Evolution.....	14
1.1.1 Currently known rules of evolution	16
1.2 Self-organization.....	21
1.2.1 Currently known rules of self-organization	21
1.3 Yeast polarization as model system for evolution and self-organization	24
1.3.1 What is yeast polarity?	25
1.3.2 Supporting evidence for yeast polarity as model system.....	26
1.4 Physical interaction network representation	28
1.4.1 A Venn-network diagram: Non-modular and hierarchical	29
1.4.2 Reaction-diffusion (green circle Venn diagram)	29
1.4.3 Actin (red circle).....	33
1.4.4 Timing (purple region)	33
1.4.5 Mating (orange circle).....	38
1.4.6 Bud scar (blue circle).....	39
1.4.7 What are the advantages of such a network?.....	40
1.5 Experimentally promoting modularity	41
1.5.1 Solution I: Mating (orange).....	41
1.5.2 Solution II: Bud scar (blue)	42
1.5.3 Solution III: Reaction-diffusion (green).....	42
1.5.4 Solution IV: Actin (red).....	43
1.5.5 General design considerations	43
1.6 Conceptual representation.....	44
1.7 Research questions.....	46
2 Experimental validation of a molecular model for yeast polarity.....	47
2.1 How to reverse-engineer the correct system	47
2.2 Overview of preparatory experiments for Cdc42-probe.....	50

2.3	Growth assay for Cdc42 threshold determination	51
2.3.1	Growth assays shows predicted trends while varying Cdc42 expression	51
2.3.2	Cdc42 thresholds are subject to growth conditions.....	55
2.4	Discussion	57
2.5	Acknowledgements	60
2.6	Supplementary information on methods.....	61
2.6.1	Cloning	61
2.6.2	Strain construction.....	61
2.6.3	Experimental assay protocols	61
2.6.3.1	Microscopy.....	61
2.6.3.2	Flow cytometry	62
2.6.3.3	Volumetric assay	63
2.6.3.4	Growth assays	64
2.6.3.5	Statistical methods.....	67
3	Genotype to phenotype model.....	69
3.1	Genotype, phenotype, ... mesotype	69
3.2	Mesotypical integration into a growth model.....	72
3.2.1.1	Cell growth	72
3.2.1.2	Coarse-grained polarity.....	75
3.2.1.3	Protein turnover.....	78
3.3	Model validation.....	83
3.3.1	Bem1 evolution data set.....	83
3.3.1.1	Model implementation	83
3.3.1.2	Fitness descriptions and NRP1 link	84
3.3.1.3	Cell cycle time descriptions.....	86
3.3.1.4	Cell size descriptions	88
3.3.2	Cdc42 expression sweep data set.....	89
3.3.2.1	Model implementation	90
3.3.2.2	Doubling times comparisons.....	91
3.3.3	Cla4 mutant data set.....	92
3.3.3.1	Model implementation	92
3.3.3.2	Doubling time comparisons	94
3.4	Model extensions	95

3.4.1	Different perturbation (media types)	95
3.4.2	Epistatic interactions (beyond polarity).....	96
3.5	Discussion	99
3.5.1	Outlook	104
3.6	Acknowledgements	104
3.7	Supplementary information on methods.....	105
3.7.1	Computational matters.....	105
3.7.1.1	Specifics on implementation.....	105
3.7.1.2	Initialization.....	106
3.7.1.3	Calibration.....	107
3.7.1.4	Model flexibility.....	108
3.7.2	Model validation	108
3.7.2.1	Bem1 evolution data methods	108
3.7.2.2	Cdc42 expression sweep methods	110
3.7.2.3	Cla4 mutant data methods	112
3.7.3	Model extension methods.....	114
3.7.3.1	Media types.....	114
3.7.3.2	Epistatic interactions.....	114
4	Evolutionary roles of noise and transgenerational feedback	115
4.1	Consequences of protein expression noise.....	115
4.2	Minimal model for epistasis	117
4.2.1	“Cdc42/GAP” test example	117
4.2.1.1	Model definitions.....	117
4.2.1.2	Epistasis definitions.....	119
4.2.1.3	GAP epistasis description is not dependent on expression noise distribution....	121
4.2.1.4	Noise can benefit steady state growth for the GAPs, not for Cdc42.....	123
4.2.1.5	Noise increases likelihood of negative epistasis mutants	124
4.2.1.6	Protein expression noise as evolutionary capacitor	125
4.2.1.7	Sharp lower mesotypical bound induces diminishing adaptive returns.....	127
4.2.2	General case	129
4.2.2.1	Model definitions.....	129
4.2.2.2	Noise shapes observed fitness curves, not the other way around.....	131
4.2.2.3	Noise can form the basis for global adaptive coupling.....	132

4.3	Transgenerational feedback as a noise mechanism.....	135
4.3.1	Transgenerational feedback effect on doubling times (Cdc42 case)	135
4.3.2	Transgenerational feedback triples epistatic region (Cdc42 case).....	137
4.3.3	Transgenerational feedback promotes evolutionary capacitance (Cdc42 case).....	138
4.3.4	Transgenerational feedback (general case).....	140
4.3.5	Transgenerational feedback promotes evolvability of mainly essential genes.....	142
4.4	Experimental design for transgenerational feedback	143
4.4.1	Preliminary assay	143
4.4.2	Simple assay design	145
4.4.3	More advanced assay design.....	146
4.5	Discussion	150
4.6	Supplementary information on methods.....	156
4.6.1	Epistasis fraction explained and as function of noise level	156
4.6.2	Fitness/survival contours as function of noise level and MI-point location.....	157
4.6.3	Epistasis generic mutations figure.....	157
4.6.4	Fitting empirical fitness landscapes.....	158
4.6.5	Simulated noise levels and correlations in feedback experimental design.....	160
5	Conclusions and discussion	161
	Bibliography	165
	Appendix	181
A	Material list.....	181
B	Strain and plasmid list	183
C	Preparatory experiments with a Cdc42-probe.....	185
C.1	Visible, inducible and functional Cdc42-probe.....	185
C.1.1	Cdc42 copy number control through the Gal-promoter	185
C.1.2	Fluorophore fusion.....	185
C.2	Probe validation.....	187
C.2.1	Fluorescence assays.....	187
C.2.2	Volumetric assay	192
C.3	Rough growth assay in suboptimal growth conditions	195
D	Membrane growth rate considerations for calibration	197
E	Reconstruction of burst parameters from FACS data	199
F	Effect of stochastic GAP production on Cdc42/GAP ratio.....	203

G	Convergence and error estimation	204
H	Single cell test case.....	206
I	Relaxation of model assumptions	207
J	Epistasis prediction from growth model	209
K	Effect of noise level in the two-state model	210
L	Two-state model fitness considering noisy GAPs in a $\Delta bem1$ background	211
M	Derivations volume-free two-state model.....	214
M.1	Concentration thresholds	214
M.1.1	Sharp boundaries, with feedback, log-normal approximation	214
M.1.2	Sharp boundaries, no feedback, log-normal approximation	214
M.2	Epistatic region width approximations	215
M.2.1	Sharp boundaries, with feedback, log-normal approximation	215
M.2.2	Sharp lower boundary, no feedback, log-normal approximation.....	216
M.3	Number of survivable generations approximations	217
M.3.1	Sharp lower boundary, with feedback, log-normal approximation.....	217
M.3.2	Sharp lower boundary, no feedback, log-normal approximation.....	218
M.4	Relative fitness approximations.....	219
M.4.1	Sharp lower boundary, no feedback, log-normal approximation.....	219
M.4.2	General case, with feedback.....	220
M.4.3	General case, no feedback.....	220
M.4.4	General case, feedback never decreases fitness	221
N	Preliminary experiment on Cdc42 distribution statistics	222
	Acknowledgements.....	223
	Curriculum Vitae	225
	Personal information.....	225
	Working experience	225
	Education.....	225

Samenvatting

Een van de grootste wetenschappelijke uitdagingen die deze eeuw onder handen zullen worden genomen is hoe eigenschappen van levende organismen hun oorsprong vinden in genen, de zogeheten genotype-fenotype kaart. Tevens kunnen eigenschappen op hun beurt genen beïnvloeden in een proces dat evolutie heet. De oplossing voor dit probleem zal grote maatschappelijke invloed hebben, met toepassingen in de voedselketen (bijv. het creëren van gewassen met een hoge droogte-tolerantie), industrie (bijv. materialen maken met micro-organismen) en gezondheidszorg (gepersonaliseerde medicijnen). De complexiteit van de genotype-fenotype kaart wordt veroorzaakt doordat deze doorgaans vele, sterk verweven schalen omvat (bijv. in lengte). Dit proefschrift borduurt voort op de ambitie dat uiteindelijk de oplossing gevonden wordt door lessen van simpele systemen te veralgemeniseren. Daarom onderzoeken we hier een overzichtelijk voorbeeld, polarisatie in bakkersgist, en maken inzichtelijk hoe evolutie op de kaart terugkoppelt.

Tijdens polarisatie in bakkersgist, een eencellig organisme, kiest de cel een richting waarin zij zal delen. Vele eiwitten, en in het bijzonder Cdc42p, organiseren zichzelf naar één plek op het celmembraan. Hoewel dit proces begint op een moleculair niveau, is dit uiteindelijk merkbaar op de populatie-schaal, in eigenschappen als dubbelingstijd. Om in detail te doorgronden hoe deze overgang in niveaus werkt, beginnen we van onderaf met experimentele tests van de moleculaire theorie achter succesvolle polarisatie in verschillende genetische achtergronden. Het theoretische model behandelt onder andere eiwitten die Cdc42 activeren en voor het eerst mechanistisch worden beschreven. Wij nemen hieruit voortvloeiende voorspellingen omtrent Cdc42p concentratie grenswaarden onder de loep in groeitesten met gistlijnen wiens Cdc42 productie we kunnen controleren. De experimenten hebben de theorie bevestigd waarmee moleculaire mechanismes die ten grondslag liggen aan polarisatie, zijn verklaard.

Om de sprong naar populatie eigenschappen te maken, heb ik een overzichtelijk groeimodel gemaakt, gevoed door simpele regels vanuit de eerdergenoemde theorie (waardoor moleculaire informatie dus slechts indirect gebruikt hoeft te worden). In het kort beschreven wordt Cdc42 geproduceerd in een kansproces, verdund door simpele volumegroei en bepaalt een concentratie grenswaarde of deling plaatsvindt. Ondanks dat veel details ontbreken, zijn opmerkelijke, experimenteel gemeten eigenschappen uit de literatuur in overeenstemming te brengen met model simulaties.

De eenvoud van de modelaannames gaf ook ruimte voor nieuwe inzichten op het gebied van evolutie. Ik behandel theoretisch hoe cellen die toevallig het geluk hebben meer eiwit dan gemiddeld te produceren, meer kans hebben te overleven en zo de populatie naar hun hand zetten. Op deze manier past de eiwitpiegel zich snel op een automatische en niet-genetische manier aan en kan het reageren op omgevingsveranderingen, maar dit blijft ook omkeerbaar. Gebaseerd op bestaande experimentele data verwacht ik dat dit op ruis gebaseerde

mechanisme merkbaar het aanpassingsvermogen van essentiële genen vergemakkelijkt (in gist voor 25-60% van deze genen). Vanwege de simpele aard voorspel ik dat dit ook in veel andere organismen gevonden moet kunnen worden.

Samengevat hebben we een succesvolle strategie gevonden om overzichtelijk de genotype-fenotype kaart te analyseren voor polarisatie in gist. Deze kaart kan worden uitgebreid naar andere functies, zolang er genoeg bio-functionele informatie beschikbaar is. De analyse liet ook een nieuwe evolutionaire koppeling naar deze kaart zien. Op een niveau boven het genetische, kan ruis in eiwitproductie ongehinderd worden gebruikt voor korte-termijn aanpassingen. Het experimenteel bevestigen van dit evolutionaire mechanisme in andere modelsystemen, waar we ook de eerdergenoemde strategie op kunnen uittesten voor het voorspellen van eigenschappen, zal een completer beeld geven over hoe eigenschappen in levende systemen worden gevormd en bijgesteld door evolutie.

Abstract

One of the biggest scientific challenges to be tackled this century is how traits of living organisms originate from genes, the so-called genotype-phenotype map, and conversely how traits influence genes through a process called evolution. The solution will yield a large societal impact, with applications in food (e.g., engineering drought-resistant crops), industry (e.g., material production through microorganisms) and health care (e.g., personalized medicine). The complexity of the genotype-phenotype map lies in how it typically spans multiple, interwoven scales (e.g., in size). This dissertation builds on the ambition that ultimately, a solution is found by generalizations of simpler systems. Therefore, we unravel here the map for a tractable example, polarization in budding yeast, and make insightful how evolution can couple to the map.

During polarization, the unicellular organism budding yeast chooses a direction in which it will divide. This involves self-organizing many proteins, in particular Cdc42p, to a single region on its cell membrane. While starting on the molecular scale, the process ultimately affects population traits such as doubling time. To understand the transition in scales in detail, we start bottom-up by experimentally verifying the molecular theory behind polarity success for different genetic backgrounds. The theoretical model treats, amongst others, proteins that activate Cdc42p, which are mechanistically included for the first time. Concretely, we test resulting predictions on sharp lower Cdc42p concentration bounds for viability using, inter alia, growth assays on strains variably producing fluorescent Cdc42p. The experiments confirmed the theory that allows reconstitution of molecular mechanisms underlying polarity establishment.

To advance to population traits, I constructed a tractable growth model, fed by simple rules emerging from the aforementioned theory (only implicitly encompassing the molecular information). Essentially, Cdc42p is stochastically produced, diluted by basic volume expansion, and must exceed a concentration threshold to divide. Despite disregarding many details, quantitative agreement between unintuitive, experimentally validated traits documented in literature and those from model simulations is reached.

The simplicity of the model assumptions also allows new insights in evolution. I elaborate theoretically how lucky cells that by chance produce above average amounts of protein, proliferate better to bias the observed population. Therefore, protein levels promptly adapt non-genetically, also in response to e.g., environmental changes, in a reversible and almost automatic manner. Based on existing experimental data, I predict this noise-based mechanism to notably expand the ease of evolution for essential genes (in yeast for 25%-60% of these). Due to its simple nature, I conjecture that it should be found in many organisms.

In conclusion, we find a successful strategy to tractably analyze the genotype-phenotype map in yeast polarity. The map can be expanded to other functions than polarity, provided that sufficient bio-functional information is available. The analysis also elucidates a new evolutionary coupling to this map. At a step above genes, noisy protein production can freely be utilized for short-term adaptation. Experimentally confirming the presence of this evolutionary mechanism in other model systems, and applying to these the same strategy to predict traits, will generate a completer picture of how traits of living systems are formed and shaped by evolution.

1 Introduction

Human kind has pondered for millennia on existential questions regarding inter alia the definition of life, our uniqueness, our relation to other organisms or objects, the fate of humanity and emergent properties such as free will. Initially, scientific attempts at resolving these matters were reserved for philosophers. For example, the ancient Greek philosopher Thales of Miletus believed ordinary matter was alive as well (hylozoism), while the ancient Roman philosopher Lucretius argued free will could originate from random swerves of atoms. Over the past few centuries, technological advancements allowed the ever-growing use of the scientific method to challenge hypotheses about life with empirical evidence. The efforts to answer the existential questions has transcended the realm of philosophy, and shaped the current state of the life sciences such as biology.

The key to biology is the characterization of life. In absence of consensus on the precise definition of life as described in e.g., [1], we will focus on the most important recurring criteria, namely (self-)sustenance and *evolution*. These criteria are to some extent intertwined around the concept of 'dynamical equilibrium'. Any useful definition of life must include relatively stable features, but any sensible definition must exclude fully static structures from being alive. This requires the system to incorporate a degree of flexibility, which can be exploited to evolve, while on the other hand, an equilibrium must be temporarily maintained. For the latter to occur, the arrow of time from the second law of thermodynamics is continuously countered. This loosely states that the maintenance of order consumes energy. A system accomplishing this transient equilibrium through dissipation of energy is also known to be *self-organized* [2].

In this dissertation I aim to shed light on the rules of evolution (section 1.1) and self-organization (section 1.2), two ubiquitous properties of life. As there is extensive literature on the topic, I aim to make a summary appropriate for this dissertation. As will be apparent, the omnipresence of these properties is a double-edged sword; while any living system could theoretically be used for their study, is it possible to find suitable model systems from which lessons can be drawn that have any reasonable generality? In section 1.3, I present the polarization in *S. cerevisiae* as a model system for evolution and self-organization.

The ample literature on this system required some tailoring for the need of this dissertation. By means of a Venn diagram, the knowledge is conveniently distributed in sub-categories, as done in section 1.4. Additionally, this allows discussing the experimental possibilities to dive deeper into the yeast polarity details (section 1.5), which will prove necessary for answers on evolution and self-organization. For this purpose, I advocate a novel view on the polarization network (section 1.6), which maximizes the generality of these answers, and ultimately, provides my research questions (section 1.7) with the necessary tractability.

1.1 Evolution

Generally, evolution occurs when an agent can in a memorable way adapt its functionality to affect its proliferation. Acquired traits that are useful for survival will have an increased chance of selection (*natural selection*) and agents with these traits are ultimately the most likely to remain (often dubbed *survival of the fittest*). The selection phenomenon is extremely universal and therefore applicable in many fields. Computer program code fighting for CPU time exists as a form of in silico evolution [3], there are socio-economic generalizations [4] of evolution and one could even envision the current state of the universe as resulting from cosmological evolution [5]. In this vastness of implementations of this concept, we recognize its importance.

If we constrain ourselves to biological examples, we can be more specific about what happens during evolution. Functionality is encoded in organisms in various ways, most notably genetically through its DNA, but also epigenetically (i.e. non-genetically) through e.g., histones curling the DNA, and is generally referred to as its (*epi*)*genotype*. Thus, when memorable changes are required, these need to be embedded into the (*epi*)genome, the whole of the stored information. This “database” cannot be a completely permanent record, and in nature replication or repair errors allow diversity that can be exploited for natural selection. This selection acts on the quality of the resulting trait, or generally called *phenotype*. As the phenotype results from the genotype, there is hence a reciprocal influence through natural selection, see also Figure 1.

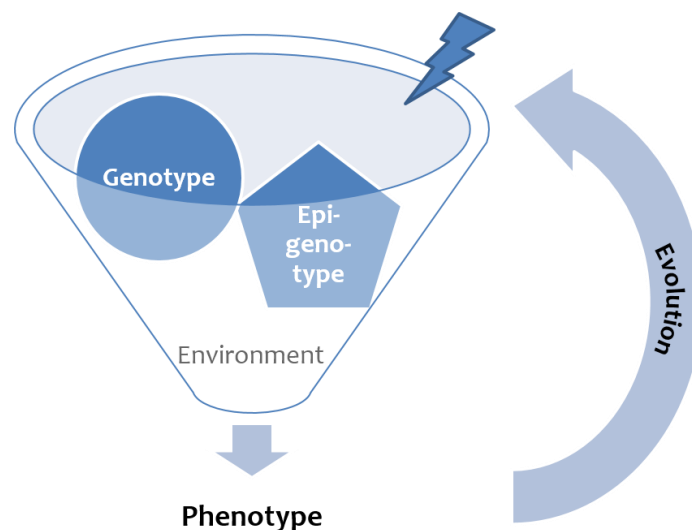


Figure 1 Conceptual representation of evolution. The information encoded in the genotype and epigenotype form components that, through the funnel of the environment, cause a particular trait or phenotype to occur. The success and ease of this process relies on its details/shape. Evolution in turn relays information of the phenotype back to the start of the funnel, where it acts (the lightning bolt) to alter the memorized information, which alters the shapes of the (*epi*)genotype. This again affects how the information flows through the funnel, and consequently the phenotype emergence, in a continuous process.

The topic of evolution is extensively covered in literature. Before diving deeper into this coverage, it is convenient to describe important terms and concepts, see Box 1. These will also be of use in the context of self-organization.

Box 1 Terminology in evolution and self-organization

Self-organization

Interacting agents (genes, proteins, organisms, etc.) which automatically order themselves at the continuous expense of energy

Streamlining

Reduction of genomic size after a period of expansion and adaptation

Neutral mutations

Mutations that in a particular background have (almost) zero fitness effect and cannot be subject to selection.

Mutational neighbourhoods

Genetic backgrounds that can be reached within a few mutations

Robustness

Maintenance of function under genetic or environmental perturbations

Neutral theory

The idea that most mutations fix due to random drift, not selection

Epistasis

Deviation from the expected result of combining mutations, based on the individual effects

Evolvability

The potential for further adaptation given the current genetic state

Pleiotropy

Multiple phenotypes being influenced by mutations in a single gene

Modularity

Degree with which interacting agents (genes, proteins, organisms, etc.) can be grouped based on their interactions, due to the lack of connectedness between groups.

Complexity

Different formulations exist, such as ease of replacing components of a system, or its combinatorial likelihood. I advocate defining complexity as the reciprocal of modularity.

Network motifs

Parts of the network where outputs follow non-linearly from inputs, such as positive/negative feedback, feed forward loops or mutual inhibition

1.1.1 Currently known rules of evolution

Despite the stochastic nature of the genotypical variation underlying evolution, rules that evolution tends to obey are described in literature. This section does not aim to give a complete overview of these rules, but states the most important ones for the context of this dissertation. Concretely, what is known regarding the *interplay between epistasis and evolvability* must be discussed, since Chapter 4 describes a mechanism which affects this interplay. The evolvability in Chapter 4 will be seen within the relevant range for *genetic innovations*, namely from deletions to duplications (a statement motivated below), and as how long *environmental changes* can be withstood. To arrive to the mechanism of Chapter 4, we first have to reconstruct the *genotype-phenotype map from multiple intermediate levels* (Chapter 3). The experimental validation (Chapter 2) of these level definitions relied on *modularity* of our biological system of interest, which is further elaborated upon in this chapter. Therefore, for the concepts *in italic* (see also Box 1 for definitions) the most relevant rules are highlighted below:

- **A mechanism for genetic innovation is starting out with a modification/refinement of an unspecialized piece of DNA, which has originated from a duplication or movement of an existing piece. For very large pieces, evolution streamlines after expanding.**

Such a process is described in [6], where authors propose the innovation, amplification and divergence model. This entails that the gene to evolve first duplicates (amplification), a process which does not particularly promote its original activity. Yet, it allows minor improvement in another function (innovation), a property acquired by chance. Having two copies for the original activity, there is room for adaptation of the second copy, which then specializes in the function which initially played a minor role (divergence). In the famous Lenski experiment, similar behavior was also observed [7], when a promoter duplicated and moved in front of a previously silent citrate transporter, allowing after some refinement a new function, namely firm growth on citrate.

Such a behavior can even be forced by removing an important function and providing the cell with very rough working material. In [8], relatively generic sequences were added through a plasmid on auxotrophic *E. coli* cells in nutrient-poor conditions. Some of these sequences already yielded a small, but essential piece of functionality for survival. Following up on this, in [9] it was shown that these generic sequences can also form a basis for evolution.

That this mechanism is not rare or with marginal effects over the course of evolution, is suggested in [10], where not a gene is duplicated, but the whole genome. There, tetraploid yeast, which predominantly occur as either a haploid or diploid organism, may have suffered from the many copies of genes in its genome, but remained viable. Gene regulation is typically not correcting the excess amount of protein [11]. Instead, the excess provided the

yeast with room for many beneficial mutations, and increased its *evolvability* compared to the standard yeast. The fitness benefit of polyploidy is not found in all environments and return to diploidy is also relatively rapid, yet emergence of polyploidy is recurring [12] because it may allow easier immediate adaptation in non-standard circumstances.

Not only in the lab, but also historically, genome duplications are known to have a large impact. This ranges from yeast undergoing a genome duplication and then diversifying (leading to speciation, including *S. cerevisiae*) [13] to flowering plants [14]. Additionally, in silico experiments also show the feasibility of this evolutionary mechanism, as described in [15], where genomic expansion followed by the opposite, genomic streamlining, is a common route towards adaptation.

- **Epistasis and neutrality in conjunction generate genetic evolvability while energetic considerations are typically negligible.**

The existence of the genome duplication/refinement mechanism just discussed, teaches us two things. Firstly, apparently the maintenance cost of genes, even genomes, is not so large, which is supported by energetic calculations [16]. Secondly, perhaps many mutations are *neutral*. Authors in [17] provide an experimental exploration of neutral mutations, and illustrate how mutational neighbourhoods of genotypes can be far from random. How networks containing many neutral mutations can arise has also been investigated in silico [18]. The abundance of neutral mutations, which promotes the *robustness* of the system, has also been the foundation of neutral theory which states that most diversity is actually caused by random drift instead of selection [19]. This could mean that the danger of crippling mutations when duplicating a gene is generally not so large.

However, the appearance of neutrality can be on occasion deceiving. Authors in [20] show that in yeast varying the dosage of 81 genes typically affects fitness when done one at a time. This would make it almost unimaginable that yeast could survive a genome duplication. A solution to this paradox may lie in the concept of *epistasis*. This entails that the phenotypic (e.g., fitness) outcome of changing multiple genes may be different from what would be expected based on the individual outcomes of every gene change. If the total outcome exceeds expectations, this is positive epistasis, if the expectation was superior this is negative epistasis. Moreover, if a mutation switches from deleterious to beneficial or vice versa conditional on another mutation, this is called sign epistasis. For reciprocal sign epistasis, the other mutation must also switch in influence (see example in Figure 2).

Epistasis is widespread in yeast with multiple known molecular mechanisms, as described in [21]. In this paper, examples include mutations in two genes both essential to the function of a complex or pathway, and mutations that repress the destabilizing mutation of a gene product through targeting protein/mRNA degradation. Also, many statistical origins of

epistasis exist from gene regulation [22]. Despite the short duration compared to the historical evolution, it is also seen in experimental evolution (e.g., [23], [24]).

Epistasis and neutrality have important general consequences on evolvability. This can be well envisioned by imagining the optimization of a function or fitness in general by genetic changes, as trying to reach the highest point in a landscape, a comparison dating back almost a century [25]. The horizontal movements represent genetic changes, the vertical movements those in fitness. The idea is that evolution favors the path that moves upward the fastest, although in a large population size other paths become available if the end point is higher [26]. Existence of epistasis is then represented by a rugged landscape (authors in [27] show it requires epistasis), while neutral mutations are flat paths. A conceptual, very simplified depiction can be found in Figure 2.

Intuitively, a rugged landscape with pervasive (fitness) valleys would diminish the one-step evolvability, as the landscape becomes filled with local maxima from which an upward trajectory is not possible. However, neutral mutations allow drifting towards other locations on the map that are more favorable to changing a particular gene. In fact, ample sign epistasis would make it likely that somewhere in the genome, there exists an upward trajectory again. Previously deleterious mutations may become improvements after a “potentiating” near-neutral mutation (akin to the emergence of the citrate usage phenotype in [7]). So, epistasis does not necessarily create local maxima which trap the system and negatively affect evolvability, at least not in conjunction with the availability of neutral networks.

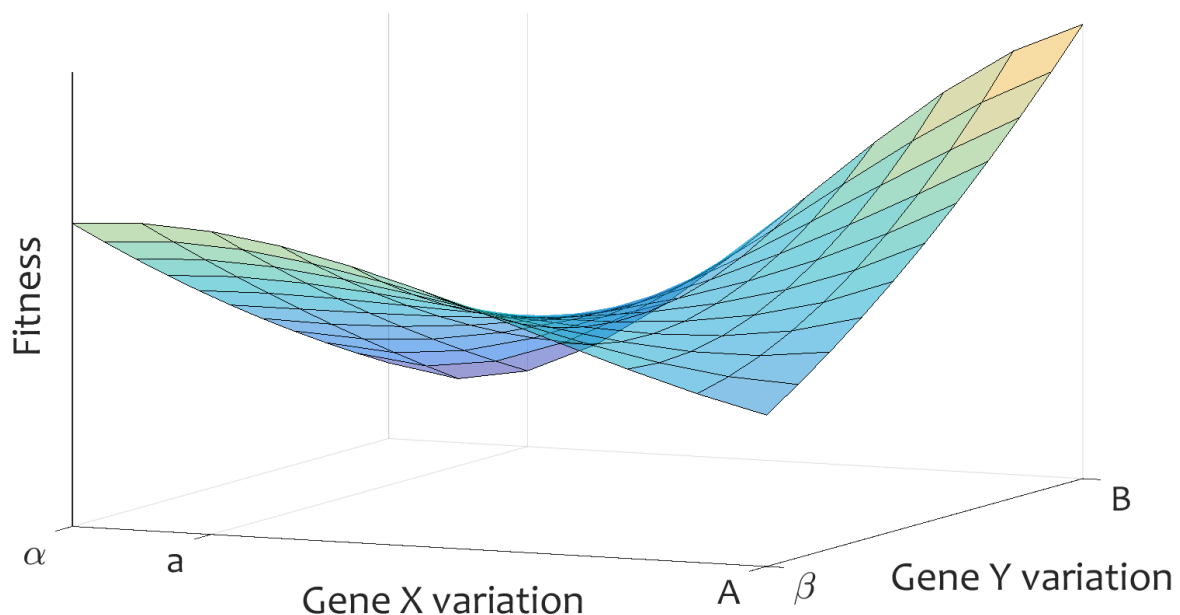


Figure 2 Hypothetical simplified view of a three-dimensional fitness landscape, where the height (z-direction) shows the fitness, dependent on the variation in genotype space of gene X and gene Y, with alleles α , a and A for gene X and β and B for gene Y. The mutation from α to a can be considered neutral when gene Y has allele β . Mutations $\alpha \rightarrow A$ and $\beta \rightarrow B$ exhibit reciprocal sign epistasis. Mutations $\alpha \rightarrow a$ has positive epistasis with mutation $\beta \rightarrow B$ (yet negative with $B \rightarrow \beta$).

The interplay of robustness and evolvability is not straightforward. As described in [28], the outcome of robust genetic networks can be beneficial or deleterious for adaptability, depending on selection parameters such as population size. Naively, one would imagine a large probability to remain stable (robust) in function as contradictory to the need for the ability to change and adapt (evolve). Yet, authors show how modest robustness to allow more variation can enhance evolvability compared to zero robustness. Furthermore, combining epistasis with neutral mutations in the right proportions may give fitness landscape plateaus, as well always an escape route off the plateau. Still, this mechanism suggests that a balance or trade-off is needed between robustness and adaptability. Analogously, authors in [29] find a trade-off between environmental noise and adaptability. In Chapter four of this dissertation, we will go further into this, and determine whether such a trade-off is always needed.

- **Epistasis, neutrality and pleiotropy cause the genotype-phenotype map to be non-trivial. Modularity mitigates complexity to some degree, but shifts the problem to the best modular level choice.**

As explained previously, evolution forms a feedback loop from phenotype back to genotype. In order to understand evolution, we must determine how genotypes are connected to phenotypes. This would have been much more tractable if the effects of individual mutations simply add up, but as mentioned, epistasis is commonplace and prohibits such simplicity.

Moreover, multiple genotypes can map to the same phenotypes because of neutral networks. But conversely, it is also possible for a single genetic mutation to alter multiple phenotypes, which is called pleiotropy. This may not always be very common [30], but shows that the *genotype-phenotype map* is highly complex.

One way to at least isolate part of this map is to dismantle the map's architecture into pieces. If this is possible, the map or network is said to be modular. For example, in yeast some functional *modularity* has been observed in the paper of [31], and in subfunctions [32]. It is important to realize that the classifications of genotypes or phenotypes into modules is always fuzzy rather than binary, and secondly that using functional classes may not necessarily be the optimal segregation method. How to appropriately define a module has been a long outstanding question, see e.g., [33].

Mathematically inclined approaches also exist, where a suitable level definition (trophic level) and the extent to which this segregates the network (trophic coherence) is shown to explain stable, robust system behavior [34]. Yet, there is no complete, sound bottom-up approach to biophysically understand how genotypes lead to phenotypes. The benefit would be an improved insight in how evolution couples back to the lower levels. This approach is further discussed and pursued in Chapter 3 of this dissertation.

Ultimately, different segmentation methods (level definitions) may elucidate different kinds of modularity, that may all have their own advantage for depicting particular features of the genotype-phenotype map. In this light, I see modularity as the reciprocal of another famous concept in evolutionary literature, namely *complexity*. Notably, no definition of the latter will satisfy everyone, since many definitions exist, see e.g., [35].

- **Evolution can act on multiple levels simultaneously, also affecting evolvability. Levels exist in various ways, such as (length) scale and time.**

Analogous to the genotype-phenotype transition, the path back to genotype through evolution is also level-dependent. A multi-level approach is also suggested for example in [36], where levels are interpreted thermodynamically. Irrespective of level choice, it is clear that selection can act on multiple levels. In a simple case of defining levels through size, on the nanoscale a gene encoding for an essential protein will undergo positive selection. More extremely, on the astronomical scale orbital variation may have influenced mammalian evolution through climate changes [37]. When simultaneous selection on levels occurs, this can lead to interesting dynamics. In the cell to populations scales one can observe selfish cells ruining the population by only doing what is good for the cell [38]. Clearly, presence of multiple levels is important in describing evolution in general.

Interestingly, one could also add a temporal level interpretation to evolution, in order to also take into account epigenetic changes. In [39], epigenetic modifications are viewed as an evolutionary precursor for genetic changes and an intermediate in the relay race of possible adaptations. One can envision many forms of epigenetic inheritance, see e.g., [40] for an extensive review on epigenetic mechanisms. For a generic mechanism, it has been mathematically studied how adaptation to environmental cues takes place through transgenerational feedback [41]. In Chapter 4, a concrete example of this feedback through protein copy number is described, although it is the “genetic environment” that is sensed.

- **Fluctuating environments can facilitate evolution, and can in itself even influence the probability of certain genetic variation to occur**

Generally, the environment has a profound effect on evolution, because systems are never in complete isolation. Systems aim to adapt to a certain environment, and its fluctuations may facilitate the speed of evolution. Such a process is described in [42] as a trade-off ratchet, where a dead-end in the fitness landscape becomes useful when the landscape changes through the environment. Fitness peaks and valleys get rearranged, promoting otherwise implausible evolutionary paths.

Surprisingly, the environment can also make its influence noticed by influencing the process at the core of evolution, namely causing genetic variation. These variations are not always random, but differ per environment [43], [44] and may even be of service for evolvability during particular stresses [45].

1.2 Self-organization

Next to evolution, the second critical property of life is self-organization. This phenomenon is seen in length scales across many orders of magnitude. Intracellularly, these can be protein complexes such as microtubules, organelles such as the Golgi apparatus, while macroscopic examples include the schooling of fish (see Figure 3). Several rules emerge when observing the variety of these systems.

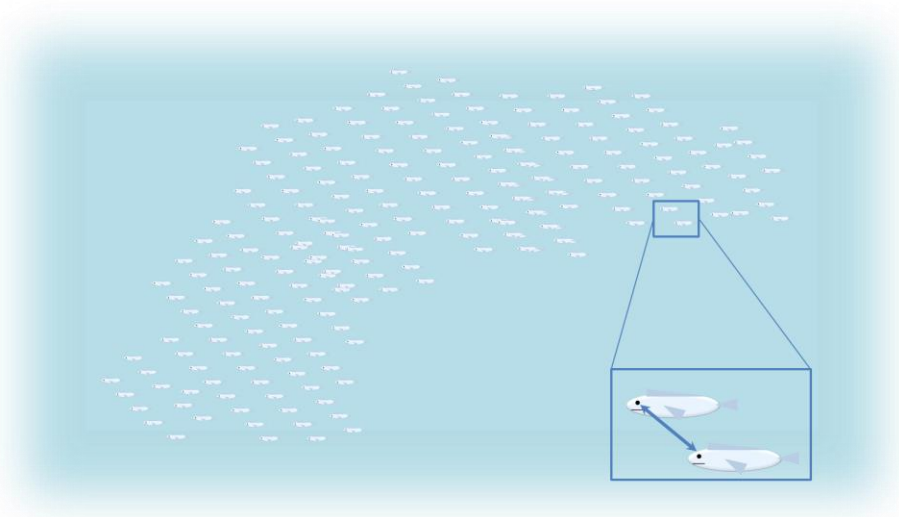


Figure 3 A school of fish as an example of self-organization. Collectively, a (curved) shape is formed, but underlying interactions are local, making it an example of a self-organizing system. The inset shows how each fish only looks at its neighbour, changing direction concordantly and possibly moving inward when having no outside neighbour, to avoid being exposed to marine dangers at the edge of the school.

1.2.1 Currently known rules of self-organization

- **Subunits interact locally (consuming energy) to form a global pattern**

Inherent to self-organization is that a global pattern forms from mere local interactions of the constituent units [46]; no fish has a notion of the shape of the school it is in, yet by only paying attention to its immediate surroundings the school as a whole exhibits complex collective motions. In the case of identical subunits (the school only contains one type of fish),

the description of the large-scale behavior seems like a small step once one understands the rules the single subunit obeys. However, once the global pattern is formed from different kinds of subunits, the behavior can appear more complex and varied, leading to the following rule.

- **Self-organization may be encoded in intricate networks that generate new, emergent properties, possibly levels higher.**

In order to maintain organization, living systems wield similar motifs as people use in electrical engineering, such as positive feedback, negative feedback and feed-forward loops. For example, a centrosome (an organelle arranging chromosomes) may use the self-organization of microtubules (cables of the protein tubulin) alongside motor proteins to position itself stably and accurately within the bounds of the cell [47]. Here, stabilizing, negative feedback on the relative position of the centrosome emerges from self-organization of simpler, smaller parts from which such an effect is not obvious. Similarly, multiple positive feedbacks may lead to stable, robust pattern formation in yeast polarity [48], another example of self-organization of proteins. Moreover, neurons are suggested to be able to self-organize into feed-forward loops [49]. Synthetic biologists also make convenient use of building self-organizing systems in terms of network motifs, as is done for artificial polarity in yeast [50].

Because of the leap in scales from subunits to the global structure, the emergent properties often seem unpredictable. Dissection into network motifs is one possible manner to deal with this. Regardless of whether such a mathematical/engineering approach is used or another, defining intermediate scales to predict behavior is necessary to get a grasp on the emergent properties and patterns. This chops the path to the emergent properties as depicted in Figure 4.

This level definition problem is the same as encountered in the previous section concerning evolution and predicting the genotype-phenotype leap. In section 3.1, a more detailed overview is given regarding existing level definitions, their benefits and shortcomings. Consequently, a novel definition, better suited to understand evolution in a self-organized system, is provided and put to the test for the model system of choice, yeast polarity.

In short, the idea is that a *mesotype* can be introduced, which coarse-grains molecular information of the gene product interactions to simple rules, such as concentration thresholds to polarize. While it can be biophysically justified from the bottom-up why the mesotype exists, it also helps for understanding evolution. For example, evolution favors lucky cells, where luck is measured at the mesotype level, which has lasting effects on the population level resulting in e.g., higher expression than initially expected (see Chapter 4).

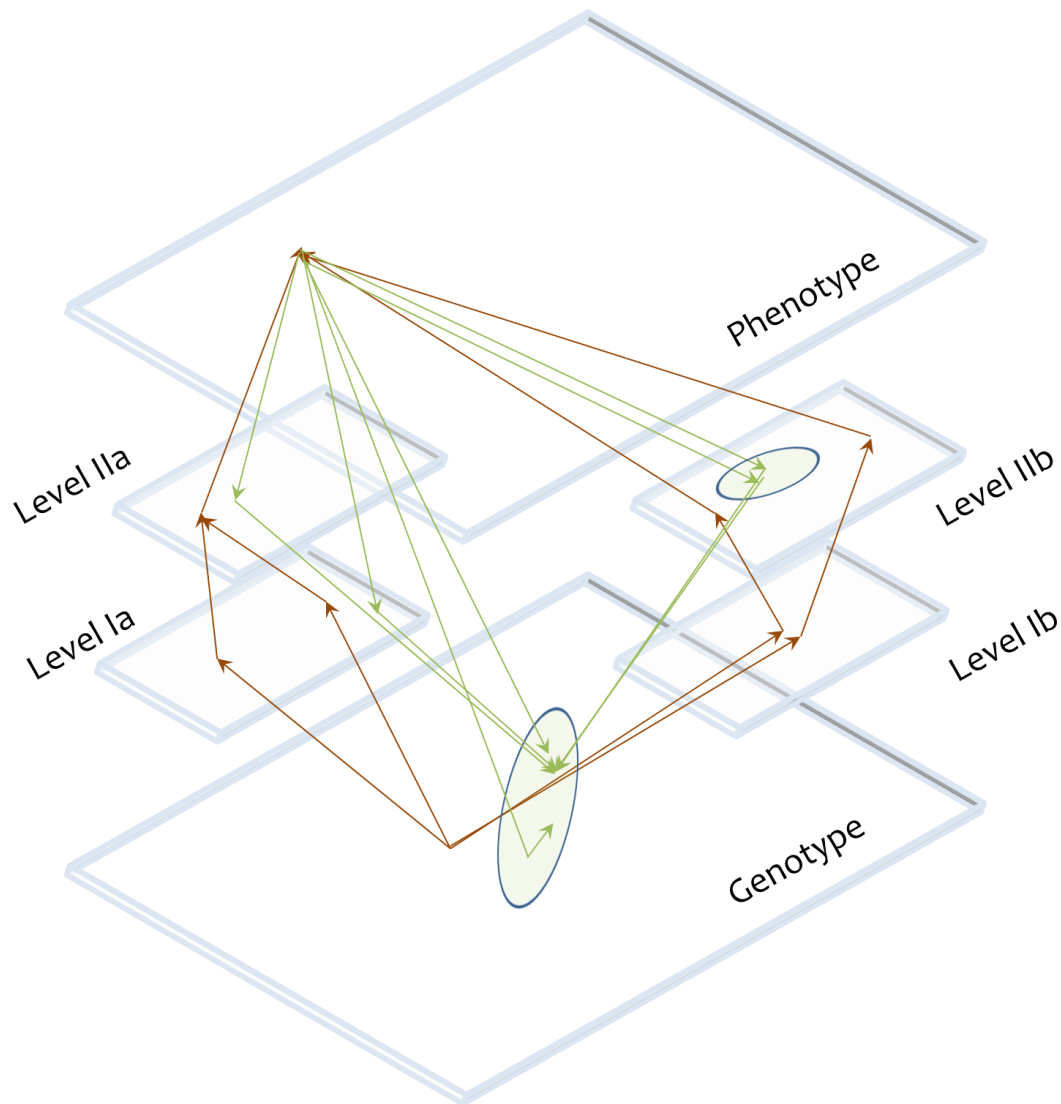


Figure 4 Conceptualization of level definitions (blue plateaus) within the genotype-phenotype map. Dark orange arrows indicate a function affecting a fitness-influencing phenotype encoded in the genome, which can also be subject to self-organization. Within the genotype space, the current gene set comprising the phenotype (starting point of the arrows in the genotype plateau) yields different trajectories, depending on the level definitions (a or b). For definition a, the trajectory spreads out more at level I yet converges at level II, the opposite of definition b, symbolizing non-modular (and more complex) interpretations at level Ia and IIb, and more modular (and less complex) interpretations are level IIa and Ib. As an example, suppose definitions a are constructed from GO annotations, with gene *CLA4* mapping in level Ia to polarization and cytokinesis, which both map onto cell division in level IIa, converging in that level. The modularity is different in set b, where different polarity mechanisms (level IIb) arise using the functional subunits concept [51] (see also 1.6), forming level Ib. *Cl*a4 is a member of the transport subunit, but also involved in the polar activation unit. The role difference becomes more pronounced at the mechanistic level (WT/rescue), although with the same effect on cell growth. Ultimately, any level definition results in the same phenotype, which influences the next genotype through evolution (green arrows). Within the genotype space, there exist a neutral neighbourhood (green ellipse) within which the evolution arrows may land, and multiple endpoints within this area exist. The number of arrows indicate the evolvability of the previous genotype. The most southern landing arrow cannot immediately access its endpoint, it must bend by finding an appropriate neutral mutation first, denoting epistasis. The best (most modular) level choice for build-up of self-organization phenotype is not necessarily the same for evolutionary predictability; in definition a, evolution turns out to act on multiple levels, while in definition b effects are more predictable, skipping one level.

- **Self-organization can be steered/overwritten/masked by the presence of more dominant spatio-temporal cues**

While self-organization may work autonomously from external signals, it can be redirected in the presence of external cue. An example of this is the formation in zebrafish of the posterior lateral line, a sensory organ, from primordial cells (see e.g., this review [52]). About 125 cells move along a line defined by external chemical signaling from other cells. This chemical cue affects the leading primordial cells during migration, but not the cells that follow. Even when inhibiting the external chemical cue and through laser ablation physically separating the leading cells from the bunch, the primordial cells can still rejoin [53]. The leading cells signal towards the trailing cells and reconstruct their original state through self-organization. *While external cues may facilitate this, they are not essential.*

Similarly, the direction of division in budding yeast (polarization), which will be further discussed in the next section, can also depend on chemical signaling in the form of a mating cue. However, this cannot be the only source the cells rely on for determining directionality. Suppose surrounding, competing cells give out a false mating cue, causing other cells to prepare their mating machinery although no such mating attempt will be made. Affected cells should be able to overwrite this attempted deception and find their own direction to divide, which is indeed the case [54].

This suggests that self-organization can be interpreted in two ways. On the one hand, self-organization may act as a back-up mechanism, in case external signaling remains absent, is deceptive or otherwise insufficient for correct function. Alternatively, it may serve as a quick basal manner of organization, after which evolution tailors it towards perfect functionality.

1.3 Yeast polarization as model system for evolution and self-organization

In order to add to the known patterns of evolution and self-organization as described in the previous sections, we will turn to a system that is exemplary, a model system. Technically, a model system is part of a model organism with desirable practical properties. These include easy growth conditions, fast generation times, and availability of tools for (genetic) manipulation.

The latter is in line with the properties required for a good evolutionary study. Ideally, abundant knowledge of the underlying genes is available. The genotype-phenotype map is in general quite complex, as previously mentioned, even when it is clear to which function proteins connect. Thus, the genotype must be as well defined as possible.

Furthermore, it must be shown that evolutionary solutions to imposed perturbations, that are needed to probe the systems, are relatively reproducible. A problem in the interpretation of evolutionary trajectories during experimental or historical evolution commonly lies in the multitude of paths that can be taken. The same starting points in the same experiments hence yields different results, which can signal a level of complexity (or absence of modularity) not suited for a model system.

Additionally, in terms of generalizability for self-organization, we ideally like to see multiple modes of transport (of the underlying subunits) that are commonly employed. These can be categorized as active and passive transport. A passive mode entails that net transport results from diffusion, while a series of reactions takes place in the process. Directed movement of cargo to a predefined spot is an example of active transport, such as movement along the cytoskeleton.

Furthermore, a tractable number of components causing the local interactions is desired, which equates to at most about ten different subunits generating the global pattern. Thirdly, modularity is important, so ways to experimentally promote this in the model system must be known. Conversely, within the module we seek a rich interaction network for emergent properties (e.g., robustness and redundancy), Finally, the model system must allow exertion of controlled perturbations with measurable effects.

What I will argue is that polarization in budding yeast qualifies a suitable model system for evolution and self-organization. After elaborating on the general idea behind yeast polarity, this section will show supporting arguments for this claim.

1.3.1 What is yeast polarity?

The unicellular, approximately spherical organism *S. cerevisiae* is well-known for its contribution to the food industry, yielding products such as beer, wine and bread. It must break its internal symmetry in order to generate a new daughter cell (bud) from itself, starting from one point in the plasma membrane (bud site). The process of directing cell growth, involving dozens of proteins that condense onto one point on the membrane loosely forming a modular protein network, into the direction of the bud site is called polarization (see Figure 5 for an example). This occurs at the end of the G1 phase of isotropic growth.

The first step in this process is choosing the bud site, which is the step where the internal, spherical symmetry is broken. External cues, such as pheromone gradients or 'historical landmarks' (proteins deposited at an earlier cell division), can impose the symmetry breaking, but in that case the global pattern is not formed by local interactions alone. However, symmetry breaking can also spontaneously occur by amplifying noise in the spatial distribution of the involved signaling proteins, leading to a random directionality in the accumulation of proteins. It is in this case that about a dozen proteins *self-organize* to form a spot on the membrane from which further signaling for cell division may take place.

1.3.2 Supporting evidence for yeast polarity as model system

Generally, yeast has been a convenient object of study. It forms part of the main bundle of model organisms [55], and the consistent research efforts has yielded a wealth of information, accessible through online databases. It has been sequenced for more than 20 years [56] and a myriad of gene, protein, function, interaction and sequence information can be found online [57], [58], where the knowledge is continuously updated. Strains with single gene knockouts are readily available for use in the lab [59] and protocols are abundantly available as well (e.g., [60], [61]). While beyond the need of this dissertation it is worth noting is that with the advent of Crispr-Cas in yeast, the speed of genetic manipulation has sharply increased (see e.g., [62] for six manipulations at a time). In conclusion, the majority of practical considerations has been addressed when working with yeast.

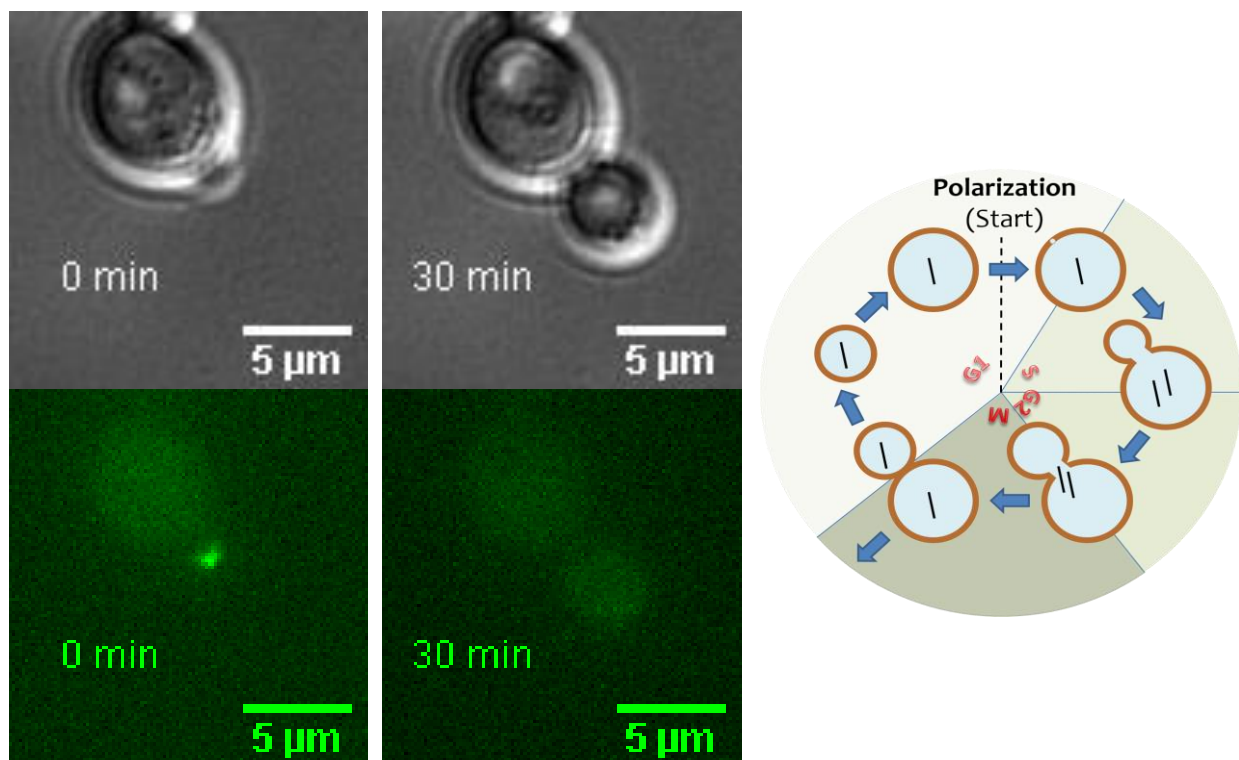


Figure 5 Example of budding yeast at two time points, first forming a bud and 30 minutes later, during a microscopy experiment (Left and center top brightfield, left and center bottom widefield fluorescence). Inside the cell, a binding partner of the most important protein during polarization is fluorescently labelled to visualize the signal that the cell uses to mark the future bud site. Minutes later, the bud expands as the cue vanishes. Right: schematic overview of cell cycle phases, where black lines inside the cell indicates ploidy and the cue for polarization establishment ('Start') is indicated with a dashed line.

For our evolutionary needs, we can note that it has been used numerous times in evolution experiments, for example evolving against salt stress [63], multiple stresses [64] or even the stress of a severe genetic perturbation [24]. One reason for this may be the tractability of the genetic architecture, which makes it well suited for analysis. It is important to make some sense of the genotype-phenotype map and yeast offers this to a certain degree. As mentioned before, it exhibits decent functional modularity [31] and within the polarity module there is more sub-functional modularity [32]. This module has another very appealing property. Under the right perturbation, it becomes possible to make experimental evolutionary trajectories of a trait reproducible, identifiable and quantifiable [24].

Going into more detail on yeast polarity, one can observe that it has been extensively studied through multiple means, so that the leap from genotype to phenotype will soon be feasible. For example, in a bioinformatical approach the evolutionary history of yeast polarity has been studied in [65] (among others of budding yeast), while good reviews of detailed molecular knowledge of the current system can be found in [66], [67]. Hence, not only is yeast well studied, it holds for its polarization machinery as well, which seems to have a tractable basis upon which evolution research may be conducted.

Finally, in terms of self-organization, we reencounter the desirable properties postulate beforehand. Yeast can polarize using multiple pathways, for example, proteins can concentrate on the plasma membrane through reaction-diffusion (passive transport), or by means of transport on actin cables (active) [48]. Secondly, as will be shown in the next section, the number of components within the reaction-diffusion pathway in particular is tractable. The interaction network is not fully mapped yet in literature, but will be rich given earlier studies showing clear epistasis. In section 1.5, it will be shown in detail how the modularity and prevalence of certain elements above other interactions of various pathways can be enforced experimentally, which is required as well. For example, in [24] we note how deletion of a single gene (*BEM1*) leads to promotion of the reaction-diffusion pathway, while the dependency on other proteins (more specifically named GAPs, see 1.4.2) becomes critical, something normally modelled in low detail [68].

In conclusion, polarity in budding yeast will provide an excellent platform upon which evolutionary and self-organization studies may be conducted. Analogous alternatives are either expected to be less rich in network behavior (for example, consider the minimal systems of budding yeast and *E. coli* [69]), or currently still much too elusive (human cells). This system is at the right time with the right size to study.

Now that the use of studying the spontaneous symmetry breaking during polarization in budding yeast has been established, we will piece by piece dissect and describe the whole network. This introduction features two ways to classify this network, each with their own benefits and shortcomings. Firstly, a more traditional way focused on physical interactions is described, which needs to precede the second, conceptual way suitable for generalization to other systems.

1.4 Physical interaction network representation

Studying biology bottom-up, whether the object is an organism, a single function or even one network component, undergoes multiple historical phases, in which the level of detail is steadily increased. Firstly, the relevant genes and possible alleles must be documented. Using crossing techniques or genetic manipulations, the step of finding simple genetic interactions follows, from which an interaction map can be made and hypotheses can be drafted regarding functional groupings. Yet, genetic interactions do not shine sufficient light on the associated biological processes. Therefore, relevant physical interactions must be studied, whose range covers bulk affinity assays to detailed genetic engineering.

It is at this stage where multiple important system properties become apparent. A physical interactions map will be a more reliable representation of for example (causal) hierarchy, modularity, possible redundancies and pathway groupings. Fortunately, thanks to many years of tedious work, this level of understanding has generally been reached for budding yeast. The next subsection will treat the polarization module in terms of its internal physical interactions, displaying four pathways with a clear hierarchy, although with significant overlap, and one timing pathway as a control knob.

Importantly, this overlapping, hierarchical representation allows grouping of available literature and enables us to sensibly zoom into various parts of this network. Since we wish to fully understand the self-organization core, the layers around it must be peeled like an onion. Once we zoom in, we will discover that other factors play a pivotal role in the detailed mechanistic understanding of this network. Evidence for this is already shown in the fact that we need to distinguish a timing module. Generally, precise temporal information is not included in all pathways. Secondly, the spatial confinement and distribution of components is not always represented. Sometimes it can be assumed for metabolic pathways that the protein pool is essentially well-mixed (as employed in e.g., this model in [70]), but this will not suffice for the polarity pathways.

If we could add precise spatio-temporal information to the overlapping hierarchy representation, we could discriminate polarity mechanisms, putting more order and hierarchy into the strength of each interaction. This allows predictions on many mutants, redundancies or near-neutral mutations, which may yield possible avenues for evolution. Yet, the current state of literature has not yet reached this stage for the reaction-diffusion module, which is at the core of self-organization. This dissertation aims to make this jump. The traditional, physical interaction representation helps us to understand what is needed to experimentally explore it beyond its limits, which provide the lessons in section 1.5. Then a conceptual leap has to be made, which is why the section thereafter advocates the use of mechanistically labelling subcomponents, dubbed functional subunits [51]. Still, we must first build upon the more traditional view on protein networks, which also serves as a good overview on polarity literature.

1.4.1 A Venn-network diagram: Non-modular and hierarchical

A convenient way of ordering the myriad of physical interactions is by placing these in appropriate regions of a Venn diagram. This depiction allows overlap between categories, which accompanies well the non-modular nature *within* the yeast polarity module, as will be shown. Five categories are constructed: timing (purple), mating (orange), bud scar (blue), reaction-diffusion (green) and actin (red), in that hierarchical order (see Figure 6). These correspond to five pathways, where the timing pathway acts as a control for the other four.

While the diagram is an extensive description, completeness competes with usefulness, as the purpose is to provide a clear overview. Therefore, some simplifying choices have been made. In polarity of yeast, but also of many other organisms, Cdc42 is the polarity hub [71], and should therefore have the most prominent role in the diagram. Since we wish to study self-organization, the reaction-diffusion pathway that allows polarization even in absence of external cues is essential to describe, and as a separate pathway actin will likely be a factor in polarity as well. Yet, obscuring effects by dominant external cues governing other pathways will have to be included for experimental reasons, while the timing pathway may be needed to provide an interpretation for experimental results in [24]. The bounds of the diagram will then not include expendable or infrequently encountered interactions within these polarity pathways, making polarity to a reasonable extent separable from other modules [31].

In the following subsection, each pathway will be highlighted. For three protein (classes), there is additional attention (boxes). These will be more prominently featured in experiments throughout this dissertation and require more detailed knowledge for the conclusions.

1.4.2 Reaction-diffusion (green circle Venn diagram)

The core of self-organization in yeast polarity is a pathway governed by reactions, causing e.g., (un)binding or activation of components, and diffusion of proteins, hence the name reaction-diffusion pathway. As mentioned earlier, the hub is Cdc42, a small GTPase which consequently has two states [72]; one GTP-bound (active state) and one GDP-bound (inactive state), see also Figure 7. The former is also known as its active state, with the latter being inactive. The activity refers to the capacity of a particular conformation to kickstart downstream signalling protein effectors, which relay the signal of polarity establishment towards further steps in cell division.

Furthermore, two other positional states may be identified. In itself, Cdc42 has a, post-translationally appended, hydrophobic tail¹, which makes it membrane prone. Symmetry breaking during polarization is considered established when a high concentration of Cdc42 has localized to one point on the plasma membrane. This does not require all Cdc42 to be

¹ Interestingly, it is commonly thought that the CAAX protein motif in Cdc42 causes geranyl-geranyl attachment needed for GDI interaction. However, Cdc42 in some strains (such as W303 in this dissertation) has a CTIL end, T being polar and hydrophilic instead of an aliphatic residue. This is of small importance for the post-translational modification, mostly affecting the balance between Cdc42's with a geranyl-geranyl or farnesyl tail [73].

concentrated in one spot; remaining Cdc42 can be found on other places on the membrane, but Cdc42 can be made cytoplasmic (usually in the GDP-state) when the tail is covered by Rdi1 [74], a chaperone part of the general class of GDI dissociation inhibitors (GDI). Its mechanistic functionality may resemble that of the GDI for mammalian Cdc42, since it shares large resemblance with yeast Cdc42 as it is well conserved [75]. In mammals, the GDI can only discriminate the active from the inactive form when on specific membranes, and does not facilitate Cdc42 dissociation, only blocks re-attachment by covering the hydrophobic tail [76]. Furthermore, it protects Cdc42 from misfolding and degradation [77].

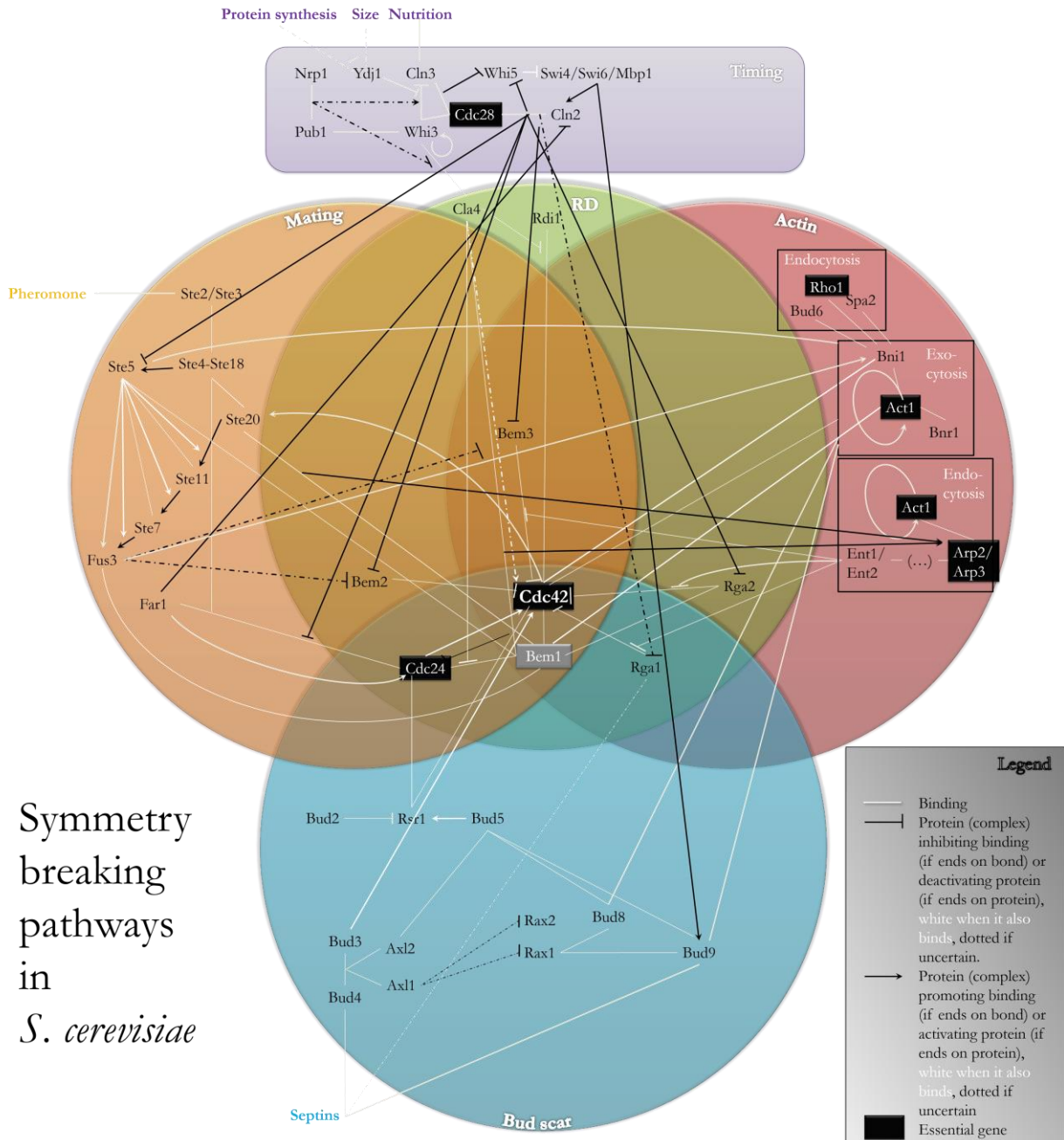


Figure 6 Venn diagram depicting four polarity pathways in budding yeast (mating in orange, bud scar in blue, reaction-diffusion in green and actin in red, in that hierarchical order), with the corresponding timing pathway (purple), based on physical interactions found in literature.

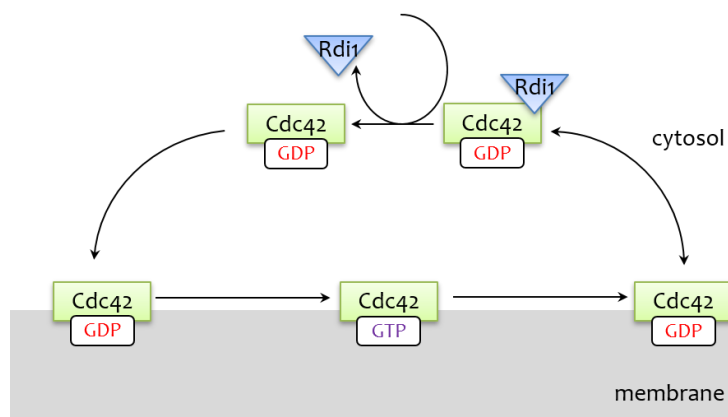


Figure 7 Simplified overview of the Cdc42 GTPase cycle. Cdc42 with an attached GTP on the membrane is considered active, otherwise it is inactive. Chaperone Rdi1 helps to maintain Cdc42 in the cytosol.

Conversely, a protein complex is normally responsible for recruitment of Cdc42 back to the membrane, with Bem1 as the glue [78]–[80] for Cdc42, Cdc24 and Cla4. Cla4 is a kinase of the PAK-family, capable of phosphorylating Rdi1 to disrupt its binding with Cdc42 [81], freeing the GTPase. Subsequently, Cdc24 can bind and effectively activate Cdc42 [80], [82], by promoting the dissociation of its GDP, which will usually get replaced by a GTP due to the typical GTP/GDP ratio in cells [83]. This is why Cdc24 is typically referred to as a guanine exchange factor (GEF), whose activity might be slightly upregulated by Bem1 [84]. Bem1 can then anchor to the membrane, guided by active Cdc42 (possibly with help of Cdc24 and Cla4 [80]), completing a positive feedback loop for activation and recruitment of Cdc42 to one spot in the membrane, as accurately modelled in [68]. Simultaneously, Cla4 (without Bem1) may act as a switch to avoid excessively long duration of the polarized state during the cell cycle, phosphorylating the GEF and priming the dismantling of the Bem1 complex [79], [85], [86].

In principle, the start of the recruitment and activation cycle is an amplification of a small noise perturbation, which suffices for swift polarization [68]. Alternatively, the initial condition may be set by asymmetric translation of Cdc42 mRNA. Before budding, Cdc42 mRNA has been shown to associate to the cortical endoplasmic reticulum [87]. Together with the bursty nature of Cdc42 production (see later on this dissertation), this may lead to significant, temporary asymmetry in Cdc42 content.

Recruitment and activation must also work in tandem with removal of unwanted active Cdc42 outside the incipient bud site. Before returning to the cytoplasm, Cdc42 must first be deactivated (GDP-bound), possibly because this increases the specificity of the GDI [76], blocking a quick return to the membrane. One class containing 4 proteins, Bem2 [88] (although most convincingly *in vitro*), Bem3 [89], Rga1 and Rga2 [90] (the latter in complex with Bem1 [91]), is responsible for deactivation: the GTPase activating (in the sense that GTP to GDP action is promoted) proteins, in short GAPs. How these mechanistically perform their action is not yet clear. In wild-type, it has been shown that these details are not of importance for polarity success [68], and therefore difficult to measure. It is at least known that Bem2 and Bem3 are cytoplasmic at first and localize at the bud site after Start [92]. Rga1 is also localized at the membrane, forming an exclusion zone around the previous bud site [93].

Box 2 Theorized influence of Cla4

In absence of Bem1, only less potent positive feedbacks than from the Bem1-Cdc24 complex remain to recruit Cdc42, which are needed to allow polarization (see e.g., [24]). Otherwise, the GTPase detachment from the membrane through Cdc42-GTP hydrolysis by the GAPs will dominate everywhere. The important role of a *rescue mechanism* as a form of alternative membrane recruitment (possibly containing Cla4) is explained in [51] and experimentally verified in Chapter 2.

Such a route requires three steps; disruption of the Rdi1 binding to cytosolic Cdc42, replacement of the GDP by GTP attached to the Cdc42 (turning Cdc42 into its active state) and preferential attachment of the then active Cdc42 to the membrane where other Cdc42 already is present. Supposedly, Cla4 plays a crucial role in these aspects. Firstly, in [81] evidence was presented that Cla4 could disrupt the binding of Cdc42 with the GDI Rdi1, but this requires the kinase activity of Cla4. Given the exposed membrane anchor (the geranylgeranyl Cdc42-tail), membrane attachment of Cdc42 without the GDI seems likely. The second step may be automatic, as the role of the GDP dissociation inhibitors (GDI) is keeping the GDP in place [94], which otherwise would get replaced by a GTP given the typical abundance of GTP over GDP in the cytoplasm [83]. Finally, it was shown in [95] that for Cla4 to properly fulfil its kinase ability, it requires binding with Cdc42 and the membrane (through its PH domain). In conclusion, Cla4 will only phosphorylate the Rdi1 and disrupt its binding with Cdc42 when already attached to other Cdc42 on the membrane, thereby leading to a positive feedback loop for membrane recruitment of active Cdc42. This process is depicted in Figure 8:

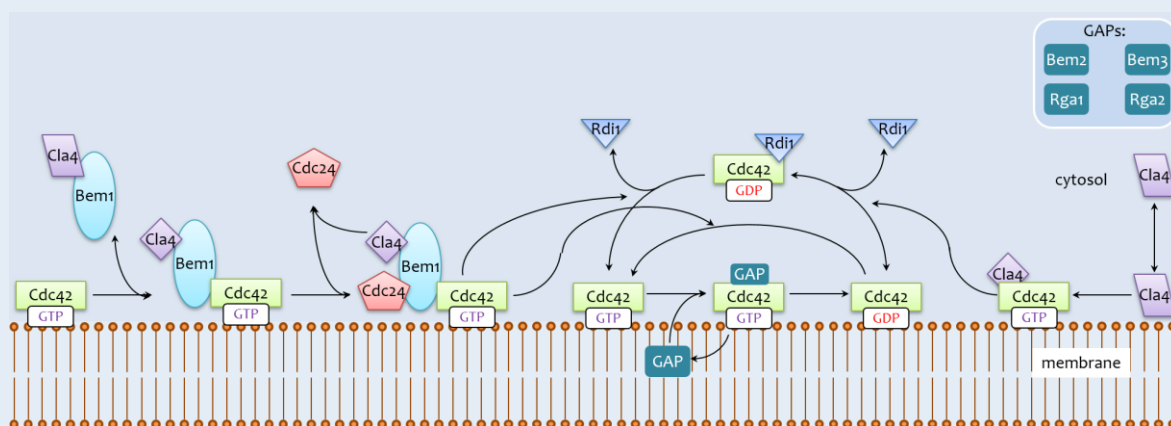


Figure 8 Bem1 and Cla4 (theorized)-mediated positive feedbacks, combined with GAP-mediated non-polar deactivation [68], [51], proteins not to scale.

In summary, simply put the Bem1-GEF-PAK complex activates and recruits, the GAPs deactivate and the GDI recycles Cdc42, forming its life cycle (see also Figure 8). In absence of scaffold Bem1, Cla4 presumably takes on another function, see Box 2. This in turn may make the role of actin more critical and visible, as explained in the next subsection.

1.4.3 Actin (red circle)

Actin transport of Cdc42 or downstream effectors may also play a role in positive feedback (e.g., [48]) leading to polarization. However, this positive contribution is under debate, other authors claiming actin may actually negatively dilute the Cdc42 accumulation [96]. Precise spatial localization of endo- and exocytosis may promote the positive contribution again [97]. Still, a shift is visible towards the belief that Cdc42 is not the crucial component in the vesicle transport [67]. Instead, another cargo may be of importance, which could contain Bem1 [98] but cannot be limited to it. Even when all known (Bem1-mediated) and hypothesized positive feedbacks of the reaction-diffusion network have been disabled (see section 3.3.3) viable genotypes still exist. Clues surface when digging deeper into the actin network.

We can distinguish two types of actin transport, both relying on fast actin (de-)polymerization [99], [100]; the formin-dependent endo-/exocytosis and the clathrin dependent endocytosis. The former relies on Bni1 (binding to active Cdc42 [101]) and to a lesser extent Bnr1 to nucleate sites of actin bundles leading to their assembly [102], mediating exocytosis. Conversely, the endocytosis is determined by [103] Rho1, which signals Bni1, Spa2 (convenient target for a clean polarity marker [24], [104]) and Bud6. Initially, Cdc42 itself was connected to vesicle delivery [105], but findings on its GFP fusion casted doubts on this [96].

Instead, there is strong evidence from literature for transport of another class of important polarity proteins, namely the GAPs. These travel using the epsin-coatings in the clathrin dependent transport route [106] (see the mechanistic explanation in Figure 9 and Box 3). The full pathway for the formation and coating of actin cables is very extensive. For conciseness, Figure 6 omits the irrelevant details for the understanding of polarity. The interested reader is referred to [107] for an elaborate review.

1.4.4 Timing (purple region)

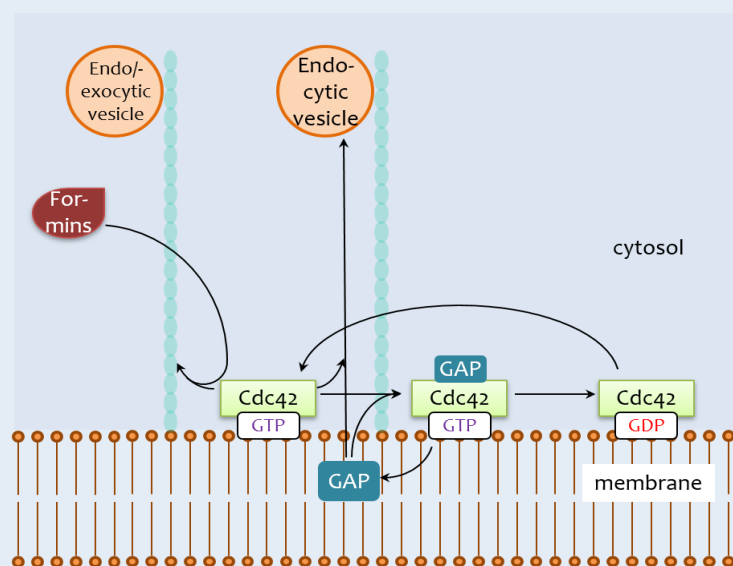
In the purple rectangle denoting timing in Figure 6, the cell finds itself initially in early G1 phase. The protein Whi3, which binds cyclin (cell cycle regulator) Cln3 mRNA [108], was initially thought to act as a cytoplasmic retention pool for Cln3 [109], which is translated but degraded very quickly [110]. More recently, Whi3 was shown to sufficiently destabilize and decrease translation efficiency of Cln3 mRNA [111], thereby limiting nuclear Cln3 levels. Instead, Cln3 remains at the endoplasmic reticulum, until levels of its chaperone Ydj1 reach high enough to trigger its release, which may correspond to the moment of sufficient cell size and relatively limited (compared to G2) protein synthesis [112]. Nutritional cues also affect Cln3 e.g., nitrogen starvation heavily represses Cln3 mRNA [113].

Box 3 Theorized polarity contribution of actin

It has been shown that a clathrin-dependent form of endocytosis, aided by epsins (coating vesicles, see e.g, [114]), targets GAPs (except Bem2) and colocalize to the polarization site [106], [115]. Possibly, these epsins facilitate polarity maintenance without displacing the Cdc42. Firstly, since Cdc42 does not seem heavily trafficked, mainly GAPs unbound to Cdc42 must be displaced. Secondly, Cdc42's downstream PAKs Ste20 and Cla4 may phosphorylate motor proteins which help to activate the Arp2/3 protein complex, nucleating actin assembly [116]. Therefore, any unbound GAPs may also be specifically removed from the bud site, allowing Cdc42 to remain active for a longer time at the polarization site.

This adds to the local depletion of the GAPs at the bud site; in the theoretical model validated in Chapter 2, free GAPs (available for attachment to Cdc42) were already locally depleted by forming a complex with active Cdc42 until there were insufficient left to stop accumulation of Cdc42 at that point. Doubling the local depletion replaces the need for the positive feedback mediated by Cla4 (in mechanistic terms shifting the polarity model from scenario E to F in [117]). Interestingly, deleting the polar activation part of the polarity mechanism through Bem1 deletion and adversely affecting the endocytosis transport through deletion of Gic1/2 is lethal [118], in line with the required set of functional subunits in [51]. Therefore, GAP depletion on the presumptive bud site through endocytosis by clathrin-binding epsins may play an important role in polarity once other mechanisms are disabled, such as in experiments in Chapter 3. A depiction of the GAP depletion from the polar zone is found in Figure 9.

Figure 9 Formin-mediated endo-/exocytosis and the hypothesized actin contribution to polarization; active Cdc42 recruits actin bundles [101], [102], along which epsin-coated vesicles can recycle the GAPs Bem3, Rga1 and Rga2 endocytically [106].



Once Cln3, bound to the essential master regulator and cyclin-dependent kinase Cdc28 through Whi3 [109], reaches the nucleus, it upregulates cyclin Cln2, which then upregulates itself through a feed forward loop to enter Start [119]. This loop consist in more detail of Cln3 binding to the kinase Cdc28, which phosphorylates Whi5, stopping Whi5 from binding and inhibiting the Swi4, Swi6 and Mbp1 protein complexes [120]. The Swi4/Swi6 complex promotes Bud9 landmark expression [121] and more importantly, Cln2 expression [122], which binds to Cdc28 [123] and as such further exports Whi5 from the nucleus [120], [124].

Then, Cdc28 bound to Cln2 phosphorylates many targets, that all work together to promote polarization. On the one hand, GAPs Bem2, Bem3 [92], Rga2 [91], [125] and possibly Rga1 [126] are phosphorylated, decreasing their activity, shifting the balance towards a larger pool of active Cdc42. Additionally, the GEF Cdc24, which is sequestered in the nucleus by Far1, is released by phosphorylation of Far1 by the Cdc28-Cln2 complex inducing degradation of Far1 [127], leading to more active Cdc42. This also removes the presumed effective inhibition of Far1 of Cdc28-Cln2 functionality by freeing it for other duties, including phosphorylating Ste5 to prime it for degradation to stop the mating pathway in absence of pheromones [124], [128], [129]. In line with the experimentally finding that expressing Cln2 under an inducible promoter and removing Cln1 and Cln3 allow control over Start [85], this means Cln2 is the cyclin that marks Start. A short mechanistic summary is depicted in Figure 10.

Despite that this timing picture has been established in relatively high detail, there is reason to assume that an important player in polarity, Nrp1, is still missing. Box 4 bundles the evidence, which while yielding an incomplete picture, provides a sufficient handle on its function for model implementation in Chapter 3.

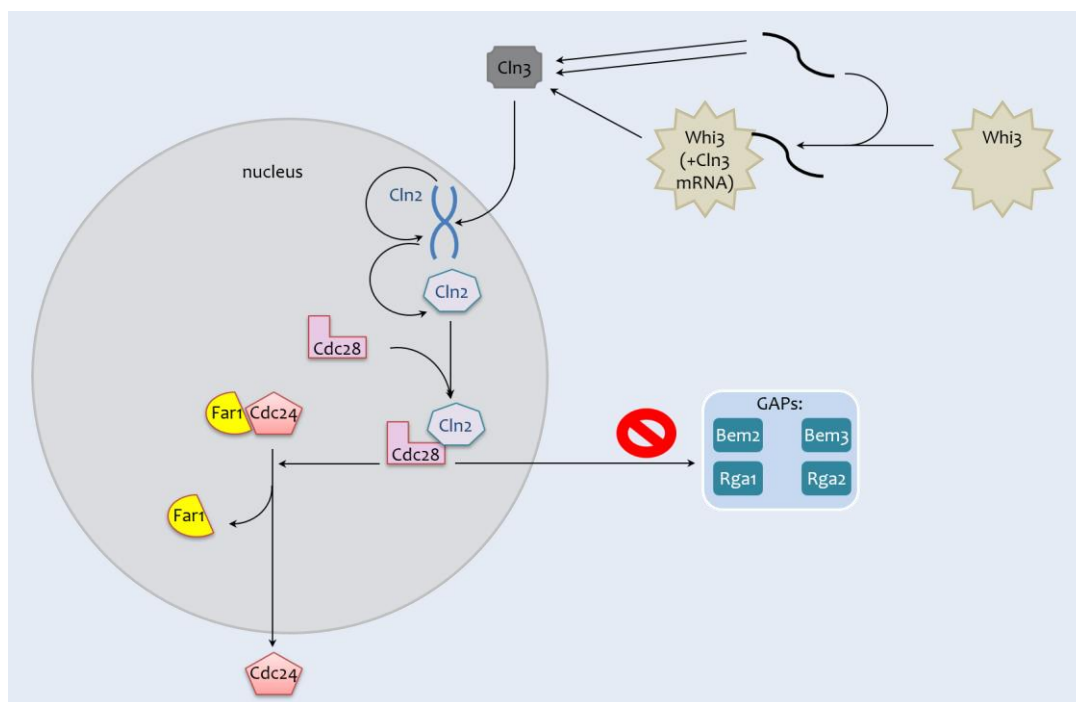


Figure 10 Brief mechanistic overview of the timing pathway, leading towards Start, by releasing Cdc24 from the nucleus and phosphorylating and inhibiting the GAPs.

Box 4 Theorized polarity contribution of Nrp1

In [24], it was found that Nrp1 also influenced polarization, often the second mutation to recover the ill $\Delta b e m 1$ after *BEM3* deletion. This is somewhat surprising, given that Nrp1 had not been implicated in polarization before, in fact it was originally not known to be of much importance. Phenotypically, it seems to shorten G1 time in the $\Delta b e m 1 \Delta b e m 3$ background, prompting the suggestion that it is involved in an earlier or more abundant release of Cdc24 from the nucleus [24].

The classification of Nrp1 as part of the timing pathway allows some discussion on the possible mechanism with which it acts. It has a known RNA-binding motif, a prion-like domain [130], [131]. Furthermore, it may have two more domain that can bind RNAs as well [132] in the form of RanBP2-type Zinc finger motifs according to www.yeastgenome.org [57] (consulted on 6/6/2019), yet that database does not relate found RNA targets in [133] to polarity, as mentioned in [24]. Thus, the function of Nrp1 is unlikely to be through a direct mRNA target.

Nrp1 is also known to form part of stress granules, a form of RNA-protein granules composed of aggregates of interaction prion-like proteins that regulate mRNAs [131], using either its prion-like domain or RNA-binding domain. This is similar to Whi3, which also localizes in stress granules, using either of those two domains [111]. In the two-step model of [131], stress granules form by proteins first binding RNAs forming complexes (e.g., Nrp1 binding mRNAs, Whi3 binding others), and these form aggregates through binding through prion-forming domains, which Nrp1 and Whi3 both have. This could establish a link between Nrp1 and the timing pathway through Whi3. Still, the exact function that Nrp1 exerts on Whi3 is unclear.

Interestingly, in an in polarity terms evolutionary related yeast [65] *Ashbya gossypii*, RNA-protein assemblies were found to be mediated by Whi3 aggregates, localizing polarization mRNAs like Bni1 and Spa2 mediated by a pumilio-family protein (Puf2) [134]. Because Whi3 has more functions, such as binding of cyclins analogous to budding yeast, the aggregation needed for polarity mRNA localization is controlled, and presumably governed by heat stress protein chaperones such as Ydj1. Possibly, if this mechanism is conserved in budding yeast as well, Nrp1 may play a role in these aggregates.

As aforementioned, Nrp1 has no polarity mRNA targets, making it an unlikely substitute for the Pumilio protein Puf2. But, the role as chaperone controlling aggregation fits

Nrp1 well. Although in *S. cerevisiae*, the role of Puf2 seems different, binding Whi3 and NRP1 mRNA rather than the proteins [135], Nrp1 could physically associate to Whi3 through stress granule protein Pub1, which binds to both [136], [137]. The role of the Pumilio protein may then be taken over by Puf4, which binds to it [137].

As a chaperone, Nrp1 could then influence which mRNA Whi3 targets, which in turn may influence the mRNA stability, half-life and ultimately protein abundance of many targets [111]. A mechanistic depiction of the role of Nrp1 is given in Figure 11. In budding yeast, notable mRNA targets of Whi3 are that of Cln3 [108], regulating G1 time which seems shorter in $\Delta bem1 \Delta bem3$ upon the deletion of *nrp1* [24], and Cla4 [138], as aforementioned important in a $\Delta bem1 \Delta bem3$ as well to contribute to the needed positive polarity feedback. This leads to the hypothesis that Nrp1 contributes to Whi3 binding and destabilizing Cln3 mRNA, allowing G1 to pass at normal speed.

As a consequence, Cln3 is faster and more heavily expressed without Nrp1, causing early G1 exit. At high Nrp1 levels, one may expect slow G1 exit due to Cln3 being lowly expressed, and it was found the Cla4 deletion becomes lethal in that circumstance [139]. This leaves the opportunity (although it is not necessary) that Cla4 is upregulated when Nrp1 is overexpressed. Possibly, Nrp1 causes less Cla4 mRNA to bind Whi3, stabilizing it. The excessive Cla4 generated may then facilitate an alternative polarization cue through its positive feedback mechanism and force G1 exit in absence of the Bem1-GEF complex mechanism, since the GEF stays sequestered due to low abundance of Cln3. This is how Cla4 might counteracting the deleterious effect of Cln3 suppression, and overexpression of Cla4 may contribute to this.

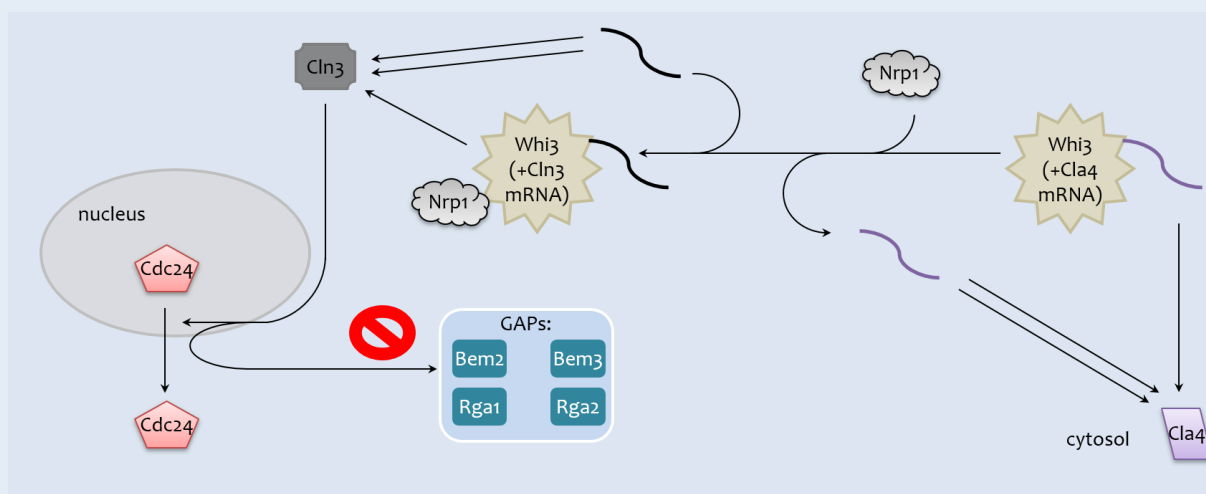


Figure 11 Hypothesized Nrp1 influence on polarization. Nrp1 works as a chaperone for Whi3, determining which mRNA becomes predominantly bound, reducing the mRNA stability and ultimately translation.

1.4.5 Mating (orange circle)

The mating pheromones α and a bind to the membrane proteins Ste2 and Ste3 respectively [140], [141]. Once bound, the Gpa1 unbinds from Ste4 and Ste18 with which it had formed a heterotrimeric complex [142]. This causes the Ste4-Ste18 complex to dissociate from the membrane [143] and relay the mating signal (although conformational change of the heterotrimeric complex could also work [144]).

Ste4 can then bind the kinase Ste20 [145] and Ste5 [146], a scaffold that binds Ste7, Ste11 and Fus3 [147]. The Ste4-Ste5 bond is essential to cause phosphorylation and activation of Ste11 by kinase Ste20 allowing propagation of the mating signal [146]. Active Cdc42 contributes by binding Ste20 to release it from its auto-inhibited form [148], but this action of Cdc42 may not be as critical [149] as its localization.

The Ste4-Ste5 bond is mediated through recruited Cdc24 [150], which is bound to Ste4 by Far1 [151]. Here, Cdc24 is exported out of the nucleus in complex with Far1 by exportin Msn5, in contrast to the situation without pheromone signalling where Far1 was degraded [127]. Evidence exists that after nuclear export, actin is important to localize the Cdc24-Far1 complex. Active Cdc42 recruits formin Bni1 [101], which allows nucleation of actin cables necessary for transport of Cdc24-Ste5 complexes to the site of free Ste4-Ste18 complexes [152]. Possibly, this actin-Cdc24 link is mediated by Bem1, which is known to co-immunoprecipitate with Act1 [98] and bind Cdc24 [78].

The mating pathway is hence extensively cross-linked with many redundant paths towards local recruitment. Bem1 can bring more Ste20 and Ste5 to the same point [98], as well as Far1 [153], which solidifies this location as the point of recruitment. A similar reinforcing effect resulting from Far1 may even activate Cdc24 upon binding to Ste4 [154]. Moreover, once this recruitment cycle starts on the membrane, it may proceed analogously to the reaction-diffusion pathway, recruiting more Bem1, Cdc24 and Cdc42 to one point in the membrane, concurrently attaching scaffold Ste5 to Ste4 complexes. For more extensive reviews of the interplay of mating and actin recruiters with accompanying literature, the reader is referred to [155], [156].

Subsequently, Ste5 can facilitate more reactions; active Ste11 can phosphorylate Ste7, and Ste7 can in turn activate and phosphorylate Fus3, with both processes promoted by their binding to Ste5 [146], [157]. Fus3 then dissociates from Ste5 [158], and can phosphorylate Bni1, necessary for its localization [159]. In the latter paper, it was conjectured that Fus3 also binds Gpa1 (formerly in complex with Ste4 and Ste18), possibly providing an alternative actin targeting route towards the pheromone signal. Finally, Fus3 may also inhibit the GAPs Bem2 and Bem3 [92], to complete the picture; both the GEF is activated and recruited, and GAPs become inactivated.

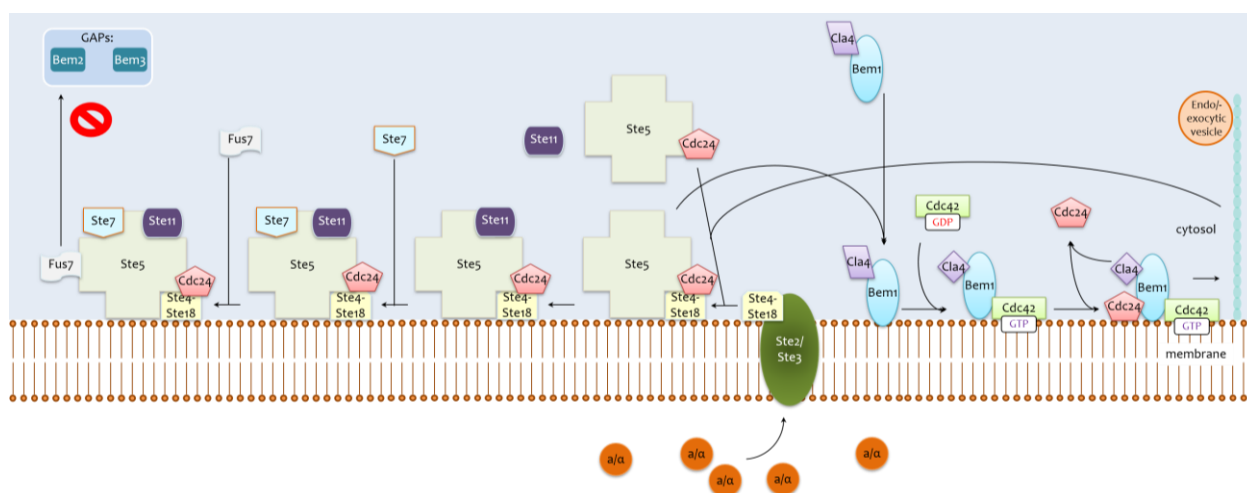


Figure 12 Overview of the mating pathway core, kick-starting the Bem1-mediated positive feedback and GAP inhibition.

1.4.6 Bud scar (blue circle)

Normally in absence of a mating cue, the bud scar cue is prevalent. The bud scar consists of septins left behind after cytokinesis in the previous cell cycle [160], [161], providing a spatial asymmetry that can be amplified. Two possible patterns exist [162], namely axial, where division takes place next to the previous division site, and bipolar, where the location opposite of the previous site is taken. The former is typically encountered in haploids, while the latter takes place in diploids, and this is regulated by exclusive expression of Axl1 in haploid cells [163].

In haploid axial budding, septins interact with Bud4, which importantly interacts with Bud3 [164], [165], a second, less-known GEF for Cdc42 earlier in G1 [166]. Bud4 also binds Axl2 and Axl1, causing Bud5, the GEF for Rsr1 [167], [168] to bind this complex [164] in haploids alone. This leads to activation and localization of Rsr1 by promoting its dimerization [169], [170] (in concordance with the GAP for Rsr1, Bud2 [167], [169]). Subsequently, recruitment results of Cdc24 [171] (GEF of Cdc42) and Cdc42 [170] itself. The reaction-diffusion pathway can then amplify this initial condition to generate a new bud site close to the previous bud site (but due to Rga1 not exactly at the site [172], possibly connecting to septins [93]).

Alternatively, in diploid cells that do not express the crucial Axl1, landmark proteins Rax1 [173] and Rax2 [174] are dominant. This prompts the hypothesis that Axl1 is also involved in inhibiting the functionality of Rax1/2 [175]. Rax1 and Rax2 are located in the bud neck ring and bud tips as landmarks [176] and cause localization of Bud8 at the opposite end of the previous division, and of Bud9 at the bud neck. Transport of Bud8 and Bud9 depends on actin (and of Bud9 also on septins) [121], linking the bud scar and actin pathway together. In turn, Bud8 and Bud9 bind Bud5 [177], localizing Rsr1 activity to generate the budding pattern at two possible poles.

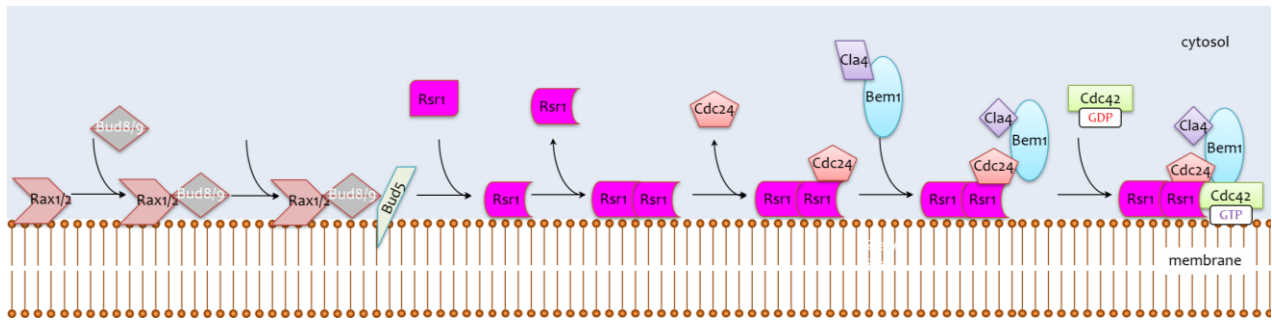


Figure 13 Core of the diploid bud scar pathway, leading towards the start of the Bem1-mediated positive feedback. In haploids, Rax1/2 are replaced by septins and Bud8/9 is replaced by the Bud4/Axl1/Axl2 complex to recruit Bud5.

1.4.7 What are the advantages of such a network?

Following-up on the hierarchical and non-modular nature of this network, one may ask why this network has been formed in this way. Naively, one pathway would be sufficient, and in a simplified form compared to those currently in place. Instead, large redundancies exist. There are relatively few essential genes in this network, 7 out of 61, where normally around 20% is essential [59]. Those that are essential are e.g., master switches like Cdc28, the GEF Cdc24 and Cdc42, and actin is essential only outside polarization (it is needed for other functions).

The existence of some inevitable degree of overlap in pathways makes sense. If the problem is to break symmetry, the answers differ only in the direction in which it is broken, so following a mating cue, bud scar cue or random cue. This means that the mechanism to activate Cdc42 can be shared, but the localization is pathway specific. Still, there is more than the minimal amount of overlap needed. One possible use of this can be exemplified by the GAPs.

Ample redundancies keep many genes free to evolve, as explained for the GAPs in [51]. In budding yeast, there are four GAPs, Bem2, Bem3 and Rga1 and Rga2. From a reaction-diffusion point of view, one GAP suffices, and the $\Delta bem3 \Delta rga1 \Delta rga2$ triple knockout is indeed viable (yet ill) under standard conditions [90]. Possibly, specializing GAPs to facilitate other functions in the cell may carry certain benefits for robustness. There is data showing that GAPs differ in their ability to rescue mutants under non-standard conditions, such as high temperatures [90].

The distribution of the GAPs across the Venn diagram also makes for a tempting suggestion. Seemingly dispersed across pathways, these GAPs might serve as a switch to maintain the correct hierarchy between pathways under different conditions. Easy switching intuitively requires components to be shared between at least two pathways but preferably not all of them. Optimizing components in a single pathway might not influence the hierarchy (just itself), but optimizing the essential components shared by all pathways might shift the relative

importance of pathways. Possibly, this could require an initially high fitness cost such as with the $\Delta bem1$.

If this notion is true, then the network is very flexible in this configuration to adapt to foreseeable situations. For example, evolution in absence of a mating cue might have *BEM2* as a logical target, or if landmarks are often not present (for example after spore formation) then the balance can shift more towards the reaction-diffusion pathway. This is not a hypothesis that will be addressed further, but may incite interesting experiments. When artificially imposing these conditions that favour dominance of a pathway, experimental evolution may reveal interpretable genetic consequences. If these involve mutations targeting the crucial components at the edge between pathways, there would be support for this idea. A positive indication is the multitude of ways to artificially set the dominant pathway, which is further discussed in the next section.

1.5 Experimentally promoting modularity

Inconveniently, the previous analysis of the physical interactions shows a certain degree of non-modularity and hierarchy within yeast polarity. Typical experiments entail recording the response to an imposed perturbation, for example determining the speed of polarization when a gene is knocked out. The hope is that it becomes possible to reverse-engineer the system of interest, yet this depends on the interpretability of the experimental results. Non-modularity complicates this, as it becomes more difficult to attribute results to roles of proteins within the system. Similarly, the hierarchical structure may cause insensitivity to particular perturbations.

For this purpose, care must be taken to maximize modularity and sensitivity to the perturbation where possible. This is not straightforward; the effects of the required preparatory modifications must be clear, with minimal off-target effects. Fortunately, depending on the parts of yeast polarity upon which one wishes to focus, several solutions in literature exist to conveniently increase or decrease importance of the various pathways.

1.5.1 Solution I: Mating (orange)

Given the default hierarchy, it is easiest to study the mating pathway. Pheromone gradients are the dominant cues, obscuring somewhat the self-organization. The simplest solution to decrease importance of the mating pathway is to restrict experiments to one mating type at a time. More solutions exist, such as what yeast utilizes to continue polarization along other pathways, when supplied with a “deceptive” pheromone signal [54].

In that experiment, yeast cells were faced with a constant pheromone signal in absence of an actual mating partner. This initially kept cells waiting in G1, but not indefinitely (up to 3 hours). Authors discovered how yeast downgrades the mating pathway. They show how prion-forming domains in Whi3 control Cln3 expression and hence the bypass from the mating cue. It is possible to delete these domains from Whi3 and these would be an unexpected upstream way (located in the timing pathway) of cutting off the mating pathway.

1.5.2 Solution II: Bud scar (blue)

Promoting the bud scar pathway seems straightforward, especially when restricting to experiments of the same mating type. However, the reaction diffusion network on its own is quite fast (also minute time scale) as it relies on the same core components (Bem1-GEF-Cdc42). Thus, the hierarchy is only slightly in favour of the bud scar pathway. This means care must be taken when handling bud scar mutants while studying the pathway they are in.

Literature does not provide easy ways to further emphasize the bud scar pathway. As the reaction-diffusion mechanism starting from an intrinsic noise cue requires amplifying only a small fluctuation/cue, it will be hard to suppress this. It requires very careful tuning of the reaction-diffusion network to make it only sensitive to an imposed cue and insensitive to noise fluctuations. Observing the physical interactions network and the distribution of the GAPs across pathways, I expect this to require modification of the GAPs. Yet, it requires theoretical modelling and more experimental data to make an accurate prediction.

1.5.3 Solution III: Reaction-diffusion (green)

To study the reaction-diffusion pathway independent of the bud scar cue (and provided the mating cue has been taken care of), people have tried various approaches. Most commonplace has been the *RSR1* deletion, forming a major link to Cdc24 (shared by the reaction-diffusion pathway), which can be supplemented with the *BUD8* for complete landmark cue removal [178]. However, it has been shown in [84] that a more elegant solution exists, which does not have an influence on the reaction-diffusion network especially when the actin pathway is suppressed by a drug. The more upstream approach that those authors take, namely to remove two landmark proteins, seems without fitness effect or epistatic interactions on the reaction-diffusion proteins. This was the conclusion based on observations of growth on a plate and polarization success with and without actin.

Remarkably, an entirely different approach has landed success as well. Rather than fine tuning the network by precise modifications, it is possible to turn the hierarchy upside down by a single, heavy perturbation. The logic is that in this network, besides the essential Cdc42, Bem1 is the other protein shared amongst all pathways, so one mutation there will affect the balance between these. As expected, the *BEM1* deletion is very deleterious but fortunately, still viable. In [24], polarization in this background seems quite random (not relying on spatial

cues), while the evolution experiment with this Bem1 mutant shows strong reliance on the reaction-diffusion pathway. It harboured two out of three recurring recovery mutations, with the other one presumably in the timing pathway. Further analysis of the evolved genotypes shows symmetry breaking behaviour that allows for theoretical modelling with minor considerations on actin [51]. Therefore, this experimental perturbation approach will be taken when studying the self-organization of the yeast polarity reaction-diffusion pathway in this dissertation.

1.5.4 Solution IV: Actin (red)

In practice abolishing the mating pathway and depreciating the bud scar pathway suffices to allow good visualization of the reaction-diffusion pathway, but a concern can still be the actin pathway. While there are multiple possibilities to disturb the influence of actin, there are problems associated with this course of action. The most obvious one is the general essentiality of actin for progression through the cell cycle, so bluntly knocking out genes is risky.

Therefore, typical ways to perturb the actin transport network rely on first keeping the cells in G1 through cyclin [85] or pheromone arrest. Only then latrunculin (an actin-binding agent that causes depolymerization) is added or a mutant thermosensitive formin (actin nucleation point) [179] is activated. The side effects and stress response when using the latrunculin are detectable [180], all the way upstream to Whi5 delaying polarization in the range of 15 minutes [179]. Furthermore, care must also be taken when using different versions of latrunculin, as lat-B may only abolish the cables but not the patches which govern endocytosis [181]. Since the clathrin-dependent endocytosis pathway has the possibility of influencing the spatial distribution of the GAPs [106], [115], there is in that case no definitive independence of actin from the reaction-diffusion pathway.

1.5.5 General design considerations

What one can learn from this, even in other model systems, is that it is typically non-trivial to choose the exact perturbations needed to isolate the pathway of choice, or to make interpretation of the data easier. Sometimes these perturbations occur upstream of pathways, sometimes inside the pathway and sometimes the best strategy is to make a big shock to the system and see what happens. The latter option becomes more tempting and viable the deeper down the hierarchy one goes, as more and more well-chosen mutations are required to make a subsystem modular. In any case, it must be stressed that all strategies require very good understanding of the underlying interactions, even with the last option of hit-and-hope, or the risk of misinterpreting experimental data becomes very high.

1.6 Conceptual representation

As the previous section has shown, we can shape our network of interest, yeast polarity as an example of self-organization, to quite some extent. Importantly, the modularity can be artificially promoted, placing different pathways in the limelight. This greatly facilitates interpretation of experiments, as the system appears to have layers that can be peeled off like an onion.

Still, the thus far final layer containing a small version of the reaction-diffusion pathway, with still some actin influences, is not readily understood. As mentioned earlier in this introduction, a conceptual step beyond physical interactions must be made to promote further knowledge, taking into account spatio-temporal interactions. This process is not obvious and heavily depends on the reduction of the polarity network to a small, near minimal state. From then onwards, we can hope to extent towards polarity as a whole, and generalize across modules.

In [51], this first step in conceptual progress is taken. There, the reaction-diffusion pathway has been dismantled in multiple mechanisms: the wild-type (WT, Bem1-mediated), rescue (in absence of Bem1), and an immobile Cdc42 variant. At the sub-mechanistic level, the origin of the prevalence and existence of these mechanisms lies in what is dubbed functional subunits. Proteins can be classified (non-exclusively) to participate in these functional subunits, which theoretically can generate any of these mechanisms. In this case, three subunits exist; Cdc42 transport, polar activation and non-polar deactivation. Which proteins are present in a particular genotype, their interaction ability (in case of mutants) and protein copy number ultimately determine which mechanism exists and dominates. A graphical summary of this hierarchical concept is depicted in Figure 14.

The framework of functional subunits may turn out to be powerful. Every subunit (combination) yields quite tractable rules for the overarching mechanism to work. For example, the combination Cdc42 transport and non-polar deactivation, where the rescue mechanism prevails, dictates strong dependence on Cdc42 copy number. Below a certain threshold, no polarization is possible, while immediately above it polarization can be quite effective. It is the outcome of these simple rules that suggest the formation of a network level between genotype and phenotype different from functional grouping, as frequently attempted. As further explained in Chapter 3, this leads to defining the *mesotype*, which will prove pivotal to show how genotypes generate otherwise surprising population phenotypes.

Whether this has generality beyond yeast polarity depends largely on analysis of other, self-organizing systems to a degree of detail and tractability with which yeast polarity has been studied in literature. As shown, mechanistically deciphering a system is only feasible when the depth of the physical interaction map is such that we know how to make the system experimentally tractable. When we have reconstructed the functional subunits for multiple

systems, the hope is that there are recurring themes, i.e. that evolution converges to the same solutions more often than not, yielding a limited set of rules to be incorporated in the mesotype. Whether this will be true remains to be seen, but currently I see this as the best attempt at a level-based understanding of the genotype-to-phenotype map.

Finally, there is one important, unavoidable contribution to the genotype-phenotype map left thus far unmentioned in this introduction, which is noise. As aforementioned, the framework of functional subunits depends not only on the proteins present at the level below, but also on the quantities. In biological systems, these are inherently noisy. So even when conceptually the system is fully determined by a hierarchical tree as in Figure 14, there is still some room for phenotypical variation across clones in the same environment.

A good study on the propagation of noise in copy number to the cell as a whole and its consequences can be found in [182]. There, variation in expression of one gene which influences growth, can influence the concentration of other proteins. Therefore, the growth can be intricately influenced by cross-correlated noisy gene expression. This can even have effects on evolution, as is shown in [183]. There, noisiness of expression can become a tweakable property for adaptation to environmental changes. In other words, the goal that was established for this dissertation, to elucidate the genotype-phenotype map structure and evolution as its reciprocal path, cannot be achieved without taking noise into account. Therefore, much of Chapter 3 and Chapter 4 is dedicated to provide some deeper understanding to the implications of noise in terms of phenotypical variation and evolution.

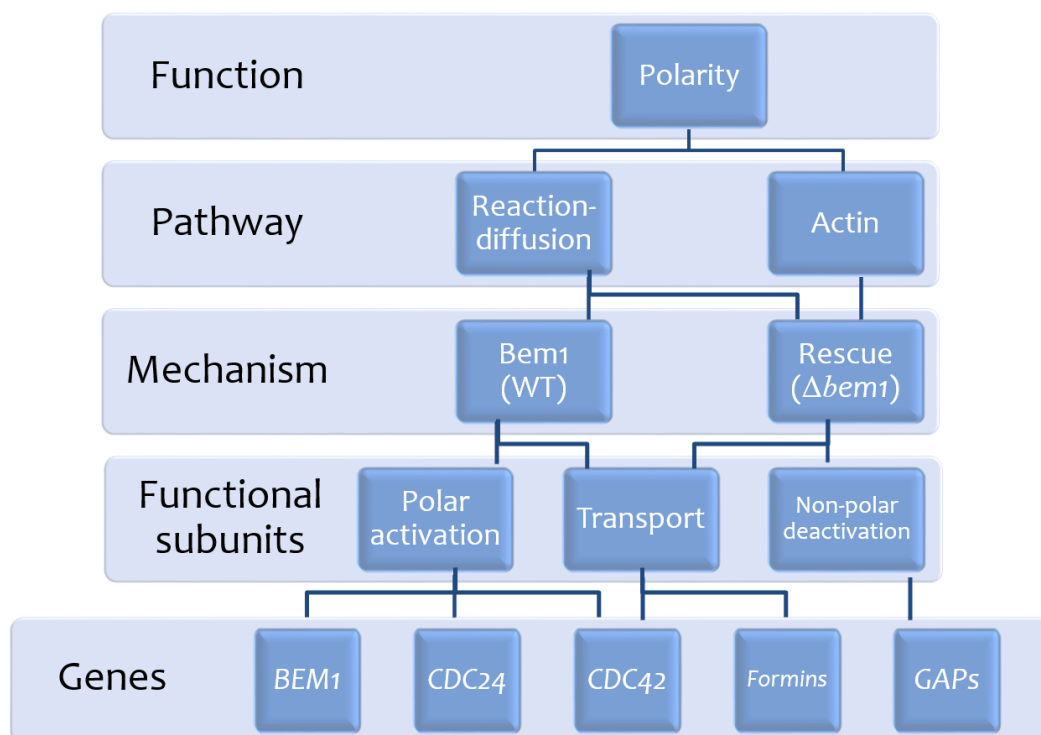


Figure 14 Gene to function hierarchy including the concept of functional subunits [51]. At the lowest level (genes), only the most important players are mentioned for clarity.

1.7 Research questions

Based on the rules in literature concerning evolution and self-organization, we can establish that there are some gaps in the knowledge regarding these fundamental properties of life. While it is not possible to address all of these, some key problems will receive the necessary attention in the next chapter of this dissertation. The following enumeration has at the end of each item in parentheses the number of the chapter where the question is addressed.

- What is an appropriate level definition to study evolution of self-organizing systems?
 - Can the theoretical framework of functional subunits be experimentally validated? (2)
 - Which theoretical predictions are best suited for experimental validation in budding yeast polarity? (2)
 - What experimental data on polarity is required? (2)
 - How do we construct the probes for the experimental validation? (2)
 - How can functional subunits be integrated into a bottom-up, biophysically sound level definition for predicting population phenotypes from genotypes?
 - Can a model based on this level definition describe the experimental genotype-phenotype map in polarity of budding yeast? (3)
 - How do we construct the experimental probe for high-resolution genotype-phenotype data? (2)
 - What are the evolutionary lessons learned from a bottom-up, biophysically sound level definition, in terms of e.g., the information requirement to predict phenotype from genotype or the environmental influence on evolution? (3)
- What are mechanistic rules for epistasis and indirectly evolutionary adaptability and robustness?
 - What mechanism makes (genetic) epistasis so pervasive? (4)
 - How important is protein copy number for fitness? (4)
 - How do selective forces shape protein copy numbers across the population to generate epistasis? (4)
 - When does selection on protein copy number cause epistasis between seemingly unrelated proteins and how strong is this effect? (4)
 - How do we design an experiment to illustrate this epistatic mechanism in e.g., budding yeast polarity? (4)
 - Does epistasis imply a tradeoff between adaptability and robustness? (4)
 - How does an epigenetic epistatic mechanism balance adaptability and robustness? (4)

Finally, a recapitulation of the results in the context of answering the main research questions will be given in Chapter 5.

2 Experimental validation of a molecular model for yeast polarity

2.1 How to reverse-engineer the correct system

In the introductory chapter 1, the case has been made to choose establishment of budding yeast polarity as a suitable model system for learning about self-organization and evolution. In this process, proteins localize to one point on the plasma membrane, breaking the internal, spherical symmetry of budding yeast. The overview of literature (1.4) revealed that even this specific subfunction in the life cycle of budding yeast uses a long chain of hierarchically ordered and interconnected reactions. While it is possible to make our system of interest simpler (1.5), the essential core still contains open questions.

Despite experimental ingenuities e.g., the ability to relocalize Cdc42 binding partners at will to the membrane or cytoplasm through a light sensitive tail [184], we are left at a point where a conceptual step is needed to advance. For example, why does the $\Delta bem1$ have such a stochastic phenotype, what is the relative strength of the different GAPs and how do they mechanistically act on Cdc42? Without mechanistically understanding the molecular model underlying the core of polarity, we cannot improve on methods to more generally describe the transition from genes to traits (phenotypes), or how evolution affects the opposite path.

Until now, the best experimentally validated model for budding yeast polarization has been that of [68], focusing on the role of the Bem1-mediated positive feedback (left side of Figure 8). There, a simple, constitutive view on the role of the GAPs suffices as long as Bem1 is present. However, when this feedback falls away, polarization cannot take place anymore without assuming other interactions, as can be envisioned through the interpretation in [117], defining 6 scenarios for polarization (labelled A to F). In that paper, the Bem1-feedback that allows both a net self-recruitment of Cdc42 and GEF recruitment, is needed for polarization (scenario A). As the absence of Bem1 triggers the emerging importance of the GAPs [24], the sensible addition would be an interaction between Cdc42 and GAPs (scenario F), yielding again a polarizing system. This is not very trivial, as this requires a feedback inhibition of GAP action by active Cdc42. However, if a simple self-recruitment of Cdc42 can be conjectured (without GEF recruitment), only a simple interaction between GAPs and Cdc42 is required (scenario E). In any case, polarization entails active Cdc42 somehow disabling GAP activity at the polarization site.

The simplest way to accomplish this inhibitory effect on the GAPs is by making the second step in the GAP cycle of attachment to Cdc42, nucleotide hydrolysis and dissociation, rate-limiting as done in the model of [51]. Intuitively, the result is as follows; active Cdc42 (with GTP) temporarily sequesters GAPs by binding to it, rendering the GAP unable to act on other Cdc42-GTP for a period of time. In a membrane location with relatively low (compared to the

amount of GAP available) active Cdc42 this does not matter; sufficient GAPs remain to deactivate and ultimately dislodge all Cdc42 present there. However, when Cdc42 has localized to such an extent that it outnumbers the GAPs, Cdc42 recruitment locally triumphs over dissociation. As can be expected, copy numbers of Cdc42 relative to GAPs will have to be in the right regime in order to restrict the latter behavior to only one possible membrane location in the cell, such that polarization may be established.

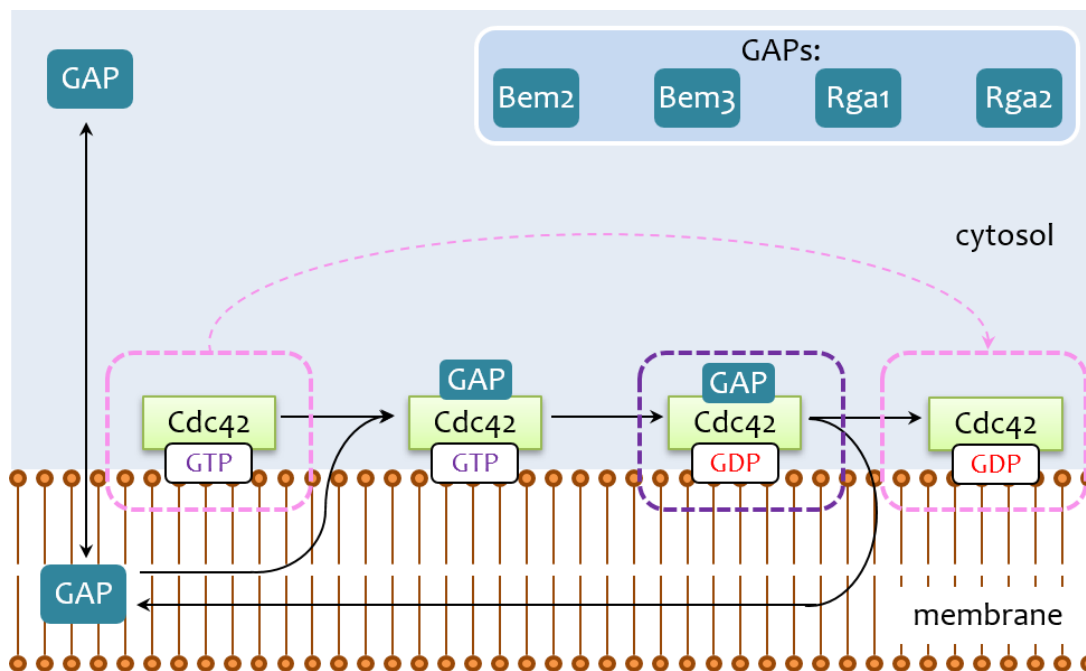


Figure 15 Two models for effective GAP action. The two states enclosed in pink connected by the dashed pink arrow represent the GAP action as modelled in [68], the full reactions are covered in [51], although the state enclosed in purple is implicit.

Conveniently, the mechanisms behind polarization in the complete model of [51] allows dissection into “functional subunits”. This is an abstract building block, which can serve as a level definition between genotype and phenotype (as in Figure 4 and Figure 14 of Chapter 1). More precisely, this level is positioned beneath mechanisms and modules. Therefore, validating this model means an important conceptual step in completing the genotype-phenotype map. As described in Chapter 1, finding intermediate levels between genotypes and phenotypes facilitates prediction of the latter, and makes concrete where the feedback from phenotype to genotype runs to cause adaptation.

More concrete predictions for validation of this model follow from numerical simulations, where the surprising result was established that symmetry breaking relies very sharply on the ratio of total Cdc42 and GAP proteins [51]. Beneath a certain value, polarization is impossible (see also Figure 16), while just above it the time to polarize time is minimal, increasing slowly

again for increasing $[Cdc42]/[GAP]$ ratio. More concretely, the prediction is that a Cdc42 concentration threshold to polarize exists for every background, and that this is highest for a $\Delta bem1$, followed by a $\Delta bem1 \Delta bem3$ and then WT. Such a striking feature is theoretically quite suitable for experimental validation. If validated, there is strong evidence for the polarity picture as depicted in Figure 15.

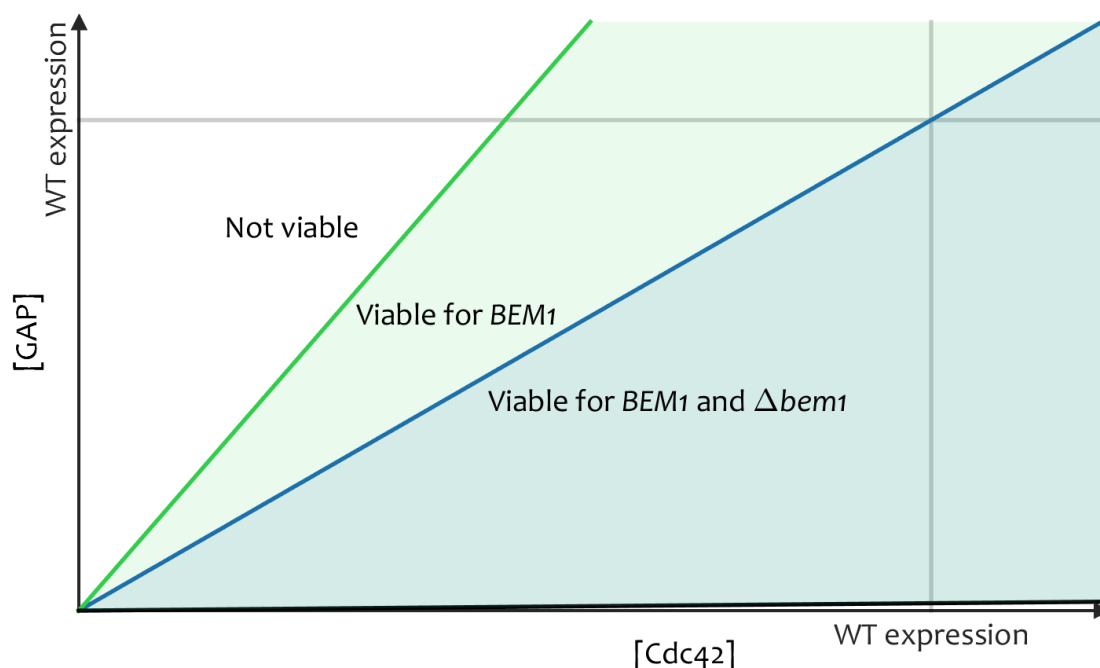


Figure 16 Schematic polarity cones for yeast strains with and without BEM1, marking the viable regions as a function of total Cdc42 concentration and GAP concentration in the cell, as under the model of [51]. Under WT expression, the $\Delta bem1$ is at the edge of viability. At very low GAP levels, there is a marginal region which does not allow sustained growth either.

As Cdc42 concentration is most important to control in any assay validating the remarkable dependence of doubling time on Cdc42/GAP concentrations, it is convenient to cross-validate this control. Ideally, Cdc42 is fluorescently labelled, which presents the first experimental obstacle. As the C-terminus contains the membrane anchor for Cdc42, fusing GFP to its N-terminus instead proved a common recipe for its labelling (e.g., [105], [97], [179]), conveniently also under inducible expression. However, it has become apparent in recent years that this construct suffers from artefacts, such as increased temperature sensitivity [48] and notably decreased Cdc42 association with vesicles secreted from the Golgi apparatus [96]. Clearly, an alternative, less invasive way of tracking Cdc42 needs to be devised.

Fortunately, much progress has been made for a related organism, fission yeast [185]. Strikingly, careful placement of the fluorophore (sfGFP) in the middle of the Cdc42 sequence yielded a fully functional fluorescent version of Cdc42. Given that conservation of this GTPase extends even to humans [75], it seems obvious to try the same approach on *S. cerevisiae* Cdc42. The construction and subsequent validation form the content of Appendix C.1, which

is summarized in the next section. Incidentally, as its human analog can be considered an important target in oncology studies [186], [187], the success of transferring this technique to other species may hold a relevant promise for its application to human biology as well.

Subsequently, assays concisely described in the next section (and extensively in Appendix C.2) demonstrate the functionality of the fluorescent Cdc42 under inducible expression. This allows us to proceed to the original experimental purpose, namely the detection of the aforementioned sharp Cdc42 concentration boundary for polarity. The growth assays at the heart of section 2.3 will highlight the boundary, albeit at a qualitative level. As we will see, more intricate modelling as done in Chapter 3 is required to understand the precise, quantitative consequences of the sharp Cdc42 boundary. Nevertheless, the assay suffices to draw conclusions on the detailed mechanics underlying the Cdc42 polarity model including the GAP details and its level interpretation in terms of “functional subunits”.

2.2 Overview of preparatory experiments for Cdc42-probe

Using methods as extensively described in section 2.6, sfGFP was successfully sandwich fused to Cdc42 in *S. cerevisiae* (following [185] in *S. pombe*). Additionally, the new *sfGFP-CDC42^{SW}* construct was placed under a Gal1-promoter, and inserted in strains with various polarity backgrounds, such as the $\Delta bem1$ and the $\Delta bem1 \Delta bem3$ background. Flow cytometry experiments on the more traditional N-term fusion of GFP to Cdc42 showed that we are able to vary the expression of Cdc42 under the Gal1-promoter from at least an order of magnitude beneath, to about an order of magnitude above endogenous expression. This yields sufficient dynamical range to probe the cone from Figure 16.

The single-cell microscopy experiments brought a minor artefact of the fluorescent sandwich fusion to our attention. Particularly aging cells accumulate bright fluorescent spots, consistent with the notion that sfGFP remnants withstand degradation [188]. These spots are brighter than the polarity spots located at the future bud sites, which are very difficult to see. This construct is hence not suitable for Cdc42 localization measurements. However, flow cytometry experiments on strains containing the sandwich fusion showed that the artefacts only had minimal influence on the total fluorescence distribution across the population. This caused this distribution to remain as theoretically expected [189] and as with the N-term fusion. Consequently, the construct serves to quantify Cdc42 copy number distributions on a population scale.

If we also take into consideration that the Gal1-promoter has a large variability of expression (i.e., is noisy), we can see the first indication that Cdc42 copy number influences fitness and is subsequently under selection. This is because even at a single induction level, the Cdc42 expression can still span more than one region of Figure 16 in terms of expressed protein

copies. How noisy expression exactly couples to population phenotypes is the topic of Chapter 4. Also, a proposal for a more sophisticated experiments is presented there.

Fitness is not only affected by the Cdc42 expression we induce, but also by the environment. This is well visualized in an assay that determines cell volumes under different inducer (galactose) levels. Presumably due to osmolarity effects, cells expand by almost 1 μm in diameter when varying galactose from zero to the maximum of 2% in the medium. Correcting for osmolarity removes the corresponding fitness effects. Secondly, the richness of the medium influences the results significantly. Remarkably, slow growth medium (less rich in amino acids) acts to equalize fitness (and volumes) of polarity backgrounds with low and high Cdc42 expression levels. This phenomenon is further explained in 2.3.2 and 3.4.1. For now, it suffices to state that we require rich medium (particularly rich in amino acids) to exacerbate phenotypical difference between strains with distinct polarity backgrounds and/or that differentially express Cdc42.

Conveniently, suboptimal medium conditions allow us to better determine that the sfGFP construct itself exhibits very little fitness effects. Volume effects (indicative of fitness effects) of this construct are within 10% of WT with unmodified *CDC42*. Hence, in terms of the goals in this chapter to validate the molecular polarity model including the GAPs through a fitness assay, this construct is fully functional and appropriate here.

2.3 Growth assay for Cdc42 threshold determination

2.3.1 Growth assays shows predicted trends while varying Cdc42 expression

Ultimately, our goal from the start of this chapter has been to test the molecular model describing the GAPs. For this to be confirmed, we restate that a Cdc42 concentration threshold to polarize exists for every background, but highest for a $\Delta bem1$, followed by a $\Delta bem1 \Delta bem3$ and then WT. As failure to meet the threshold negatively impacts fitness, measuring this along a gradient of Cdc42 expression for these backgrounds is required.

By distributing biological samples in a 96 well layout, it is possible to do the necessary bulk measurements in automated form. The protocol is further described in the supplement to this chapter (2.6). In short, growth of cells in a well can be effectively detected through transparency measurements of the well, expressed as the optical density (OD at 600 nm wavelength light). At low OD, the rise in biomass is proportional to the rise in OD. This means that e.g., if strain 1 causes a doubling of the OD within a time T , this may be because of a doubling in cross-sectional area of the cells, because the number of cells has doubled, or a combination of both. Experience with the growth model of Chapter 3 learns that the size distribution across the population equilibrates within a few-fold of the doubling time into the experiment. This means we can attribute the rise in OD to an increase in number of cells, and

consequently, to fitness. The latter step follows from assuming exponential growth, which translates to a linear fit in log-space. By using a self-made GUI, weighted least squared regressions are performed on all growth curves, resulting in growth rates with (fitting) error estimates. An example of how this works is depicted in Figure 17.

A first, low-resolution check of the aforementioned model hypotheses is given by the frequency with which growth is observed across technical replicates of the same genotype, see Figure 18. Seemingly, we observe some indication of thresholds. Wild-type with *GAL1pr-sfGFP-CDC42^{SW}* always grows, except at the lowest level of induction. The threshold for growth is more diffuse for the $\Delta bem1$ backgrounds. Likely, we are not getting binary results for replicates in the same medium as the production of Cdc42 is stochastic. Initial conditions will become important, such as the saturation duration prior to inoculation for measurements (also considering Appendix C.2.1.3). Therefore, we only qualitatively state that the threshold for the $\Delta bem1$ backgrounds from this figure is around 0.06%. This is consistent with the earlier rough estimate of endogenous expression equivalence between 0.05% and 0.2% (see Appendix C.2.1.1, Figure 59), at which both the $\Delta bem1$ and $\Delta bem1 \Delta bem3$ should be viable.

However, we can pinpoint the Cdc42 barrier more precisely by considering growth rates. As aforementioned, the colony equilibrates quickly in biomass growth rate, and initial conditions become irrelevant. If we only consider the (maximum) growth rates and therefore only the replicates where growth actually took place, we get a good measurement on the behavior of a strain background, cleared of initial condition-specific effects. Ideally, the data would then be as reproducible as possible, with the measurement errors as the dominant noise source.

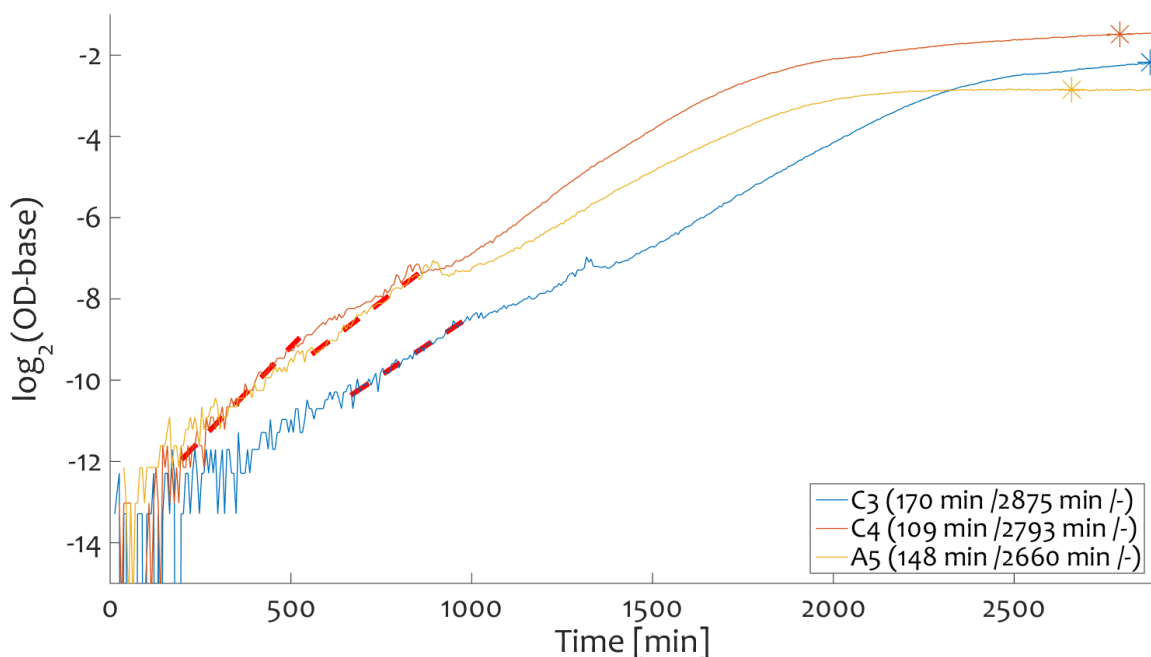


Figure 17 Growth curve GUI with curves (colored lines) from different wells (coded in the legend) with doubling time estimates (the first values in parentheses in the legend) resulting from linear fits (red dashed lines) on the log OD (background correction is called 'base' in vertical axis label).

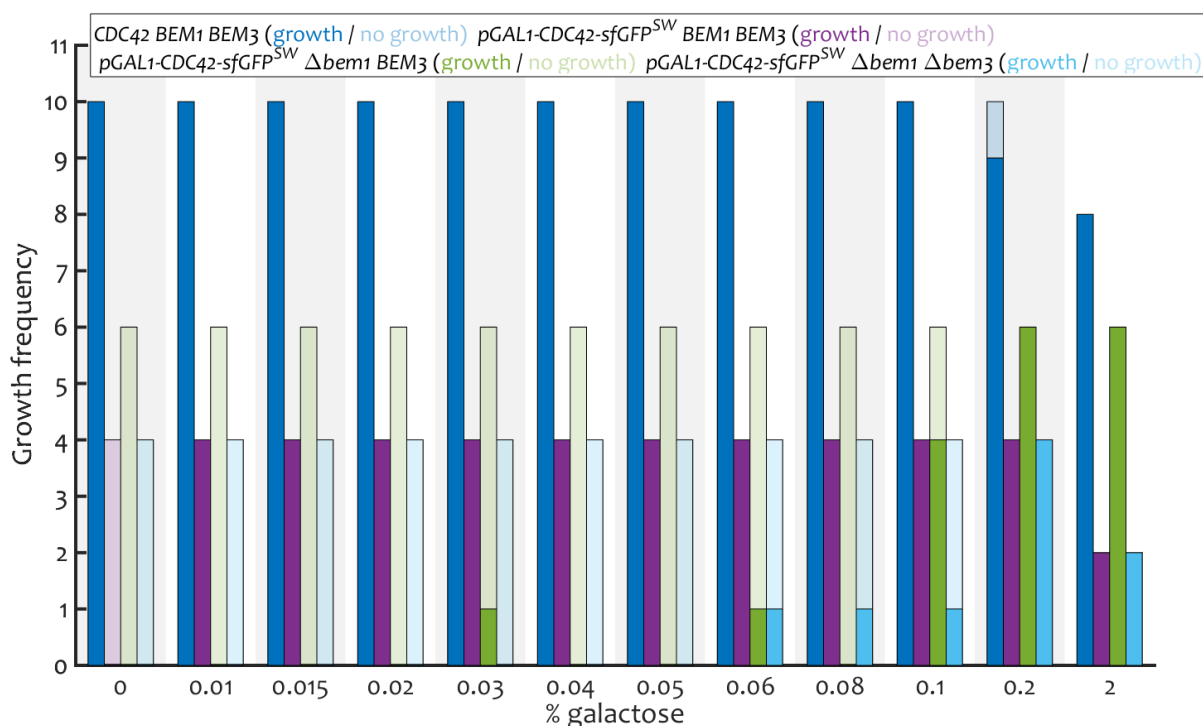


Figure 18 Bar plot of the frequency of growth (opaque) and no growth (transparent) as a function of galactose content in the medium. Backgrounds comprise of pure WT (dark blue, YldB001), and those with $GAL1pr$ -sfGFP- $CDC42^{SW}$ as the only source of Cdc42 in an otherwise WT (purple, YldB003), $\Delta bem1$ (green, YWKD069a) and $\Delta bem1 \Delta bem3$ (light blue, YWKD070a) background.

When performing multiple technical replicates of the same genotype (typically 10 for pure WT), it is observed that the fitting error is not the dominant error. A clear overdispersion is seen in the data, so we need to incorporate the uncontrolled source of noise into the data. As the number of data points for bootstrapping is low, one parametric solution is inflating the variance of each point by the same numerical constant, assuming all points are equally affected. This factor known as the modified Birge ratio [190] has been estimated in pure WT, averaging across media, and applied to all other backgrounds to account for overdispersion.

Secondly, using pure WT as a reference, we find that it is influenced by the medium, possibly due to osmolarity effects, as experienced in Appendix C.2.2. To de-trend all growth rates, the relative fitness is considered, the ratio of the growth rate of a strain to that of WT in the same medium. This gives the plot of Figure 19.

Probability distributions for the growth rates follow from Bayesian analysis. Its advantage over the frequentist approach is that it allows answering the question whether the hypothesis is true (instead of whether we can reject it), as explained in e.g., [191]. Bayes factors (posterior odds ratio) can be constructed (as shown in Table 1) denoting the probability ratios of one growth rate being larger than another, to the opposite scenario. These form a means to statistically test whether one rate can indeed be considered larger than the other. As a guide, we adhere to the convention mentioned in [192], where an odds ratio of greater than 10 provides strong evidence that the hypothesis of 'rate 1 larger than rate 2' is true.

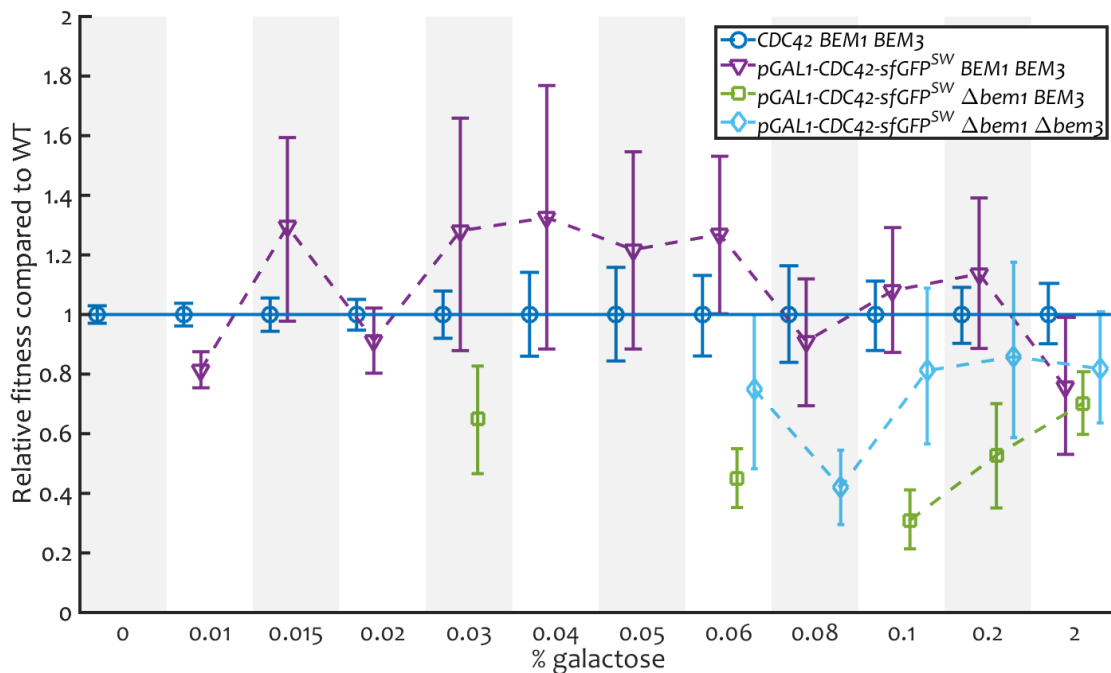


Figure 19 Relative growth rates (fitness) of four strain backgrounds as a function of galactose content in the medium. Individual measurement errors have been corrected for overdispersion through the average Birge ratio in pure WT [190]. Growth rate distributions (the error bars indicate the 68% credible interval) have subsequently been constructed by using the Metropolis-Hastings algorithm [193] with a prior assuming doubling times of >70 min. while otherwise uninformative. Backgrounds comprise of pure WT (dark blue circles, YldB001), and those with $GAL1pr-sfGFP-CDC42^{SW}$ as the only source of Cdc42 in an otherwise WT (purple triangles, YldB003), $\Delta bem1$ (green squares, YWKD069a) and $\Delta bem1 \Delta bem3$ (light blue diamonds, YWKD070a) background. Dotted lines serve as a guide to the eye.

Strikingly, what can be seen from the growth rates is mostly as predicted. The trends in growth rates suggest indeed a cut-off for Cdc42 copy number, which is different across genetic backgrounds. In the otherwise WT background, only very low expressions of Cdc42 are insufficient for viability, and at 0.01% induction a negative fitness effect is still significant. However, above 0.01% there is some evidence that the sandwich fusion somehow confers even a slight advantage, as also noted in Appendix C.2.2. In any case, this shows that the $sfGFP-CDC42^{SW}$ can adequately replace endogenous Cdc42 in terms of fitness as well, without negative consequences.

Moreover, the growth rates are significantly lower than WT (with construct) for the $\Delta bem1 \Delta bem3$ below 0.1% and the $\Delta bem1$, except at maximum induction. The negative effect of the higher Cdc42 barrier relative to WT in these backgrounds is problematic for the cells, unless expression is high, in which case all growth rates converge. Similarly, rates of the $\Delta bem1 \Delta bem3$ are always higher than those of the $\Delta bem1$ (sometimes significantly, except at maximum induction), which points towards a higher concentration barrier for the $\Delta bem1$, as hypothesized.

The question remains why the boundaries are perhaps less sharp as one may have naively foreseen. More concretely, extrapolating the $\Delta bem1 \Delta bem3$ curve indeed suggests it halts at a lower induction level than the $\Delta bem1$. Yet, the occasional single growth detection occurred for the $\Delta bem1$ at a rather low level. The intuition behind this is as follows.

The promoter, either the endogenous or the *GAL1*, is intrinsically very noisy. This causes part of the population to acquire a large excess of Cdc42. If sufficient cells have this lucky fate, the population can grow for a substantial period of time. Particularly the first phase at very low OD is crucial to survive and most subject to chance, which makes the critical Cdc42 threshold seem less sharp. Additionally, the observed Cdc42 distributions imply a monotonic, but very non-linear relation with induction and galactose content. Therefore, the sharpness of the cut-off can only be reconstructed taking into account the population protein distributions.

Table 1 Posterior odds ratio for the hypothesis that one strain has a higher growth rate than another, for the strains of Figure 19 which follow the respective shorthand notation WT, WT (sfGFP), $\Delta bem1$ (sfGFP) and $\Delta bem1 \Delta bem3$ (sfGFP). Cells in green represent values for which there is strong evidence (>10 , [192]) that this hypothesis is true.

Galactose \ Strains	WT/ WT (sfGFP)	WT/ $\Delta bem1$ (sfGFP)	WT/ $\Delta bem1$ $\Delta bem3$ (sfGFP)	WT (sfGFP)/ $\Delta bem1$ (sfGFP)	WT (sfGFP)/ $\Delta bem1$ $\Delta bem3$ (sfGFP)	$\Delta bem1$ (sfGFP)/ $\Delta bem1$ $\Delta bem3$ (sfGFP)
0%	-	-	-	-	-	-
0.01%	293	-	-	-	-	-
0.015%	5	-	-	-	-	-
0.2%	4	-	-	-	-	-
0.03%	3	67	-	12	-	-
0.04%	3	-	-	-	-	-
0.05%	3	-	-	-	-	-
0.06%	5	4999	5	∞	19	7
0.08%	2	-	356	-	129	-
0.1%	2	∞	4	∞	4	146
0.2%	2	399	3	63	4	6
2%	5	113	5	1	2	3

2.3.2 Cdc42 thresholds are subject to growth conditions

In order to reconfirm our predictions on the Cdc42 thresholds, we turn to a repetition of this experiment, but this time under slower growth conditions. The idea would be that if growth is slower, there should be more time for the cells to accumulate Cdc42 or more time to be lucky, as the Gal-promoter is very noisy. Under these circumstances, even mutants as the $\Delta bem1 \Delta bem3$ with relatively low expression from the *GAL1pr-sfGFP-CDC42^{SW}* construct (given Figure 19 below 0.06%) may then survive.

Since changing sugars is problematic with regards to the use of the Gal-promoter, the poorer growth conditions are accomplished by a deficiency in amino acids (at least 4 times less than in Figure 19). Unfortunately, due to a problem with medium evaporation (up to 50%), galactose concentrations can be up to twice as designed. Still, the resolution in expression is sufficient for our conclusions.

As can be seen from Figure 20, this time strains with the *sfGFP-CDC42^{SW}* sandwich construct as their sole Cdc42 source grow at every run of every non-zero inducer concentration measured. In the negative control condition of 0% galactose, these strains indeed do not grow, as before (only WT with endogenous *CDC42* grows here). More details on the doubling times are found in Appendix C.3. As expected, the poorer medium has shifted the minimum Cdc42 concentration threshold downwards, to below the level induced at 0.04% galactose concentration (based on measurements designed at 0.02% galactose, with the factor of 2 taking into account possible evaporation).

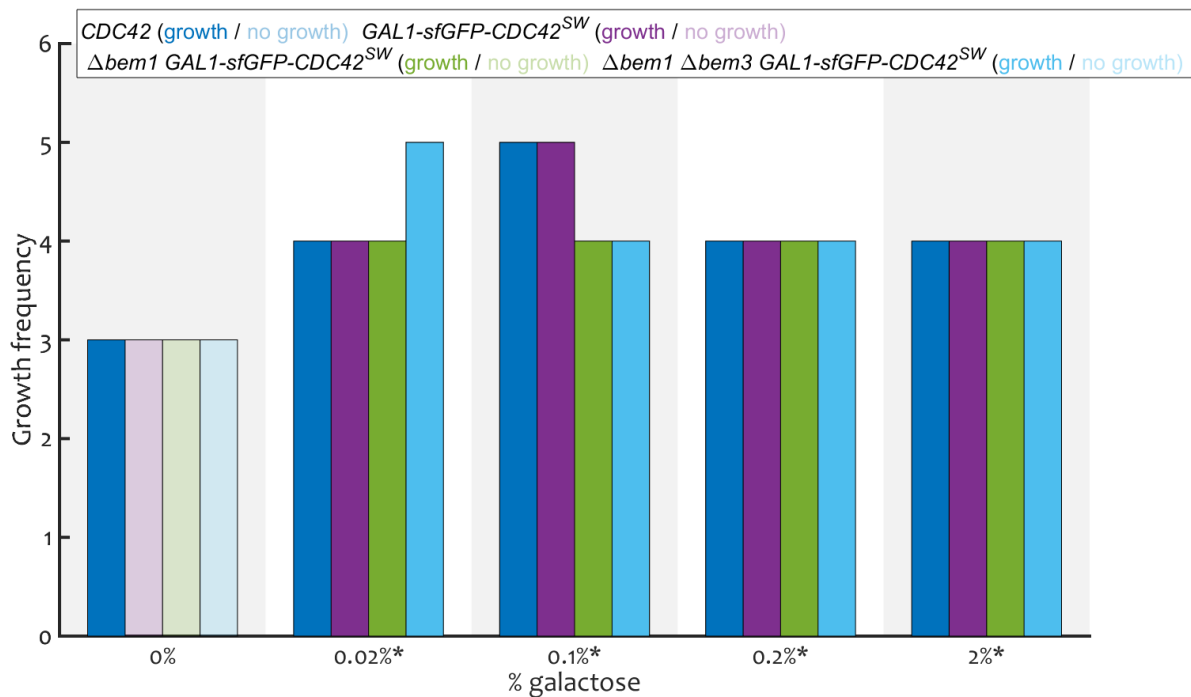


Figure 20 Poor (low amino acid containing) medium promotes survival of $\Delta bem1$ backgrounds. Bar plot of the frequency of growth (opaque) and no growth (transparent) as a function of galactose content in the medium. Backgrounds comprise of pure WT (dark blue, YWKO62c), and those with *GAL1pr-sfGFP-CDC42^{SW}* as the only source of Cdc42 in an otherwise WT (purple, YWKO65a), $\Delta bem1$ (green, YWKO69a) and $\Delta bem1 \Delta bem3$ (light blue, YWKO70a) background.

2.4 Discussion

Determining how self-organizing systems evolve is a crucial part in understanding life as a whole. The spontaneous symmetry breaking during cell polarity for budding yeast provides a unique opportunity for this. Here, a self-organizing system can be reverse-engineered to such detail, that constructing the bridge between genes and traits, one level at a time, becomes within reach. The functional subunits concept, which aims to summarize molecular details regarding pattern formation, forms a stepping stone. However, this concept first had to be tested by validating the mechanistic model it forms for this polarity process.

The central model prediction is the existence of minimum concentration thresholds (for polarity success) of Cdc42, the key polarity player. These boundaries differ per genotype, e.g., it is very low for WT, but higher for the $\Delta bem1 \Delta bem3$ and even higher for the $\Delta bem1$. By replacing the endogenous Cdc42 by an inducible, fluorescent variant, we were able to test this hypothesis, removing some of the remaining mechanistic unknowns of this system. The following was found (conclusions in **bold**, hypotheses or claims in *italic*):

- **The novel sandwich fusion of Cdc42 with sfGFP in *S. cerevisiae* (following [185] in *S. pombe*) exhibits essentially no fitness effects. While the cell volume effects of this construct are within 10% of WT (indicative of fitness effects), the growth rate measurements seem consistent with no significant fitness effects.**

There might be some size effects of the addition of sfGFP on the diffusive constant and degradation of Cdc42 as explained at the end of Appendix C.2.2.2, as volumes with the sfGFP sandwich are slightly smaller. This may even suggest a positive fitness effect for the sfGFP construct, and from Figure 19/Table 1, it is indeed closer to being beneficial than deleterious. However, at every single galactose concentration above 0.015%, the fitness difference is not significant. As far as fitness is concerned the construct is therefore fully functional, prompting:

*The GAL1pr-sfGFP-CDC42^{SW} construct is the first fully functional fluorescent Cdc42 in terms of fitness for *S. cerevisiae*.*

- **Fluorescent microscopy on Cdc42-sfGFP sandwich fusions shows artefacts in aging cells which leads to small (on the population scale) artefacts. These interfere with Cdc42 localization, but not with the determinations of copy number. The artefacts are brighter than the polarized Cdc42-sfGFP spots, which are on the edge of detectability.**

This is arguably related to sfGFP enduring degradation, which may be mitigated by mNeonGreen [188]. As authors there state this fluorophore can be up to 5 times brighter, this will likely provide the signal-to-noise ratio jump needed for localizing Cdc42. We thus expect:

The use of mNeongreen in a sandwich fusion with Cdc42 will likely mitigate the artefacts in Cdc42-fluorophore sandwich fusions, allowing clear visualization of Cdc42 localization.

Follow-up work conducted in the Laan Lab also provides further support for this claim, yielding sharp images of Cdc42 localization using confocal microscopy (personal communication).

- The *GAL1pr-sfGFP-CDC42^{SW}* construct as the only source of Cdc42 demonstrates minimum viable Cdc42 thresholds. These are lowest for WT (below 0.015% galactose induction) followed by those of $\Delta b e m 1 \Delta b e m 3$ and $\Delta b e m 1$ mutants. As growth rates of the $\Delta b e m 1 \Delta b e m 3$ are always higher than those of the $\Delta b e m 1$, the latter presumably has the highest threshold. WT outgrows the $\Delta b e m 1 \Delta b e m 3$ and $\Delta b e m 1$ mutants unless expression well exceeds the respective threshold of the mutants.

This conclusion is arguably the most important one, as it demonstrates how a theoretically sharp boundary can still be observed across genotypes, although heavily subject to noise in Cdc42 copy number. The resolution of the experiment is also high enough to decisively state that WT has the lowest bound. The $\Delta b e m 1$ backgrounds could benefit from more statistical power with more data points, as the evidence is compelling but still influenced slightly by uncertain growth around the threshold. It still seems certain enough to claim the hypotheses from the molecular model which we wanted to investigate are true, which then implies:

To polarize in absence of the strong positive feedback loop mediated by Bem1, local depletion of GAPs available for deactivating Cdc42 at the polarity site is required.

Following further experimental evidence from [51], we accept this hypothesis definitively. Additionally, as explained there, it fits neatly within the context of functional subunits, which seems a useful and tractable molecular level definition.

- The fitness dependency on Cdc42 is subject to growth conditions, which can be manipulated in a predictable manner. Slower growth conditions (poorer amino acid content) equalize fitness of otherwise ill mutants, facilitating their growth and lowering the minimum Cdc42 expression for viability (e.g., for the $\Delta b e m 1 \Delta b e m 3$ with inducible Cdc42, from 0.06% inducer concentration to at least below 0.04%).

For our understanding of the system's evolvability, it is not only of importance how flexible this self-organizing system is in terms of genetic alternatives. Routes towards adaptation occur within an environment, and generally this environment influences the fitness resulting from a genotype. Disentangling the intertwined influences of genotypes and environment proved critical in determining the mechanistic details of polarity. We established in the growth assays in Figure 20 (see also Appendix C.3) that poorer medium lowers the Cdc42 thresholds and hence mitigates the effect of low Cdc42 expression, which was most apparent in the assay of 2.3.1. We influenced the fitness of struggling polarity mutants with different media types, although their mutations belonged to a module seemingly unrelated to

metabolism. These findings allow further speculation on the emergence of certain genotypes, and the influence the environment may have had on the evolution of this particular system.

While the aforementioned Cdc42 threshold was found particularly relevant in the $\Delta bem1$, it becomes less relevant as the growth is slowed down due to another cause (low amino acid content in this case). As growth conditions might rarely be optimal, determining what genotypes are better is not so self-evident. More paths of evolution may become feasible this way. This medium effect is further quantified through modelling in the next chapter. Furthermore, the differential roles of noisy Cdc42 expression during times of fast and slow growth on evolutionary capacity will be treated in 4.2.1.6. For now, it is tempting to formulate the following hypothesis:

Evolution of the yeast polarity system is subject to a trade-off ratchet (as the lac operon [42]), using environmental fluctuations to land on non-obvious genotypes.

With the green light on the verification of the mechanistic details of the model, as a next step I will construct the genotype-phenotype map. We have seen that there is an intricate interplay between genes and environment. Furthermore, there is by no means a single Cdc42 concentration at an expression level, not in the induction case nor in the endogenous case. These effects prevented us thus far from a full quantitative understanding of the system at hand, with volumes and fitness amongst the phenotypes concerned. Building upon the molecular model, and the accompanying functional subunits concept that forms a level above proteins, we must model our way to the phenotype level. This will be the focus of the next chapter, and brings us closer to the ultimate goal of discovering new evolutionary rules.

2.5 Acknowledgements

The following people (unless otherwise annotated, affiliated with the Laan Lab at the TU Delft at the time) contributed to the data in this chapter:

- Fridtjof Brauns (Frey Group, LMU Munich) for development and explanation of the theoretical molecular polarity model and fruitful discussions on the experimental design to validate this.
- Marit Smeets, which as a student under my supervision, acquired movies and the flow cytometry data of *GFP-CDC42* strains.
- The Wedlich-Söldner Lab (WWU Münster) for providing the strain RWS116 [85] and *GFP-CDC42* strains RWS119 and RWS1421 [48].
- Felix Meigel with whom I have worked in tandem to generate *GAL1pr-CDC42* and *GAL1pr-sfGFP-CDC42^{SW}* haploids (and performed the tetrad dissection) used throughout this chapter, and to acquire flow cytometry and volumetric data on these. Also, he performed the growth assay in slow growth conditions (Appendix C.3).
- Ilse de Bruin, which as a student under my supervision performed growth assays on the *GAL1pr -CDC42* and *GAL1pr-sfGFP-CDC42^{SW}* haploids.
- Liedewij Laan for diploids from [24] I used as starting points for Cdc42 modifications.
- Esengul Yildirim for guiding me through the *GAL1pr-CDC42* and *GAL1pr-sfGFP-CDC42^{SW}* strain construction design.
- The Rong Li Lab (currently JHUSOM, US) for plasmid pRL368 [194] I used as a starting point for cloning the CDC42 constructs.
- Sophie Martin (UNIL, Switzerland) for her personal communication regarding her advice on the fluorophore for the *S. cerevisiae* sandwich version.
- Nicole Imholz (Bokinsky Lab, TU Delft) for the sfGFP-template plasmid from iGEM I used for cloning.
- Diego Gomez Alvarez (Youk Lab, TU Delft) for the protocols used during flow cytometry measurements.
- Melanie Wijsman (Daran-Lapujade group, TU Delft) for instructions for the transformation protocol used.
- IMB group at TU Delft for use of their facilities for the volumetric assay and the corresponding protocol.
- Laszlo Balkay (Unideb) for the Matlab code *fca_readfcs* used to read-out flow cytometer data.

2.6 Supplementary information on methods

As a general notice, all details of used materials (brands, Appendix A) and strains/plasmids (Appendix B) are enumerated separately in the Appendices. For conciseness, these are generally not named throughout this chapter.

2.6.1 Cloning

Starting with pRL368, which has *Gal1pr-GFP-CDC42* and the *URA3* marker, the goal is to obtain a plasmid which has the sandwich fusion *GAL1-sfGFP-CDC42^{sw}* instead of the N-term GFP fusion. This plasmid would have to be integrated replacing the endogenous *CDC42*, so homology regions have to be inserted into the backbone as well. Homology regions are designed to be separated by a unique restriction enzyme cut site (for EcoRI), so that the regions form the flanks of plasmid when linearized, which can then integrate into the genome.

Using standard molecular biology techniques (e.g., [195]), pRL368 is amplified and ligated to remove the GFP, yielding pWKD006. Using Gibson assembly [196], aforementioned genomic homology regions are inserted to yield pWKD010, Similarly, the sfGFP was added to obtain pWKD011 (see Figure 58).

2.6.2 Strain construction

After cutting pWKD010 and pWKD0111 with EcoRI, yeast transformation then followed the main line of [197], using diploid YWKD045 (heterozygous for *BEM1* and *BEM3*), yielding YWKD054 and YWKD055. These were sporulated (using standard methods, see e.g., [198]) to generate the polarity mutants haploids with and without sfGFP used throughout this dissertation².

2.6.3 Experimental assay protocols

2.6.3.1 Microscopy

For the details on microscopy hardware, see Appendix A. Several *GAL1pr-sfGFP-CDC42^{sw}* strains were measured under the microscope. The time lapse movie underlying Figure 60 was constructed by measuring YWKD065a at 30 °C for 16 hours every 4 minutes (60x objective, 1x zoom) brightfield (LED at 5% power, exposure 30 ms) and wide field fluorescence images (200 ms, 485 nm excitation, Spectra at 10% power). Wide field fluorescence images were taken at 11 z-planes (600 nm apart) and then projected using the maximum intensity at each pixel.

² The strain YFM007a resulted in a similar manner, only instead of sonicating the spores to separate them, the tetrads were dissected (see acknowledgement at 2.5).

Cells were pre-grown in a rotating wheel at 30 °C in CSM-URA + 2% raffinose and 0.1% galactose, then 500 times diluted into CSM-NF +2% raffinose and 0.1% galactose. Cells were shaken loose by vortexing and pipetting up and down with 10 µl pipette. Then cells were placed into 96 well plate for imaging, by first coating the wells with 70 µl 0.1 mg/ml concanavalin A (incubated 30 min. at room temperature) with the lid on. After removing it, 200 µl of 500x diluted cells were placed in the wells, and the same amount of water was placed in surrounding wells to counter evaporation. Additionally, the plate is sealed by surrounding the edge with parafilm. The cells are then centrifuged stuck to the bottom of the well (1000 rpm, 3-5 minutes), at which time the plate is ready for imaging.

2.6.3.2 Flow cytometry

2.6.3.2.1 GFP

Using a BD FACScan flow cytometer, the fluorescence intensity measurements concerning all strains with the N-term GFP fusion to Cdc42 were performed (see 2.5). The protocol encompassed the following:

- Turning on machine and allow 5 minutes of idle time to warm-up lasers
- Fill the sheath tank with at most 3L of PBS
- Empty the waste tank
- Flip the valve at the sheath tank, and if there are any, dislodge air bubbles from the vent line
- To remove bubbles from the sheath line, drain (10") and fill (20") three times, the first two times without a tube, the last time with 3 mL MilliQ in a tube submerging the probe ("Sample Injection Port"). Continue the fill until the sheath to waste line is completely filled.
- Flip to standby, and the system is ready for measurements (put probe sample tubes).

Strains measured were YMS02-04, replicates of RWS116, 119, and RWS1421 from the Wedlich-Söldner lab, see also Appendix E: Reconstruction of burst parameters from FACS data for subsequent data processing.

2.6.3.2.2 sfGFP

After three days, strains grown in a 96 well plate were diluted in the morning in fresh media, and then measured in the afternoon (in a well plate). Measurements took place using the following protocol on a FACSCelesta (see 2.5):

- Turning on machine and allow 30 minutes of idle time to warm-up lasers
- Fill the sheath tank with at most 10L of PBS
- Empty the waste tank
- Prime system: press run button and HTS/prime in menu (~3 min.) and repeat

- Calibrate using beads (1 drop in in 150 μ l PBS) in well A1 by clicking Cytometer/CST in the menu, select flat bottom and supply lot ID of beads.
- Set to standby, and the system is ready for measurements
- For measurements, set threshold to remove debris, set plate to 96 well flat bottom.
- Then select how to measure, premix well before measuring, sampling volume/well (with 50 μ l of dead volume) and amount of washing volume in 'Standard-Throughput' mode
- In Experiment/Experiment lay-out/Acquisition, select which wells to measure, stopping gate, and do not record more than 2500 events/s (then well too dense).
- Afterwards, use a plate for cleaning, where 3 wells contain FACS Clean, 3 wells FACS Rinse and 6 wells MilliQ (all 200 μ l, with 50 μ l dead volume).

2.6.3.3 Volumetric assay

2.6.3.3.1 Measurement

Strains used for the Coulter Counter measurements were inoculated in 5 mL CSM-low fluorescence + sugar incubated at 30 °C. After two days, these were diluted in fresh 5 mL of the same medium, with the dilution factor depending on how dense the tube appeared. Tubes were categorized in four groups, which were diluted up to 1000 times. The next day the densest samples were diluted a factor of 5. After sonication (6 min. on time with 15" off every 30" on at 70% of 500 W), the cells were then brought to the Coulter Counter, and measured using the following procedure (see acknowledgement 2.5):

- The probe is submerged in Coulter CLENZ Cleaning Agent. Remove this cuvette and replace by cuvette with Isoton II Diluent. Flush five times.
- Measure blank (new cuvette with Isoton II Diluent) to get an indication of the background reads, which stabilize around 50.
- Samples were diluted to about 100 000 cells/ml by eye, sometimes overdiluting first and from the counts preparing a new sample diluted in Isoton II with the appropriate density in 10 mL volumes.
- Usually, all samples are measured five times
- Run a blank (cuvette with only Isoton II) between different samples and at the end to confirm all cells are flushed

Not all strains at all concentrations were ultimately sufficiently grown for the measurements, leaving the strains shown in Figure 65 and Figure 66.

2.6.3.4 Growth assays

2.6.3.4.1 Media

There are several considerations for medium choice. Firstly, we would like to have our medium composition very controlled and reproducible. Therefore, we use synthetic medium, consisting of yeast nitrogen base (0.69% w/v), CSM amino acid mix (0.79% w/v) and sugar. To maintain compatibility with the medium of the microscopy assays, we choose the nitrogen base without the autofluorescent components riboflavin and folic acid. The amino acid mix is four-fold higher in concentration for the assay in 2.3.1 (and filter sterilized, to preserve amino acids) compared to the assay of 2.3.2 (see also Appendix C), which used autoclaved amino acids. The former assay ensures that strains do not run out of amino acids early in log phase, causing a small shift in growth rate within which cells proceed to make these amino acids themselves. The latter can illustrate the effect of poorer medium on Cdc42 thresholds.

As the *GAL1* promoter is used, we cannot use dextrose as our sugar (suppresses expression [199]), but also not only galactose either as this sugar should be the inducer and not the metabolite. Without dextrose, the Gal promoter should eventually turn on for the percentages used in this dissertation ($\geq 0.01\%$) [199]. For the metabolite, we therefore use 2% raffinose. To avoid osmolarity differences (see the growth assay in Appendix C.3) the varying galactose can be buffered by the for budding yeast non-metabolizable sorbitol, which has comparable molecular weight. Only in that assay, x% galactose was therefore supplemented with (2-x)% sorbitol.

Finally, evaporation of media during the experiment turned out to be a potential problem. While plates used for the growth assay in 2.3.2 (see also Appendix C.3) did have additional water between wells and at the sides of the plate to counter evaporation, even that does not stop media from evaporating significantly. This also creates an evaporation bias (up to nearly 50%) for strains on the edge of the plate compared to the centre of the plate.

In this assay, this has the consequence that galactose concentration may be up to twice as high as intended (as the water evaporates). Secondly, OD values may be overestimated, as the wells become denser with cells, mainly due to growth but partially due to evaporation. The latter will not be a severe effect though, since a doubling of OD in 2 days due to evaporation in strains which typically double every 2 hours is only an effect of a few percent. Still, the quantitative growth assay where galactose concentrations are most critical, in 2.3.1, mitigates these problems by not trying to block the water with parafilm on the sides, but by a transparent sticker on top of each well.

2.6.3.4.2 Strains and layout

To test the Cdc42 dependence in different backgrounds, strains with *GAL1-sfGFP-CDC42^{sw}* in either a WT (YidB003/YWKD065a), Δ *bem1* (YWKD069a) or Δ *bem1* Δ *bem3* (YWKD070a) background are placed in a 96 well plate layout. Additionally, all runs featured YidB001, a *CDC42* strain in a WT polarity background, as a reference. Concretely, four growth

measurement runs were done with the layout as in Table 2. The assays of 2.3.2 / Appendix C had the layout of Table 3, where the pure WT was YWKD062c this time.

Table 2 Media and strain position in a 96 well layout for the four growth assay runs of 2.3.1.

Column (every run)		Row	Run 1/2	Run 3/4
1	0% Gal	A	YldB001	Only medium
2	0.01% Gal	B	Only medium	YldB001
3	0.015% Gal	C	YldB001	YWKD065a
4	0.02% Gal	D	YWKD069a	YWKD070a
5	0.03% Gal	E	YldB001	YWKD065a
6	0.04% Gal	F	YWKD069a	YWKD070a
7	0.05% Gal	G	Only medium	YldB001
8	0.06% Gal	H	YWKD069a	Only medium
9	0.08% Gal			
10	0.1% Gal			
11	0.2% Gal			
12	2% Gal			

Table 3 Distribution of strains and media for the growth assay runs of 2.3.2. All media has 1x autoclaved CSM + 2% raffinose, the third run also has (2-x) % sorbitol, where x is the galactose percentage. Wells marked with an asterisk were switched to 0% galactose in the measurement plate.

Strain	% Gal	1st run				2nd run		3rd run (Sorbitol added)		
CDC42	0					D2*	B4*	F5	G5	G6
CDC42	0.02	B1	C3			C6	D6	B2	C2	D2
CDC42	0.1	C10	B12			C7	D7	B5	C5	B6
CDC42	0.2	G1	F3			E6	F6	B11	C11	D11
CDC42	2	F10	G12			E7	F7	E11	F11	G11
GAL1-CDC42	0.02	A4	B4	C4		C3	C4			
GAL1-CDC42	0.1	A9	B9	C9		C9	C10			
GAL1-CDC42	0.2	F4	G4	H4		F3	F4			
GAL1-CDC42	2	F9	G9	H9		F9	F10			
GAL1-sfGFP-CDC42 ^{SW}	0					D4*	B5*	F6	F7	G7
GAL1-sfGFP-CDC42 ^{SW}	0.02	D3	D4			C5	D5	B3	C3	D3
GAL1-sfGFP-CDC42 ^{SW}	0.1	D9	D10			C8	D8	C6	B7	C7
GAL1-sfGFP-CDC42 ^{SW}	0.2	E3	E4			E5	F5	B10	C10	D10
GAL1-sfGFP-CDC42 ^{SW}	2	E9	E10			E8	F8	E10	F10	G10
Δ bem1 GAL1-sfGFP-CDC42 ^{SW}	0					D3*	B6*	E8	F8	G8
Δ bem1 GAL1-sfGFP-CDC42 ^{SW}	0.02	B3	B6	C6	D6			B4	C4	D4
Δ bem1 GAL1-sfGFP-CDC42 ^{SW}	0.1	B7	C7	D7	B10			B8	C8	D8
Δ bem1 GAL1-sfGFP-CDC42 ^{SW}	0.2	G3	E6	F6	G6			B9	C9	D9
Δ bem1 GAL1-sfGFP-CDC42 ^{SW}	2	E7	F7	G7	G10			E9	F9	G9
Δ bem1 Δ bem3 GAL1-sfGFP-CDC42 ^{SW}	0					G9*	E10*			
Δ bem1 Δ bem3 GAL1-sfGFP-CDC42 ^{SW}	0.02	C2	B5	C5	D5					
Δ bem1 Δ bem3 GAL1-sfGFP-CDC42 ^{SW}	0.1	B8	C8	D8	C11					
Δ bem1 Δ bem3 GAL1-sfGFP-CDC42 ^{SW}	0.2	F2	E5	F5	G5					
Δ bem1 Δ bem3 GAL1-sfGFP-CDC42 ^{SW}	2	E8	F8	G8	F11					

2.6.3.4.3 Protocol

First, a 96 well plate is inoculated with about 5 μl of cells into 100 μl medium per well in a predefined layout. In addition, sterile MilliQ is put in the four reservoirs in the perimeter (4 times 2 mL) and 77 times 100 μl in the reservoirs between wells. The wells are then covered by a sticker and with the lid on, it is placed shaking slowly (100-150 rpm) for 2 days at 30 $^{\circ}\text{C}$.

Then, using a multi-pipette, 10 μl of cells of this plate are transferred to a new plate which has 90 μl of media in the same layout, which dilutes the cells a factor of ten. This is then repeated for a new plate, where 5 μl of cells are transferred to 95 μl media, to ultimately accomplish a 200x dilution compared to the original plate. When making the dilution, wells are gently mixed before dilution by pipetting up and down. The 200x diluted plate is then sealed by a transparent sticker, and covered with the lid to minimize evaporation (and water is placed in the reservoirs as before).

The diluted plate is placed in a TECAN Infinite 200 PRO for about 48 hours at 36 $^{\circ}\text{C}$ for measurements. The first 1000s linearly pre-shakes (amplitude 1 mm) the plate, before shaking just before each measurement round in a 380 second interval. These rounds start with 90" linear shaking (amplitude 2 mm), 90" orbital shaking (amplitude 1.5 mm), 90" linear shaking (amplitude 1 mm), waiting 10" and then OD-600 +/-9 nm measurements (25 flashes, 5ms settle time). Rounds were executed in two consecutive loops of 24 hours, after which OD values were documented in a excel sheet.

2.6.3.4.4 GUI

A graphical user interface (GUI) in *Matlab* (designed in R2014b) was made to facilitate the analysis of the many growth curves measured by the plate reader. The doubling times, which are most important outputs from the GUI, were found as follows. The output of the TECAN plate reader is in an excel format. These are read into Matlab with *xlsread*. As time loops run at most 24h, multiple loops were required to get the data in the time span of the experiments of Chapter 2. The time stamps of the second loop were offset by the mean time difference between the iterations of the first loop, as the second loop restarts at time stamp 0.

Then, the background of the signal must be subtracted, as there is an OD-value even for wells without cells (about 0.08), which differs slightly per well. For every well the average of ODs at the first ten time points provides an estimate of this basal background value. The instrument error is estimated as the standard deviation of these points. In the event that all values are the same (because of rounding of the data before storing the data in the excel file), the error is then taken to be $10^{-\text{digit}}/2$, where digit is the lowest number of figures after the decimal point for which rounding to this number does not alter the number. In other words, the error is the precision implied by the rounded number.

After background subtraction per well, the log (base 2) of the OD values were taken. Still, because of the noise in the machine, some values are below the background, of which the log

is not defined. Therefore, the part of the data is only retained which constitutes the longest stretch of points above the background with a user-specified signal to noise ratio. By default, this is twice the instrument error. Furthermore, the data stretch must also satisfy the user-specified lower- and upper-time bounds. By default, these are such that every window along which will be fitted (as explained in the next paragraph) can symmetrically accommodate a user-defined number of points.

Along this data stretch weighted least squares regression is done, which is explained in e.g., [200]. Normally in linear space the WLS weights are chosen inversely proportional to the instrument error. As an approximation in log space, the reciprocal of the difference between the log2 of the background subtracted values +/- the instrument error are taken. These fits are done along a moving, symmetrical window of a user-defined size, which by default is 51 points (must be uneven). The window centre location is chosen such that linear fit with the highest R-squared is found, provided it exceeds a user defined minimum (by default 0.9), otherwise no fit is given.

The WLS parameter fits, including the growth rates, with standard errors are then calculated per well, where the user can modify the user-defined fitting constraints per well. In >95% of the cases, the default settings work fine, so manual improvements are rare. This results in the growth rates given in 2.3.1. For the data in Appendix C.3, there is the slight modification that the OD of the stretch of data considered must also exceed 0.1. This was to find the advanced growth phase instead of the maximal growth phase, which better mimics conditions in the related volume experiment.

2.6.3.5 Statistical methods

Error bars in Figure 59, Figure 61 and Figure 62 are fitting uncertainty following from maximum likelihood estimation (*mle* in Matlab R2014b) from gamma distributions on flow cytometry data, see also Appendix E: Reconstruction of burst parameters from FACS data. At most 3000 iterations and 6000 function evaluations were allowed, with the lower bounds of parameters at zero and upper bounds at infinity, except the fraction of the off-state which is at most 1. Promoter off-state fraction values, combining replicates, were constructed by weighted averaging (using the errors in the weights).

For Figure 63 and Table 9, only data was considered for wells with the number of data points exceeding 1000. The median sample distributions follows from Laplace's result [201], where the density estimate at the median position was approximated by using Matlab's *ksdensity*, finding the first point at which more than half the probability mass is left of that point. The error on the standard deviation follows from 1000 bootstrap samples from the *bootstrap* function in Matlab. Correlation coefficients from the green and red channels including uncertainties resulted from *corrcoef*. Values combining for replicates were again constructed by weighted averaging. When reporting the relative medians, standard deviations and correlation coefficients, errors were processed through propagation of uncertainty [202], assuming no cross-correlations.

Volume medians and standard deviations in Figure 65 and Figure 66 were estimated setting all counts in the measured bins in the Coulter Counter at the centre of the bins. Uncertainties were somewhat crudely estimated using the *bootstrap* function, using 100 samples. Only data from strain/medium combinations that (with all runs combined) has more than 500 counts in total were considered, as some are background counts. Again, when combining replicates, weighted averages are taken, and propagation of uncertainty [202] assuming no cross-correlations yielded the errors of the relative values.

The growth rate data from Figure 18 and Figure 19 were generated as through a GUI as described in the previous subsection 2.6.3.4.4. When a control well (only medium no cells) exhibited growth, all samples in the same medium in that run were omitted from analysis (discarding one column and one row out of 4 runs). Data was further processed by observing that wells with and without growth could typically be very well discriminated by having an OD rise above or below 0.02. This helped to objectify the judgements in the difficult cases where growth is marginal.

By weighted averaging, runs of the same background and medium were combined. As clearly the fitting errors were underestimating the error across replicates, all the errors were increased by the same modified Birge ratio [190], estimated in pure WT, averaged across media. The distribution type of the errors remained the same, so this means the growth rate values minus their means, divided by the corrected errors are still distributed as a Student's t-distribution with the fitting window size minus 2 as the degrees of freedom, as before [200].

We then refine this estimate with Bayesian analysis, defining the likelihood of the mean value as the product of the t-distributions of the points. When multiplied by the uniform prior between a rate of 0 and 1/70 [1/min.], a distribution proportional to the posterior is obtained. The posterior samples can then be retrieved by use of the Metropolis-Hastings algorithm [193], implemented in Matlab as *mhsample*. A normal distribution centered around the mean with 10 times the corrected error was used as the proposal distribution, while the random number generator makes numbers uniformly across the range [sampling point \pm half the WT error at 0% galactose]. A burn-in period of 1000 points was used, leaving 50000 sample points from the posterior distribution.

Monte Carlo simulation of the ratio distribution (10000 samples) of the rates divided by the pure WT rates, and then taking the mean of those for the estimate of the mean and the 0.16 and 0.84 quantiles, gave the corresponding 68% credible intervals shown in Figure 19. Similarly, Monte Carlo simulation of the difference distributions (10000 samples) between two rates gave probability estimates of one rate being larger than the other (number of times simulated data is greater or smaller than zero). The highest chance of the scenarios 'value 1 is larger than value 2' and 'value 1 is smaller than value 2' is taken, and those odds are documented in Table 1. The convention for strong evidence found in [192] is adhered.

3 Genotype to phenotype model

3.1 Genotype, phenotype, ... mesotype

The previous chapter showed the experimental validation of a detailed molecular model for the polarization mechanisms in budding yeast, the case study for self-organization in this dissertation. In particular, growth rate measurements on polarity mutants were generally in line with expectations on their viability as a function of Cdc42 concentration, which is central to polarity success. However, we already observe some shortcomings if we reset our goal to predict properties as growth rate more quantitatively. Even in a clonal population, polarity success turns out to be subject to stochasticity. When genotypes do not behave so discretely, this complicates our understanding of evolutionary targets in the genome, and the efficacy of evolution.

Given the level-based idea from the introductory chapter 1 to make trait generation of evolving, self-organizing systems tractable (as visualized in in Figure 4), the need of another intermediate level becomes apparent. Whereas the molecular mechanisms for self-organization emerge from the genotype, that can be modified through evolution, the selection criteria for evolution are phenotypes contributing to fitness, on for example a cellular, population, or organismal level. This means that to understand how evolution can shape the genotype as a function of phenotype, first the path from genotype to phenotype must at least to some extent be understood.

In itself, the prediction of phenotypes from genotypes is a central problem in many fields, such as medicine [203], [204], synthetic biology [205] and forensics [206]. This has by no means a trivial solution; for example the common disease-common (genetic) variant hypothesis as tested by genome-wide association studies typically does not hold [207], [208], partly due to ubiquitous epistasis [209] and pleiotropy (although one gene affecting many traits is not universally found, see [30]). More comprehensive modelling is hence necessary, with methods ranging from statistical/phenomenological [210], [211] to biophysical [212].

Much recent interest has risen in statistical methods, benefitting from the increasing availability of large genotype/phenotype data sets (e.g., [213], [214]). The downside is that this 'big data' approach will not provide mechanistic understanding of the emergence of phenotypes and subsequently, not of potential new evolutionary paths either, contrary to the potential of the biophysical approach [215], [216]. However, the latter approach in turn requires much detail, which together with the difficulty to interpret outcomes in more general terms is very (computational) time consuming. Furthermore, there is no guarantee that reverse-engineering genotype to phenotype maps is possible when this map is much

more complex than one-to-one. Hence, conceptual progress is required to make a justifiable bottom-up model to predict phenotypes feasible.

It has been suggested in e.g., [217] that introducing intermediate levels between genotype and the trait of interest could improve tractability of the prediction process, and yield general insights [22]. Several types of levels have been defined in literature, e.g., endophenotypes [218], ontotypes [219]. While used with success in situations when pure predictability of phenotypes is demanded, previously posed questions regarding evolution and evolvability remain unanswered due to a lack of explicitly incorporating underlying molecular mechanisms. As an illustration of how combining interpretable parameters and a biophysically sound justification remains a shortcoming, the principles behind endophenotypes and ontotypes are briefly described here (see also Figure 21 for a schematic overview).

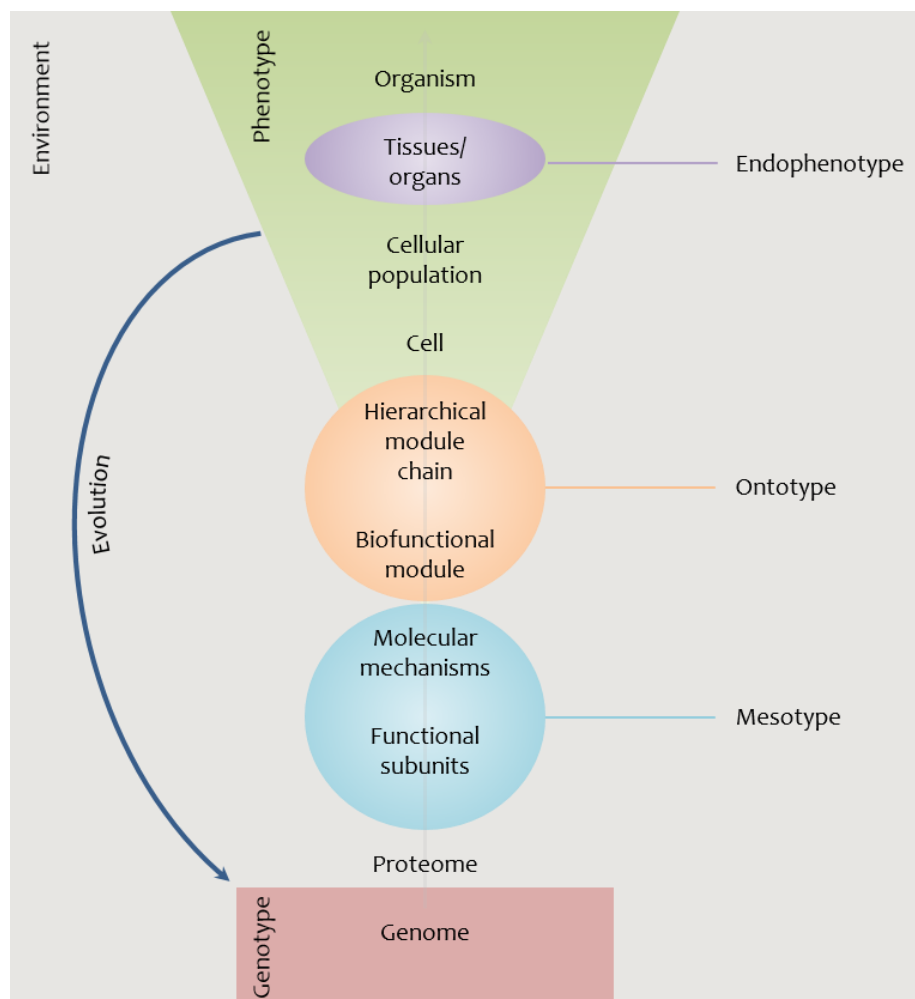


Figure 21 Schematic overview of the generation of phenotypes, starting from the bottom-up from the genotype/genome (red box), in a particular environmental background (grey). Evolution (dark blue arrow) couples phenotype (green cone, depicting the expansion of possible phenotypes across higher-order structures) back to genotype. Light blue, orange and purple ellipses depict the approximate realms of mesotypes, ontotypes and endophenotypes respectively.

Endophenotypes are now in common use for neurological diseases, as an observable internal risk factor that correlates with (future) diseases. Diverse examples exist, leading to diffuse definitions that are not always met [220], and its interpretations as risk factors or very early manifestations of the disease itself are also up for debate [221]. The ontology comprises a state defined by the effect on the hierarchical chain of GO annotations (starting from the genes of interest) upon which phenotypes can be predicted through machine learning. While quite a successful method for predicting growth rates when ample data is available, this statistical scoring method explains the emergence of phenotypes from genotypes only in terms of biofunctional classifications. While a part of the biophysics involved will have been integrated through the GO annotations, phenomena granting insight in the question why phenotypes occur are bypassed.

As a solution, in [70] a chemically justifiable approach is taken that modularly encompasses three main processes in budding yeast: metabolism, the cell cycle and growth. This makes good use of the functional grouping tendencies [31] in *S. cerevisiae*, yielding a coarse-grained model that remains tractable. Yet, while the underlying Michaelis-Menten dynamics may be a sufficiently accurate approximation in terms of metabolites, this relies on well mixed reagents and a negligible role for diffusion. In a crowded cell this is often not the case (as explained in e.g., [222]), and such an approach also ignores the coupling between reaction and diffusion that can generate pattern formation as described in [223]. There is still a need for a multi-level modelling approach that coarse-grains the spatio-temporal features of cellular processes in a way that applies to a broader class of cellular modules.

In Chapter 2, the experimental side of [51] has been expounded, validating a theory on the level 'above' genotype in *S. cerevisiae*. The analysis of the mass-conserving reaction-diffusion equations governing symmetry-breaking during cell polarity revealed a grouping of protein interactions into "functional subunits". Combinations of these constitute symmetry breaking mechanisms with well-defined properties, such as the requirement of a minimum Cdc42 concentration in order to polarize. Considering the generality of pattern formation through these equations (see [224], [225]), it seems an appropriate emergent level upon which to build a model attempting to describe higher order, population phenotypes.

That this is not straightforward is illustrated as follows. Theoretically, the boundary between insufficient and sufficient Cdc42 to polarize in a $\Delta bem1$ mutant (using the rescue mechanism) is sharp; above this boundary polarization is fast, below this it is not possible. Paradoxically, the experimental fitness of the $\Delta bem1$ mutant with WT Cdc42 expression, which is expected to be around this boundary, is neither high nor zero, but very low instead. This genotype cannot be mapped onto the appropriate phenotype yet. But coarse-graining principles derived from the functional subunits can be integrated into a higher-order, population dynamics model to complete the mapping, which will be the content of this chapter.

To avoid confusion, we define the set of coarse-grained rules under a new name, the *mesotype*. The goal is to integrate this in a population model; its validity will then be tested on

experimental data of the model system of choice, yeast polarity. The goal is to describe and predict phenotypes of polarity mutants, such as cell volumes, doubling times and cell cycle times. This approach is applicable to other systems provided sufficient knowledge is present of the underlying protein networks to characterize the resulting physical mechanisms. Once this modelling approach to bridge the gap from genotype to phenotype is validated, we zoom out to discuss general lessons on how evolution influences this bridge. For example, the tractable nature of the calculations reveals the hypothesis that transcriptional noise appears to be a driver of epistasis emergence. These discussions form the foundation for understanding the evolutionary coupling to this self-organizing system in Chapter 4.

3.2 Mesotypical integration into a growth model

To understand how the mesotype affects population phenotypes, or even a single cell throughout its life, we must model its coupling to other existential processes in the cell. For the purpose of tractability, these are summarized into a model, after having identified the main phenomena that should be taken into account. As we saw in the previous chapter that the ability to polarize depends strongly on the amount of Cdc42, an obvious inclusion is a sound description of the dominant factors affecting concentration. This leads to three branches in the population growth model tree that incorporates polarity: cell growth, polarization dynamics via coarse-grained rules (mesotype), and protein turnover.

One of the strengths of the model is that different polarity mutants can be represented in the simulations by simple changes in parameters, as will be described in the next subsections. All GAP mutants in [24] require four fitting parameters, the *nrp1* mutant one and extension to incorporate the *CLA4* mutation one more. The resulting model keeps track of the Cdc42 protein concentration $[Cdc42]$ that varies stochastically across time and population. This in turn influences the outcome of the coarse-grained polarity rules, which influences the growth phases of the cell. Conversely, cell growth affects the protein concentration as well, leading to interesting behaviour that would have been difficult to predict beforehand.

Figure 22 shows a graphical summary of the integrative cell growth model, while Table 4 provides a summary of the parameters. Details on the model implementation in Matlab can be found in 3.7.1.

3.2.1.1 Cell growth

In the cell cycle of yeast, two types of growth can be identified: isotropic growth, which roughly corresponds to G1 phase and where only the mother cell grows, and polarized growth, which roughly corresponds to S, G2 and M phase and where only the bud grows. There is no consensus in literature (see e.g., [226]) on how to characterize these growth

phases. More specifically, authors in [227] identify four regimes wherein volume scales linearly with time t , although one may have chosen to fit only two regimes of constant membrane area growth rate on the data. On the other hand, two regimes of approximately constant volume growth are found in [228], which are then integrated in an detailed model for growth.

Hence, in absence of a standardized model and as the exact nature of the growth is likely to be condition or even strain dependent, a parsimonious model is chosen here. However, as demonstrated in Figure 77 of Appendix I and the description of epistasis in a more minimalistic model (Chapter 4), the exact nature of the growth assumptions with the exception of the mesotype coupling will prove cosmetic for the most important conclusions.

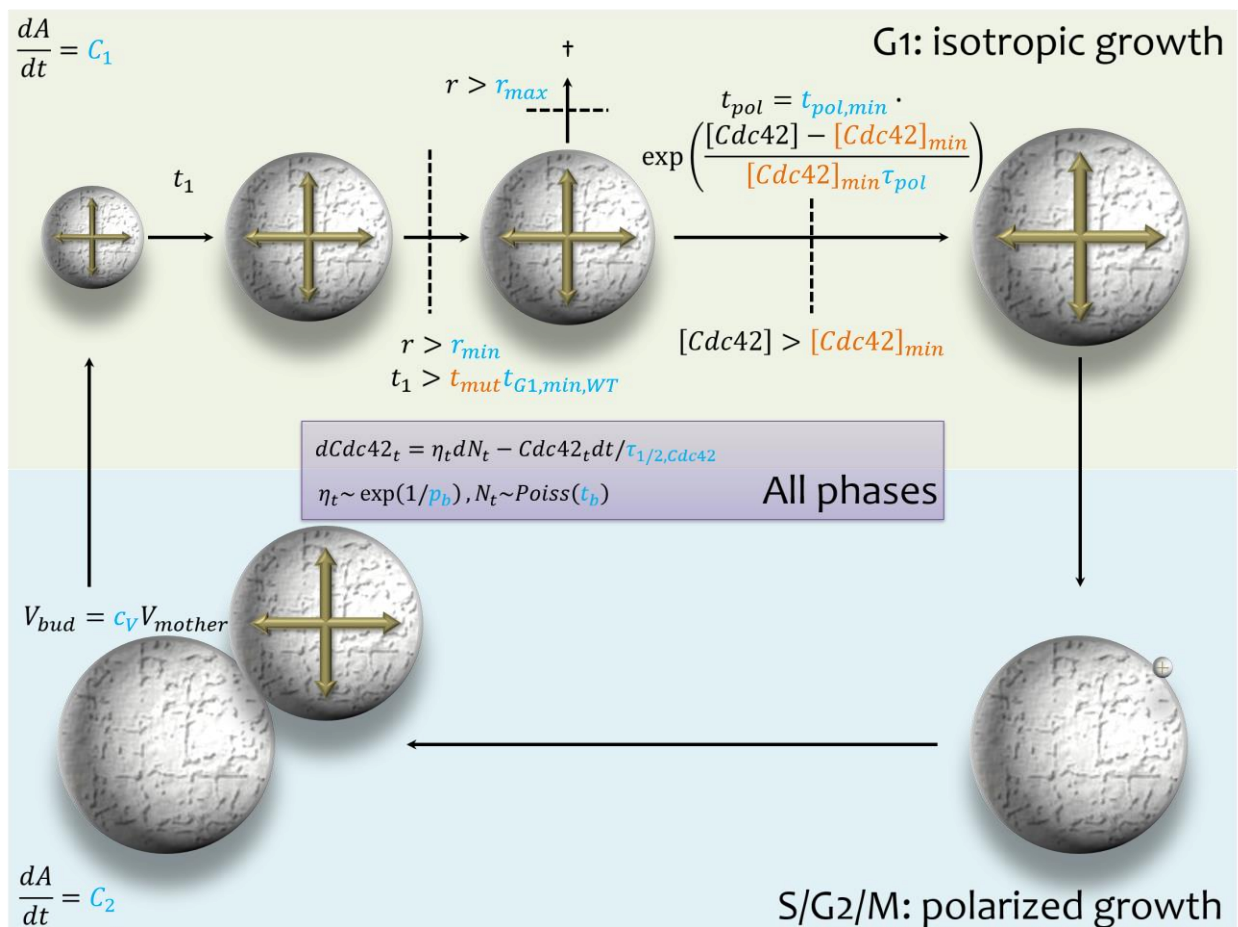


Figure 22 Graphical overview of integrative growth-polarization model with stochastic Cdc42 production, with parameter values in light blue and fitting parameter values in orange. Top half represents the isotropic growth phase, the bottom half the polarized growth phase.

We start by assuming a constant cell membrane area ($A = 4\pi r^2$ with cell radius r) growth rate dA/dt (see left corners of Figure 22). This rate follows the commonly observed trend, that the polarized growth (rate C_2) occurs faster than the isotropic growth (rate C_1). This factor will be between 2 and 3, inferring from fig. 7 of [227] and fig. 1c of [228] respectively.

From analytical work, we can also get a fair estimate for the value of membrane growth rate during isotropic growth, if it is assumed that WT is relatively optimized such that for optimal media, the minimum size to polarize r_{min} is not an important factor. Then, lower rates slow WT down, but higher rates do not confer much of an advantage in terms of doubling time. In this case, the rate C_1 is defined as (see derivation in Appendix D: Membrane growth rate considerations for calibration):

$$C_1 = \frac{4\pi \left(1 - 0.7^{\frac{2}{3}}\right) r_{min}^2}{t_{G1,min} \left(1 + 0.7^{\frac{2}{3}} \frac{t_{pol,min}}{t_{G1,min}}\right)} \quad (3.1)$$

where $t_{pol,min}$ and $t_{G1,min}$ are the minimal polarization time and time spent in G1 respectively. This value of C_1 is independent of the exact value of C_2/C_1 , which is determined from calibrating the speed of growth by the wild-type doubling time (taken to be 83 min. in YPD, from data of [24]). This restricts the model sufficiently to make deviations from this idealized behaviour inconsequential (see Appendix I: Relaxation of model assumptions).

Now that isotropic and polarized growth are defined, the transition points between the two phases must be set. Bud emergence, the start of polarized growth, must be preceded by polarization, which still occurs during isotropic growth. In turn, polarization cannot occur under certain circumstances, which can be absorbed into the following three conditions: i) there exists a minimum cell size to polarize, ii) there is a minimum time already spent in G1 before polarization, iii) polarization proteins must be at appropriate concentrations.

The first condition will only be an average resulting from a distribution, with 2 μm for the minimum radius as an appropriate value given data of [228]. The second condition aims to summarize gene expression cues for cell cycle progression. As explained in 1.4.4, the timing of the polarization event requires a cascade of upstream effects, which will require a minimum time ($t_{G1,min}$) for completion, which the time spent so far in G1 (t_1) must exceed. For WT, the mean mother G1 time value of 15.6 minutes from table S7 of [228] is taken as a proxy. Still, this $t_{G1,min}$ may be affected by certain mutations. In particular, we phenomenologically model the $\Delta nrp1$ mutation to reduce this minimum time, as reasoned in Box 4, by a factor t_{mut} . *The inclusion of Nrp1 hence requires one parameter.*

Finally, the last condition reflects the symmetry breaking model from [51] validated in chapter 2. Essentially, this links the most important hallmark of the functional subunits observed in budding yeast [51] (recall the polarity cone of Figure 16) in a coarse-grained fashion; a minimum Cdc42 concentration ($[Cdc42]_{min}$) must exist before pattern formation can succeed and downstream effectors are recruited. This minimum will depend on genotype, most notably on GAP genes present for the rescue mechanism, and on GEF, Bem1 and GAP concentrations in a more relaxed way in the WT mechanism (see 3.2.1.2). Therefore, the minimum will be a genotypic fitting parameter to let mutant doubling times match experimental results.

After all conditions are fulfilled, the cell proceeds to polarized growth after a polarization time specified in the next subsection. If these conditions are not met before the cell exceeds the prespecified radius r_{max} , the cell is pronounced dead. This is motivated by e.g., movies of [24], where $\Delta bem1$ cells are not capable of halting growth if polarization is delayed, ultimately leading to their demise. Here, a feasible maximum size of 6 μm is taken.

Having specified the conditions to complete the transition to polarized growth, after some time the reverse transition must be made. Again, the exact details of the timing for this event are debatable. From [228], daughters are predicted to have a constant, small starting volume V (25 μm^3), regardless of the mother size. By contrast, movies of the for this model critical $\Delta bem1$ mutant appear to show a different picture, with large cells giving birth to large daughters [24].³ Interestingly, a simple daughter volume (V_{bud}) scaling of 70% ($c_V = 0.7$) of the mother volume (V_{mother}) also yields a volume reasonable consistent with the aforementioned 25 μm^3 daughter volume when assuming a typical young mother with a radius of about 2 μm (still close to the minimum to bud r_{min}). Therefore, by default the volume scaling is chosen, although the algorithm accommodates the fixed daughter size option. Finally, once returned to isotropic growth, the bud is defined as a new cell, inherits all proteins proportional to its volume with respect to the mother and the cycle can restart.

3.2.1.2 Coarse-grained polarity

The polarization times that fall in the isotropic growth phase form together with the aforementioned Cdc42 concentration ($[\text{Cdc42}]$) threshold to polarize, the two links between growth and mesotype. In short, the polarization time t_{pol} depends exponentially on the relative Cdc42 amount in excess of the threshold. This is a simplified approximation for tractability of more thorough reaction-diffusion equation computations [51]. How such a simple functional form for t_{pol} encompasses the molecular information sufficiently can be motivated as follows, also considering Figure 23 and Figure 24.

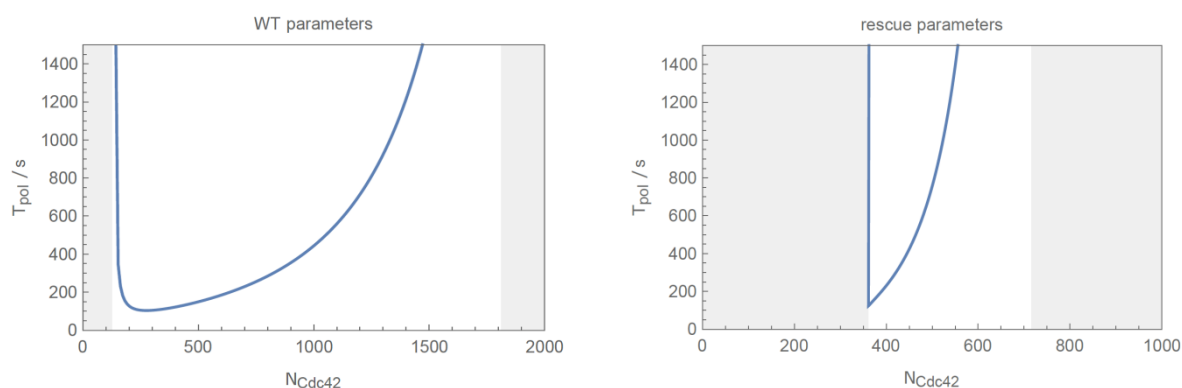


Figure 23 Figure courtesy of Fridtjof Brauns. Qualitative plot for polarization times with respect to Cdc42 copy number (based on parameters of [68]), for the Bem1 (left) and rescue mechanism (right).

³ It may also be the case that daughter size is genotype dependent. It should be noted that both *BEM1* and *BEM2* have been implicated in G2/M morphogenesis checkpoint regulation [88].

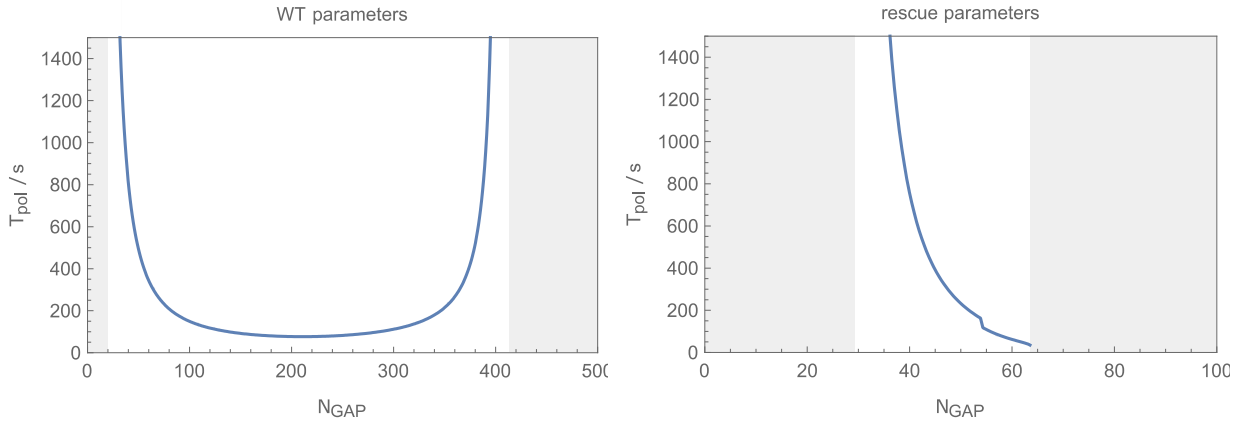


Figure 24 Figure courtesy of Fridtjof Brauns. Qualitative plot for polarization times with respect to GAP copy number (based on parameters of [68]), for the Bem1 (left) and rescue mechanism (right).

What is apparent is that regardless whether the Bem1 or rescue ($\Delta bem1$) mechanism is active, a minimum Cdc42 concentration always exists below which polarization is impossible. There is no strict upper bound on Cdc42, although polarization will slow down. Conversely, there is a maximum GAP concentration, and there is almost no lower bound. Again, polarization may still slow down for very low GAP concentrations. For any GAP concentration, the minimum Cdc42 concentration is always smaller for the Bem1-mechanism than for the rescue mechanism. Similarly, the maximum GAP concentration is always higher for the Bem1-mechanism than for the rescue mechanism (see Figure 16 of section 2.1).

For various reasons, the Cdc42-axis is the most important one to retain in detail for the model. Firstly, the experiments conducted in this dissertation only take large, discrete steps in GAP space, whereas Cdc42 concentration is gradually changed. Secondly, we learned on the importance of noise throughout this dissertation for the fitness of strains, which is much lower for the GAPs than it is for Cdc42 [229], see also the discussion in 3.2.1.3.1. Finally, it is convenient to reduce dimensionality and complexity, both for computational as for interpretability reasons. Therefore, the behaviour of numerical, reaction-diffusion computations can be summarized as follows:

$$t_{pol} = \begin{cases} \infty & , \quad [Cdc42] < [Cdc42]_{min} \\ f\left(\frac{[Cdc42]}{[Cdc42]_{min}} - 1\right) & , \quad [Cdc42] \geq [Cdc42]_{min} \end{cases} \quad (3.2)$$

where $f\left(\frac{[Cdc42]}{[Cdc42]_{min}} - 1\right)$ is some function of the relative excess Cdc42 above the minimum. $[Cdc42]_{min}$ is linearly dependent on GAP concentration, only the slope is shallower for the Bem1 mechanism than for the rescue mechanism and zero for $[GAP] = 0$. Hence, $[Cdc42]_{min,WT} < [Cdc42]_{min,\Delta bem1}$ and because of the cone structure, the following identities hold:

$$[GAP]_{BEM2BEM3} = a_w [Cdc42]_{min,WT} = a_r [Cdc42]_{min,\Delta bem1}$$

$$[GAP]_{\Delta bem2BEM3} = a_w [Cdc42]_{min,\Delta bem2} = a_r [Cdc42]_{min,\Delta bem1\Delta bem2}$$

$$[GAP]_{BEM2\Delta bem3} = a_w [Cdc42]_{min,\Delta bem3} = a_r [Cdc42]_{min,\Delta bem1\Delta bem3}$$

$$[GAP]_{\Delta bem2\Delta bem3} = a_w [Cdc42]_{min,\Delta bem2\Delta bem3} = a_r [Cdc42]_{min,\Delta bem1\Delta bem2\Delta bem3}$$

$$[GAP]_{\Delta bem2\Delta bem3} = [GAP]_{BEM2BEM3} - ([GAP]_{BEM2BEM3} - [GAP]_{\Delta bem2BEM3})$$

$$-([GAP]_{BEM2BEM3} - [GAP]_{BEM2\Delta bem3}) = [GAP]_{\Delta bem2BEM3} + [GAP]_{BEM2\Delta bem3} - [GAP]_{BEM2BEM3}$$

where a_w and a_r are the conversion factors between GAP and minimum Cdc42 concentration thresholds for the WT ($BEM1$) and $\Delta bem1$ background respectively. So then:

$$[Cdc42]_{min,\Delta bem2\Delta bem3} = [Cdc42]_{min,\Delta bem2} + [Cdc42]_{min,\Delta bem3} - [Cdc42]_{min,WT} \quad (3.3)$$

$$[Cdc42]_{min,\Delta bem2\Delta bem3} = [Cdc42]_{min,\Delta bem2} + [Cdc42]_{min,\Delta bem3} - [Cdc42]_{min,WT} \quad (3.4)$$

$$\frac{[Cdc42]_{min,WT}}{[Cdc42]_{min,\Delta bem1}} = \frac{[Cdc42]_{min,\Delta bem2}}{[Cdc42]_{min,\Delta bem1\Delta bem2}} = \frac{[Cdc42]_{min,\Delta bem3}}{[Cdc42]_{min,\Delta bem1\Delta bem3}} \quad (3.5)$$

The final results are four identities, with the latter two ratios being smaller than unity. Having eight unknowns, this shows that *only four parameters are required to set the thresholds for all the GAP mutants from [24]*.

Looking at the single Cdc42 dimension for the Bem1 and rescue mechanisms, we see that polarization times look like an exponentially increasing function of excess Cdc42. This suggests defining polarization time as the following functional form:

$$t_{pol} = \begin{cases} \infty & , \quad [Cdc42] < [Cdc42]_{min} \\ t_{pol,min} \exp \left[\left(\frac{[Cdc42]}{[Cdc42]_{min}} - 1 \right) / \tau_{pol} \right] & , \quad [Cdc42] \geq [Cdc42]_{min} \end{cases} \quad (3.6)$$

The minimum polarization time $t_{pol,min}$ can be estimated for WT from experimental data [230] (fig. S1B) to be about 5 minutes. It seems reasonable to assume a similar “top speed” for the rescue mechanism with perfectly balanced concentrations, given the similar performance of the rescue and WT mechanism at the end of an evolution experiment [24].

The scaling parameter τ_{pol} represents the sharpness of the exponential, i.e. how much the polarization time suffers from having excess amounts of Cdc42. This is always in relation to the GAP concentration. For both mechanisms in Figure 24, behaviours are quite similar in the

low GAP-regime, with the lower bounds seemingly on the same order of magnitude. There, $t_{pol} \approx t_{pol,min} \exp\left(\frac{[Cdc42]}{[Cdc42]_{min}\tau_{pol}}\right)$, implying $\frac{[Cdc42]}{[Cdc42]_{min,WT}\tau_{pol,WT}} \approx \frac{[Cdc42]\tau_{pol,rescue}}{[Cdc42]_{min,\Delta bem1}\tau_{pol,rescue}}$. We find that a $[Cdc42]_{min}$ of an order of magnitude lower for WT than for the rescue background to be self-consistent. Therefore, τ_{pol} will be an order of magnitude higher for WT, yielding the expression:

$$\tau_{pol,WT} \approx \frac{[Cdc42]_{min,\Delta bem1}\tau_{pol,rescue}}{[Cdc42]_{min,WT}}$$

Furthermore, experimental data from section 2.3 does not show a strong upward trend in doubling times when overexpressing Cdc42 about an order of magnitude in a rescue background, making a value of e.g., $\tau_{pol,rescue} = 25$, consistent with the data.

3.2.1.3 Protein turnover

As shown in the previous two subsections, $[Cdc42]$ is a critical quantity. In cells, concentrations such as that of Cdc42 are not at all constant. On the one hand, the processes of DNA transcription and subsequent RNA translation to protein contain steps that use finite resources (such as RNA polymerases) and take variable amounts of time (see e.g., in humans [231]). On the other hand, protein degradation is also not a fully deterministic process, occurring at variable speeds (see e.g., [232]). Hence, protein copy number noise in the population and across time is quite common (e.g., [233]). Even cells with the same genes and the same size in the same cell cycle phase have fluctuating amounts of Cdc42.

To adequately model its time-dependence and cell-dependent variations, assumptions on protein production and degradation need to be made. The latter is relatively straightforward. The WT estimate for average number of Cdc42 proteins is considerable, about 8700, only one order of magnitude lower than Act1 [234]. Assuming no cell cycle dependent degradation targeting, there will be no relevant stochastic effects in terms of degradation, so the Cdc42 decay will have the following simple form:

$$\frac{d[Cdc42]}{dt} = -\frac{[Cdc42]}{\tau_{1/2,Cdc42}} \quad (3.1)$$

The recent literature value for Cdc42 protein half-life $\tau_{1/2,Cdc42}$ is about 8 hours [235], as measured by non-radioactive amino acid labelling.

On the other hand, modelling production requires more attention. Given the short Cdc42 mRNA half-life (10-12 minutes [236]) compared to its protein half-life, it may be possible to integrate out the mRNA level in the gene to protein process as shown in [189]. After all, from the point of the protein production time is instantaneous compared to its overall life, so production occurs in effective bursts. Further assuming memory-less properties for

polymerase and ribosomal arrivals on DNA and mRNA respectively (they occur independently from each other) cause the times between bursts and the burst sizes to be exponentially distributed.⁴ Both a rigorous derivation and an intuitive explanation for this is found in [237]. This results in a gamma distribution for protein copy number across the population [189].

Remarkably, it is possible to infer microscopic details such as average time between bursts and average burst size from bulk measurements on a population of yeast cells. Concretely, the two parameters of the gamma distribution, inferable from flow cytometry measurements of a fluorescent version of the protein of interest, relate in a straightforward manner to these burst characteristics. Given this gamma probability density function for the concentration of protein c_p across the population:

$$f(c_p; k, \theta) = \frac{c_p^{k-1} e^{-c_p/\theta}}{\theta^k \Gamma(k)} \quad (3.2)$$

the shape parameter k and scale parameter θ can be viewed in light of average time between bursts $\tau_{b,av}$ and average burst size $p_{b,av}$, using [189]:

$$\tau_{b,av} = \frac{\tau_{doubl}}{k} \quad (3.3)$$

$$p_{b,av} = c_{sc} \theta \quad (3.4)$$

with c_{sc} as a scaling constant⁵ related to an average volume, connecting protein number to concentration distributions, required to interpret θ in terms of burst size.

It may also be the case that Cdc42 expression is regulated, for example by temporal regulation and/or positive/negative feedback. While from [189] it is even then still possible to infer average burst times and sizes, it complicates the matter for this model as e.g., it would require investigating at what point in the cycle temporal regulation is occurring. Fortunately, as shown in the analysis of Appendix E, the population protein distribution is fairly accurately described by a single gamma distribution, although a bit better by a double gamma. This

⁴ An analogous mathematical representation is given in the center of Figure 22. Within a time-increment dt , the number of bursts that are exponentially spaced in time is Poisson distributed. This means production is mathematically described by stochastic compound Poisson process $\eta_t dN_t$, where η_t reflects the exponentially distributed sizes of bursts given by Poissonian jump process N_t . In total, the process for Cdc42 copy number $Cdc42_t$ consists of this production term, minus the drift term representing degradation and scaling as $1/\tau_{1/2,Cdc42}$.

⁵ In [189], c_{sc} is straightforward to calculate for bacterial growth displaying exponential growth of cell volume during the cell cycle, namely $1/\ln 2 \approx 1.44$. For the integrative growth model in yeast, the factor c_{sc} of 0.8133 was needed to center the protein copy number distribution at 8700 copies [234] for observed k and θ , following data processed as in Appendix E. This multiplicative constant does not affect the gamma nature of the resulting distribution, as θ is its scale parameter.

suggests minor regulation in the form of a switch-like expression where a transcription factor can assist or repress transcription mildly.⁶ However, the single gamma approximation appears sufficiently good to neglect this minor regulation.

The analysis of that data yielded burst property values of $\tau_{b,av} = 57$ min. and $p_{b,av} = 4940$ proteins. It should be noted that these are effective numbers, resulting from integrating out the mRNA step from DNA to protein. Presumably this is the result of a burst of transcription yielding a few dozen mRNAs that survive on average 10-12 min., although it has been noted for actin that every mRNA molecule seems to get transcribed about a 1000 times [238].

With protein production defined, the final criterion for the isotropic to polarized growth transition can be executed. When this condition is not fulfilled, growth continues isotropically until the next burst of transcription, at which point this condition is re-evaluated. This suits an implementation of the Gillespie algorithm [239], which runs until the condition is fulfilled, or until the cell is too large. As aforementioned, the critical radius is 6 μm , although it is possible for a cell to momentarily exceed this size during the isotropic growth in the next iteration, as the volume check is only done right before polarization.

3.2.1.3.1 Note on exclusion explicit GAP turnover

It is possible to justify the omission of stochastic GAP expression by looking at literature. The data of [229] gives the mean μ and standard deviation σ of the fluorescent intensity distributions of many protein, including all GAPs. This can be translated to coefficient of variations V ($=\sigma/\mu$). This leads to an estimate for the average number of bursts per cell cycle $k = 1/V^2$, which is 62, 46, 54 and 53 for *BEM3*, *BEM2*, *RGA1* and *RGA2* respectively. This is more than an order of magnitude larger than for *CDC42* (for endogenous Cdc42 this value inferred from flow cytometry was 3.48, see Table 4) and equates to bursts on a minute time scale, effectively yielding constitutive expression. The distribution of the Cdc42/GAP ratios will then also be dominated by the stochasticity from *CDC42*, retaining approximately the same form as *CDC42* alone (see Appendix F: Effect of stochastic GAP production on Cdc42/GAP ratio). The protein number scale (parameter) is affected, shifting all minimum concentrations by the same factor, but relative GAP strengths are expected to stay the same. Furthermore, it can be shown that one source of stochasticity suffices to explain e.g., the epistasis amongst GAPs, in a minimal model with fairly robust relative GAP contributions, see Figure 39. Hence, only stochasticity in Cdc42 production is considered here for the sake of simplicity.

⁶ The poor functionality of the measured N-term GFP fusion of Cdc42 excludes the possibility that fitness effects warp the protein distribution as in Chapter 4 to provide an alternative explanation for the deviation from the gamma distribution.

Table 4 Growth model parameter summary, with symbol, description, value and source, color coded by category: orange for fitting parameters, red for polarization parameters, green for growth parameters, blue for protein turnover parameters and grey for technical parameters for simulations.

Parameter	Description	Value	Reason/source
$[Cdc42]_{min,\Delta bem1}$	Minimum Cdc42 concentration to polarize for $\Delta bem1$ background	Positive number	Fitting parameter
$[Cdc42]_{min,\Delta bem1\Delta bem3}$	Minimum Cdc42 concentration to polarize for $\Delta bem1\Delta bem3$ background	Positive, smaller than $[Cdc42]_{min,\Delta bem1}$	Fitting parameter
$[Cdc42]_{min,\Delta bem1\Delta bem2}$	Minimum Cdc42 concentration to polarize for $\Delta bem1\Delta bem2$ background	Positive, smaller than $[Cdc42]_{min,\Delta bem1\Delta bem3}$	Fitting parameter
$[Cdc42]_{min,WT}$	Minimum Cdc42 concentration to polarize for WT (<i>BEM1</i>) background	Positive, smaller than $[Cdc42]_{min,\Delta bem1}$, close to minimum in $(t_{doubl}, [Cdc42]_{min})$ space	Fitting parameter
t_{mut}	Mutant and WT minimum G1 time ratio	1 for <i>NRP1</i> , between 0 and 1 for the $\Delta nrp1$ background	Fitting parameter
$t_{pol,min}$	Minimum time to polarize	5 [min.]	[230]
$\tau_{pol,rescue}$	Scale parameter rescue fitness landscape	25	Consistent with data and reaction-diffusion equation solutions using parameters from [68]
$\tau_{pol,WT}$	Scale parameter WT fitness landscape	500 (very roughly $\frac{[Cdc42]_{min,\Delta bem1}\tau_{pol,rescue}}{[Cdc42]_{min,WT}}$)	Consistent with data and reaction-diffusion equation solutions using parameters from [68]
$t_{G1,min,WT}$	WT minimum G1 time before polarization phase	15.6 [min.]	[228]
c_V	Daughter to mother volume ratio	0.7	Consistent with [228]
c_p	Ratio between polarized and isotropic membrane growth rates	2.1324	WT calibration

C_1	Isotropic membrane growth rate	$\frac{4\pi(1 - c_V^{2/3})r_{min}^2}{t_{G1,min,WT} \left(1 + c_V^{2/3} \frac{t_{pol,min}}{t_{G1,min,WT}}\right)}$	Analytical work assuming optimized WT
C_2	Polarized membrane growth rate	$c_p C_{1,WT}$	Assumption consistent with [227]
T_{WT}	Target doubling time for WT calibration	83 [min.]	[24]
$T_{WT,FACS}$	Doubling time for the WT measured in the FACS	200 [min.]	See 3.6
P_{WT}	Target average Cdc42 copy number for WT calibration	8700	[234]
$r_{min,WT}$	Minimum radius before polarization phase	2 [μ m]	[228]
k_{endo}	Endogenous Cdc42 shape parameter (mean number of bursts / cell cycle [189])	3.48	Inferred from FACS (see Appendix E)
$k_{endo,WT}$	Endogenous Cdc42 shape parameter, rescaled to target WT	$k_{endo} \frac{T_{WT}}{T_{WT,FACS}}$	Inferred from FACS
$t_{b,WT}$	Average time between Cdc42 bursts	$T_{WT}/k_{endo,WT}$	Inferred from FACS and [189]
$p_{b,WT}$	Average size of Cdc42 bursts	$0.8133 P_{WT}/k_{endo,WT}$	Inferred from FACS and [189], corrected by WT calibration
$\tau_{1/2,CDC42}$	Half-life Cdc42 protein	$7.9 \cdot 60 / \log(2)$ [min.]	[235], scaled to have exponential base
N_{init}	Number of cells at the start of the simulation	1000	For doubling time convergence
r_{init}	Size of the cells at the start of the simulation	$1.1 r_{min,WT}$	For doubling time convergence
P_{init}	Number of Cdc42 proteins at the start of the simulation	0	For doubling time convergence
N_{target}	Minimum colony size to reach	$5 \cdot 10^6$	Sufficiently large for convergence within 1-2%

In the next subsection, the model will be used in the following ways: Firstly, it is put to the test by attempting to describe the experimental data of 16 polarity mutants from [24], which represent all possible genetic variants with *BEM1*, *BEM2*, *BEM3* and *NRP1* either present or deleted. This experimental data encompasses volume distributions, doubling times and cell cycle times. Additionally, the model is extended towards several polarity mutants where the experimental focus is on doubling times; 4 Δ *cla4* strains and three strain backgrounds where the endogenous *CDC42* is replaced by one under the variable GAL1-promoter. Finally, we explore the limits of the model by extending beyond polarity, to determine the minimum information required to make sensible predictions on epistasis.

3.3 Model validation

With the mesotypical inclusion into the cell growth model in place, it must be determined how useful this model is in describing polarity phenotypes, starting from genotypes. In particular, can this model reproduce observed features that are otherwise difficult to explain? For this purpose, we match the simulations to diverse experimental data of polarity mutants across the board; those range from supposedly well understood (e.g., GAP) to enigmatic (e.g., *nrp1*) mutants, from strong (deletions) to subtle (varying expression) perturbations and from coarse (doubling times) to detailed (G1 times) phenotypes. Once the value of the model has been established within polarity, the section thereafter covers what to learn from this model beyond polarity.

3.3.1 Bem1 evolution data set

The first goal is to get an accurate description of the genotypes present in [24]. There, authors show evolutionary recovery of the very ill mutant Δ *bem1* by successive deactivation of two GAPs and *NRP1*. The experimental data set provided some puzzling aspects: while from the biophysical analysis of the polarity mechanisms (recall Chapter 2) we expect a sharp survival cut-off for the Δ *bem1* when too many GAPs are present, in reality the fate of the pure Δ *bem1* is very uncertain. It has a broad range of volumes and mysteriously, while ill, it has faster G1 times than the much fitter Δ *bem1* Δ *bem3*. Also, there has been no clear link between *Nrp1* and polarity, while these experiments showed the influence of that gene deletion.

3.3.1.1 Model implementation

To provide some clarity for these curious phenotypes, the genotypes of our interest (16 in total) are incorporated into this growth model. This requires fitting of the doubling time data of [24] using the free parameters that are connected to the various gene deletions. The minimum *Cdc42* concentration to polarize is the main fitting parameter, and represents different GAP deletion phenotypes. In particular, when more GAPs are deleted, the minimum

Cdc42 concentration will decrease. The detailed model description in section 3.2 showed that four parameters describe all GAP mutations present.

Additionally, the *nrp1* deletion is hypothesized to reduce the minimum required G1 time, as explained in 1.4.4 (Box 4). The idea is that Nrp1 causes Whi3 to bind Cln3 mRNA, and consequently absence of Nrp1 frees Cln3 mRNA. This increases the amount of Cln3, causing faster cell cycle progression through G1. Additionally, there are indications that connect Nrp1 to Cla4. As opposed to its effect on Cln3 mRNA, Nrp1 may cause Cla4 mRNA to detach from Whi3, leading to more Cla4 and as explained in 3.3.3.1, lower Cdc42 concentration thresholds. However, this hypothesized effect of Nrp1 is not incorporated in the model, so that Nrp1 only adds one free parameter (smaller G1 time) to the model. The system remains well overdetermined as five mesotypical fitting parameters are needed to describe the 16 mutants of [24].

3.3.1.2 Fitness descriptions and NRP1 link

Figure 25 and Figure 26 show the comparison of the experimental data (symbols) to the simulated data (lines) for 16 genotypes. Conveniently, when placing the GAP variation on the (bottom) horizontal axis, there is a linear mapping to the minimum Cdc42 concentration to polarize as just described, allowing the placement of genotype (below), (part of the) mesotype (top) and phenotype (doubling times, vertical) in a single plot. Note that the simulated data can be generated for any GAP concentration, providing more resolution to the genotype-phenotype map than when based on experiments.

As seen from Figure 25 and Figure 26 (see Table 5 in the Appendix for the numeric representation of the data), fits are successful for gap deletions with and without the *bem1* deletion. Incorporation of the mesotypical rule of a hard polarity cut-off as a function of GAP (or Cdc42) concentration into this population growth model translates to a much softer transition between viability and unviability. As long as a sufficient part of the population accumulates enough Cdc42 relative to the number of GAPs, this part of the population continues to divide, allowing the population to survive. This makes the fate of the *bem1* deletion background rather diffuse.

Moreover, other subtle observations on $\Delta bem1$ become apparent by observing the curves of the rescue mechanism. For example, from [24] the viability of spores with the genotype $\Delta bem1 SPA2-GFP-URA$ is about 10%, while this was a factor 1000 lower for spores with endogenous *SPA2*. This was unexpected, given that the fluorescent marker on *Spa2* did not show an observable phenotype before [104], suggesting that the effect is weak under normal circumstances. However, from the steepness of the curves (blue lines in Figure 25 and Figure 26) where the $\Delta bem1$ genotype is located, even small effects on e.g., Cdc42 recruitment causing slight shifts in $[Cdc42]_{min}$ (its minimum concentration to polarize) have relatively large results. This explains the sharp increase in viability, despite a minimal invasiveness of the GFP.

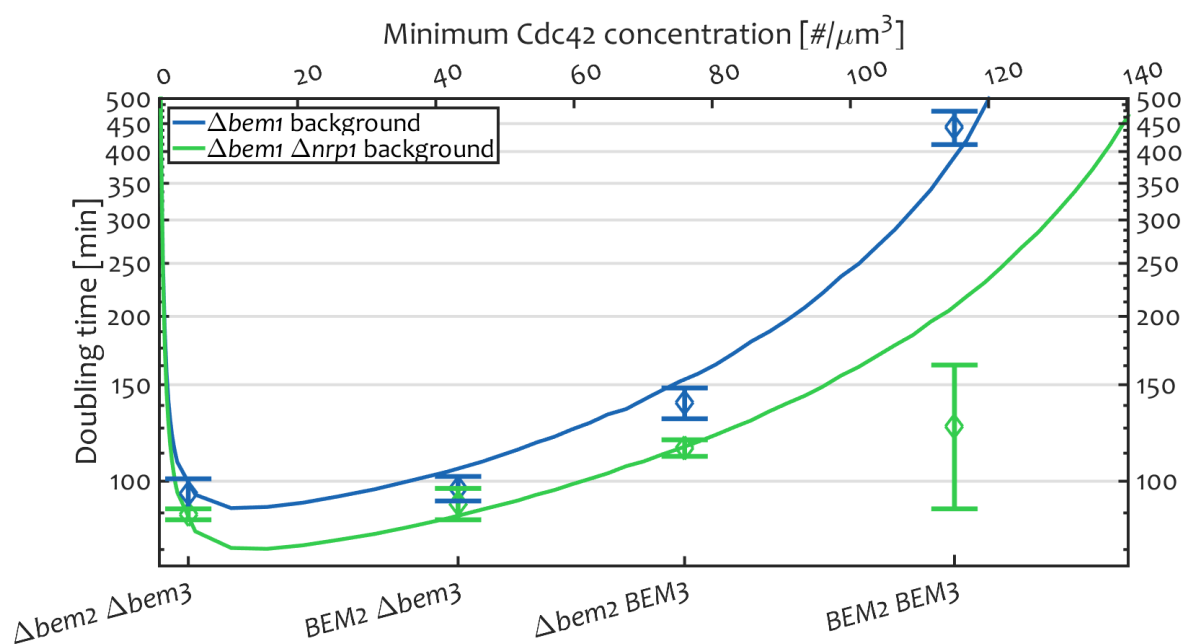


Figure 25 Simulated doubling times for various minimal Cdc42 concentrations to divide (top-axis) for the $\Delta bem1$ (blue line) and $\Delta bem1 \Delta nrp1$ (green line) background. The position of various GAP deletions on the x-axis is determined by fitting the experimental values (diamonds with error bars) to these lines.

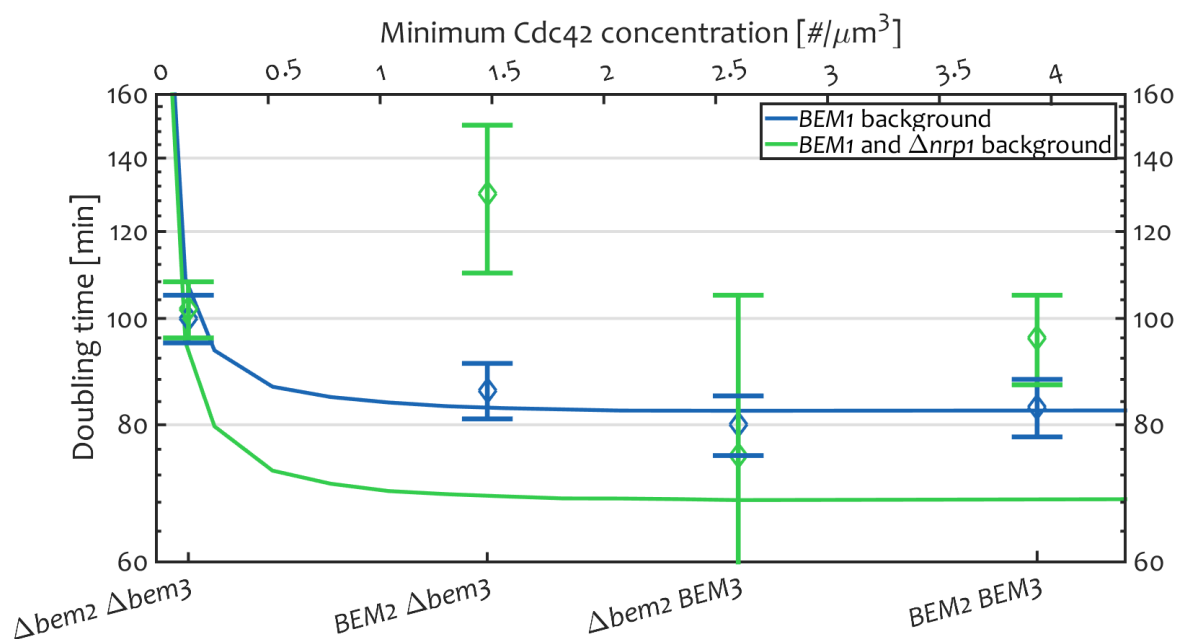


Figure 26 Simulated doubling times for various minimal Cdc42 concentrations to divide (top-axis) for the $\Delta bem1$ (blue line) and $\Delta bem1 \Delta nrp1$ (green line) background. The position of various GAP deletions on the x-axis is determined by fitting the experimental values (diamonds with error bars) to these lines.

Moving the analysis from the GEF to the GAP reveals one of the most robust model outcomes (see also Appendix I: Relaxation of model assumptions): the ratio between relative GAP contributions of Bem3 and Bem2, which is about 1.8 (x-axis tick locations of Figure 25 and Figure 26). Given that Bem2 is almost twice as abundant as Bem3 [234], this means that the GAP potency of Bem3 for polarization is 3 to 4 times higher than that of Bem2. This provides a convenient reverse-engineered measure of GAP strength *in vivo*, which yields a less extreme result than what was observed *in vitro*, where the contribution of Bem2 was negligible [89].

While the *bem1* and GAP deletions are easily incorporated into the model, more care must be taken with analysing the *nrp1* deletions (green lines in Figure 25 and Figure 26). Doubling time descriptions are qualitatively correct, but not always quantitatively correct. The experimental doubling time of the $\Delta bem3 \Delta nrp1$ is higher than expected, although the large experimental error bars may indicate unpredictable effects of disrupting the exact timing of Start. On the other hand, the beneficial effects of the *nrp1* deletion for a pure $\Delta bem1$ are underestimated.

The simulations suggest that an additional, unmodelled interaction of *NRP1* exists. The doubling time discrepancy for the $\Delta bem1 \Delta nrp1$ would not be solved by incorporating the omitted hypothesized Nrp1-Cla4 interaction described in 3.3.1.1. This would increase the minimum Cdc42 thresholds, which makes particularly the $\Delta bem1 \Delta nrp1$ even less fit. Hence, from the literature overview in Box 4, it is not apparent on the mechanistic level how to specify the last unmodelled Nrp1 interaction. This fits the view seen later in this chapter, that generally mutant incorporation into this model benefits from including functional information (typically the minimal information needed for the ontotype route as well [240]), which for Nrp1 is currently uncertain. Still, in absence of a direct link of Nrp1 with polarity, its influence on the ability to polarize in the $\Delta bem1$ background can to a decent extent be reproduced. This is supporting evidence for the hypothesized Nrp1 mechanism on Cln3 (Figure 11).

3.3.1.3 Cell cycle time descriptions

The next peculiar observation in [24] is that the $\Delta bem1$ delete actually has a smaller G1 time until the first polarization spot appears than WT. This seems very counterintuitive. WT should be fairly optimized in terms of speed, so it is unlikely that minutes can be gained from the time to polarize alone. This suggests that the time spent in G1 is shorter for the $\Delta bem1$. Based on the model simulations with parameters from the doubling time fit of Figure 25 and Figure 26, the G1 times can be generated. Those simulated times correspond to the average time in isotropic phase until $t_{mut} t_{G1, min, WT}$ is exceeded and size r_{min} is exceeded, see Figure 22.

Given the membrane growth rates as given in Table 4 (such that WT has a doubling time of 83 minutes), the shorter G1 time of $\Delta bem1$ compared to WT is not seen (see Figure 27). However, a fair comparison between the experimental data necessitates remarking one subtle difference. The fluorescence measurements on which observed G1 times were based

require synthetic media, which slightly slows the WT growth down. This influences times spent in G1 by the following reasoning.

Both the paper of [24] and simulations (see later on in 3.3.1.4) show fitness relates to size, such that WT cells are typically smaller than the $\Delta bem1$. Under the assumptions that daughter size scales with mother size, WT daughter cells may spend more time in G1 than the $\Delta bem1$ daughters to reach radius r_{min} . For fast growth as in YPD, this is not the case, as the growth during the minimum G1 time guarantees reaching the minimum size. However, for slower growth medium, WT is more frequently limited in G1 by the minimum size requirement than its $\Delta bem1$ counterpart, as the growth during the minimum G1 time can be insufficient.

Figure 28 shows the cell cycle times when accounting for the change in media (see 3.4.1 for details, effectively C_1 and C_2 from Figure 22 are reduced by a factor 0.5, yielding a WT doubling time of 109 min.). As under this experimental condition it is not possible to reconstruct the exact experimental doubling time, those times have been rescaled to match the simulated WT time to first spot to make a fairer comparison. The trends show good accordance with the experimental data from [24]. Indeed, the shorter G1 time property of the $\Delta bem1$ is described (comparing the top two red bars of Figure 28).

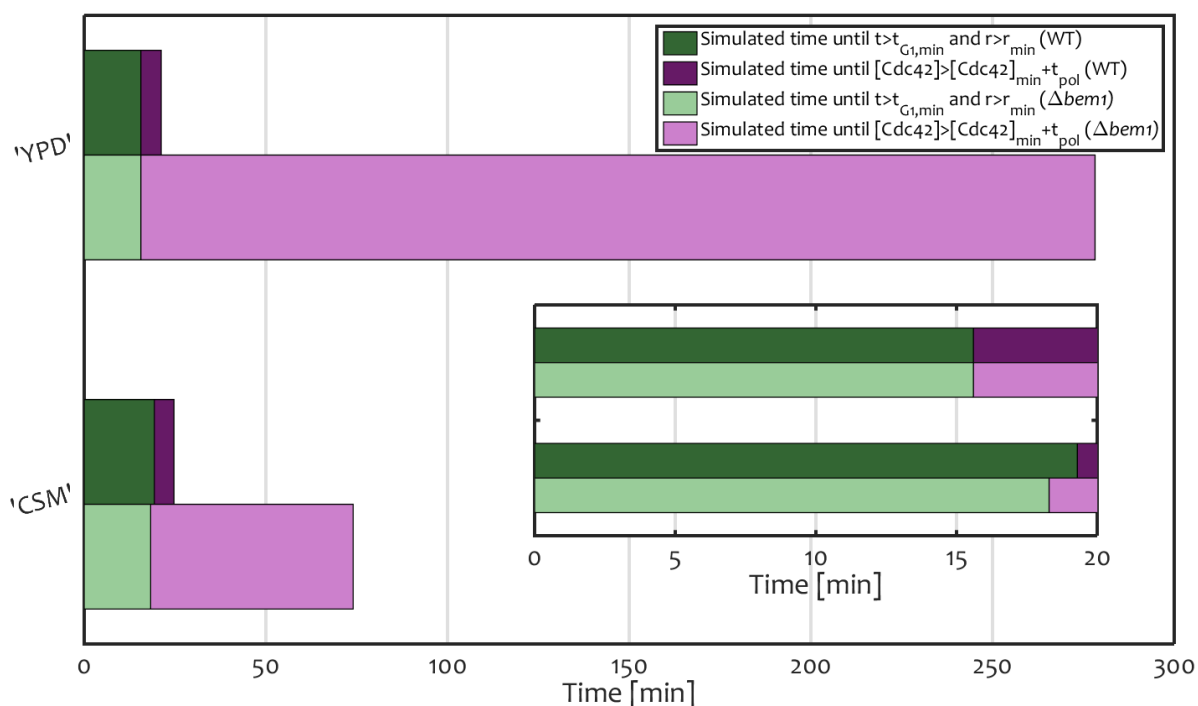


Figure 27 Bar plot from simulations of WT and $\Delta bem1$ mutants of [24]. In dark/light green are the simulated values for the average times until both the minimum G1 time and the minimum size to polarize are exceeded per mutant for WT/ $\Delta bem1$ respectively. In dark/purple are the simulated values the average time it takes after the conditions in red are fulfilled to surpass the Cdc42 threshold, added to the polarization time. Top bars show simulations under conditions reflecting YPD as in Table 4, bottom bars with half the membrane area growth rates, reflecting synthetic medium (CSM). The inset shows a zoomed version of the figure.

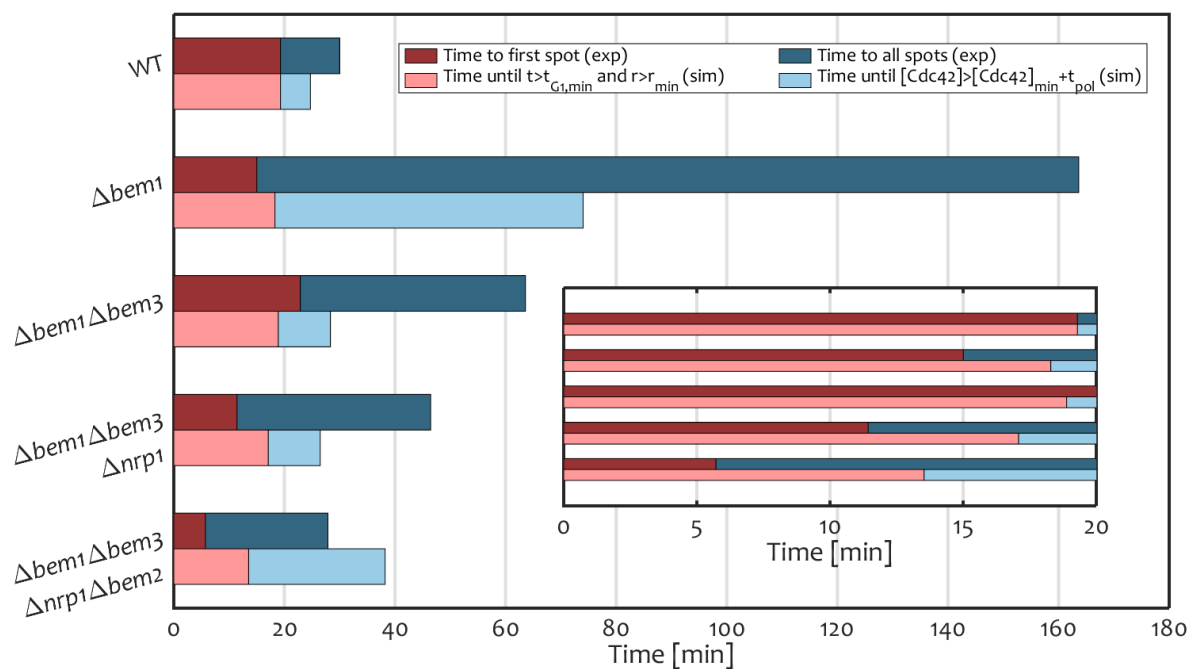


Figure 28 Bar plot from simulation of several mutants of [24]. In dark red are the experimental values for the time to first spot, in light red the simulated average times until both the minimum G1 time and the minimum size to polarize are exceeded per mutant. In dark blue are the experimental values for time to all spots, in light blue are the average times it takes after the conditions in red are fulfilled to surpass the Cdc42 threshold, added to the polarization time. The inset shows a zoomed version of the figure, to see the differences in times to first spot. All experimental values are rescaled from [24] to match the WT time to first spot value with the simulated minimum time before polarization for WT.⁷

An alternative explanation of the G1 time paradox is that the minimum G1 time, set by time required for e.g., cyclins to express downstream pathways, is in reality not constant for every cell, but a distribution. This may benefit the $\Delta bem1$ relatively more. These cells continuously need to counteract the degradation and dilution that cause lucky cells with above average Cdc42 concentration, to equilibrate again. Particularly the dilution is a problem (see Appendix H: Single cell test case), making every minute count. This makes cells with a less than average minimum G1 time more likely to survive, leading to a smaller average time to first spot, as the cells with higher than average minimum G1 time get purged out. This is also consistent with fig. 1D of [24], which shows the $\Delta bem1$ either forming a bud fast, or not at all. Since the exact minimum G1 time does not matter much for WT, this accounts for an apparent shorter G1 time for the $\Delta bem1$. This effect is not included in the model, but can be integrated if desired.

3.3.1.4 Cell size descriptions

The final comparison of the model with the literature comes in terms of size data. As seen from fig. 2 of [24], polarization time scales with size. This also follows from the simulations, see Figure 29. Regardless of the relative location of the Cdc42 thresholds, slower growing mutants increases their size approximately proportionally.

⁷ Legend created using: Adrian Cherry (2016), gridLegend – a multi column format for legends (latest version on <https://www.mathworks.com/matlabcentral/fileexchange/29248-gridlegend-a-multi-column-format-for-legends>), Matlab Central File Exchange, version 1.4, 20/1/2016.

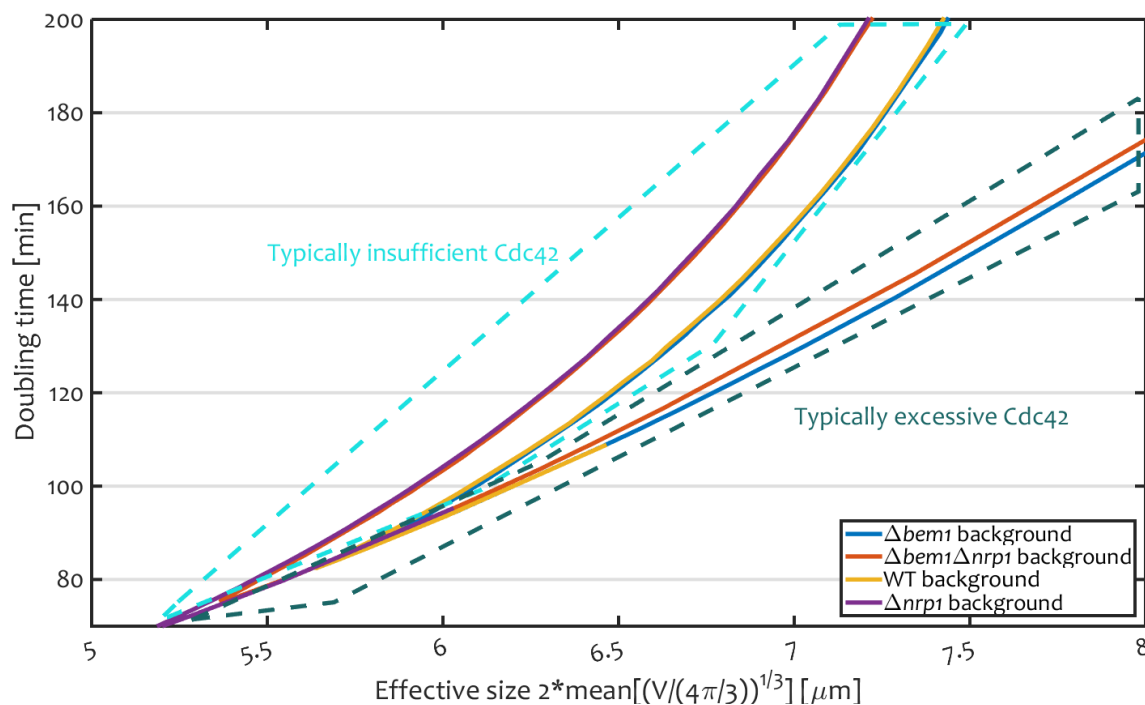


Figure 29 Doubling time as a function of the cell size, which is twice the mean of the effective radius (volume from simulated Coulter measurement (only live cells), converted to a radius) for different strain backgrounds. Polygons in cyan and dark green represent the simulated populations in which cells typically suffer from insufficient amounts of Cdc42 or excessive amount of Cdc42 respectively, slowing their growth relative to the optimum.

When the concentration threshold is challengingly high, as it is for e.g., the $\Delta bem1$, most cells typically suffer from insufficient amounts of Cdc42 to progress through the cell cycle as fast as possible (cyan polygon in Figure 29). Analogously, when due to the exponential nature of the polarization time as function of Cdc42 concentration (recall Figure 22) so much Cdc42 is present that most cells start to slow down, size also grow accordingly (dark green polygon). It must be noted that as seen from Figure 25 and Figure 26, the experimental mutants of [24] are usually far from this domain (where the curve steeply rises at the low x-axis end). Instead, the simulations forming this part of Figure 29 represent for example (unmeasured) further GAP deletions or Cdc42 overexpression in the $\Delta bem1 \Delta bem3 \Delta bem2$ background.

Additionally, the trends visible in the size distribution data of [24] are also seen in the model, see Figure 30. The $\Delta bem1$ is characterized by a large size, with a relatively broad distribution. The fitter mutant $\Delta bem1 \Delta bem3$ improves in that aspect, but it still not as small and homogeneously distributed across the population as the WT, showing that size and fitness are seemingly inversely related to each other.

3.3.2 Cdc42 expression sweep data set

In the previous subsections it was established that the model incorporates features of polarity mutants under strong perturbations, and for a wide array of phenotypes. To define the value of this model, it is necessary to determine what its resolution is. In other words, what level of detail is still accurately described?

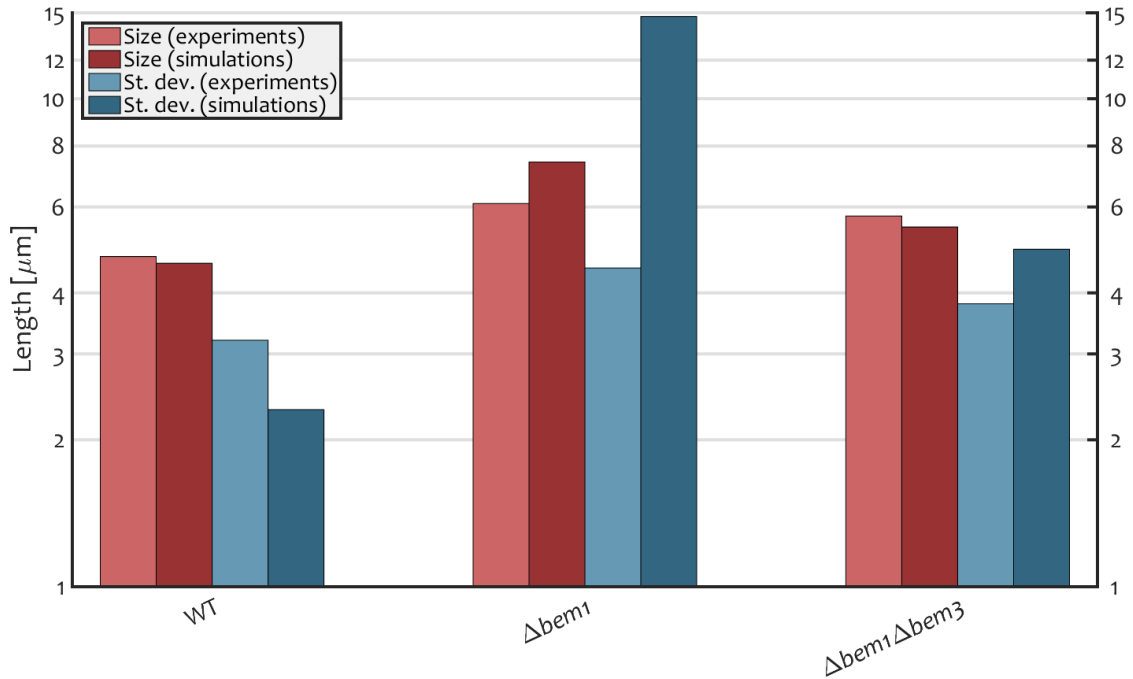


Figure 30 Comparison of experimental data from [24] (combining information of fig. 2E,F and 3B to determine the time point of the mutants in the evolution experiment from which to read out size properties) and simulations. Size is defined as twice the mode, found by the maximum value of the simulated volume data covering only live cells, converted to an effective radius. The size standard deviation is twice the sample standard deviation of the effective radii.

For that purpose, we zoom into the polarity cone that theoretically describes under which protein concentrations polarization still takes place, using the experimental data from Chapter 2 used to validate that theory. Previously discussed mutants probed in a step like fashion the GAP axis, but as mentioned there, a gradual sweep is more appropriate on the orthogonal Cdc42 axis, as this involves varying only one protein. Also, restricting the comparison of experimental and simulated outputs to doubling times alone will provide the most telling variation with the different genetic backgrounds. This experimental data is the same as presented in Chapter 2, to validate the theoretically proposed polarity mechanisms, but details on the protocols are found in the supplement to this Chapter, section 3.7.

3.3.2.1 Model implementation

The experimental conditions of the data from Chapter 2 first needs to be translated into the model. Given that the model takes into account stochastic production of Cdc42, changing its expression levels through the galactose induced *GAL1* promoter requires changing the two parameters that govern production in the model: the average Cdc42 expression burst time and size t_b and p_b respectively (see also Figure 22 or Table 4). From the theory of [189], it was shown how to extract the parameters from the population distribution of protein copy numbers. Typically, such distributions are measured by flow cytometry, with the protein of interest labelled using a fluorescent marker. If it is possible to translate the flow cytometry data from fluorescent counts to actual number of proteins, one obtains the burst parameters.

In this dissertation, we have measured Cdc42 with either GFP and sfGFP attached (see section 2.2). Paradoxically, while results in e.g., 2.3 proved the sfGFP sandwich fusion was fully functional with respect to fitness, we would prefer to use the GFP flow cytometry results for the collection of burst parameters. As seen in 2.3.1, varying expression leads to large fitness effects, which warp the fluorescent protein distribution through *transgenerational feedback*, an effect discussed in detail in Chapter 4. Therefore, a non-functional (but not deleterious) tagged Cdc42 version, like the N-terminally GFP fusion to Cdc42 (see discussion in 2.1), would allow more accurate reconstruction of Cdc42 burst dynamics.

A description of the flow cytometry experiments is found in Appendix C.2.1.1, while Appendix E shows the reconstitution of burst parameters. In short, the distribution of *GFP-CDC42* under various expression levels (through different galactose concentrations) was normalized to represent actual protein numbers by comparing to the reference distribution of *GFP-CDC42* under the endogenous promoter. This yielded Cdc42 copy number distributions under 0.05%, 0.2% and 2% (galactose) induction levels. It requires interpolation to obtain values for the inducer levels used for the doubling times measurement to correct for the different media used in the flow cytometry and growth assay experiments. This is explained in more detail in the supplement of the chapter (3.7.2.2). Ultimately, we are able to crudely estimate the expression burst parameter values for the inducer conditions used in the growth rate assays.

3.3.2.2 Doubling times comparisons

Figure 31 shows the simulated growth rate values based on the Cdc42 expression inferred from flow cytometry, compared to the experimental doubling times as previously shown in Figure 19. Here, the membrane growth rate differences across media are accounted for by matching the WT doubling times, which gives a bandwidth of membrane area growth speeds. Consequently, simulated growth rates for different expression levels are also spread across a bandwidth, contained within the dashed and dotted lines.

An important deficiency in the flow cytometry data is the lack of coverage particularly at the low expression levels. As growth rates were mainly measured in this regime, at first sight agreement is poor. Yet, the main cause of uncertainty at low expression does not stem from unknown membrane growth speed, but lack of burst parameter fits (particularly as we require interpolation, see 3.7.2.2). Indeed, quantitative agreement improves at high induction levels.

Overall, the trends described in 2.3 are well captured. The $\Delta b e m 1 \Delta b e m 3$ background survives for lower expression than the $\Delta b e m 1$, predicting viability thresholds at 0.06% and 0.1% inducer concentration respectively. WT survives even better, down to 0.04% induction concentration, although this threshold is quantitatively higher than experiments show. At the highest expression levels, the $\Delta b e m 1 \Delta b e m 3$ starts to suffer slightly from excessive Cdc42, which is less of a problem for the $\Delta b e m 1$. This is however a small effect, falling within the plotted bandwidth. Indeed, the conclusion based on experimental data was that there is no strong evidence that these doubling times are different.

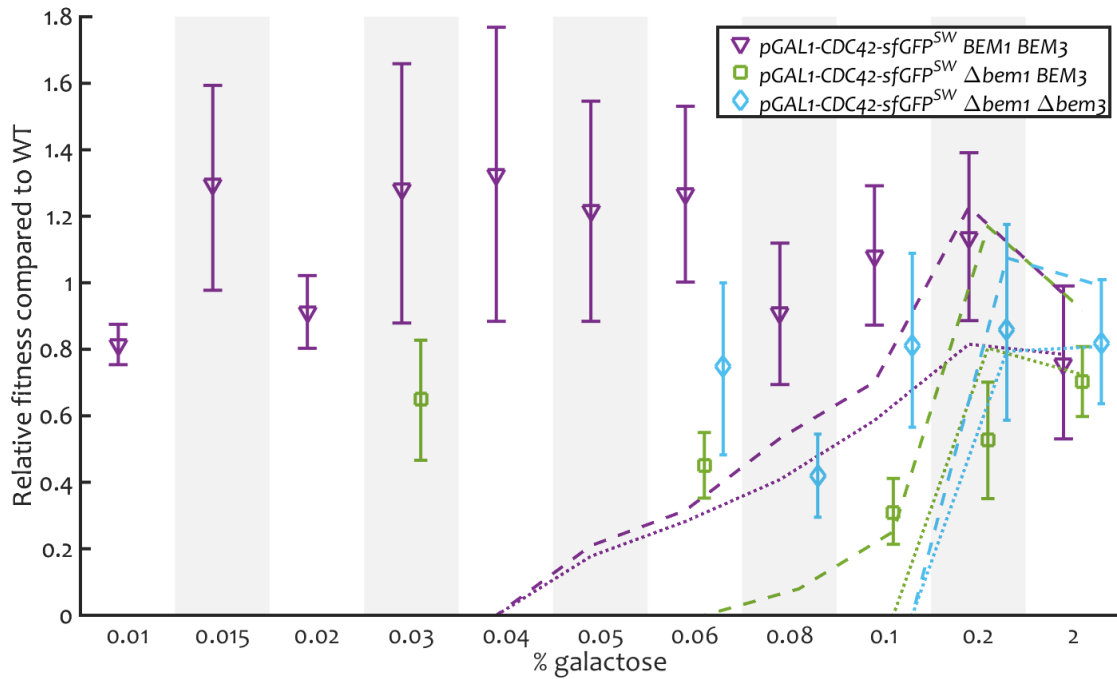


Figure 31 Simulated growth rates (upper limit dashed line, lower limit dotted line, based on uncertainty of the pure WT growth rate) and experimental growth rates for strains carrying $pGAL1$ -sfGFP- $CDC42^{SW}$ as their sole $Cdc42$ source (dashed lines with error bars), as a function of galactose concentration. In the simulations, membrane area growth rate is corrected for growth in synthetic medium, using the doubling time of WT with endogenous $CDC42$ at every concentration. The colors purple, green and blue indicate otherwise WT, $\Delta bem1$ and $\Delta bem1 \Delta bem3$ background respectively.

Ultimately, we gained upon the understanding of the experimental data in the following aspect. In 2.3.1, we needed to settle with the observation that despite a sharp threshold on the molecular level to polarize, the doubling times as a function of mean expression levels are diffuse. We have gained ground on how the sharp threshold quantitatively translates to the growth rate curves observed. This inspires trust that the simulated data fits benefit if more flow cytometry data were to be included, as that of [51].

3.3.3 Cla4 mutant data set

In the previous subsections it was established that the model incorporates peculiar features of polarity mutants when the molecular mechanism dominant in WT, the positive feedback mediated by Bem1, is removed. Based on this foundation, we aim to experimentally characterize the remaining polarity mechanism further by addressing more mutants in the $\Delta bem1$ background, and try to interpret this using the model. A logical starting point is adding the $\Delta cla4$ mutation, as it is thought to have a strong interaction with the $bem1$ deletion (even synthetic lethality [241]). We will see the result is not simply the same for all $\Delta cla4$ mutants.

3.3.3.1 Model implementation

In Box 2, the hypothesized influence of Cla4 in absence of Bem1 was motivated. After disrupting the binding with Rdi1 [81], and subsequent activation of Cdc42 by natural release

of its GDP, Cdc42 would move to the membrane. Cla4 then binds Cdc42, specifically where it is bound to Cdc42 [95]. This process was depicted in Figure 8. The additional hypothesis of Cla4 in relation to Nrp1 (see Figure 11 in 1.4.4) is not considered, as results in 3.3.1.2 did not provide supporting evidence for this claim. Furthermore, only mutants with the rescue ($\Delta bem1$) mechanism are taken, as the hypothesized role of Cla4 in the WT mechanism has not been considered in detail in terms of theoretical reaction-diffusion modelling (see e.g., [68]).

As explained in more detail in the methods of section 3.7.2.3, this makes the numerical implementation of the Cla4 deletion relative straightforward when Bem1 is not present. As graphically shown in Figure 32, the polarity cone narrows, leading to a linear shift in minimum Cdc42 concentrations for different genetic GAP backgrounds. This adds another degree of freedom to the model, namely the slope of the cone, from which all remaining model parameters for the different observed *cla4* mutants are deduced.

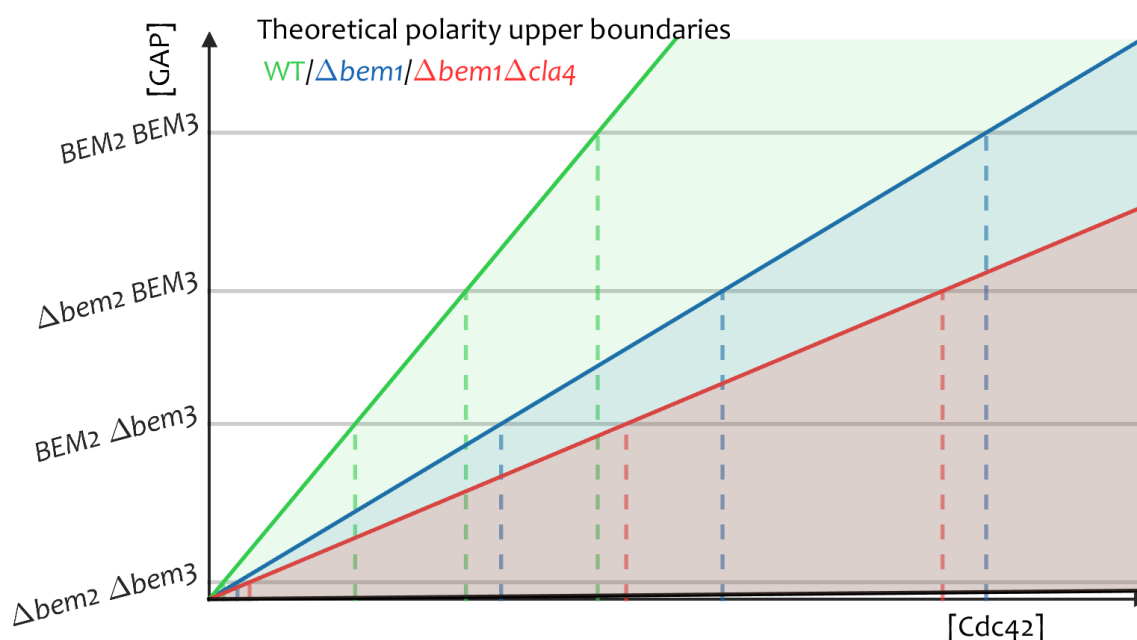


Figure 32 Polarity cone in the phase diagram of GAP concentration and minimum Cdc42 concentration. Horizontal grey lines represent various GAP mutants, whereas the colored vertical lines originating from the intersection with the horizontal grey lines and the edge of the polarity cone for WT (green), the rescue mechanism (blue) and the rescue mechanism including a *cla4* deletion (red) represent minimum Cdc42 concentration for those backgrounds respectively, given a particular gap mutant. At very low GAP levels, there is a marginal region which does not allow sustained growth either.

The experimental data set to which simulations can be compared also requires some attention. It consists of the doubling times of the $\Delta cla4$ in various mutants (see 3.6). that are the arithmetic mean of the doubling times across biological replicates placed in a 96 well plate which is subject to automated OD measurements. These times are in turn the weighted average of three technical replicates. By the method of generation of these diverse genotypes (3.7.2.3), potential background and unidentified suppressor mutations may arise, thwarting immediate comparison of the resulting doubling times. Therefore, the on paper identical

mutants that are also measured before in [24] are rescaled to those values, and the new mutants times are also rescaled by the same factor as an approximate correction.

3.3.3.2 Doubling time comparisons

Figure 33 shows the comparison of the experimental doubling times and simulated doubling times for the $\Delta bem1\Delta nrp1\Delta cla4$, $\Delta bem1\Delta nrp1\Delta cla4\Delta bem3$, $\Delta bem1\Delta cla4$ and $\Delta bem1\Delta nrp1\Delta cla4\Delta bem3$ mutants, adding one free parameter to the model fit (which required five for the 16 mutants of [24]) as previously explained.

Remarkably, the GAP *BEM3* and *NRP1* rescue the synthetic lethality of the $\Delta bem1\Delta cla4$. Immediately, this suggests another remaining form of positive feedback, but this discussion is left for 3.5. Secondly, the $\Delta bem1\Delta cla4$ is located outside the bounds of the plot, as this is a lethal combination, both in reality as in the simulations. Only the $\Delta bem1\Delta nrp1\Delta cla4$ is misfitted, because $\Delta bem1\Delta nrp1$ was quantitatively off in the previous section due to the decent yet incomplete description of *NRP1* in the model⁸. This means that overall, there is good quantitative agreement between model and experiments.

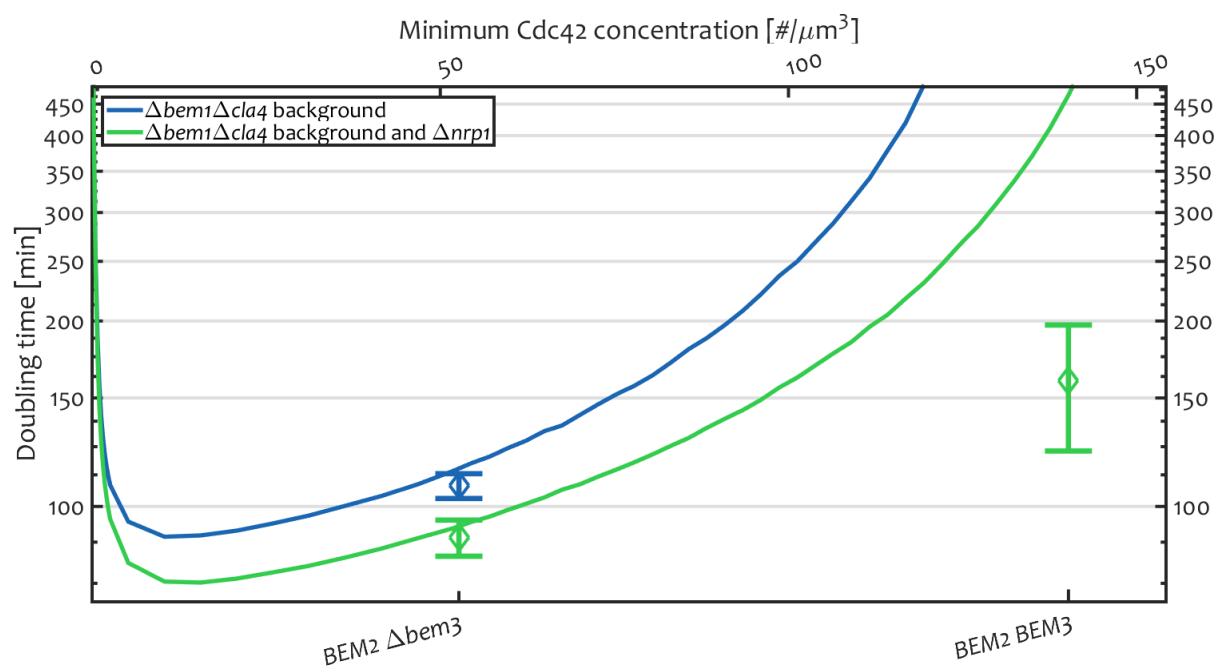


Figure 33 Simulated doubling times for various minimal *Cdc42* concentrations to divide (top-axis) for the $\Delta bem1\Delta cla4$ (blue line) and $\Delta bem1\Delta cla4\Delta nrp1$ (green line) background. The x-axis position of various GAP deletions is determined by fitting the experimental values (diamond with error bars) to these lines, while maintaining the relative GAP contributions of the fit in section 3.3.1.2. Experimental values for the *cla4* mutants are rescaled to match doubling times of *CLA4* strains with those of [24] (and have their error bar as well). The $\Delta bem1\Delta cla4$ falls off the graph, depicting synthetic lethality.

⁸ If we would artificially 'correct' the $\Delta bem1\Delta nrp1$ discrepancy in Figure 25 by also letting the *nrp1* deletion reduce the minimum *Cdc42* concentration drop to about 100 [#/ μm^3] the full deviation of $\Delta bem1\Delta nrp1\Delta cla4$ is also explained without compromising the other points. This points towards incomplete *Nrp1* inclusion as the cause of the $\Delta bem1\Delta nrp1\Delta cla4$ doubling time discrepancy.

3.4 Model extensions

The validation of the model in the previous section shows the ability of the model to deal to some extent with diverse polarity genotypes, even those not fully theoretically comprehended, and allows the generation of an array of phenotypes, up to the population level. While the last subsection investigated the resolution of polarity gene variations, the model can also incorporate other perturbations, both in origin and function. Concretely, the former is exemplified by the cells responding to a different environment, such as different media type. The latter claim is elaborated through the statistical analysis of epistasis predictions of polarity genes with genes of arbitrary function.

3.4.1 Different perturbation (media types)

Phenotypes are not dependent on genotypes alone; this dependence occurs within the context of an environment. Changes in environment are known to influence for example evolvability [42]. There, media changes discriminate between various genotypes in their effect, opening up new evolutionary avenues. By the same token, we may wish to examine the influence of an environmental perturbation in this polarity system, and determine what this means for e.g., the fit WT versus the ill $\Delta bem1$.

A preview of the media effect was already given in 3.3.1.3 when validating modelled G1 times with experiments. There, the $\Delta bem1$ only displayed shorter average G1 times than WT in the simulated equivalent of synthetic medium, not in YPD (see Figure 27). The translation of medium to model parameters was simply made by halving the membrane area growth rates C_1 and C_2 (see Figure 22). One may think that suboptimal medium also slows down Cdc42 production, but there is no clear experimental indication for this [242]. This leaves membrane growth speed as the main knob to accomplish lower maximum growth rates and to reflect the internal process of slower membrane phospholipid production.

As an example, the effect of halving the membrane growth rate is shown in Figure 34. This demonstrates the reduced deleterious effect of the *bem1* deletion when grown in less rich medium. This is because slower membrane growth rates give more time to generate enough Cdc42 to get over the minimum threshold to polarize, which is otherwise more problematic for the $\Delta bem1$. Analogously, this also shifts the minimum amount of galactose to be added in a *GAL1-CDC42* system that replaces endogenous *CDC42* expression, to remain viable.

This was also observed in the growth assays in this dissertation. When optimizing the richness of the synthetic medium (by adding more amino acids and sterilizing through filtration) as done for the assay in Figure 19 (2.3.1), the $\Delta bem1$ and $\Delta bem1\Delta bem3$ with inducibly expressed Cdc42 are severely affected below 0.08% inducer concentration, and not growing beneath 0.06% with one exception. In poorer synthetic medium (see Appendix C.3, Figure 67), down to 0.04% inducer concentration (0.02% + at worst 50% evaporation there) these backgrounds remained clearly viable. Hence, the richness of the medium determines the strength of Cdc42 concentration barrier. If the media is richer, there is less time to be lucky with expression.

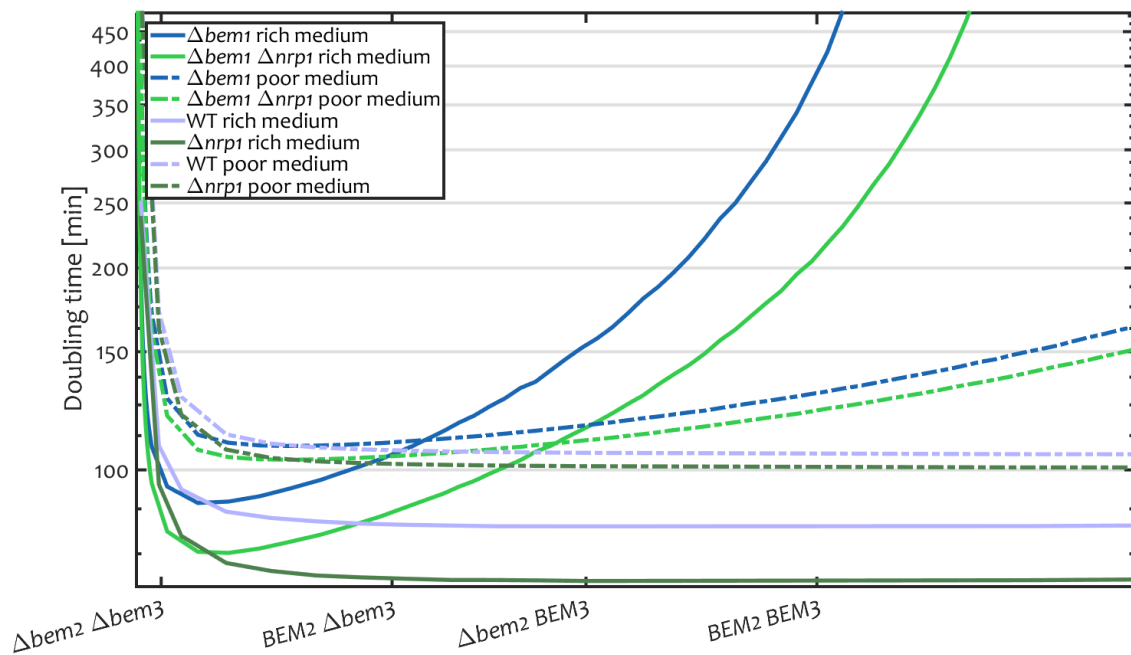


Figure 34 Simulated doubling times as a function of GAP concentration for the $\Delta bem1$ (blue), $\Delta bem1 \Delta nrp1$ (green), WT (purple) and $\Delta nrp1$ (dark green) background, for rich (filled line) and poor (dashed-dotted line) medium (defined as yielding half the membrane growth rate of rich medium).

A critical reader could note the polarization threshold depends on the GAPs as well as Cdc42, given the polarity cones in Figure 16 and Figure 32. This can give the impression that the GAP expression noise is relevant to the aforementioned argument. The most relevant GAP to include is Bem3 (from Appendix: Relaxation of model assumptions, the dominance of the Bem3 contribution to the effective GAP action is quite robust, and at least 60%). A model extension explicitly including Bem3 as the major factor in GAP concentration as a stochastic variable would then also predict more Bem3 over time, possibly keeping the Bem3/Cdc42 ratio relevant for polarization more robust than when only Cdc42 is considered.

However, the Bem3 contribution to the noise (and of other GAPs for that matter) is limited (see 3.2.1.3.1). Inferring *BEM3* burst parameters from expression assays in [229] yields 62 bursts per cell cycle (46 for *BEM2*), showing that GAPs are approximately constitutively expressed. By contrast, Cdc42 is expressed a handful of times per cell cycle from our flow cytometry experiments (see value for k in Table 4). The noisy burst behaviour required for ill-growing mutants can then almost entirely be attributed to Cdc42. So, the beneficial effect of less rich medium persists in an extended model explicitly including GAPs expression noise.

3.4.2 Epistatic interactions (beyond polarity)

After including environmental effects, we aim to assess the incorporation of other modules than polarity into the growth model. The phenotype used to evaluate the performance of the model is epistasis of genes in the polarity module, with genes outside polarity. This makes the best use of the myriad of information available from high-throughput studies on the effects of individual mutations. Here, BIOGRID [58] and SGD data [57] are used, as explained in 3.7.3.2.

As the polarity gene we consider only *BEM1*. The null mutation has a large effect, brings yeast close to inviability, and consequently effects of other genes on *bem1* mutants are most likely to have been picked up and documented in literature. We categorize the incorporation of Bem1-interacting mutants into the model predictions as: coarse phenotypical effects, refined phenotypical effects or by functionality. In this order, the mutant categorization requires progressively more information about the genes/proteins, whose interactions with Bem1 are addressed. A quantitative assessment of the predicted epistatic effects is found in Appendix J.

Starting from the broadest category, there are many mutations that mildly decrease or increase fitness, However, if these mutations decrease or increase growth rate in a manner (generally) unrelated to polarization, they may interact with a *bem1* deletion in a similar way as with a medium switch. More concretely, if a mutation has a negative fitness effect on its own (slowing growth), this may more often allow the time needed to elevate Cdc42 production to above the threshold within a $\Delta bem1$ background. Thus, naively we may expect more positive epistatic interactions than usual for the *bem1* deletion, with those deleterious mutants, and more negative interactions for beneficial mutants.

However, it may be more appropriate to zoom in for the second mutant category and consider only mutants that have smaller/larger size at start or are fast/slow in G1, as the threshold concentration (problematic for the $\Delta bem1$) is relevant at that time. Analogously, more positive epistatic interactions are then expected for the *bem1* deletion with those mutants that have smaller size at start or are faster through G1. Similarly, more negative epistatic interactions are expected for those that are larger at start or have slow G1.

Still, considering all these mutants ignores the ample functional information that is present for many genes. So, if this is included in the way the mutants are incorporated into the model, better predictions may be made. Since we generally expect improved performance of the $\Delta bem1$ with more Cdc42, we may get more positive epistatic interactions for proteasome mutants (that may delay Cdc42 degradation⁹). Analogously, the $\Delta bem1$ benefits from more time to make Cdc42, so positive epistasis is also expected for mutants related to phospholipids (so that membrane expansion may get slowed down). By contrast, more negative epistatic interactions are expected for translation mutants (ribosomal-related)¹⁰, as then less Cdc42 expression is expected, which was already problematic for the $\Delta bem1$.

The conversion of mutant categories to changes in growth model parameters are also concretely stated in Appendix J. The subsequent predictions on epistasis signs with $\Delta bem1$

⁹ Since from [235] it is known that the half-life of Bem3 is more than 20% longer, a model extension explicitly including Bem3 as the major factor in GAP concentration as a stochastic variable would still predict a higher, but smaller benefit of less degradation for Cdc42 than for Bem3.

¹⁰ From [236], the half-life of mRNA of Bem3 is a factor two larger than for Cdc42. Therefore, a ribosomal deficiency will affect Cdc42 mRNA more heavily, as proportionally more ribosomes may be hitting the mRNA half-time with their minimum translation time needed. A model extension explicitly including Bem3 as the major factor in GAP concentration as a stochastic variable would hence still predict the same direction for the epistasis, but the effect is more diluted.

should be interpreted as expecting more/less positive epistasis than normal, as we ignore many details. Here, more/less than normal is relative to the situation considering all mutants with $\Delta bem1$ interactions, where 16% has positive epistasis with $\Delta bem1$ (see dotted line and most left point in Figure 35). The observed epistasis within a group is thus compared to 16%.

Because of the limited sample size, the estimates of the actual percentage of positive epistasis within a group has notable uncertainty. Following Bayesian analysis on the data, we obtain credible intervals for the positive epistasis fraction in Figure 35 (see for more details 3.7.3.2). Given this uncertainty, we calculate whether the deviation from the 16% is meaningful or not. This is assessed by means of the posterior odds ratio or Bayes factor, which reflects the probability that the deviation in positive epistasis fraction exists in the hypothesized direction, divided by the probability that the deviation is in the opposite direction. The values in each group are shown in the bars (right axis).

Using the typical interpretation of Bayes factors [192], there is generally no added value for the model when only considering phenotypical evidence (factors in bars typically $\lesssim 3$, except small at start/ fast in G1 phenotype). However, when functional information is used, positive evidence exists that this model in part describes the expected epistatic relations (factors $\gtrsim 3$, strong if $\gtrsim 10$). This was also encountered for the $\Delta nrp1$ mutant description: under the hypothesis that Nrp1 mechanistically affects the timing of the Start transition, there is enough information to describe a substantial part of the Bem1 interaction. The value of biofunctional information encountered here also seems in line with the theory behind ontotypes [240].

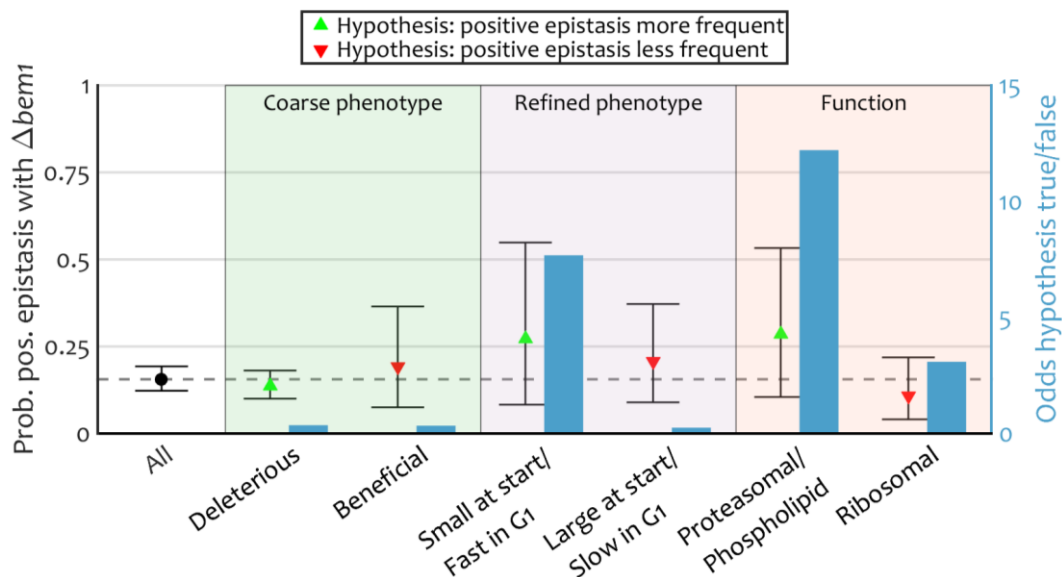


Figure 35 Epistasis predictions require functional information. Plotted are the most probable values and 95% credible intervals (left y-axis) for the fraction of the epistatic interactions with $\Delta bem1$ that is positive for different mutant types, categorized by accommodation into the model through a coarse phenotype (green background), more refined phenotype (purple), or through functional information (orange). The corresponding model hypothesis is shown through the plot symbol of the most probable values, which is either a green triangle (up-facing) for more frequent positive epistasis than expected from considering all mutants (dashed line), or red triangle (down-facing) for less than expected. The blue bars (right y-axis) show the Bayes factor (posterior odds ratio) for the corresponding hypotheses.

3.5 Discussion

As explained in the introduction of this chapter, a key problem in biology is the prediction of phenotype from genotype. Evolution uses phenotypes that determine fitness as criteria for selection, so understanding of the possibilities and limitations of evolution relies on a detailed grasp of how phenotypes are generated by genotype and/or the environment. Particularly the former and to some extent the latter path of generation have been explored in this chapter through a population dynamics model, aiming to describe phenotypes of polarity mutants. This model incorporated cell growth, protein turnover and polarization through what is defined as the mesotype. It is now possible to recapitulate the resulting conclusions (**bold**), and discuss their implications, summarized in hypotheses or claims (*italic*):

- **The cell growth model fits all doubling times and epistasis of 4 strains with a $\Delta bem1$ *NRP1* and 4 *BEM1 NRP1* background using 3 fitting parameters to integrate 8 genotypes. Qualitatively, predicted trends in G1 times for the typical evolutionary rescue path of the $\Delta bem1$ are in accordance with literature. Similarly, trends in cell size are also followed.**

It is not straightforward for any model to capture the ample epistasis between Bem1 and the GAPs Bem2 and Bem3. In any case, molecular mechanisms must be included in some way to account for the much-increased importance of the GAPs in the $\Delta bem1$ background. Bio-functional information, used in the ontotype approach, alone is not sufficient for this complex network (e.g., outcomes depend strongly on even concentrations), nor is disregarding diffusion and merely examining reactions in an Michaelis-Menten approach. Even reaction-diffusion equation results are in itself not enough; a cell having insufficient Cdc42 (or too many GAPs), as is the default for the $\Delta bem1$, does not necessarily die, leaving a smoother transition between viability and unviability than expected from the mesotypical rules summarizing molecular interactions alone. That is where this model has its added value:

The mesotypical inclusion into a growth model with protein turnover is both new and necessary to successfully describe phenotypes from the bottom-up.

Given the rather parsimonious approach taken from all growth details, yet the more detailed description of protein turnover, it seems that the latter is the key addition towards the success of the model. So, while simple in the sense that only few fitting parameters were required, it may be possible to strip this model further to its essence. Therefore, in Chapter four the hypothesis is put to the test that:

Protein expression noise coupling to a simple mesotype (minimum protein level to polarize) drives epistasis in the yeast polarity system.

- The model fits 5 out of 8 doubling times quantitatively correct (within experimental uncertainty) for the $\Delta bem1 \Delta NRP1$ (with 1 qualitatively off), adding 1 parameter (reduced minimum G1 time) for the Nrp1 action.

In absence of solid evidence of the exact molecular mechanism of Nrp1, it is not straightforward to make any prediction on what Nrp1 will do. Another convenient feature of this model that in absence of such clear information, phenomenological integration of phenotypes (here, reduced G1 time) is easily facilitated. That alone accounts for most of the doubling time variation in 8 polarity mutants, arguing that:

Nrp1 acts on polarity mainly through influencing its cue at Start (GEF release/GAP inhibition).

Still, the description is far from perfect, and this may in part be due to the difficulty of measuring $\Delta nrp1$ (large experimental errors), but also because functional information may have some added value. The latter is in line with the ontotype approach. As the additional hypothesized Nrp1-Cla4 interaction (without Nrp1 less stable Cla4 mRNA, less Cla4 and higher Cdc42 concentration thresholds) would e.g., rather deteriorate the fit of the $\Delta bem1 \Delta nrp1$, the solution is not in this direction. Instead, deletion of Nrp1 may not only affect the average G1 time, but breaking the Nrp1 regulation can also increase the dispersion of G1 time distributions. If G1 times are not considered as a constant but as a distribution of times, particularly the $\Delta bem1 \Delta nrp1$ benefits from this, even accentuating the already present smaller mean G1 time with respect to WT found in simulations (see end of 3.3.1.3).

- Doubling times of Cla4 mutants in the $\Delta bem1$ background are accurately described using one added fitting parameter

The accurate description of Cla4's function is interesting given the uncertainty of its molecular mechanism. Despite the theoretically inherently forgiving nature of the positive feedback (in terms of needing detailed information), it was possible to infer a good detailed suggestion for the action of Cla4 from a combination of literature sources. This model combined with the doubling time data provide supporting evidence for this mechanism:

Cla4 effectively decreases the minimum Cdc42 to polarize in a $\Delta bem1$ background through anchoring to the membrane, activating where Cdc42 is present, and dislodging Rdi1 from cytosolic Cdc42 which can then attach to the membrane, completing a positive feedback loop for Cdc42.

As the $\Delta cla4$ mutation is not always synthetically lethal for every $\Delta bem1$ mutant, a new missing link surfaces in the polarity puzzle. The fitted 20% decrease in minimum Cdc42 concentration that the Cla4 addition causes in the $\Delta bem1$ background suggests that 80% is still unknown. While sensible from a theoretical point of view (the nature of the positive feedback can be very generic), it is experimentally not clear which protein(s) should be

responsible for this. For example, candidates Ste20 and Skm1, PAKs related to Cla4, cannot dislodge Rdi1 from Cdc42 [81], so these independently do not suffice.

Actin transport of Cdc42 or downstream effectors may be an answer, yet the effect on Cdc42 is controversial [97], [243]. In an attempt to reconcile opposing views, we turn to the clathrin-dependent form of endocytosis, which has not received as much attention in polarity with the discovery of the formin-dependent mode of endocytosis [103]. As explained in the introduction (Box 3 in 1.4), free GAP removal from the bud site by epsin-mediated endocytosis adds to the local depletion of the free GAPs in the theoretical model [51] validated in Chapter 2.

There, GAPs were depleted by forming a complex with active Cdc42 until there were insufficient left to stop accumulation of Cdc42 at that point. Alternatively, in the endocytosis case active Cdc42 recruits the formation of actin cables and facilitates endocytosis. This transport along endocytic vesicles removes free GAPs away from locations where active Cdc42 is abundant. Effectively, insufficient GAPs remain at the future bud site to deactivate all the Cdc42 on the membrane, establishing pattern formation. At this moment, this appears the most credible alternative for the known polarity mechanisms, leading to the statement:

The positive feedback of Cdc42 in a $\Delta bem1$ background is not mediated by Cla4 alone, one or more hidden players account for 80% of the positive feedback. The likeliest scenario is local free GAP depletion on the presumptive bud site through endocytosis by clathrin-binding epsins.

For a follow-up experiment to shed more light on the remaining positive feedback, one can think of a renewed evolution experiment. Starting out with the evolved mutant from [24], a further subtle epsin-related mutation from [106] can be added which negatively influences its GAP interaction. A possibility is removing both *ENT1* and *ENT2* and replacing these by a weakly expressed ENTH-domain from Ent1, which contains the presumptive binding site to the GAPs. This way, the GAP depletion at the presumptive bud site is weakened.

- **When varying Cdc42 copy number through gradually inducing its promoter (12 values), the measured changes in doubling times are to a certain degree in line with model predictions based on expression levels obtained by flow cytometry**

To make accurate model predictions for the fitness dependency on protein copy number, the mechanistic theory, overarching growth model and experimental coverage of expression data must be adequate. A limiting factor here is the flow cytometry data available, where expression levels from 3 values (plus one literature value) must be interpolated to generate the set of 12 levels used in the growth assay.

Where the latter assay's resolution matches most closely the flow cytometry assay's resolution (0.2%-2%), the model fits the experiments well. This is in spite of some side notes

on the direct comparison between the strains from the growth and flow cytometry assay, such as that both use different versions of fluorescent Cdc42. The discrepancy becomes larger once expression at more inducer values must be interpolated, suggesting that the limiting factor is not the underlying mechanistic model for polarization, nor the growth model in itself, but availability of experimental data with good coverage to feed into the model.

This puts the difficulty to quantitatively match the predictions of the mechanistic GAP model based on functional subunits (see Chapter 2) into context. There, even the most prominent feature, the sharp Cdc42 concentration boundary to polarize, becomes diffuse in terms of population phenotypes that are acquired in bulk, such as doubling times, which we use to test many genetic scenarios. This preliminary data set addresses this issue and gives a hopeful outlook, namely:

The growth model allows the observation of fine features of the underlying mechanistic model for polarization, which would otherwise be impossible to retrieve from experimental data.

- **Halving the membrane area growth rate as a proxy for poorer medium increases doubling time of WT by 30%, yet reduces that of $\Delta bem1$ by 65%, bringing them within about 20% of each other.**

A rather surprising find is that simulations let the $\Delta bem1$ notably benefit from poorer medium. As this background suffers from shortage of Cdc42, a slowdown of dilution and more time to be lucky with a large expression burst turn this ill strain into a decently viable mutant. WT on the other hand, already in the sweet spot with regards to protein concentrations in its fitness landscape, can only deteriorate from its optimal state. The two backgrounds are brought close together without having to delay WT to an unrealistic extent. In an attempt to generalize this reasoning, one will always see an optimal mutant suffer from suboptimal media, while expression bursts always favor the ill-adapted, because they are content with all the time they get in poorer medium. This suggests that:

Less rich medium may act as an equalizer for competitive fitness between diverse genetic backgrounds, offering ill-adapted mutants more time to maintain and improve itself through evolution, possibly facilitating evolution by shallowing fitness valleys.

This also has implications for the hypothesis how this polarity network may have ended up this way. It is rather remarkable that symmetry breaking is already accomplished by a much simpler system (without Bem1, Bem2, Bem3, Nrp1 and Cla4) using completely different mechanisms (GAP depletion, possibly actin, instead of the positive feedback by the Bem1 protein complex). The current, WT system is more complex, with a small gain in doubling

time, and one can speculate how a system could have evolved to such an advanced state, despite ample epistasis and corresponding fitness valleys.

One conjecture [51] is that at first without Bem1, GAP concentration is critical. The next step in evolution would be a protein (the future Bem1) that binds the GEF, lifting its autoinhibition, increasing its potency and conferring an advantage. At this point the optimal GAP concentration may increase slightly, which may be another adaptation step. This protein could then evolve to bind Cdc42, yielding the more robust Bem1-mediated positive feedback loop, after which GAP concentration becomes less critical, and free to evolve along the lines of the innovation, amplification, duplication model [6] to the state it is now.

However, after having illustrated the effect that the environment can have on shallowing fitness valleys (here through poorer medium), the focus may shift to the clear selective fitness advantage every adaptation must have. Under slow growing conditions (considering the original territory of budding yeast is fruit outside of a laboratory), more avenues may have been available to reach to same goal, which is a fast, yet robust polarity mechanism. In the next chapter the mechanism behind shallowing fitness valleys is uncovered, to further understand how evolutionary restrictions are relaxed.

- **Extending this growth model's predictions, namely description of epistasis between the $\Delta bem1$ and other mutations, requires functional rather than mere phenotypical information.**

The model has coarse-grained many molecular details related to growth, focusing mostly on protein turnover and polarization (mesotype), as polarity mutant phenotypes were the main objective. If one however wishes to extend or adapt this model to other goals, this is in theory well possible as there are many handles in this model to accommodate changes. Ideally, this newly incorporated module shares the same level of detail that polarization has received.

In practice, this is often not possible for systems not so well biophysically characterized as the polarization network. Consequently, in the case of epistasis between a mutation from the polarity module and one from another module, the assertion of fitness relies on the coarse description of the second module and of the growth module, only having a rigorous justification for the polarity module. While arguably far from perfect, it is reassuring that even then some lessons can be learned, if one at least has added functional information for its new mutants. This requirement seems reasonable; it is in line with the usage of ontotypes [240]. Function may not tell the whole story, but it gives a notion of the story line.

Accurate phenotypical description is ideally based on an appropriately coarse-grained bottom-up description of molecular details, but cross-module descriptions already benefit from functional information.

3.5.1 Outlook

This chapter provided lessons on how genotypes (and lesser extent environment) give rise to (population) phenotypes. This provides some handles to then understand how evolution acts to select upon the appropriate phenotypes. In that light, the most interesting hypothesis has been concerning the influence of protein copy number noise on epistasis.

Therefore, the next chapter aims to expand further on the implications for the general emergence of epistasis. While it may be difficult to predict specific epistatic interactions without functional information or physical interaction knowledge, interesting assessments can be made for the prevalence of epistasis, in particular the existence of negative epistasis. As will be shown in the next chapter, a general evolutionary mechanism is conjectured, which may *inter alia* facilitate the generation of epistasis. This acts far beyond the characteristics of this growth model, but the polarization case serves as a good illustration of this principle to understand evolvability better.

3.6 Acknowledgements

The following people (unless otherwise annotated, affiliated with the Laan Lab at the TU Delft at the time) contributed to the data in this chapter:

- Leila Iñigo de la Cruz, who as a student under my supervision, produced an indicative doubling time measurement of the WT used in flow cytometry, and most importantly, the strains and growth rate data used in 3.3.3, making use of analysis software of Enzo Kingma.
- Matthias Peter Lab (ETH Zürich) [85] for the strain YMP203.
- Fridtjof Brauns (Frey Group, LMU Munich) for the use of his figures: Figure 23 and Figure 24.

3.7 Supplementary information on methods

3.7.1 Computational matters

3.7.1.1 Specifics on implementation

The integrative growth model was written in *Matlab* and run on version R2014b. The same random number generator settings, the default for Matlab R2014b namely Mersenne twister with random seed 0, are used throughout this thesis, except for the plot regarding the effect of different seeds on reproducibility of the doubling times in Appendix G: Convergence and error estimation.

The implementation uses a while loop whose iterations correspond to generations in the colony. The stopping criteria are that both the minimum population size must be reached and the last simulated protein copy number ('FACS') and volume ('Coulter') measurement time point (if supplied beforehand by the user) must have passed in order to stop. At every iteration inside the while loop, every cell attempts to go through four phases. In the first phase isotropic growth occurs during a deterministic amount of time per cell, either the time needed to reach the minimum size to polarize or the minimum time needed in G1 to polarize (whichever time is greater). Both volumes and Cdc42 amount (through random draws of the appropriate exponential and Poisson distributions to match dynamics of [189]) are adjusted.

The second isotropic growth phase is an implementation of the Gillespie algorithm [239] for Cdc42 production, advancing time per cell one Cdc42 expression burst at a time. As concentration only increases after a burst and a minimum must be exceeded, at the start of the phase and after every subsequent burst the concentration is evaluated. If this exceeds the minimum threshold to divide, the cell moves to the next phase, otherwise it remains in this phase through a while loop. If the maximum cell size is exceeded the cell also exits this while loop, but is annotated as death and will not participate in any further calculations. Time of death relative to the start of the simulation is stored.

The third phase is deterministic as well, and volumes and Cdc42 amount are adjusted for an isotropic growth time for the duration of the polarization time which depends on Cdc42 concentration (see equation 3.2). Finally, bud growth also occurs for a deterministic amount of time, until the daughter size is as specified, encompassing the fourth phase.

As time since ancestors increases, so do the differences between these times for every cell, e.g., the fourth generation of cells is *timewise* mixed with the third and fifth generation. The algorithm computes one generation at a time, yet the relevant output such as doubling times, FACS and Coulter measurements are only valid when for a specific time it is certain no more cells can contribute to these outputs. This phenomenon can unnecessarily extend the computation time. Therefore, it is possible to temporary (for one iteration) pronounce cell

belonging to an upper quantile (used here 0.5) of time to ancestors dead so that computation only progresses for the cells that timewise lag behind in the colony. Every iteration the list of cells that progress one generation is then updated.

While this computational trick also accommodates many artificial dilutions of the colony to speed up calculations, the added level of stochasticity introduced into the outputs is generally not advised. However, robustness is obtained when diluting once, after reaching the critical colony size for the first time, and using a sample of 1000 live cells at that moment as the initial conditions for the second simulation round. When the colony size reaches the critical colony size for the second time, equilibrium has been reached, see Appendix G : Convergence and error estimation. Results also proved robust against ‘technical’ replications of the simulations, i.e. using different random seeds.

Doubling times are determined by averaging the ordinary least squares fits on the log cell count versus time, with time discretized to a one-minute resolution, of the final hundred moving windows of two-hundred-minute size. Simulated FACS and volume measurements are simultaneously performed at time points just before (0.1 min.) the end of the second run, unless otherwise specified. Only for the volume measurements the dead cells are passed on to the output volumes as imaginary numbers, so that the user can decide whether to include these into the simulated measurement (for the FACS they are discarded). A single cell time trace (of the copy number and the concentration, which includes the volume information) is found in Appendix H: Single cell test case), to illustrate the proper functionality of the model.

3.7.1.2 Initialization

A sensible choice must be made for the initial conditions. This is relevant, as if only one cell is taken, cell death is risked before the colony can reach a substantial size (though always 5 runs are done before giving up if all five fail). While an interesting case for e.g., studying spore germination efficiency leading to colony formation on plates, for this dissertation the equilibrium properties of the colony are most interesting in terms of observations. The starting population has no proteins and consists of 1000 cells, with cell size at 110% of the minimal cell radius to polarize. The starting colony is asynchronized by setting the starting time stamp for all cells evenly spaced between minus the doubling time of WT to 0.

Ultimately, the results of interest are equilibrium properties, such as the maximum doubling times. As aforementioned, to improve convergence but also to minimize the influence of the initial conditions, a dilution step was added after reaching the critical colony size. Resampling 1000 cells provided a new starting point for the next iteration of the simulation (much like the procedure for the growth rate assays) to attain convergence.

3.7.1.3 Calibration

After specification, the model is calibrated to accommodate existing data. More specifically, the WT should follow these ‘ground truths’, sorted in decreasing order of importance:

- Doubling time in YPD is 83 minutes [24]. This fixes the ratio C_2/C_1 .
- In the fitness landscape of doubling time as a function of minimum Cdc42 concentration to polarize, WT is close to the optimum, as it would have evolved towards there. This sets $[Cdc42]_{min,WT}$ needed to evaluate the previous statement.
- Mean population number of Cdc42 is 8690 [234]. This corrects $p_{b,av}$ slightly, previously inferred from FACS data.
- As a relatively loose condition, we should have approximately $[Cdc42]_{min,\Delta bem1}\tau_{pol,rescue} = [Cdc42]_{min,WT}\tau_{pol,WT}$, as we expect reasonably similar behaviour for low GAPs.

Initially, average time between bursts $\tau_{b,av}$ and average burst size $p_{b,av}$ are taken from FACS data on *GAL1-GFP-CDC42* normalized to the endogenous *CDC42* expression as found from *pCDC42-GFP-CDC42* (see Appendix E: Reconstruction of burst parameters from FACS data). Also, $C_2/C_1 = 2$ is assumed, as that value seems to follow from literature data that best matches a constant area growth rate for isotropic and polarized growth. Then a first sweep is done for the value of $[Cdc42]_{min,WT}$. An apparent plateau in doubling times as a function of $[Cdc42]_{min}$ runs from roughly 3 to 25 [# proteins/ μm^3] with a small minimum close to 7. The $\Delta bem1$ threshold appears to be located around 130 to 160 [# proteins/ μm^3]. So, there is likely a factor 20 difference between $[Cdc42]_{min,WT}$ and $[Cdc42]_{min,\Delta bem1}$. For arguments given in the coarse-graining section, this should also be reflected in τ_{pol} , so that $\tau_{pol,WT} = 20\tau_{pol,rescue} = 500$ to get approximately the same behaviour for low [GAP].

Then a new sweep at low $[Cdc42]_{min,WT}$ can be done for WT parameters (so using $\tau_{pol} = \tau_{pol,WT}$ and a range of $[Cdc42]_{min}$ close to 7. The new target $[Cdc42]_{min,WT}$ follows from finding the minimum of doubling time with respect to $[Cdc42]_{min}$, which does not shift much thanks to the plateau around that value. Then the ratio C_2/C_1 is adjusted such that the doubling time is 83 min., and $C_2/C_1 \approx 2.13$ is found.

Now we supply the time to reach approximate equilibrium at 5 million cells to the algorithm to probe the final population distribution of protein copy numbers, yielding the mean Cdc42 amount. $p_{b,av}$ is then adjusted accordingly to match 8690 from [234] as close as possible. It turns out a correction factor from the FACS data of 0.8133 is needed to match this value.

If needed, this round of calibration can be repeated to get a completely self-consistent system where $[Cdc42]_{min,\Delta bem1}\tau_{pol,rescue} = [Cdc42]_{min,WT}\tau_{pol,WT}$, $2 < C_2/C_1 < 3$ such that $\tau_{doubl,WT} = 83$, $p_{b,rel}$ such that $\langle Cdc42 \rangle \approx 8690$ and $[Cdc42]_{min,WT} \approx \underset{[Cdc42]_{min}}{\text{argmin}} \tau_{doubl}([Cdc42]_{min})$. In this dissertation, $[Cdc42]_{min,WT} \approx 4$ was found, with $C_2/C_1 \approx 2.13$, $\langle Cdc42 \rangle \approx 8646$ and $\tau_{doubl,WT} \approx 82.4$.

3.7.1.4 Model flexibility

Tests for relaxing assumptions for this model can be found in Appendix I: Relaxation of model assumptions. These show that the calibrations provide quite stringent conditions, in the sense that the main results do not depend very much on exact assumptions underlying these parameters. The robustness of the results hence suggests identifiability problems (when doubling times are the main source of observations) for some non-fitting parameters if these would be attempted to be fitted. This provides positive evidence for the generality of the model, as these exact values of these parameter values are expected to vary between environments, details in (full) genetic background, etc.

Regarding the number of fitting parameters, one must note that this does not equal the actual degrees of freedom. Table 4 shows all the boundaries on the five fitting parameters, and given the number of data points upon which these will be fitted (e.g., [24] already contains doubling times for 16 genotypes, several mutant sizes and time until first polarization spots), this system is safely in the overdetermined regime. Including observations of ill mutants, such as $\Delta bem1$ data, makes the system well-conditioned (the fitting parameters are sensitive with respect to change in the parameters) and identifiable (the value of each parameter can be uniquely determined).

3.7.2 Model validation

3.7.2.1 Bem1 evolution data methods

The goal is to get an accurate description of the genotypes present in [24], particularly for strains subject to the rescue mechanism ($\Delta bem1$ background). The minimum Cdc42 concentration to polarize is the main fitting parameter, and represents different GAP deletion phenotypes. In section 3.2, it was shown that four parameters describe all GAP mutations present. Additionally, the *nrp1* deletion is hypothesized to reduce the minimum time required to spend in G1, as explained in Box 4.

Using the other parameters from Table 4, a $[Cdc42]_{min}$ sweep was done, ranging from normalized values $\left(\frac{[Cdc42]_{min}}{P_{WT} / (4/3\pi r_{min}^3)}\right)$ of 10^{-4} , from 0.001 to 0.01 (step 0.001), and from 0.02 to 0.2 (step 0.2) and 0.21 to 0.65 (step 0.01) for the WT and $\Delta bem1$ background, setting t_{mut} to unity. To avoid fitting on data with high experimental errors the *nrp1* deletion effect is chosen by manual fit to match the data well that has the lowest standard errors, also reducing the dimensionality of the fitting procedure. It was found that $t_{mut} = 0.75$ works well, and for this value the $[Cdc42]_{min}$ was repeated. This yielded the lines in Figure 25 and Figure 26.

The bottom x-axis in these figures resulted from the curves to the experimental data of [24]. The concentration thresholds are fitted numerically on the 16 mutants. Using Matlab R2014b *fminsearch*, the minimum of the objective function was found, which is the sum of scores to the fourth power (to penalize deviations very heavy), between model and experimental values, having the four $[Cdc42]_{min}$ parameters as variables, restricted as explained in section

3.2. Model doubling times for a particular $[Cdc42]_{min}$ in between or beyond those calculated in the $[Cdc42]_{min}$ sweep, were linear interpolated or extrapolated respectively using Matlab's *interp1* function. This completed Figure 25 and Figure 26, whose data is also represented in Table 5.

Table 5 Mutant doubling times of [24] in YPD, compared with the fit of the growth model, together with corresponding parameters. $[GAP]$ refers to the effective GAP concentration. The experimental error of $\Delta bem2\Delta nrp1$ was not measured but set at a conservative, high value of 30 minutes.

Genotypes	Doubling time (min.)	Standard error (min.)	Model doubling time (min.)	$[Cdc42]_{mi}$ ($\#/\mu m^3$)	$\frac{[GAP]}{[GAP]_{all}}$	$\frac{t_{G1,min}}{t_{G1,min,WT}}$
$\Delta bem1$	443	31	391	115	1	1
$\Delta bem1\Delta bem3$	97	5	105	43	0.38	1
$\Delta bem1\Delta bem2$	139	9	153	76	0.66	1
$\Delta bem1\Delta bem3\Delta bem2$	95	6	100	4.2	0.04	1
$\Delta bem1\Delta nrp1$	126	37	209	115	1	0.75
$\Delta bem1\Delta bem3\Delta nrp1$	91	6	87	43	0.38	0.75
$\Delta bem1\Delta nrp1\Delta bem2$	115	4	116	76	0.66	0.75
$\Delta bem1\Delta bem3\Delta nrp1\Delta bem2$	87	2	87	4.2	0.04	0.75
WT	83	5	82	3.9	1	1
$\Delta bem3$	86	5	83	1.5	0.38	1
$\Delta bem2$	80	5	82	2.6	0.66	1
$\Delta bem3\Delta bem2$	100	5	107	0.10	0.04	1
$\Delta nrp1$	96	9	68	3.9	1	0.75
$\Delta bem3\Delta nrp1$	130	20	69	1.5	0.38	0.75
$\Delta bem2\Delta nrp1$	75	30	68	2.6	0.66	0.75
$\Delta bem3\Delta nrp1\Delta bem2$	102	6	94	0.10	0.04	0.75

For the colony time at the end of the simulation, all volumes are stored and converted to an effective size (diameter), assuming spherical shapes. For all values in the $[Cdc42]_{min}$ sweep and for all backgrounds, statistics such as the mode (peak of the kernel smoothed probability density estimate, from Matlab's function *ksdensity*) or the mean were calculated, yielding Figure 29. For the modes, the bin edges from 0 to 10 μm in steps of 0.01 μm were used, as also in the modes from Figure 30, which also displays the standard deviation from the effective sizes. These are compared to the experimental values from [24].

Setting C_1 and C_2 to half the value from Table 4 (to mimic less rich medium) and using the fitted $[Cdc42]_{min}$ and t_{mut} , WT and $\Delta bem1$ mutants are simulated once again. In the last iteration, the times of the first, second, third phase of the iteration are stored (see 3.7.1.1 for the phase description). These are averaged across the population. The first phase corresponds to the cell isotropically growing while no polarization can take place. This hence relates to the time to first spot from [24]. The sum of the second and third phase corresponds

to the cell trying to meet the Cdc42 threshold to polarize, and then polarizing. This can hence be correlated to the time to all spots from [24]. Because it is difficult to reconstruct exactly what the growth conditions were, all experimental values were rescaled by the same factor such that for WT, time to first spot matches the simulated equivalent, yielding Figure 28.

3.7.2.2 Cdc42 expression sweep methods

Using the flow cytometry data for mutants with *pGAL1-GFP-CDC42* normalized to the endogenous expression *pCDC42-GFP-CDC42*, galactose concentration in the medium was related to Cdc42 expression. This expression is represented by burst parameters under various levels of galactose induction (see Appendix E: Reconstruction of burst parameters from FACS data), using the identities (given how shape and scale parameters relate to average burst size and times, [189]):

$$\tau_{b,rel} = \frac{k_{Gal}}{k_{endo}}$$

$$p_{b,rel} = \theta_{Gal} k_{endo}$$

The results are summarized into Table 6. Substituting these burst parameters into the model gives doubling time predictions. Ideally, the doubling times are compared to experiments on strains carrying *pGAL1-sfGFP-CDC42^{SW}* but the measured galactose concentration do not always match. This requires interpolating for the measured, albeit limited number of values.

Table 6 Results from fitting gamma distributions on the data of chapter 2, see also Appendix Reconstruction of burst parameters from FACS data, giving shape parameter k and scale parameter θ . Average times between bursts $\tau_{b,rel}$ and average burst sizes $p_{b,rel}$ are all relative to WT, such that endogenous expression would have values of 1.

Concentration	k	θ	$\tau_{b,rel}$	$p_{b,rel}$
0.05% Gal	0.03	0.22	121	0.76
0.2% Gal	2.7	1.2	1.37	3.88
2% Gal	10.1	0.72	0.34	2.51

Apparently, most variation exists in the average time between bursts. The average burst size must be dominated by the number of ribosomes present (translation) and therefore we will make the very rough approximation that it is constant at 2.5 relative to endogenous expression (smoothing the other data points). We can also use the information that the Gal1-promoter dynamical range is about 1000 [244]. So, given that we fix the burst size, the average burst time at 2% must be a factor 1000 more than the minimum burst size, so 198 min. at very low gal, i.e., 0.01%.

The average burst time is inter- and extrapolated linearly on a log scale (given the seemingly order of magnitude difference in the few data points present) to the galactose concentration values upon which the doubling times are experimentally measured. This yields the values from Table 7:

Table 7 Interpolated values for average times between bursts, relative to endogenous expression, for different galactose concentrations.

Gal Concentration	$t_{b,rel}$ [min.]
0.01%	198
0.015%	174
0.02%	160
0.03%	141
0.04%	129
0.05%	121
0.06%	67
0.08%	26
0.1%	13
0.2%	1.37
2%	0.34

Because the growth rates for the strains in 2.3 have been measured in synthetic medium (filter sterilized 4xCSM + 2% raffinose and 0-2% galactose instead of YPD), growth rates are different for the WT in this medium than for the WT used for calibration of the model from [24]. Therefore, the membrane area growth rate is adjusted to reach the WT doubling times in each medium, as an approximation of the effect of synthetic medium.

It is assumed that this is the only effect of synthetic medium. One could hypothesize that the burst properties get affected, for example because less proteins are produced in that situation. In the paper of [242], no change in Cdc42 expression was found if the sugar was changed from dextrose to ethanol (which likely leads to slower growth rates). Of course translation may still be affected, but this is not expected if nitrogen is not limiting [245]. Possibly, larger cells may induce more protein production (as a general effect, there could be more of all components) and Appendix C.2.2 shows variation in volumes. Volume-dependent expression is not taken into account in this model.

C_1 and C_2 are both reduced by a factor ranging from 0.1 to 1 (step size 0.1), with otherwise WT parameters (see Table 4). This yields WT doubling times at different membrane area expansion speeds. Through linear interpolation (extrapolation uses default C 's) the precise factor that needs to be multiplied to C_1 and C_2 to match the doubling times of WT in Figure 19 is determined (for the minimum, mean and maximum values within the credible interval).

Incidentally, similarly interpolating the mean effective radius and standard deviation of these, values increase from 0.02% to 2% galactose from about 2.2 to 2.8 μm , and 0.2 to 0.36 μm .

These increments are roughly consistent those observed in Figure 65. Apparently, the fitness effects of adding galactose to the medium that already has 2% raffinose, can readily be absorbed into the model as a generic influence on the membrane growth rate, in line with the presumed osmolarity effects.

The membrane area growth rates are in turn used to calculate the doubling times at each medium (minimum, mean and maximum) for the sfGFP strains, taking into account to altered production at each medium. Backgrounds (presence or absence of Bem1 and Bem3) are represented by the threshold values fitted in 3.3.1 to approximate the *GAL1pr-sfGFP-CDC42^{SW}* in an otherwise WT, $\Delta bem1$ and $\Delta bem1 \Delta bem3$ background. The lower and upper model doubling times bounds are then converted to relative fitness through dividing by the mean WT doubling times. These are then plotted on top of the experimental data (Figure 31).

3.7.2.3 *Cla4* mutant data methods

The *Cla4* mutants were created and measured by Leila Iñigo de la Cruz, as attributed in 3.6. What follows is a summary of her protocols in her Master thesis. All mutants were generated by first crossing yMP203a (a *cla4::LEU2, trp1*) with pure WT (yLL3a, *leu2,3-112 TRP1*) (yields yLIC08) and sporulating to retrieve a $\Delta cla4$ mutant (yLIC09) that has mating type α (spores that do not survive on a medium without leucine and tryptophan, but survive in a medium without only leucine. This haploid is then crossed with yLL131a ($\Delta bem1 \Delta bem3 \Delta nrp1 SPA2-GFP$ mutant) to yield yLIC1051. Through subsequent sporulation and selection on different media, different polarity mutants are obtained. Those growing on added (1 mg/ml) G418 have the *bem1* deletion, on added (0.1 mg/ml) clonNAT have the *bem3* deletion, on -uracil have the *SPA2-GFP* mutation, on added (0.45 mg/ml) Hygromycin have the *nrp1* deletion and on -leucine have the *cla4* deletion.

Doubling times were measured as follows: after first inoculating in a 96 well plate, on the next day a 10x dilution is placed into 100 μ l YPD in a 96 well plate inside a TECAN Infinite 200 PRO for about 48 hours at 30°C. The first 1000s the plate is linearly pre-shaken (amplitude 1 mm), before measurements (6-7 min. interval) and the same shaking for 260 or 330s (OD-600 +/-9 nm measurements, 25 flashes, 5 ms settle time). Weighted linear regression on background-subtracted OD values, smoothed with Matlab's *rlowess* (10 points) yielded doubling time estimates for this data set.

Ultimately, the arithmetic mean is taken of the doubling times across biological replicates, which are the weighted average of three technical replicates. The crossings could have introduced unidentified mutations, adding uncertainty to the doubling times. Therefore, by the logic of generic mutations as seen later in 4.2.1.5, the mutants also measured by [24] are rescaled to those values, and the new mutants times are rescaled by the same factor as an approximate correction, see Figure 33 and Table 8.

For the rescue mechanism, Cla4 is hypothesized to mediate the positive feedback in absence of $\Delta bem1$, see Figure 8 in Box 2. Schematically, the influence on the ability to polarize was found in Figure 32. This leads to the following identity:

$$\frac{[Cdc42]_{min,WT}}{[Cdc42]_{min,\Delta bem3}} = \frac{[Cdc42]_{min,\Delta bem1\Delta cla4}}{[Cdc42]_{min,\Delta bem1\Delta cla4\Delta bem3}} \quad (3.5)$$

From fits in 3.3.1 the left-hand side is known, so it shows that the $cla4$ deletion can be modelled using only one additional parameter, which given the role of Cla4 must also satisfy:

$$[Cdc42]_{min,\Delta bem1\Delta cla4} > [Cdc42]_{min,\Delta bem1}$$

Note how we cannot determine $[Cdc42]_{min,\Delta bem1\Delta cla4}$ exactly from that genotype (that combination of mutations is synthetically lethal) but $[Cdc42]_{min,\Delta bem1\Delta cla4} = [Cdc42]_{min,\Delta bem1\Delta nrp1\Delta cla4}$ and we can determine the latter.

The normalized $[Cdc42]_{min}$ sweep $\left(\frac{[Cdc42]_{min}}{P_{WT} / (4/3\pi r_{min}^3)}\right)$ was extended from 0.61 to 1.2 (step 0.01) for the $\Delta bem1$ background. Fits are performed similarly to those in the Bem1 evolution data set, this time minimizing the objective from the four doubling times of the $cla4$ mutants in Table 8.

Table 8 Mutant doubling times in YPD, rescaled to match isogenic mutants of [24], compared with the fits of integrative growth model, together with corresponding parameters. $[GAP]$ refers to effective GAP concentration. Experimental errors of $cla4$ mutants were assumed to be the same as for corresponding $CLA4$ strains of [24].

Genotypes	Doubling time from [24] (min.)	Standard error (min.)	Unscaled new doubling time (min.)	Rescaled new doubling time (min.)	Model doubling time (min.)	$[Cdc42]_{min}$ ($\#/\mu m^3$)	$\frac{[GAP]}{[GAP]_{all}}$	$\frac{t_{G1,min}}{t_{G1,min,WT}}$
$\Delta bem1$	443	31	-	-	399	115.3	1	1
$\Delta bem1 \Delta cla4$	-	-	∞	∞	∞	140.2	1	1
$\Delta bem1\Delta bem3$	97	5	121	97	105	43.2	0.38	1
$\Delta bem1\Delta bem3 \Delta cla4$	-	-	135	108	115	52.7	0.38	1
$\Delta bem1\Delta nrp1$	126	37	86	126	209	115.1	1	0.75
$\Delta bem1\Delta nrp1 \Delta cla4$	-	-	109	160	465	140.2	1	0.75
$\Delta bem1\Delta bem3 \Delta nrp1$	91	6	93	91	87	43.2	0.38	0.75
$\Delta bem1\Delta bem3 \Delta nrp1 \Delta cla4$	-	-	89	87	93	52.7	0.38	0.75

3.7.3 Model extension methods

3.7.3.1 Media types

To mimic the effect of poorer medium, this is reflected in the model as before, by reducing C_1 and C_2 . The same threshold sweep was conducted as for the Bem1 evolution data set, only this time halving both the membrane area growth parameters, yielding Figure 34.

3.7.3.2 Epistatic interactions

BIOGRID data [58] (as of 08-03-2018) was taken to obtain all interactions with *BEM1* that are either positive genetic, synthetic rescue, negative genetic, synthetically lethal. The former two are considered “positive”, and the latter two “negative”. For those mutants with multiple, conflicting interactions the majority (of the low throughput data, if available) vote prevails, and in the case of a tie the mutant is discarded from further analysis. Interactions with *BEM2*, *BEM3*, *CDC24*, *CLA4* and *NRP1* are deemed to be known with such certainty that they are manually overwritten to positive, positive, negative, negative and positive respectively, regardless of other literature.

For the effects on other mutants, data of SGD [57] was used, more specifically from <https://www.yeastgenome.org/observable/APO:0000110> (date of access 06-03-2018), <https://www.yeastgenome.org/observable/APO:0000308> (date of access 06-03-2018), <https://www.yeastgenome.org/observable/APO:0000141> (date of access 08-03-2018), <https://www.yeastgenome.org/observable/APO:0000255> (date of access 08-03-2018), and through the search button and genes subsection on the SGD website with queries proteasome, ribosome and phospholipid. For the (fermentative) fitness, only null mutants in YPD are considered. To maintain the sample size as large as possible, for the size and G1 time mutants not only null mutants are considered, but also those with reduction of function and overexpression mutants.

These mutants are then all correlated with the positive and negative interactions documented for Bem1. The posterior density distribution for the fraction of positive epistatic interactions compared to the total epistatic interactions for every class of mutants is then calculated, using a uniform, uninformed prior. Using the binomial distribution as a logical likelihood function for how the data follows from the positive epistatic fraction parameter, this leads to a beta distribution for its posterior density, from which the most probable value and 95% credible interval is plotted in Figure 35. By differencing Monte-Carlo samples of each mutant case and the total Bem1 epistatic interaction set as a reference, the posterior odds of the aforementioned model hypotheses can be constructed.

4 Evolutionary roles of noise and transgenerational feedback

4.1 Consequences of protein expression noise

In the previous chapter, several non-intuitive cell polarity phenotypes of budding yeast, such as epistasis between GAPs within a *Δbem1* background, were successfully described using a growth model. This involved making assumptions on membrane expansion, protein turnover and the mesotype, the coarse-grained rule set emerging from molecular mechanisms. In order to understand what exactly is crucial in order to describe e.g., the epistasis (the phenotype most transferrable and comparable to other systems), the first goal of this chapter is to further reduce the model. Consequently, once the minimum requirements have been identified, it becomes possible to improve our mechanistic understanding of the origin of epistasis in more general terms (adding to those known [246]), beyond yeast polarity.

As discussed in section 3.5 and considering results in Appendix I (insensitivity of epistasis to certain model assumptions) and in literature (e.g., degradation only matters for 15% of the proteins [235]), at first sight the critical components are noisy protein expression and the mesotype. This implies discarding, among others, protein degradation and cell volume effects. The importance of noisy gene expression influencing phenotypes under selection has been featured in literature. On the one hand, noise is known to be an integral part of evolutionary strategies with various benefits, see e.g., [233], [247]. As described in [233], a heterogenous population for example in terms of metabolism, can help the colony anticipate environmental fluctuations or survive antibiotics. Authors in [247] show how higher noise levels in gene expression can improve Darwinian fitness in yeast and even speed up genetic adaptation.

On the other hand, noise is also considered a process against which cells intend to buffer, whether this is variation in protein concentration or environments [248]. The buffering results in widely observed mutational robustness, often referred to as canalization [249]. Particularly when fitness depends sharply on protein concentrations, low expression noise seems critical, explaining the typically quiet, observed expression distributions in this case [20]. In this paper, authors reason that this observation suggests evolution shapes promoter noise to favour robustness, as population fitness otherwise suffers from noise-driven deviations of protein copy numbers to suboptimal levels. However, in this chapter I will argue that stochastic variability can be the source of robustness, rather than the objective to be minimized.

Essential for this claim is the expected memory feature in the population protein distribution, which is caused by expression noise. Simply put, a population where the average (stochastic) protein production is suboptimal given the mesotype, will improve (or deteriorate less) its point of departure amongst the most prolific survivors with every generation, compared to the situation without expression noise. This changes the protein distribution across the population in a non-trivial way. As the mesotype leaves an inherited, non-genetic imprint on

the population, this epigenetic mechanism can be considered a *transgenerational feedback*, different from other known epigenetic mechanisms in [40] and more along the lines of [41]. The important difference is that the selection takes place with respect to suboptimal genetic environments instead of the “tangible” environment.

Box 5 Terminology in Chapter 4

Term	In this context	Symbol
Cell cycle time	Fixed time step between division decision events	T_0
Fitness	Population fitness, reciprocal of doubling time	ω
Relative fitness	Population fitness scaled by the maximum fitness attainable in the model, set by the cycle time	ω_r
Protein concentration	Number of proteins inside the cell normalized by the mean expressed number of proteins, equivalent to normalized concentration as volume is constant here.	x
Mesotype	Progeny as function of protein concentration	$g(x)$
Mesotypical inflection (MI) point	Inflection point of mesotype function g , loosely the concentration ‘half-way’ between best and worst outcome in terms of progeny	c_i
High/low protein state	Discrete state where cells have higher/lower amounts of protein relative to the MI point at the moment before division.	$f(x)$ where $x = 4c_i/3$ and $2c_i/3$ right before division, $2c_i/3$ and 0 afterwards
Expression	Added protein concentration during a cell cycle time. As concentration states are discrete, expression is described by a probability of state change following evaluation of a cumulative distribution function.	$F_V(x)$ or shorthand $F(x)$
Noise level	Coefficient of variation associated with the expression	V
Mesotype-preserving mutation	Mutation that is represented in the model solely by a shift in c_i , such as the GAP deletions which shift the minimum Cdc42 threshold (MI-point)	Mutation size Δc_i
Epistasis generating function	2 nd derivative of relative fitness w.r.t. the MI-point, whose sign represents positive (negative) epistasis for same (opposite) sign mesotype-preserving mutations.	η
Growth factor	Factor by which the population grows per generation	γ
Transition matrix	2-by-2 matrix containing all probabilities to maintain or switch protein states per cell cycle	M
Sustainably viable region	Region in concentration space outside which given enough time, any population eventually dies out	

To illustrate the noise mechanics behind epistasis and the transgenerational feedback that influences evolvability, the following section (4.2) describes a model where fitness emerges solely from a simple mesotype (sharp concentration threshold) and noisy Cdc42 expression. In order to help the reader through the vocabulary used in this chapter, Box 5 describes some commonly used definitions.

Once the familiar Cdc42 case has been described, the model is generalized to incorporate other mesotypes. This will prove sufficient to provide more insight into existing experimental bulk studies on epistasis. Section 4.3 will then treat the transgenerational feedback, quantifying this effect and generating concrete predictions on its relevance. For conciseness, the interested reader is referred to Appendix M for mathematical derivations. Finally, an experimental design for the validation of the feedback is expounded in 4.4, making use again of the well-studied yeast polarity network.

4.2 Minimal model for epistasis

4.2.1 “Cdc42/GAP” test example

4.2.1.1 Model definitions

The goal is to describe the epistasis of the GAPs in the $\Delta bem1$ accounting only for the absolute minimum number of processes in the cell. Therefore, the Cdc42 mesotype is reduced to only a sharp lower concentration boundary in order to successfully divide. This means that either the progeny of a cell is 0 or 2, depending on protein concentration (see also Figure 36). It is therefore attractive to only consider two protein concentration states, called low and high. These are always relative to the location of the threshold, which differs per genotype as in Chapter 3.

Cell volume growth and cell cycle events are disregarded, reducing the life of a cell to only the division events. This also means that copy number and concentration can be used interchangeably, as volume is fixed. Every set time T_0 , every cell divides or dies based on the mesotype. The protein production during this time until division is stochastic, based on a sensible probability density function with coefficient of variation V , dubbed the noise level. Based on both theoretical¹¹ and experimental arguments (considering 99% of observed coefficients of variation in yeast are smaller than 1 [229]), we will further assume $0 < V < 1$.

¹¹ Fixing the cell cycle time in the context of Chapter 3 implies that the protein produced per cell cycle is a Coxian random variable [250]. This results from summation of unequal exponentially distributed random variables, generating a hypoexponential distribution (or Erlang, without degradation), with the number of exponentials also as a random variable (Poissonian number of bursts within a fixed time). Biologically, the latter implies the mRNA lifetime being notably smaller than the protein life time [189]. For the hypoexponential distribution coefficient of variation is known to be smaller than 1 [251].

The model for the Cdc42 case is graphically depicted in Figure 36. Alternatively, the two-state model is mathematically represented by a transition matrix governing the probabilities to switch states. The entries of this 2-by-2 matrix are dependent on the probability distribution for protein generation and the mesotype. In matrix form, the two-state model reads:

$$\begin{bmatrix} f(2c_i/3) \\ f(0) \end{bmatrix} = \begin{bmatrix} 2(1 - F_V(2c_i/3)) & 2(1 - F_V(4c_i/3)) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} f(2c_i/3) \\ f(0) \end{bmatrix} = M_{Cdc42,s} \vec{f} \quad (4.1)$$

where symbols are described in Box 5.

At steady state, the population grows according to the largest eigenvalue λ_{max} of transition matrix M , which we can rewrite to fitness as:

$$\omega = \frac{\log_2 \lambda_{max}}{T_0} \quad (4.2)$$

The relative fitness then becomes:

$$\omega_r(c_i) = 1 + \log_2(1 - F_V(2c_i/3)) \quad (4.3)$$

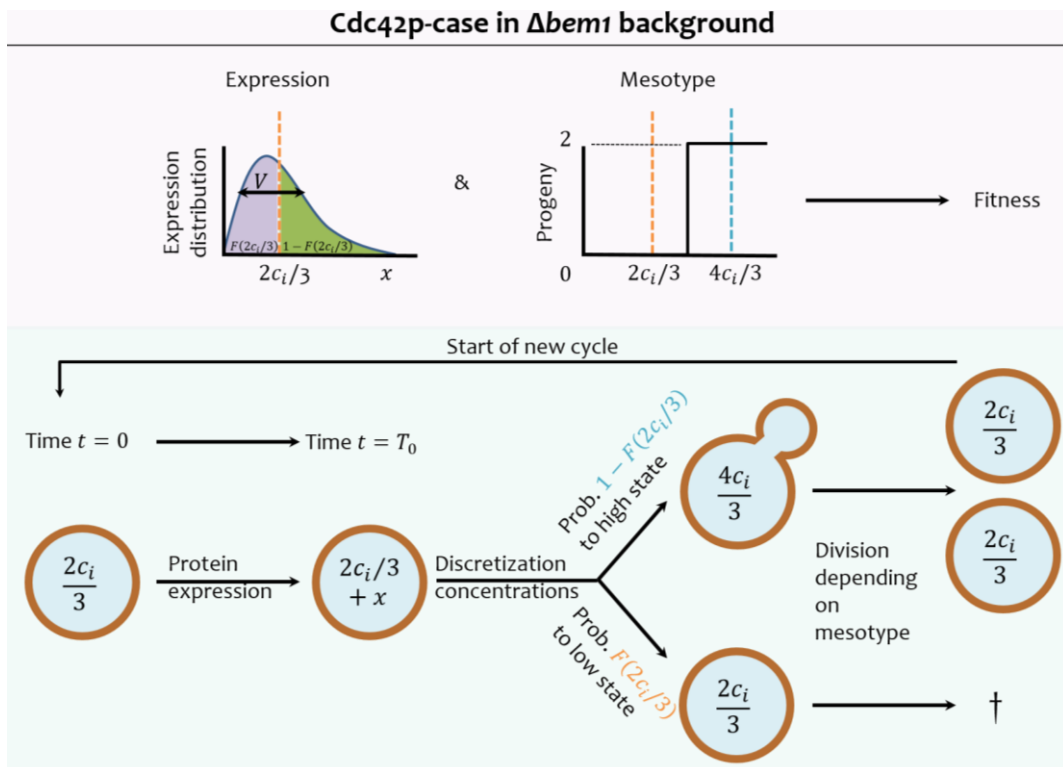


Figure 36 Two state model for fitness reflecting yeast cells polarizing with success depending solely on Cdc42 concentration. Noisy expression during cell cycle time T_0 , combined with the simplified mesotypical rule for progeny depending on Cdc42 concentration yields fitness, as stochastic switching occurs between the low and high Cdc42 concentration state, with probabilities following from cumulative distribution function F , which is a function of noise level / coefficient of variation V .

4.2.1.2 Epistasis definitions

The definition of choice for epistasis is relative multiplicative epistasis, which implies dependence of the relative fitness effect of a mutation on the genetic background. This is mathematically defined as [252]:

$$\epsilon = \log_{10} \left(\frac{\omega_r(c_1)}{\omega_r(c_4) \prod_{i=2,3} \frac{\omega_r(c_i)}{\omega_r(c_4)}} \right) = \log_{10} \left(\frac{\omega_r(c_4)\omega_r(c_1)}{\omega_r(c_2)\omega_r(c_3)} \right) \quad (4.4)$$

which e.g., for our target GAP epistasis in the $\Delta bem1$ background means that c_1 is the mesotypical inflection point of Cdc42 in the mutant $\Delta bem1 \Delta bem3 \Delta bem2$, c_2 corresponds to $\Delta bem1 \Delta bem2$, c_3 to $\Delta bem1 \Delta bem3$ and finally, c_4 to $\Delta bem1$. From the doubling times in [24], the observed epistasis is -0.49.

In this model, the most important mutations take into consideration *mesotype-preserving mutations*, i.e. those that cause a shift in concentration space of the mesotype (progeny g), so a shift in MI-point c_i . For the Cdc42 case, this includes GAP activity mutations (such as deletions), because we saw in Chapter 2 how the *bem3* and *bem2* deletions lowered the minimum [Cdc42] threshold to polarize. As concentration values such as c_i are normalized to mean expression per cycle, which remains unchanged in this example as cycle time is fixed, deletion of GAPs invariably decreases the mesotypical inflection point.

More generally, mesotype-preserving mutations include mutations that influence the copy number of the protein. Again, as c_i is normalized to mean expression, when the latter changes c_i changes accordingly (e.g., increased expression lowers c_i). Therefore, expression changes are always reflected in the MI-point, such as due to promoter or 5'UTR mutations, mRNA interactions, activity enhancers or suppressors.

Indirectly, other mutations that change fitness but do not influence the protein of interest are also felt in c_i . As these mutations can only be generically incorporated into the model by a change in T_0 , the average expressed protein per cycle changes accordingly, subsequently shifting c_i . This class of generic fitness mutations will also be considered in section 4.2.1.7. An overview of the defined mutation types can be found in Figure 37. However, mutations that change the shape of g as a whole (e.g., mutations that for example make the protein of interest completely redundant) are beyond the scope of this model.

To show how epistasis is easily generated for mesotype-preserving mutations, let us consider the special case where *no epistasis* is present. In order to have no epistasis through the definition of 4.4, the fitness must be of the form:

$$\omega_r(c_i) = e^{-\alpha c_i} \quad (4.5)$$

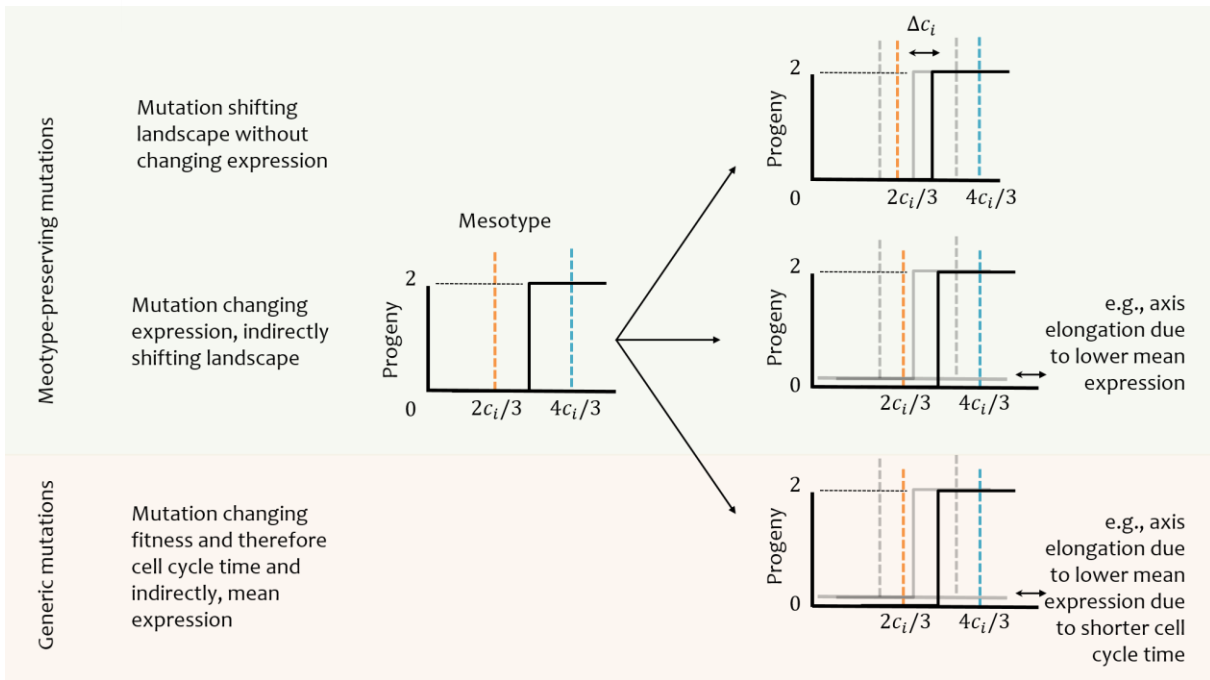


Figure 37 Graphical representation of mesotype-preserving and generic mutation definitions.

Consequently, *no epistasis implies that the slope of the log of the fitness is constant* (as this is multiplicative, not additive epistasis). In that case, the second derivative of the log fitness is constant. Conversely, non-zero second derivatives of the log fitness show regions where mutations that shifts c_i cause non-zero epistasis. This prompts the definition of an “epistasis generating function” η , as the second derivative of the log fitness:

$$\eta(c_i) \equiv \frac{\partial^2 \log_{10} \omega_r(c_i)}{\partial c_i^2} \quad (4.6)$$

Subsequently, when the log fitness is contour-plotted as a function of inflection point c_i and noise level V , the spacing between contours for constant V shows in which regions of concentration space mutations harbouring epistasis can exist. As an example, a purely hypothetical case with log fitness contours as a function of noise level and mesotypical inflection point is plotted in Figure 38. The violet arrows denote mesotype-preserving mutations, that cause a constant translation in c_i given a certain V , starting at various genetic backgrounds.

When considering the log relative fitness contours across the dashed line, which is along a constant noise level, the mutations cause the same shift in log relative fitness, regardless of the starting point (the population moves up exactly one contour). Consequently, there should be no epistasis involved with these mutations, because its relative effect is always the same. Concordantly, the epistasis generating function (lower part of the figure) is constant at zero, corresponding to no possible epistasis.

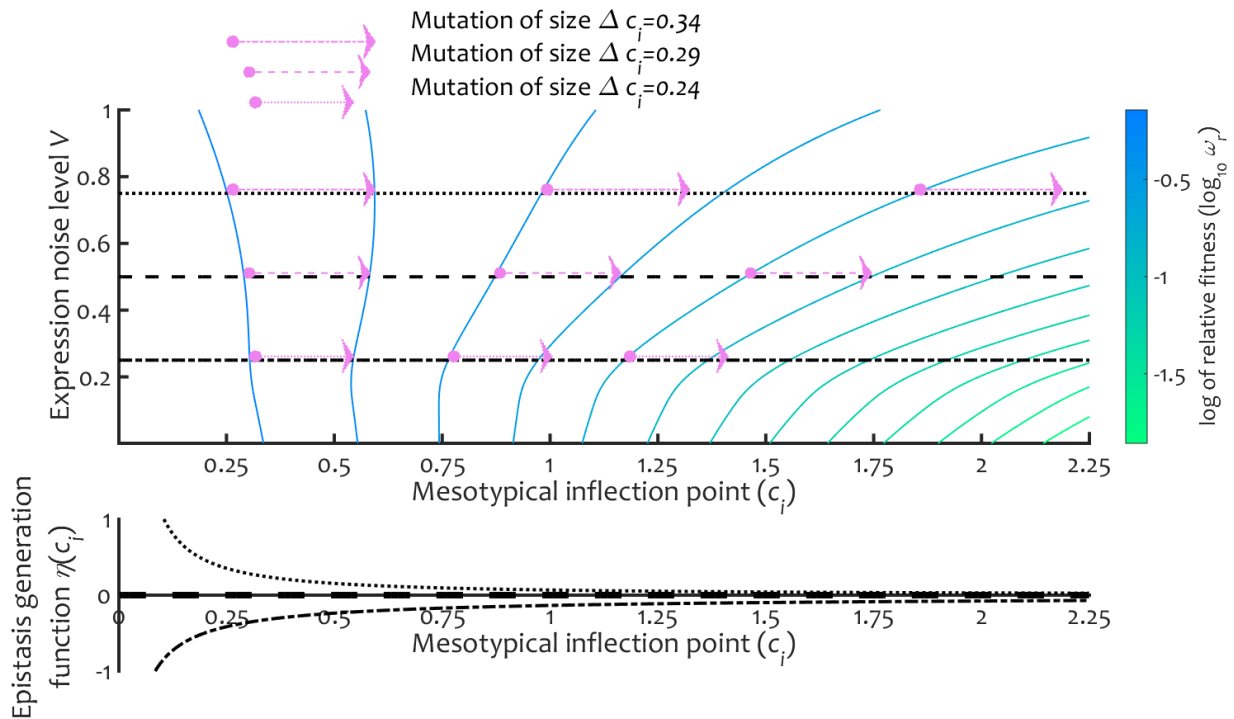


Figure 38 Visualisation of epistasis by means of fitness landscape contours. Top: Contour plot of log relative fitness as function of concentration and the coefficient of variation, for a hypothetical (non-biological) case. Violet arrows denote shifts in mesotypical inflection point c_i and are of constant size for a given V , representing mutations from the dot at the arrow base to the point in state space where the arrow head is. Bottom: Second derivative of fitness with respect to inflection point along the dashed, dotted or dash-dotted lines of the top axis of the log relative fitness. Only the dashed line is always zero, not allowing any epistasis. Positive (negative) values of the second derivative imply positive (negative) epistasis between mutations that shift c_i in the same direction.

By contrast, the mutations along the dotted line have decreasing relative impact on fitness when the c_i starting points are higher. Consequently, the genetic background matters for the mutational effect, and this is reflected in a non-zero value for the epistasis generating function η . The positive value of η represents positive epistasis between mutations that shift c_i in the same direction. By the same token, the mutations along the dash-dotted line have increasing relative impact when the c_i starting points are higher. Again, the genetic background matters for the mutational effect, and this is reflected in a negative value of η . This represents negative epistasis between mutations that shift c_i in the same direction.

4.2.1.3 GAP epistasis description is not dependent on expression noise distribution

The two-state model that includes noisy Cdc42 expression described in Figure 36 is supposed to explain the epistasis in the GAPs, defined by equation 4.4. Using the fitness definition of equation 4.3, the goal is to approach the observed value of -0.49. However, the question is which noise distribution is most appropriate.

Assuming for the noise level $V = 0.832$ as measured from data used in Chapter 3, Figure 39 shows common and less common choices of noise distributions fit at least 60% of the GAP

epistasis in a $\Delta bem1$ background (70% for more sensible production distributions). The insensitivity with regards to noise distribution suggests that epistasis is not generated by the presence of any distribution in particular, but by sufficient noise of any kind. Once the locations for the concentrations are fitted using the observed coefficient of variation, varying V greatly affects the epistasis that would have been observed (Figure 78 in Appendix K). This means the noise level is critical for the epistasis, but the exact way the noise is generated is not important. This motivates that for the remainder of the chapter we can safely alternate between a gamma (the more natural choice [189]) and log-normal distribution (more tractable mathematics), without much loss of generality.

The important message is hence that this minimal model suffices to describe the observed GAP epistasis. Even the simplistic view of binning protein levels into two states does not adversely affect this outcome. Figure 40 compares the shape of the relative fitness of this model with that of the more advanced model of Chapter 3 (the blue line in Figure 25, where the minimum Cdc42 concentration to polarize equates to the mesotypical inflection point). While the fixed volume assumption does not allow 1:1 comparison of concentration thresholds, the qualitative match for quantities as epistasis is sufficient. The two-state model assumptions such as binning of protein concentrations do not notably cause fitness artefacts. This generates trust to proceed with this model, to derive other claims regarding the polarity system and ultimately, to suggest more general mechanistic lessons on epistasis, a tractable goal given the minimal nature of this model.

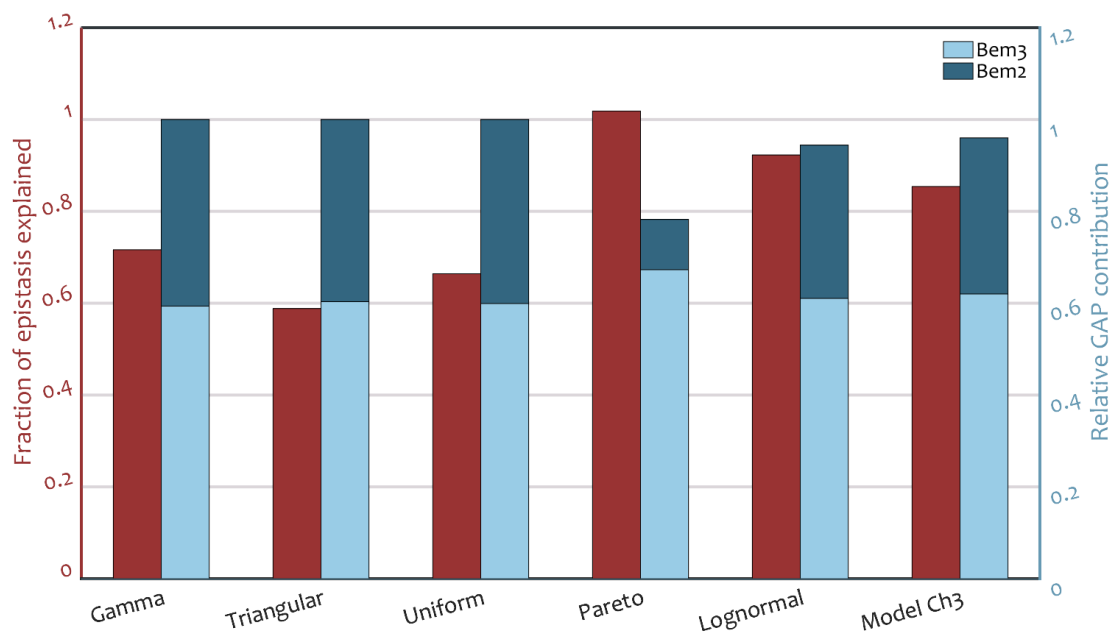


Figure 39 Distribution type of expression noise is not important for describing GAP epistasis in a $\Delta bem1$ background. Left vertical axis: epistasis explained for the GAPs in the $\Delta bem1$ background in the two-state model when different Cdc42 production distributions are assumed (last column model 3 results), with coefficient of variation of 0.832 (as observed for Cdc42), except the triangular and uniform distribution (set to the maximum CV with positive support). Right vertical axis: fraction of GAP action for Bem3 (light blue) and Bem2 (dark blue) respectively.

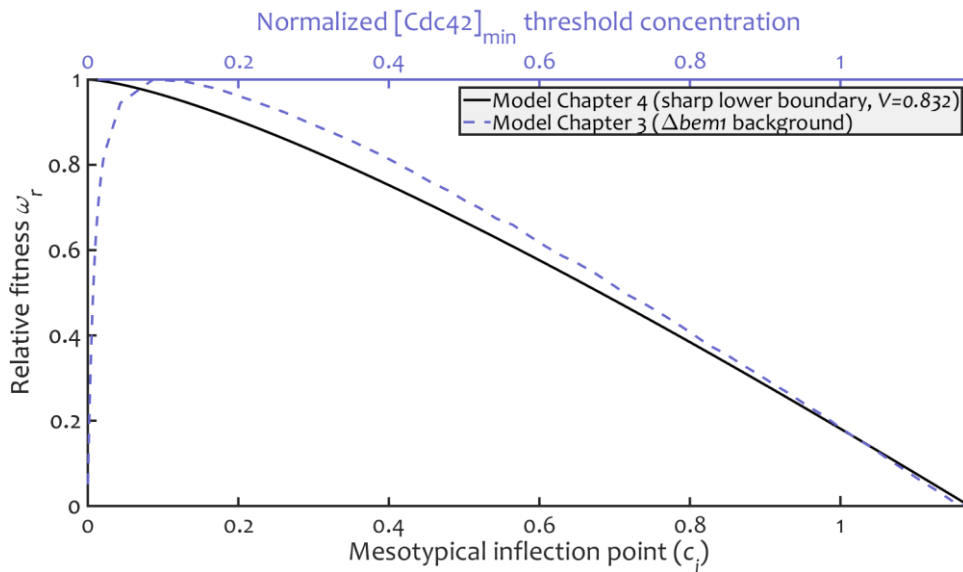


Figure 40 Comparison of fitness curves as function of *Cdc42* threshold location, for the models from chapter three (dashed blue line, top blue x-axis) and the two-state model of this chapter (black line, bottom x-axis), relative to mean expression. The curve of the chapter three model is equivalent to Figure 25, with the $[Cdc42]_{min}$ -axis normalized to match the end of the two-state model curve to allow a qualitative model comparison. The two-state model curve follows from 4.3 assuming a gamma distribution and the observed noise level $V = 0.832$ for *Cdc42*.

4.2.1.4 Noise can benefit steady state growth for the GAPs, not for *Cdc42*

As seen from 4.3, the population remains viable as long as the mesotypical inflection point c_i is less than 1.5 times the median of protein production distribution F_V . Here we also see the downside of excess noise (high coefficient of variation V). In the worst case of a maximum difference between median and mean (assuming a unimodal production distribution) [253], the maximum threshold the population can withstand decays as:

$$c_{i,lim} = \frac{3}{2} - \frac{3}{2}V\sqrt{3/5} \quad (4.7)$$

However, for the log-normal distribution (a sensible choice for the protein production distribution), the deleterious effect is less severe than the maximum, namely:

$$c_{i,lim} = \frac{3}{2\sqrt{1+V^2}} \quad (4.8)$$

An interesting note is that in a variation of this model reflecting noisy GAPs (still in the $\Delta bem1$ background, see Figure 79 of Appendix L), the opposite occurs. There, the mesotype involves a sharp upper bound, and this changes the limiting concentration to $(3/4)/\sqrt{1+V^2}$. In this case, noise is actually beneficial. Yet, the GAP coefficients of variation are so low (<0.15) [229] that the viable concentration region only expands with 1%, so no real advantage is taken from this situation. The retained potential is that a non-viable GAP genotype can be evolutionary rescued by an adaptation that simply increases the noise. This GAP case is in line with the statement in [247], that noise can sometimes increase mean fitness.

Still, this is contrary to the common opinion about noise, which is generally considered deleterious (e.g., [254]). Also, the idea (presented in [255]) that essential genes should have less noise can still hold, but can be extended. We also expect that toxic genes (for which too much gene product is deleterious) could have more noise than purely non-essential genes.

4.2.1.5 Noise increases likelihood of negative epistasis mutants

The larger the coefficient of variation, the larger also the region where epistasis may take place. This is exemplified by Figure 41, where the log relative fitness is contour-plotted for different mesotypical inflection points and noise levels (blue to green contours). As explained, the spacing between log doubling time contours illustrates whether epistasis is present. When the noise level increases, fitness contours disperse, leading to a larger region in concentration space with a significantly positive slope in log relative fitness. This is the region where epistasis is generated.

As a measure of the width in the concentration ratio space for notable epistasis, we define w_e : the difference between the maximally (or minimally) allowed mesotypical inflection point and the inflection point where the relative fitness is $\Delta\omega_r$ less than the maximum (of 1). In practice, the experimental precision for which doubling times can be measured sets $\Delta\omega_r$.

$$w_e \stackrel{\text{def}}{=} c_{i,lim} - c_{i,\omega_r=1-\Delta\omega_r} \quad (4.9)$$

For the Cdc42 case, and assuming the sensible log-normal distribution, we obtain:

$$w_e = c_{i,F_V(2c_i/3)=1/2} - c_{i,\omega_r=1-\Delta\omega_r} = \frac{3/2}{\sqrt{1+V^2}} - \frac{3}{4} \exp\left(V\sqrt{\frac{\pi}{2}} \operatorname{erf}^{-1}(1 - 2^{1-\Delta\omega_r}) - V^2/2\right)$$

Using the tanh approximation for erf [256] and assuming high precision measurements and small V (usually true in yeast, 95% has $V \leq 0.36$ [229]), it is possible to expand to yield:

$$w_e = \frac{3V}{4} \sqrt{\frac{\pi}{2}} \ln\left(\frac{1}{\Delta\omega_r \ln 2}\right) \quad (4.10)$$

So, the width in the observed fitness landscape in which negative epistasis can be found, scales mostly linearly with the amount of noise (V as a measure) and incidentally, is proportional to the log of the reciprocal of minimum relative fitness difference $\Delta\omega_r$ (e.g., 0.1) that can be measured. Analogously, for the GAP case, whose mesotype is subject to a sharp upper boundary, the width is quite similarly given by:

$$w_e = \frac{3V}{8} \sqrt{\frac{\pi}{2}} \ln\left(\frac{1 - \Delta\omega_r \ln 2}{\Delta\omega_r \ln 2}\right) \quad (4.11)$$

4.2.1.6 Protein expression noise as evolutionary capacitor

When the population is outside the viable region so when the location of the sharp boundary is at $c_i > c_{i,lim}$, noise gives the population an advantage compared to no noise. A simple example is when the population starts with very low protein concentrations (near zero). With noise, part of the population will be able to divide for another few rounds by chance, which is not possible without noise and $c_i > 3/2$. Specifically, the part of the population surviving the first round is given by $1 - F(4c_i/3)$, and every subsequent round a fraction $1 - F(2c_i/3)$ of the remainder survives. A population starting out with size N_0 will then have approximately the following number of generations to find a mutation that allows more durable survival:

$$2(1 - F(4c_i/3))(2 - 2F(2c_i/3))^{n_{gen}-1} \approx 1/N_0 \quad (4.12)$$

$$\Rightarrow n_{gen} = \frac{\log N_0 - \log\left(\frac{1 - F(2c_i/3)}{1 - F(4c_i/3)}\right)}{-\log\left(2(1 - F(2c_i/3))\right)} \quad (4.13)$$

For example, assuming once again the log-normal distribution, and expanding to first order near the boundary $c_i = c_{i,lim}$:

$$n_{gen} \approx \frac{V\sqrt{\pi/8} \log N_0 - \log 4}{c_i/c_{i,lim} - 1} - 1 \quad (4.14)$$

After an initial bump (surviving the first round of division is the most difficult), the survival time increases logarithmically with the population size, linearly with noise level and is inversely proportional to the distance to the sustainable survival limit $c_i = c_{i,lim}$.

For the average $V = 0.22$ [229] and a typical population size of a million, the population usually survives 2 generations even when $c_i = 1.7$, and 4 to 5 for $c_i = 1.6$. If the noise is relatively large, as for Cdc42, the amount of generations that will be survived becomes 11 when $c_i = 1.7$ and still 7 when $c_i = 2$. As another (slightly less extreme) example, Figure 41 contains a contour plot (red to yellow contours) which shows how many generations a population can survive given a certain c_i and V , starting with a population size of 10^6 , when assuming a gamma distribution for protein production. In the figure, the most favorable initial condition is used, to show the maximum number of generations that can be survived.

As can be seen from Figure 41, high noise can allow growth for dozens of generations, well beyond the sustainable growth limit (purple line, bottom panel). This more than makes up for the deleterious dependency of the limit on the noise (as derived in 4.8). In Appendix L, Figure 80 shows the same plot, but from the perspective of the GAPs (sharp upper boundary). Analogously, even if there are insufficient GAPs, the population survives for many generations, but the purple sustainable growth limit even favors more noise here. The epistasis generating functions are discussed in the next subsection.

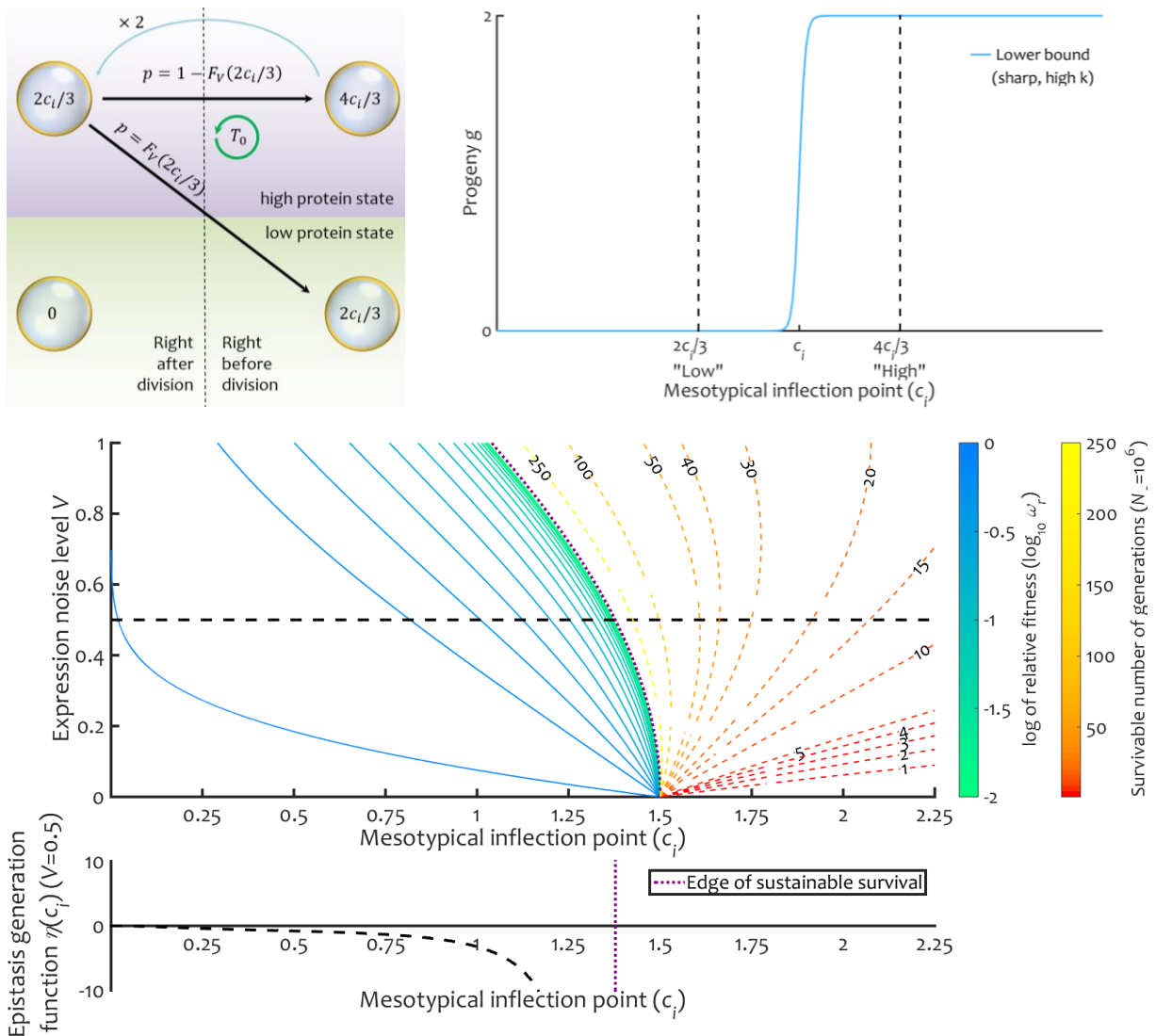


Figure 41 Noise in *Cdc42* expression limits the sustainable growth region, but extends unsustainable growth, as well as the region that harbours epistasis. Top: Two state model schematic summary (left) and mesotype (right) of the *Cdc42* case. Middle: Contour plots of log relative fitness (blue to green) and number of generations a population of a million cells survives (red to yellow) as function of mesotypical inflection point of the sharp lower boundary in the fitness landscape, and the noise level (coefficient of variation), assuming a gamma protein expression distribution. The purple dotted line indicates the survival edge. The dashed black line denotes the line $V = 0.5$, along which the plot in the bottom is valid. Bottom: Second derivative of log relative fitness with respect to mesotypical inflection point (along the dashed line of the top axis). Here, it is always negative for values of c_i that support growth, allowing negative epistasis for mutations of the same sign.

One might question how a population with heavily unfavorable initial conditions could arise with such an initial size, while clearly not being very fit (asymptotically going extinct). Yet, since yeast switches between diploid and haploid states, one possibility is not so fit haploids arising from a fit heterozygous diploid. After sporulation protein concentrations may be below par, which is a one-time hurdle that must be taken. Similarly, when in prolonged period of saturation, yeast starts to recycle its own proteins (see e.g., [257]), leaving low concentrations when new food is supplied. Noise is an easy solution to these one-time hurdles.

Moreover, in chapter three we have seen how the environment, such as not so rich media, can influence fitness. There, the ill $\Delta bem1$ profited immensely from slow growth conditions, allowing temporal coexistence with the WT, as the difference in fitness was heavily reduced. Especially from a heterozygous diploid, it is easy to see how ill haploids emerge that can grow to sufficient size when the growth conditions are suboptimal.

The added role of noise shown here is that it lets the cells stall for time until the conditions change again, allowing survival of for example another 10 generation times. This delays the purifying selection of potentially lethal mutants. Ultimately, fitness valleys can be buffered to some extent this way, increasing the evolvability beyond the expected levels (or decreasing the population size one would have expected to be needed for certain evolutionary paths). The increased adaptability due to noise is sometimes overlooked in literature [254], presumably due to neglecting the main mechanism behind this beneficial noise effect. This will be explained further in section 4.3. This is also complementary to the positive effect of noise described in [247]. In that paper, a colony with noisier expression may significantly reduce its time to genetically adapt mean expression to e.g., new environments.

4.2.1.7 Sharp lower mesotypical bound induces diminishing adaptive returns

In anticipation of generalizing this model to general mesotypes, as done in the next subsection, it is already possible to preliminary relate the Cdc42/GAP case to phenomenological rules of epistasis encountered in literature. In addition to the epistasis across mesotype-preserving mutations, it is possible to consider generic mutations, that are known to influence cell cycle time T_0 . These thereby influence ω and indirectly c_i due to the normalization of concentration to average protein production within T_0 . The combinations of mesotype preserving mutations and generic mutations yield the following scenario's:

- Two beneficial mutations, both shifting c_i , exhibit negative epistasis.

The sign of η shows this (see bottom plots of Figure 41 and Figure 80, Appendix L), when mutations occur in a direction where $\log \omega(c_i)$ contours bunch progressively closer together.

- Two deleterious mutations, both shifting c_i , exhibit negative epistasis

Similar to the previous point, without a sign change of the second derivative representing η , the deleterious effects compound to make each other worse, showing negative epistasis.

- One beneficial and one deleterious mutation, both shifting c_i , exhibit positive epistasis

Beneficial and deleterious mutations are in opposite direction in the $\log \omega(c_i)$ plots. Therefore, moving across the beneficial direction (lower c_i), and then moving to the deleterious direction (higher c_i) effectively means a sign change in the derivative of $\log \omega(c_i)$. In other words, the deleterious mutation is not so bad when the beneficial mutation has occurred earlier, showing positive epistasis. But as the amount of deleterious mutations

typically exceeds the number of beneficial mutations [258], the previous two cases generating negative epistasis form the majority of cases, compared to the rarer case of positive epistasis.

- One mutation that shifts c_i and a generic mutation that affects T_0 generate positive, negative and sign epistasis.

This case is a bit more complicated. Suppose a general mutation X_m where we can only know it has an effect on T_0 (no precise fitness landscape is known), yielding cycle time T_1 . If it slows down the cycle $T_1 > T_0$, this is a deleterious mutation when $c_i \ll c_{i,lim}$. However, as c_i approaches $c_{i,lim}$, this becomes beneficial at some point as the slow cycle allows more protein production, hence decreasing c_i again. This means that the mutation X_m can show sign epistasis, so is beneficial or deleterious depending on genetic background (location of c_i).

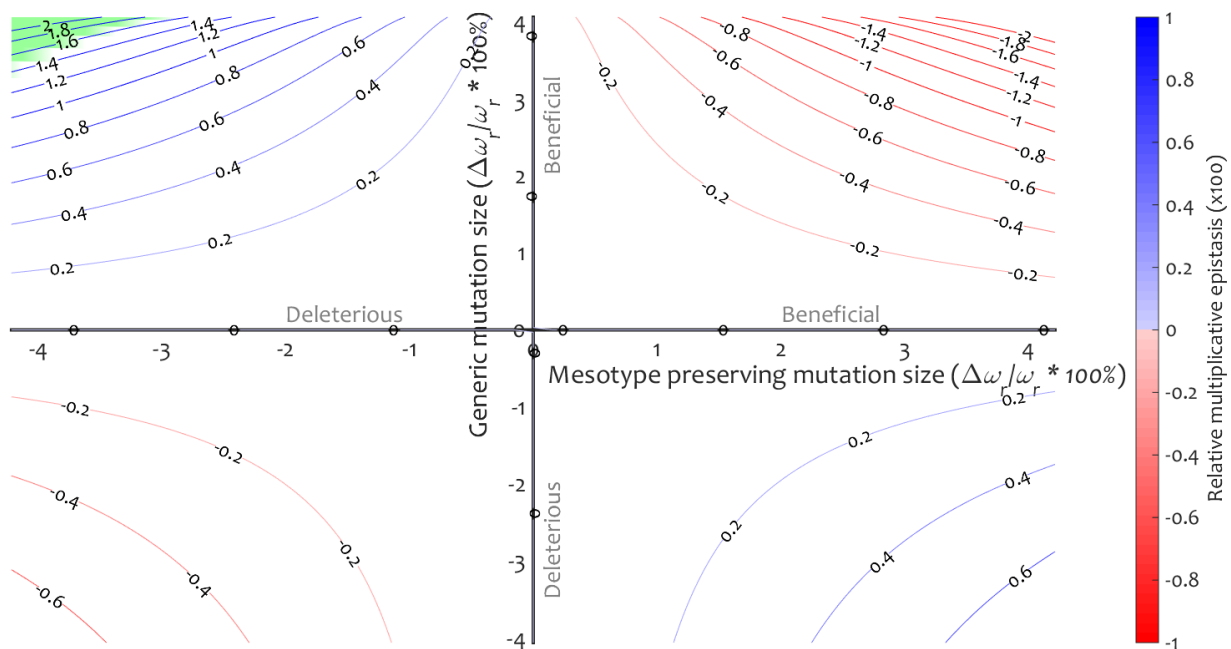


Figure 42 Epistasis exhibited in the *Cdc42* (sharp lower boundary) case. Epistasis (times 100) of typical mesotype-preserving mutations and generic mutations, affecting T_0 and subsequently fitness, in the sharp lower boundary case ($V = 0.22$). Positive (red) and negative (blue) epistasis can be seen, as well as sign epistasis of the generic mutation, marked by the green region at the left top.

As many situations are imaginable, let us restrict ourselves to the most likely case during adaptation, where beneficial mutations are typically of a <5% size [259] and a typical value of V is 0.22 [229] (see Figure 42). The combination of beneficial mutations (top right quadrant) exhibits negative epistasis. The former is in line with the diminishing returns observation during adaptation [260], which is further discussed in the next subsection on the general landscape case. Interestingly, the opposite is true for the sharp upper boundaries, but 0 will show that the chance that this case will occur is rare compared to the “*Cdc42*” case.

4.2.2 General case

4.2.2.1 Model definitions

We can generalize the previous sharp mesotypes to a smooth, monotonous¹² mesotype describing progeny as a continuous number between 0 and 2, as depicted in Figure 43. In this case, both the low or high protein state may be viable and must be considered. Assuming progeny is proportional to the protein concentration in a particular state/conformation/complex resulting from a chemical reaction, a frequent description of this concentration is through the Hill equation. This leads to the definition of g as the sigmoid:

$$g(x) = g_0 + \frac{2 - g_0}{1 + (x/c_i)^{-k}} \quad (4.15)$$

The minimum progeny in the landscape, also the “shallowness”, is modulated by g_0 . The parameter $|k|$ denotes the cooperativity (of the chemical binding) and the “sharpness” of this boundary. Positive and negative values of k reflect lower and upper boundaries respectively.

In loose terms one could view polarization as an ultrasensitive (very high-cooperativity) process (a positive feedback), which is an intuitive way of understanding why a sharp lower boundary results at the concentration where the positive feedback dominates. This holds in any background, but the boundary shifts downward for WT compared to $\Delta bem1$. Since GAPs counteract this process, we have a sharp upper boundary there instead. For other common situations in the cell where proteins participate in a chemical reaction that influence fitness, but without positive feedback, we expect smoother curves, so lower $|k|$.

The low and high state values that are used heavily here are then respectively:

$$g(2c_i/3) = g_0 + \frac{2 - g_0}{1 + (2/3)^{-k}} \quad (4.16)$$

$$g(4c_i/3) = g_0 + \frac{2 - g_0}{1 + (4/3)^{-k}} \quad (4.17)$$

Note that by e.g., letting $k \rightarrow \infty$ and $g_0 = 0$ then $g(2c_i/3) = 0$ and $g(4c_i/3) = 2$ the Cdc42 case is retrieved. The full state equation with the transition matrix now reads:

$$\begin{bmatrix} f\left(\frac{2c_i}{3}\right) \\ f(0) \end{bmatrix} = \begin{bmatrix} g\left(\frac{4c_i}{3}\right)\left(1 - F_V\left(\frac{2c_i}{3}\right)\right) & g\left(\frac{4c_i}{3}\right)\left(1 - F_V\left(\frac{4c_i}{3}\right)\right) \\ g\left(\frac{2c_i}{3}\right)\left(F_V\left(\frac{2c_i}{3}\right)\right) & g\left(\frac{2c_i}{3}\right)F_V\left(\frac{4c_i}{3}\right) \end{bmatrix} \begin{bmatrix} f\left(\frac{2c_i}{3}\right) \\ f(0) \end{bmatrix} = M\vec{f} \quad (4.18)$$

¹² For more generality, it is sometimes possible to divide more advanced landscapes into multiple regions with a single boundary. For example, a wide peak could be a region with a lower boundary plus one with an upper boundary, provided the progeny peak width as a function of concentration is large enough such that switching from below the peak to above it is not possible within a generation.

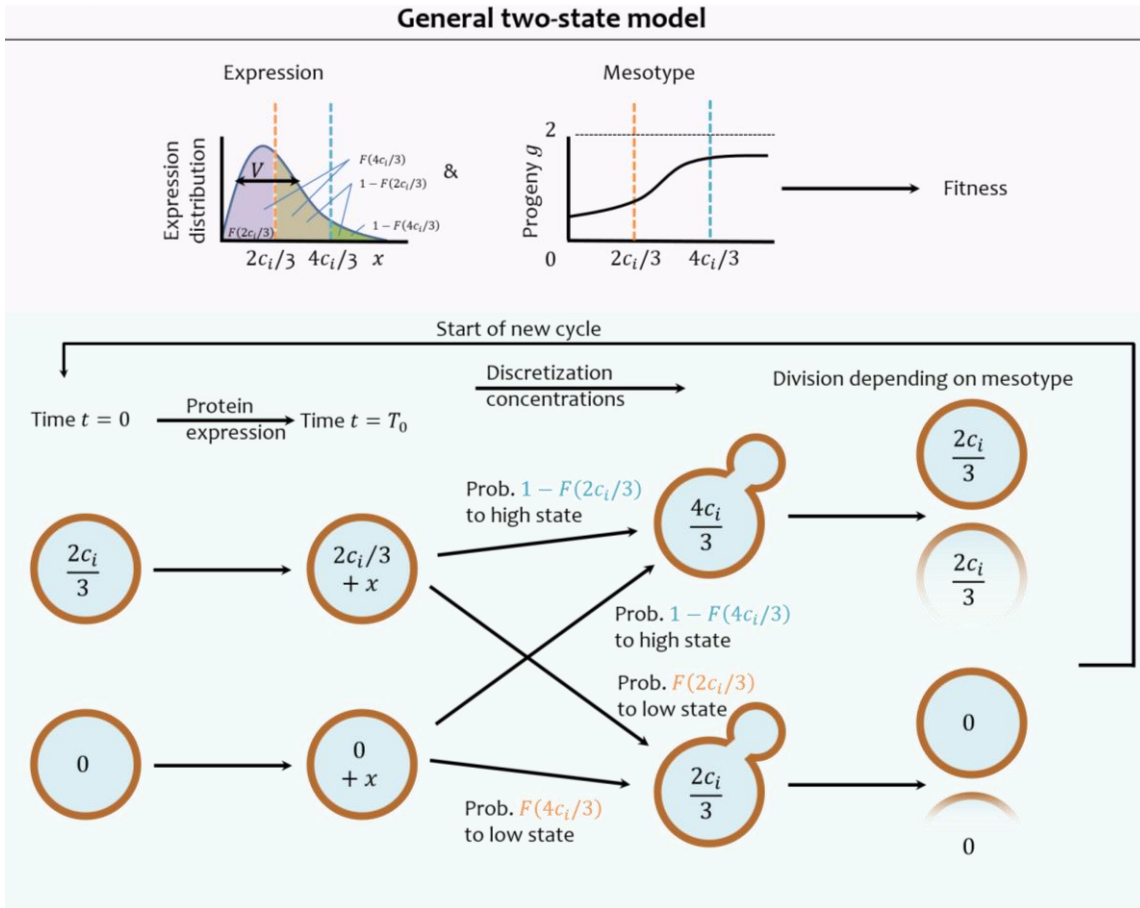


Figure 43 Two state model for fitness reflecting yeast cells polarizing with success depending solely on a general, simple mesotype function g . Noisy expression during cell cycle time T_0 , combined with the mesotypical rule yields fitness, as stochastic switching occurs between the low and high protein concentration state, with probabilities following from cumulative distribution function F , which is a function of noise level / coefficient of variation V .

Using the model definitions in 4.15 and 4.18, the general fitness, with λ_{max} as the largest eigenvalue of the transition matrix, is given by:

$$\omega = \frac{\log_2 \lambda_{max}}{T_0} \quad (4.19)$$

with

$$\lambda_{max} = \text{tr}(M)/2 + \sqrt{\text{tr}(M)^2/4 - \det(M)} \quad (4.20)$$

and

$$\det(M) = g(4c_i/3)g(2c_i/3)(F(4c_i/3) - F(2c_i/3)) \quad (4.21)$$

$$\text{tr}(M) = g(4c_i/3) - g(4c_i/3)F(2c_i/3) + g(2c_i/3)F(4c_i/3) \quad (4.22)$$

where $\text{tr}(M)$ is the trace of matrix M . In the simple case of sharp boundaries $|k| \gg 1$, then:

$$\lambda_{max} \approx tr(M) \quad (4.23)$$

which reduces to the Cdc42/GAP case when $k = \infty$ or $k = -\infty$.

For the coefficients of variation typically found in yeast and a sensible choice of the expression distribution, it can be shown (see Appendix M.4) that

$$\lambda_{max} \approx \max\left(g(4c_i/3)(1 - F(2c_i/3)), \quad g(2c_i/3)F(4c_i/3)\right) \quad (4.24)$$

$$\omega_r \approx \log_2 \max\left(g(4c_i/3)(1 - F(2c_i/3)), \quad g(2c_i/3)F(4c_i/3), \quad 1\right) \quad (4.25)$$

4.2.2.2 Noise shapes observed fitness curves, not the other way around

As λ_{max} is the growth factor from generation to generation, $\lambda_{max}/2$ can be interpreted as the mean progeny number across the population divided by that of WT (which is 2, the maximum within cycle time T_0). This is convenient in light of the relative fitness definition in [20] (to be called Keren fitness, where needed, to avoid confusion). In that paper, 81 fitness landscapes in yeast were probed as a function of protein expression level, by measuring the relative number of doublings of mutants with under- and overexpression of a gene, within the WT doubling time as a measure of fitness. Therefore, Keren fitness corresponds to $\lambda_{max}/2$.

Moreover, the Keren fitness landscapes were fitted rather ad hoc with a double sigmoid function. We can understand why this worked considering that our $\lambda_{max}/2$ is equivalent to Keren fitness. Since both g and essentially F (or $1 - F$) are sigmoids, $\lambda_{max}/2$ becomes a double sigmoid. The Keren fitness landscape data can therefore easily be interpreted in terms of noisy expression and a simple mesotype by use of the two-state model.

Therefore, the 81 genes from [20] are fitted using the two-state model to obtain a data set of inferred mesotypes. The expression levels measured in [20] are normalized to WT expression (using protein copy number data from [234]), to interpret the mesotypical inflection point as defined in Box 5. Genes whose fitness landscapes are considered are related to their own coefficients of variation from [229]. As not all genes considered in [20] are present in the data sets of [229] and [234], it was necessary to omit 22 genes from further analysis.

Performing the fits for all two-state model parameters (see also 4.6.4) improved those in [20] in 88% of the cases (metric adjusted R-squared). For the remaining cases, the observed fitness landscapes were typically beyond the defined scope of this two-state model due to multiple inflection points (recall 4.2.2.1). Therefore, these genes are excluded from the subsequent analysis, ultimately leaving 50 genes for Figure 44. This is not a large data set, but sufficient to get an idea of typical mesotypes. As seen in the pie chart, lower boundaries are the predominant mesotype. Additionally, most boundaries are relatively shallow ($1 < g_0 \leq 2$), not exhibiting gene essentiality or toxicity within the measured expression range.

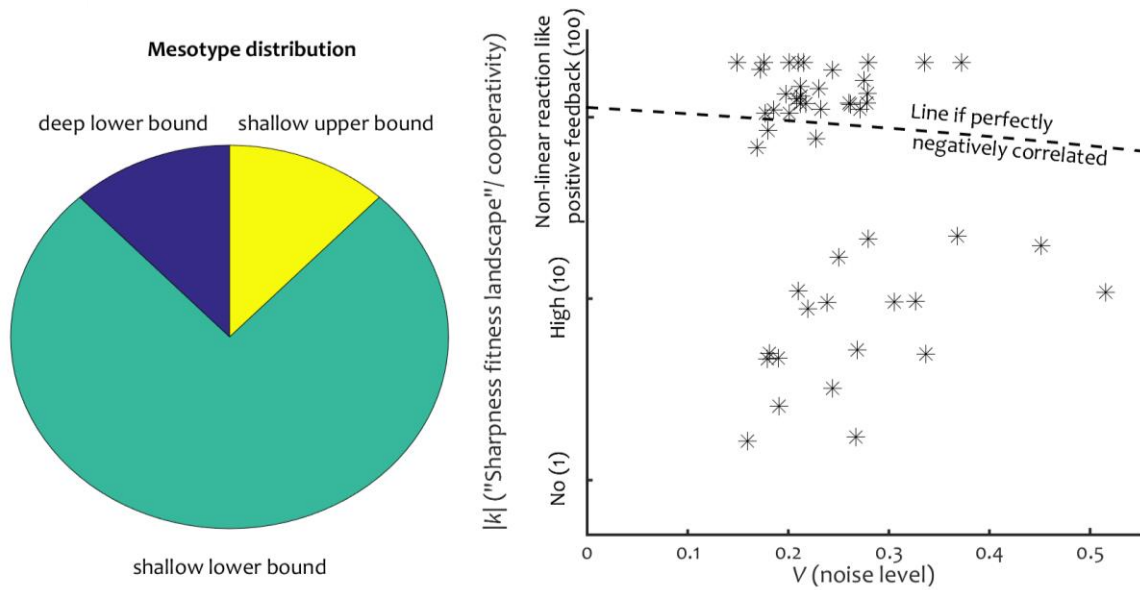


Figure 44 Abundance of lower boundaries in empirical mesotypes and absence of correlation between sharpness and noise. Left: pie chart for the distribution of 50 fitted mesotypes using the two-state model on experimental fitness landscapes from [20]. Right: scatter plot for fitted mesotypical boundary sharpness $|k|$ and noise levels from [229].

What is found regarding the relation between fitness landscapes and noise (right side of Figure 44) is not in line with the conclusions from [20]. There, it was noticed that when Keren fitness landscapes were sharp, these always corresponded with genes that have small expression noise. The Authors hypothesize that the fitness penalty for large noise in sharp landscapes causes selection for low noise in those genes. Yet, from the near-double sigmoid interpretation for the two-state model fitness described in this chapter, we can understand this correlation in a new light, which reverses the causality.

When expression of a gene connected to a sharp mesotype is noisy, the sharpness is diluted and ultimately broadens the observed fitness landscape. In other words, in order to observe a sharp fitness landscape, both a sharp mesotype and low noise are required. Sharp fitness landscapes are therefore an automatic, epigenetic consequence of low noise, and not necessarily causing the low noise by selection on the fitness cost of noise. Indeed from Figure 44, the correlation at the mesotype level between noisy genes and landscape sharpness is *not significant* (p -value 0.64, Spearman ρ). So, while a sharp fitness landscape requires a sharp mesotype, the latter is typically not accompanied by low noise, which would have been expected if fitness sharpness selects low noise.

4.2.2.3 Noise can form the basis for global adaptive coupling

The sign of the epistasis for the general model case is again reflected in the sign of η . In the case of sharp boundaries, there will always be a bunching of relative fitness contours (as seen in Figure 41 and Figure 80 for sharp lower and upper boundary respectively), showing a negative second derivative. As discussed in section 4.2.1.7, this implies negative epistasis (for mutations with the same sign in Δc_i) is generated for mutants falling into that region.

For shallower boundaries ($g_0 > 1$), the mesotype allows growth for any c_i , but in practice when the width of the probability distribution is much smaller than the total width of the mesotype space accessible through mutations. Then, there exists a region where the bunching of contours takes place between the two extreme viable states ($\lambda_{max} = g_0$ and $\lambda_{max} = 2$), leading to a sign change in the second derivative of $\log(\omega_r(c_i))$ (see the shallow upper boundary example of Figure 45). This creates a region where even for same sign mutations, positive epistasis can be generated (inside the region where $\eta > 0$).

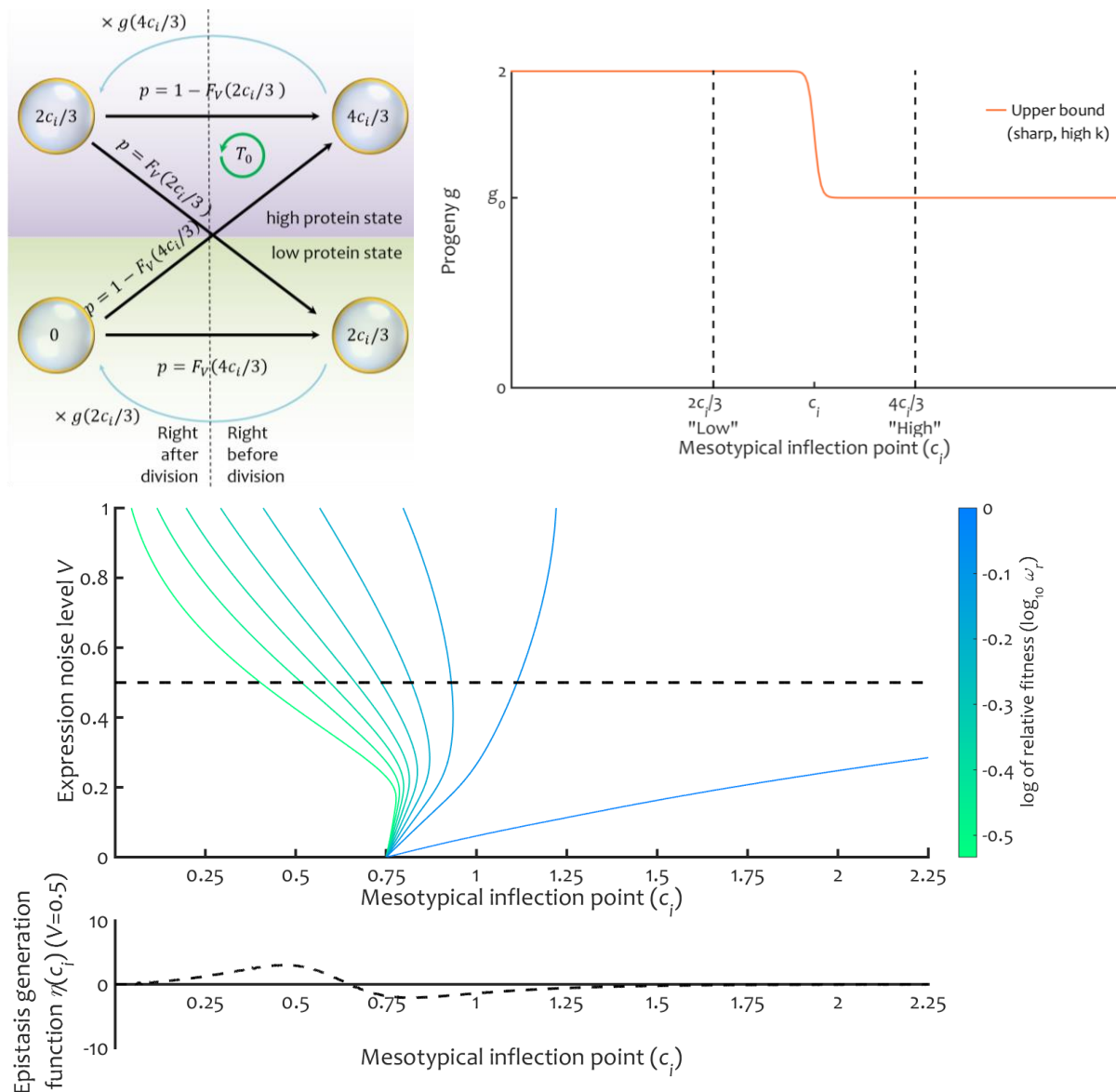


Figure 45 Noise in shallow upper boundary mesotypes extends mutational regions that harbour epistasis. Top: Two state model schematic summary (left) and mesotype (right) of the shallow upper boundary case. Middle: Contour plot of log relative fitness (blue to green) as function of mesotypical inflection point of the shallow upper boundary in the fitness landscape, and the noise level (coefficient of variation), assuming a gamma protein expression distribution. The dashed black line denotes the line $V = 0.5$, along which the plot in the bottom is valid. Bottom: Second derivative of log relative fitness with respect to mesotypical inflection point (along the dashed line of the top axis). Here, it can be positive and negative, allowing both positive and negative epistasis for mutations of the same sign.

The region with a positive epistasis generating function is in addition to the region where $\eta < 0$ (in Figure 45 when $c_i \gtrsim 0.6$), where negative epistasis is generated for same-sign mutations maintaining the MI-point in this region. It is important to note that this region with $\eta < 0$ corresponds to the mutational region near the optimum (where fitness is relatively high). This remains true with a lower instead of an upper boundary. If we then assume that adaptations are usually refinements in fitness (typical fitness gain is $<5\%$ [259]), we expect the mutational region where $\eta < 0$ to be most relevant for adaptations.

Moreover, we can combine this relevance with the empirical observation that lower bounds are most commonly encountered (Figure 44). Then, we expect that shallow boundaries exhibit similar epistatic behaviour as deep boundaries (where $g_0 \approx 0$, which always has $\eta \leq 0$) and more specifically as the Cdc42 case (lower bound case). This allows us to reformulate the conclusions from 4.2.1.7 on the prevalence of negative epistasis in more general terms.

Firstly, same sign (in Δc_i) mesotype-preserving mutations generate negative epistasis, and opposite sign mutations generate positive epistasis. The former applies more frequently from combinatorial arguments, due to the abundance of deleterious mutations over beneficial mutations [258]. This is consistent with the documented abundance of negative over positive genetic interactions (335056 against 81384, based on the interaction data incorporated into SGD (www.yeastgenome.org) from www.thebiogrid.org, date of access 09-07-2019).

Secondly, the coupling of a generic mutation to a mesotype-preserving mutation through cycle time T_0 yields negative epistasis for mutations of the same sign for lower boundaries, the most abundant boundary type (as visualized by the exclusive negative epistasis in Figure 46). This result is complementary with observations of diminishing returns among beneficial mutations, showing pervasive negative epistasis as in [260]. There, the authors pondered about a “biological basis for global coupling”. As explained, *noise generating this epistasis can be a cause for the global negative epistatic coupling between unrelated modules*.

It is also important to mark the extent of the valid predictions from this model. For mesotype preserving mutations, we expect to see a correlation between epistatic observations and noise. Combining the same aforementioned interaction data set with protein expression data from [229] does not yield experimental validation of this correlation. Instead, the number of genetic interaction partners anti-correlates weakly with coefficient of variation (Kendall’s tau = -0.05). Possibly, the data set is biased by coarse mutations that do not simply shift mesotypes (e.g., not just activity and/or copy number changes). Therefore, these mutations cannot be used to validate the prediction which holds for mesotype-preserving mutations. Alternatively, mutations may be structurally unevenly distributed inside the inflection-point ranges, i.e., a larger epistatic width is not strongly correlated to more potential mutations.

Now that the effects of noise have been determined, it is time to treat the mechanism that exploits the benefits of noise, particularly when noise is low. The mechanism is called transgenerational feedback, and will be the subject of the next section.

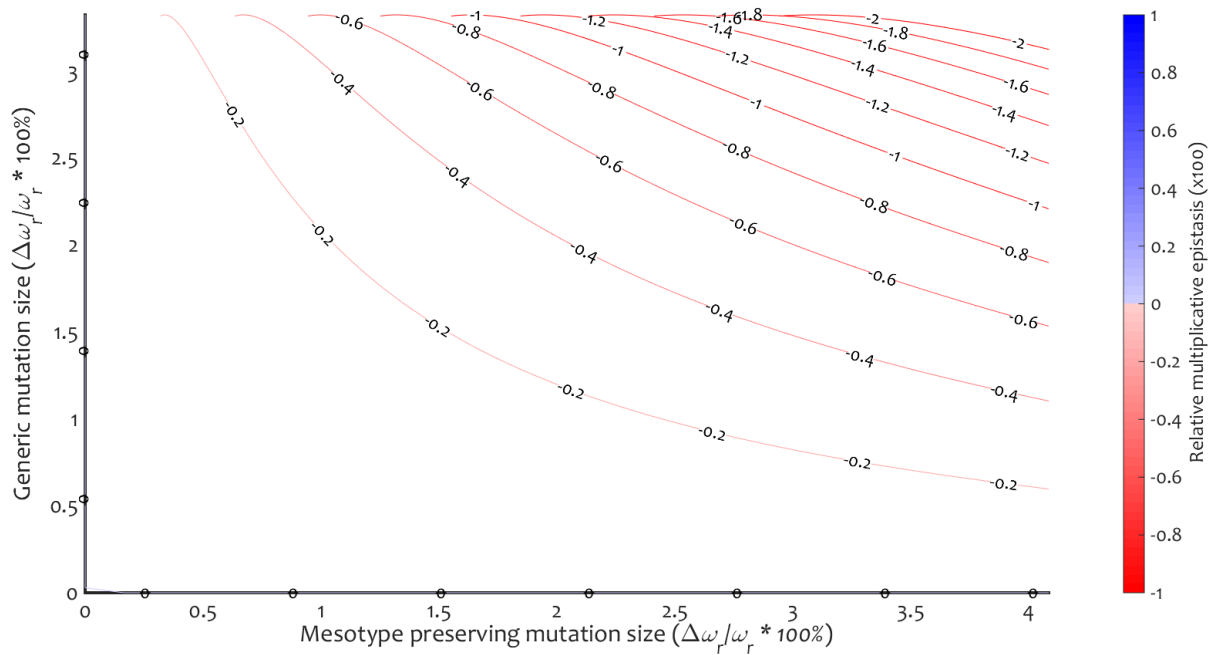


Figure 46 Diminishing returns for adaptive (beneficial) mutations in the prevalent shallow lower boundary mesotype case. Epistasis (times 100) of typical mesotype preserving mutations and generic mutations, affecting T_0 and subsequently fitness, in the shallow lower boundary case ($V = 0.22$, $g_0 = 1.2$) in the high fitness ($\eta < 0$) regime. Negative (blue) epistasis is the only form of epistasis present.

4.3 Transgenerational feedback as a noise mechanism

4.3.1 Transgenerational feedback effect on doubling times (Cdc42 case)

When protein copy numbers are subject to noise, this leads to intragenerational differences amongst cells. However, also between generations (*transgenerationally*), the effects of noise persist. Because cells with the luck to have the optimal amount of protein are more likely to survive, there is a bias towards that protein concentration at every generation. Simply put, the surviving cells have a better starting point than one would naively expect. This is a continuous epigenetic adaptation to a genetic background (e.g., Cdc42 production in a *bem1* deletion), which shifts the protein distribution away from the production equilibrium. Consequently, a “**transgenerational feedback loop**” is formed, a term coined in [41], although in the context of adaptation to environments (non-genetically).

To illustrate the importance of this effect, suppose that a protein that is normally under selection has its population distribution reset after every division to the value it has without selection. Effectively, this entails that the system loses its memory. This is best illustrated in the most extreme case of a sharp boundary in viable protein concentrations, so consider the Cdc42 case (sharp lower boundary, see Figure 47). First, the population distribution without selection must be determined (the default value to which the population now gets reset, even under selection). This means that $g = 2$ in 4.18, and that the transition matrix M now reads:

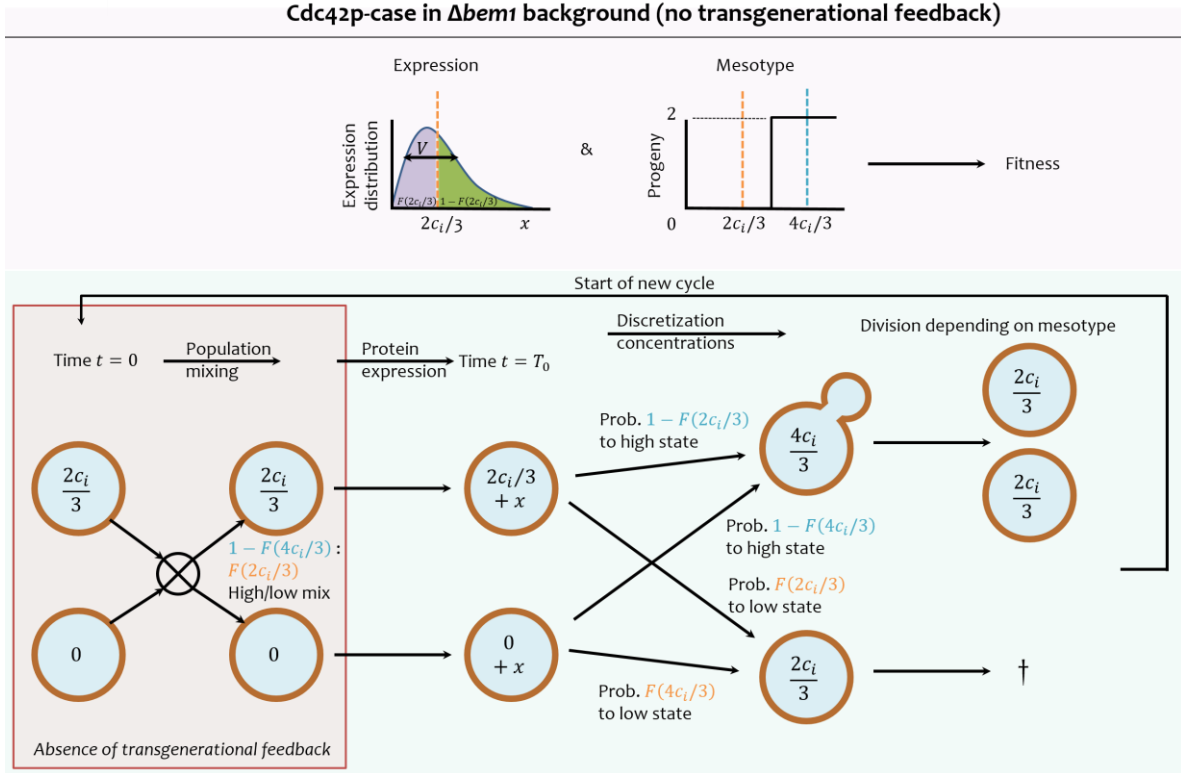


Figure 47 Two state model for fitness reflecting yeast cells polarizing with success depending solely on Cdc42 concentration. Noisy expression during cell cycle time T_0 , combined with the simplified mesotypical rule for progeny depending on Cdc42 concentration yields fitness, as stochastic switching occurs between the low and high Cdc42 concentration state, with probabilities following from cumulative distribution function F , which is a function of noise level / coefficient of variation V . Absence of transgenerational feedback is represented by a mixing step of the population distribution across states right after division, to reset the state distribution to that in the absence of selection.

$$M_{ns} = 2 \begin{bmatrix} (1 - F(2c_i/3)) & (1 - F(4c_i/3)) \\ F(2c_i/3) & F(4c_i/3) \end{bmatrix} \quad (4.26)$$

which will have 2 (number of progeny) as the largest eigenvalue, and an eigenvector (representing the division across the high and low protein state) of:

$$\vec{f}_{eq,ns} = [1 - F(4c_i/3), F(2c_i/3)] \quad (4.27)$$

This is the protein distribution (across the two bins) to which the population reverts after every division in the memoryless system. Normally, the Cdc42 case has an eigenvalue of M during equilibrium growth given by $2 - 2F(2c_i/3)$, with corresponding eigenvector $[1,0]$. However, if after every division the proportions of the population distributions are reset to $\vec{f}_{eq,ns}$, then \vec{f}_{Cdc42} cannot equilibrate to $[1,0]$, and instead the population is lost that reverts to the unfavourable low protein state. Concretely, provided that $1 - F(4c_i/3) > F(2c_i/3)$, the relative fitness with this reset ($-tgf$) changes to:

$$\omega_{r,-tgf} = \log_2 \left(\frac{2(1 - F(4c_i/3))}{1 + F(2c_i/3) - F(4c_i/3)} \right) \quad (4.28)$$

The absence of feedback changes limiting concentration $c_{i,lim}$ for the sharp lower bound to:

$$F(2c_{i,lim,-tgf}/3) = 1 - F(4c_{i,lim,-tgf}/3) \quad (4.29)$$

The relative fitness ratio with this reset ($-tgf$) and without ($+tgf$, the normal situation) is:

$$\frac{\omega_{r,-tgf}}{\omega_{r,+tgf}} = 1 - \frac{1}{\omega_{r,+tgf}} \log \left(\frac{1 - F(4c_i/3) + F(2c_i/3)(F(4c_i/3) - F(2c_i/3))}{1 - F(4c_i/3)} \right) \quad (4.30)$$

This ratio inside the logarithm is always larger than 1, as $F(2c_i/3) < F(4c_i/3)$. Hence, the transgenerational feedback *always* mitigates part of the negative influence of the noise (suboptimal growth under conditions where survival is possible without noise), and becomes stronger when fitness decreases.

Again, using the log-normal distribution as an example, the maximally allowed concentration value for the sharp lower boundary reduces to:

$$c_{i,lim,-tgf} \approx c_{i,lim,+tgf}/\sqrt{2} \quad (4.31)$$

An important note must be made here. Without noise, there cannot be any feedback (the system is deterministic, and there can be no memory). Yet in practice, there cannot be a discontinuity from $c_{i,lim,-tgf} = c_{i,lim,+tgf}$ as soon as there is minimal noise to $c_{i,lim,-tgf} \approx c_{i,lim,+tgf}/\sqrt{2}$. In reality, population size is finite and a small amount of noise will usually not yield any cell leaving the high/low bin it would have been in without noise. In other words, memory encoded in the population has to be built-up slowly, and becomes more prominent with finite population size and higher noise. Therefore, for subsequent plots without transgenerational feedback, a typical population of 10^6 cells will be assumed, and results at very low V are approximated to have continuity for $c_{i,lim,-tgf}$ as a function of noise level.

4.3.2 Transgenerational feedback triples epistatic region (Cdc42 case)

If we assume again the log-normal distribution, we can derive that the fitness shift in absence of transgenerational feedback as (assuming small (say 0.2), but not very small (<0.05) V):

$$\frac{\omega_{r,-tgf}}{\omega_{r,+tgf}} \approx 1 - \frac{1}{\omega_{r,+tgf}} \left(\frac{c_i}{c_{i,lim,-tgf}} \right)^{3/V} \quad (4.32)$$

As expected, at small c_i the transgenerational feedback has small effects. Further down the concentration space and since without feedback, $c_i < \sqrt{2}$, smaller noise increases the sharpness with which $\omega_{r,-tgf}$ rises as a function of c_i , decreasing the width of the epistatic region. Inversely, the presence of transgenerational feedback, particularly at low noise but

effectively at any level, widens the viable expression space and dampens sharp fitness features even further than noise alone does. Furthermore, it can be shown that to first approximation (for small noise), the epistatic width becomes:

$$w_e \approx \frac{V}{4} \sqrt{2} \ln \left(\frac{1}{\Delta\omega_r} \right) \quad (4.33)$$

Compared to the case with transgenerational feedback (the regular situation, (4.10)), this roughly equates to a 65% loss.

4.3.3 Transgenerational feedback promotes evolutionary capacitance (Cdc42 case)

Additionally, the transgenerational feedback also affects the always advantageous property of noise, namely the ability to survive for a few generations under conditions where the population is normally unviable. For the Cdc42 case, the equation for number of survivable generations (4.13, $+tgf$) turns $n_{gen,-tgf}$ to:

$$\frac{\log N_0 - \log \left(\frac{1 - F(2c_i/3)}{1 - F(4c_i/3)} \right) + \log \left(\frac{1 - F(4c_i/3) - F(2c_i/3)(F(4c_i/3) - F(2c_i/3))}{1 - F(4c_i/3)} \right)}{\log \left(\frac{1}{1 - F(2c_i/3)} \right) - \log 2 + \log \left(\frac{1 - F(4c_i/3) - F(2c_i/3)(F(4c_i/3) - F(2c_i/3))}{1 - F(4c_i/3)} \right)} \quad (4.34)$$

This looks like the case with feedback (4.13), only with the same term added in the numerator and denominator. Since $1 - F(4c_i/3) - F(2c_i/3)(F(4c_i/3) - F(2c_i/3)) > 1 - F(4c_i/3)$, absence of transgenerational feedback decreases the number of generations survived whenever it is larger than 1. For the log-normal case:

$$n_{gen,-tgf} = \frac{\log N_0 - \log \left(\frac{1 - F\left(\frac{2c_i}{3}\right)}{1 - F\left(\frac{4c_i}{3}\right)} \right) + \log \left(1 + \left(2\sqrt{\frac{8}{\pi V^2}} - 1 \right) \left(F\left(\frac{2c_i}{3}\right) \right)^2 \right)}{\log \left(\frac{1}{1 - F\left(\frac{2c_i}{3}\right)} \right) - \log 2 + \log \left(1 + \left(2\sqrt{\frac{8}{\pi V^2}} - 1 \right) \left(F\left(\frac{2c_i}{3}\right) \right)^2 \right)} \quad (4.35)$$

Again, the feedback effect is strongest when noise is low, but is in principle always in effect.

Figure 48 demonstrates the effect of the (absence of) transgenerational feedback. As in Figure 41, the blue to green contours denote the log fitness, while the red to yellow show the number of survivable generations (with optimal initial condition, contrary to equation above). The opaque lines show the case without feedback, the transparent lines are the same as in Figure 41, the normal situation with feedback. As derived, we see feedback improves the purple survival edge to allow more growth, and where growth is allowed in both cases, the

effect of feedback is never negative. Also, the number of generations the population can survive improves sharply, e.g., a factor of 3.5 for Cdc42's $V = 0.832$ at three times the sustainable survival limit without feedback.

In terms of epistasis, the plot indeed bunches the contours. This is in line with the reduced predicted epistatic width of the previous subsection. In terms of the epistasis sign, nothing changes. Same sign mutations (in MI-point shifts) still generate only negative epistasis.

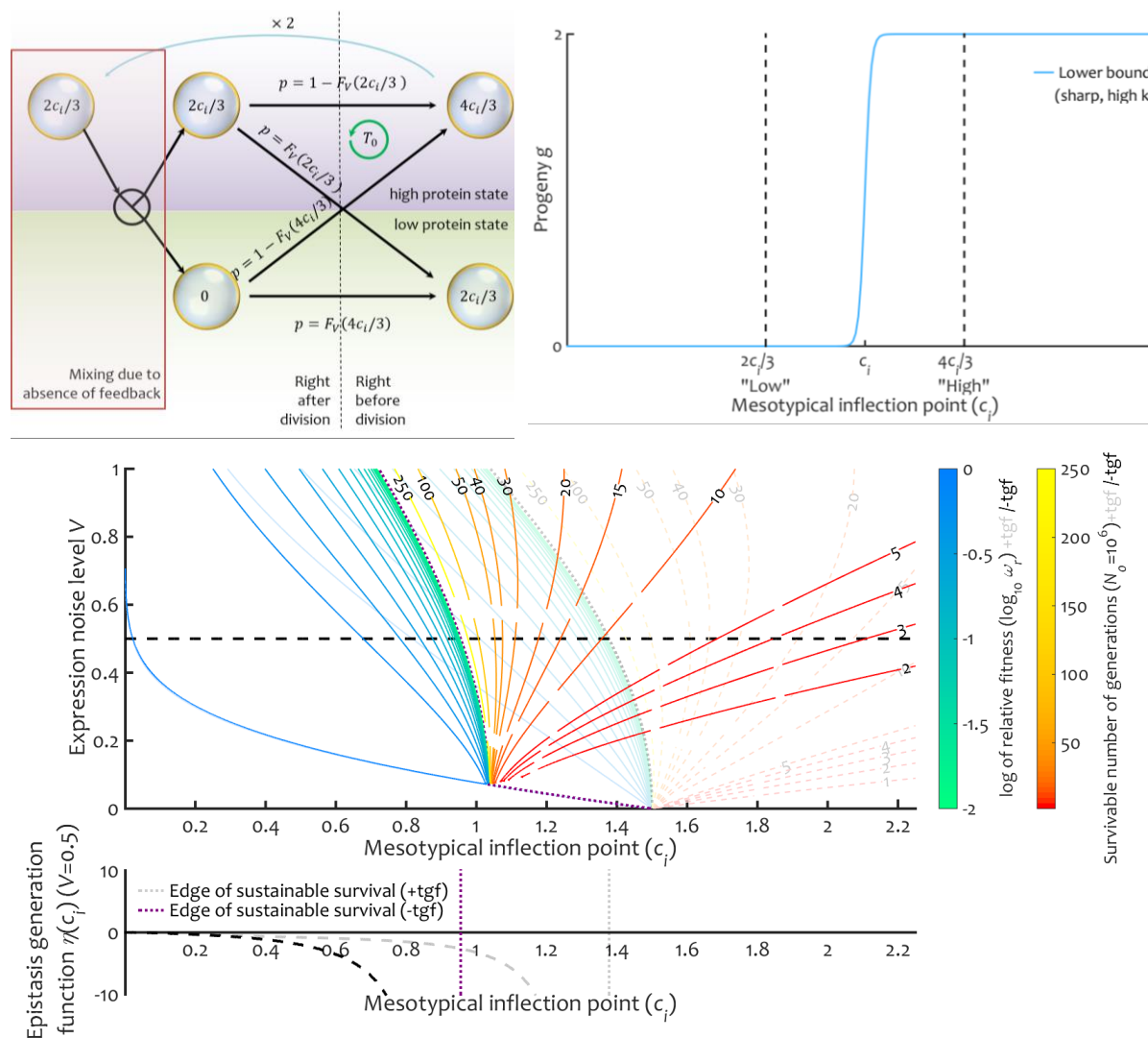


Figure 48 Absence of transgenerational feedback in the Cdc42 case is deleterious for sustainable and unsustainable growth. Top: Two state model schematic summary (left) and mesotype (right) of the Cdc42 case, without transgenerational feedback. Middle: Contour plot of log relative fitness (blue to green) and number of generations a population of a million cells survives (red to yellow) as function of mesotypical inflection point of the sharp lower boundary in the fitness landscape, and the noise level (coefficient of variation), assuming a gamma protein expression distribution, with transgenerational feedback (transparent) and without (opaque). The purple dotted line indicates the survival edge. The dashed black line denotes the line $V = 0.5$, along which the plot in the bottom is valid. Bottom: Second derivative of log relative fitness with respect to mesotypical inflection point (along the dashed line of the top axis). Here, it is always negative for values of c_i that support growth, allowing negative epistasis for mutations of the same sign.

In Appendix L, Figure 81 shows an analogous plot for the sharp upper boundary case (applicable to the GAPS), yielding similar conclusions. Again, the transgenerational feedback effect is never negative and the survivable regime expands compared to the case without feedback. Epistasis signs remain unchanged.

4.3.4 Transgenerational feedback (general case)

After the sharp boundary examples connected to Cdc42 and the GAPS in the $\Delta bem1$, the mesotypes are again generalized, as graphically depicted in Figure 49. When considering the *general* mesotypical landscape case, it is possible to derive that:

$$\omega_{r,-tgf} = \log_2 \left(\frac{g(4c_i/3)(1 - F(4c_i/3)) + g(2c_i/3)F(2c_i/3)}{1 - F(4c_i/3) + F(2c_i/3)} \right) \quad (4.36)$$

It can be proven (Appendix M.4.4) that the fitness value in the general case without feedback is never higher than with the feedback. Additionally, when outside the sustainable growth zone, the number of generations that can be survived is never larger without feedback. Consequently, the transgenerational feedback is *never deleterious*.

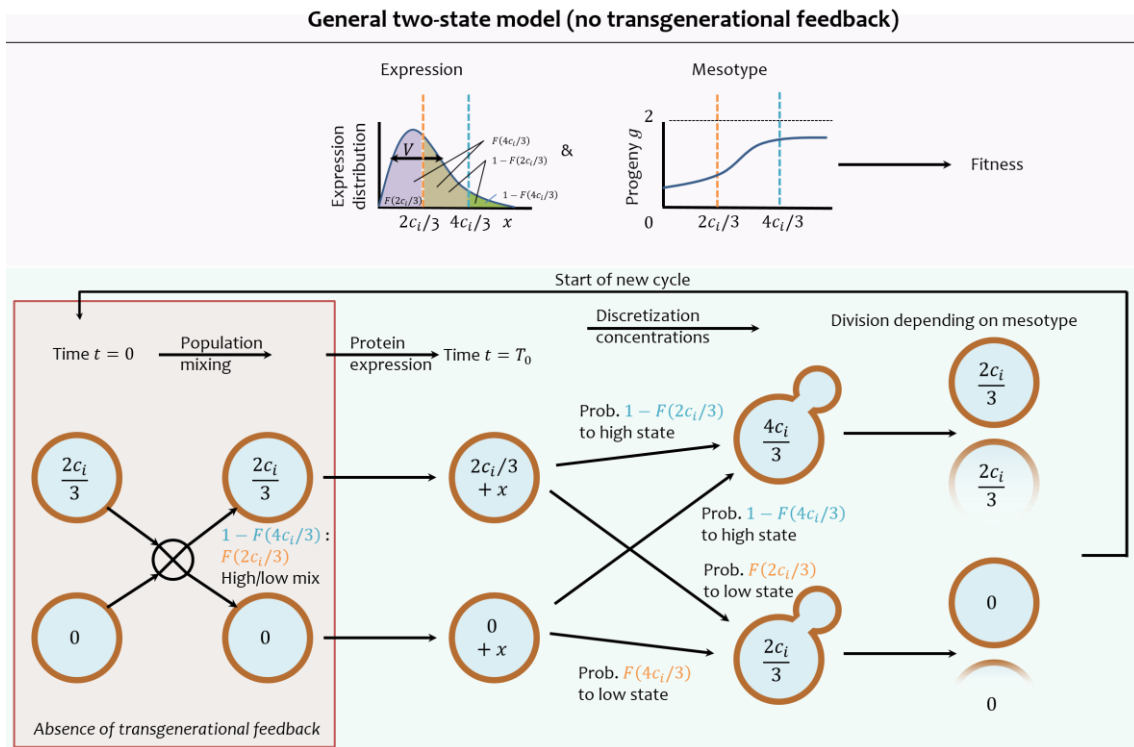


Figure 49 Two state model for fitness reflecting yeast cells polarizing with success depending solely on general, simple mesotype function g . Noisy expression during cell cycle time T_0 , combined with the mesotype mesotypical rule yields fitness, as stochastic switching occurs between the low and high protein concentration state, with probabilities following from cumulative distribution function F , which is a function of noise level/ coefficient of variation V . Moreover, absence of transgenerational feedback is represented by a mixing step of the population distribution across states after division but prior to expression, to reset the state distribution to that in the absence of selection.

Examples of the transgenerational feedback were depicted in Figure 48 and Figure 81, for sharp boundaries. As an example of a more general mesotype, a shallow upper boundary is considered in Figure 50. The transgenerational feedback becomes less pronounced, to the limit of no effect when the fitness landscape is flat. This can be seen from 4.36, as when $g(4c_i/3) \approx g(2c_i/3)$, then $\omega_{r,-tgf} = \omega_{r,+tgf} = \log_2 g(4c_r/3)$. Additionally, high levels of noise neutralize the effect of the transgenerational feedback, again showing that the feedback is most useful when noise is low. With regards to possibilities for epistasis, in absence of feedback the total epistatic region is this time larger.

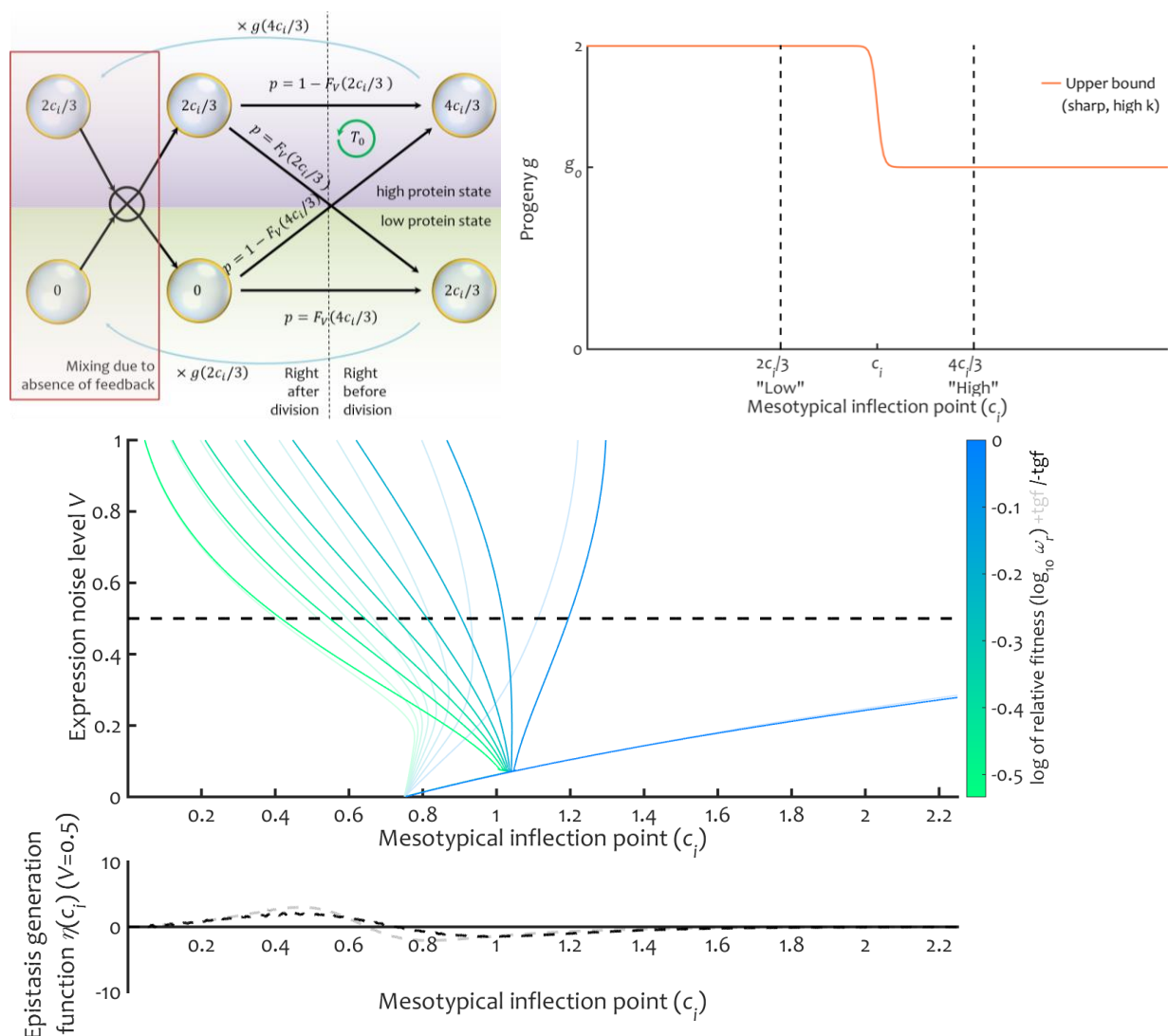


Figure 50 Absence of transgenerational feedback in the general case is deleterious. Top left: Two state model schematic summary (left) and mesotype (right) of shallow lower boundary case, without transgenerational feedback. Middle: Contour plot of log relative fitness (blue to green) as function of mesotypical inflection point of the sharp upper boundary in the fitness landscape, and the noise level (coefficient of variation), assuming a gamma protein expression distribution, with transgenerational feedback (transparent) and without (opaque). The dashed black line denotes the line $V = 0.5$, along which the plot in the bottom is valid. Bottom: Second derivative of log relative fitness with respect to mesotypical inflection point (along the dashed line of the top axis). Here, it can be positive and negative, allowing both positive and negative epistasis for mutations of the same sign.

4.3.5 Transgenerational feedback promotes evolvability of mainly essential genes

After the theory, it is time to determine how often transgenerational feedback may affect the evolvability of genes in practice. To this extent, we return to the two-state model fit on the genes from [20] as performed in the previous section. The feedback effect is then measured in terms of the size of the fitness effect with and without transgenerational feedback. As this depends on expression level of the gene, we consider mutations ranging from deletion or silencing (a mutation that has a large target size) to duplication (commonly found in adaptations during experimental evolution, e.g., with self-diploidization [261]). Another assumption is that other mutations also often encountered (synonymous and promoter mutations [260]) will not affect activity or expression beyond this range either. The relevant evolutionary range that is most often explored is then from zero to twice WT expression.

We consider a sizeable fitness decrease in absence of transgenerational feedback as $>5\%$, as from experimental evolution, this seems a typical fitness size of beneficial mutations [259]. Absence of feedback renders these mutations in that regime of expression inaccessible, influencing the evolvability of that gene. Figure 51 shows how large the region in expression space is that is affected by more than 5% in fitness due to transgenerational feedback.

If we define affected genes as genes whose expression space diminishes by $>5\%$, these occurs in 12 out of the 50 genes, and in 11 out of the 25 essential genes. This means that around 25% to 60% (95% credible interval is 26%-63%, see section 4.6.4) of the essential genes are expected to be affected in their accessible evolutionary space by at least 5% (and 15% to 50% of the essential genes at least 10%).

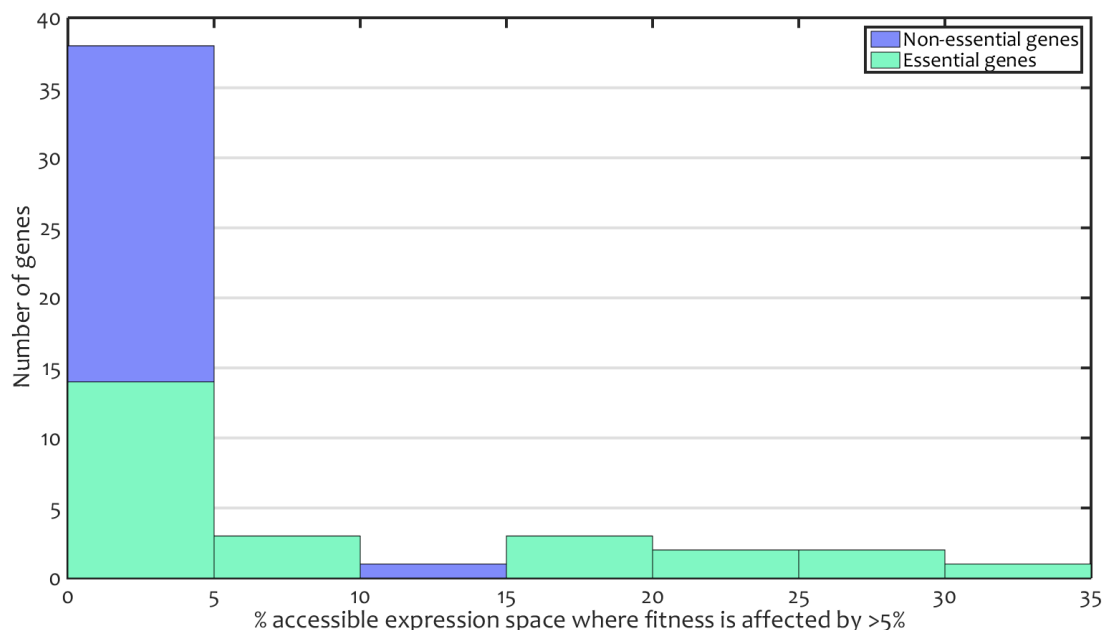


Figure 51 Transgenerational feedback contributes mostly to essential genes. Histogram of the predicted feedback effect for 50 genes whose fitness landscapes in [20] are fitted here using the two-state model. The horizontal axis shows for how much of the expression levels between gene deletion and duplication (the easily accessible expression space by mutations), the fitness is negatively affected by $>5\%$ in absence of the feedback. Blue bars represent non-essential genes, green essential genes.

For non-essential genes, transgenerational feedback is far less influential, affecting <17% of the genes. If we correct for the bias in the data for essential genes (in yeast, around 20% is essential [59]), evolvability of 8% to 25% of all genes in total should be significantly affected by transgenerational feedback.

4.4 Experimental design for transgenerational feedback

The previous section establishes that transgenerational feedback is an important noise mechanism, potentially responsible for expanding the evolvability of 25% to 60% of the essential genes. However, it has never been shown explicitly. For this purpose, a system is necessary where the mesotype can be accurately determined from molecular mechanisms, as is the case in budding yeast polarity. Yet, in experiments it is not possible to disable the transgenerational feedback in the artificial way as was mathematically done in our model. The most convincing way to see the transgenerational feedback is through a population protein distribution where the memory of the mesotype is encoded.

For this purpose, we return to the model from Chapter three, that was able to generate these distributions (the two-state model only had two bins). Unfortunately, even in the most extreme Cdc42 case (sharp lower boundary) for the ill $\Delta bem1$, the difference between protein under selection or not under selection are subtle, as they are obscured by other effects, see Figure 52. Most notably, ill cells tend to get large, which generally correlates copy number with volume, and the asynchrony of the cells (supposing they are not synchronized before measurement) also obscures the appearance of the effect in the observed protein distribution, although sizeable right at G1.

Therefore, there is a clear trade-off between proposing single-cell and bulk experiments. If protein distributions are acquired through careful analysis under the microscope, one cell at a time, one can correct for volume differences (and get concentrations, which determine the thresholds) and asynchronies. On the other hand, it is easier to get the amount of data needed to obtain the statistical power to draw conclusion through bulk studies, likely a flow cytometry experiment with fluorescent proteins. However, volumes cannot be reliably determined from flow cytometry as yeast falls into the Mie scattering domain [262], [263]. Additionally, the chemical manner to synchronize cells in a bulk fashion is likely to add artefacts to the data as well. Still, the subsequent analysis shows that it should be possible to measure the transgenerational effect combining the two approaches.

4.4.1 Preliminary assay

To obtain a first glimpse of this effect of selection on Cdc42 distributions, we can turn to the flow cytometer experiment performed in section 2.2. This experiment is further discussed in Appendix N. The idea is to compare the green (sfGFP) channel, which corresponds to Cdc42,

to the red channel which corresponds to miscellaneous auto-fluorescent proteins which are not under selection. Admittedly, this is not a cleanest or accurate assay, but a motivation for the pursuit of the more comprehensive experimental designs following in the next subsection.

In particular for the $\Delta bem1$ at low inducer concentrations (which is nearly lethal), part of the population will profit from the fat right tail in the noisy distribution (see Figure 59) generated by the Gal promoter. In this background, lucky cells become overrepresented in the data, and the typical scaling of protein abundance and volume becomes less pronounced. This bias should vanish at high expressions with healthier cells. Conversely, this effect should be minimal in WT, as it has a broader range of healthy copy numbers than $\Delta bem1$ (recall Figure 16).

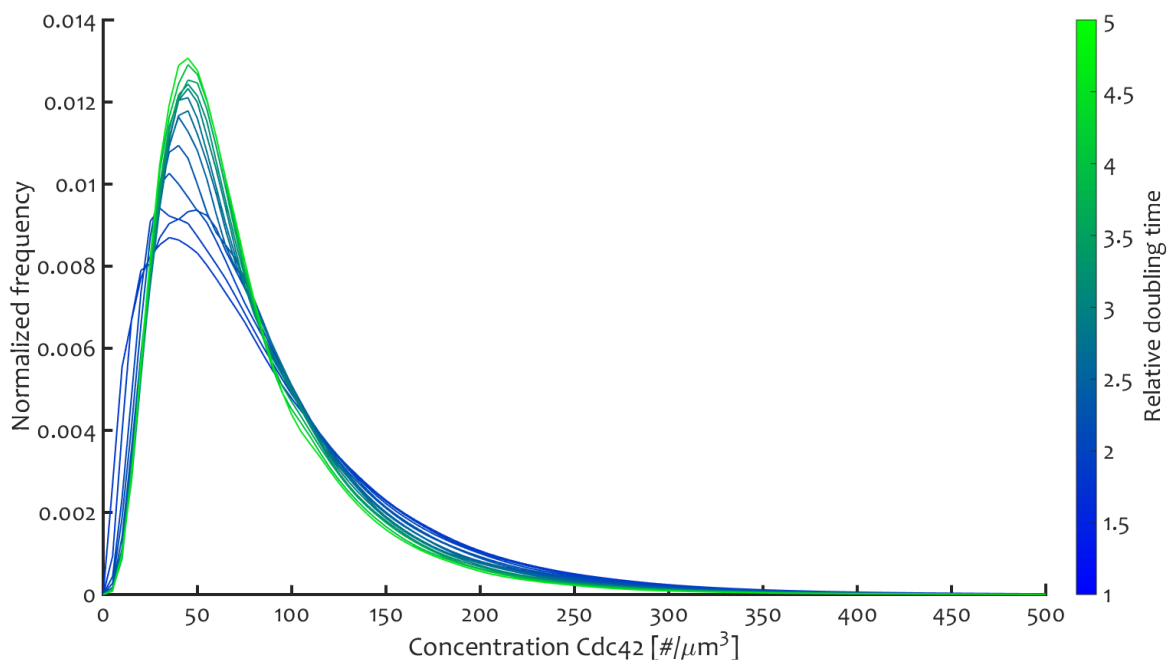


Figure 52 Concentration distributions for Cdc42 as simulated from the model of Chapter three, as a function of doubling time within a $\Delta bem1$ background relative to the fastest possible (indicated by the color bar).

What is observed (see Table 9) is less cross-correlation between the green and red channels than expected (based on volume effects between the pure WT with regular Cdc42) at low levels of Cdc42, particularly for the $\Delta bem1$ background. An explanation is that there are more smaller cells that have high Cdc42 copy numbers than expected, benefitting from the high expression tail which leads to a bias in survival of these cells. It is also stronger for the less healthy $\Delta bem1$ than for the $\Delta bem1 \Delta bem3$. This suggests selection on Cdc42 copy number for survival. This effect indeed vanishes when the induced expression of Cdc42 (in green) is so high that cells become fit, which results in a largely identical correlation between colors at high inducer concentration. These effects on the Cdc42 distributions provide some indication that Cdc42 influences fitness and vice versa, through a feedback loop.

Table 9 Observed correlation coefficients (minus the correlation in the pure WT control (YWKD062a,b,c) of $GAL1pr$ -sfGFP- $CDC42^{SW}$ expression, for a WT (YWKD065a), $\Delta bem1$ (YWKD069a, b, c combined) and $\Delta bem1 \Delta bem3$ (YWKD070a, b, c combined) background, and the pure WT control (YWKD062,a,b,c combined), for the green and red channel (the approximate standard error in parentheses), at different inducer concentration (due to evaporation uncertainty at least 0.05% galactose, 0.15%, 0.2% galactose and 2%). Values are the weighted average of mentioned biological replicates that yielded >1000 points.

Background	WT	$\Delta bem1$	$\Delta bem1 \Delta bem3$
Lower bound Inducer concentration			
0.05% Galactose	-0.38 (0.019)	-0.53 (0.024)	-0.49 (0.036)
0.15% Galactose	-0.27 (0.016)	-0.53 (0.023)	-0.41 (0.021)
0.2% Galactose	-0.18 (0.019)	-0.45 (0.018)	-0.26 (0.020)
2% Galactose	0.018 (0.014)	-0.064 (0.014)	0.0097 (0.015)

4.4.2 Simple assay design

Motivated by the preliminary results, we proceed to the new design. In chapter two, we completed a functional budding yeast Cdc42 probe, which still had some artefacts due to turnover issues of the fluorophore. Assuming these issues have been mitigated using the fluorophore mNeongreen [188], the population distribution of the sandwich fusion of Cdc42 with the fluorophore, which is under selection, must be compared to a protein not under selection to see the effects of selection. The sharp mesotype landscape leaves its imprint in the former, but not the latter. Moreover, this situation must be compared to the situation where selection is very weak as the fitness landscape is locally nearly flat, to exclude the effect of mere noise without memory. This suggests an experiment with the following strains:

- Strain 1: $\Delta bem1 pCDC42$ - $CDC42^{SW}$ -mNeongreen, $pCDC42$ -mCherry @ HO-locus
- Strain 2: $\Delta bem1 \Delta bem3 pCDC42$ - $CDC42^{SW}$ -mNeongreen, $pCDC42$ -mCherry @ HO-locus

The mCherry can be inserted at any neutral location, like the HO-locus [264]. Strain 1 and 2 are used to compare distributions of proteins under selection and not under selection, with a sharp or flat fitness landscape. We expect the variance (and subsequently, the coefficient of variation) of the Cdc42 distribution under sharp selection to decrease, as the left tail is eroded away. The ratio of coefficient of variation of Cdc42 and an unselected protein should be clearly smaller than 1, but approach 1 when the fitness landscape flattens (strain 2).

Preferably, all fluorophores have the same promoter, although this may cause some strain construction difficulties due to large homologies. Also including the ribosome binding site, we expect this to equalize production of the different fluorophores [205]. An alternative is a cistronic design of comparable strength and noise level, as done in [265]. That being said, some deviations of the average expression of unselected protein from the Cdc42 expression is not a big problem, because of the use of a rank correlation coefficient, as long as the noisiness is well matched. Most critical for the feedback is the noisiness. In addition, it is

important to note that mCherry has a long protein lifetime (~24h) [266], [267], which is convenient since short lifetimes can randomize the distributions, nullifying the feedback. Finally, we remove the C-term from Cdc42, which influences mRNA localization [87], to have more comparable circumstances for production as with mCherry.

If we measure fluorescence of single cells under the microscope and perform image analysis on those, we can determine when cells have just passed the polarity threshold. At this point, erosion of the left tail is maximal. This allows the completion of the first part of the proof for transgenerational feedback, namely that selection takes place at division. Since we can match copy number (proportional to intensity) to volume (assuming spherical cells), we can determine concentrations. However, also in the case of merely counting copy numbers, the effect is clear and reliably significant when measuring on the order of 100 cells, which is feasible even without automated tools for analysis (see Figure 53).

The second (and final) step is proving the selection has a lasting effect on the population as a whole. Therefore, the same erosion must be shown for asynchronous cells as well. In the flow cytometer, we expect to see significant erosion at a few thousand analyzed cells (see Figure 54). This is very well feasible as well (typically one gets on the order of 10000 cells). Obtaining concentrations is not particularly beneficial and also not easy to obtain.

As an indication of feasibility, the concentration and copy number distributions of Figure 53 at 100 cells and of Figure 54 at 10000 cells are plotted in Figure 55 and Figure 56. While for combining measurements, it is convenient to turn a distribution into a single metric, these plots provide some insights in the 'visual' distance between selection and no selection. Inevitably, some instrument / protocol-related obscuring factors may occur, which should be weak enough to not overshadow the simulated predictions.

4.4.3 More advanced assay design

Supposing we are not satisfied by the results of the previous assay, a further confirmation is to see the correlation between the two colors (between selection and no selection). As we also need to establish what the correlation between two unrelated proteins normally is, the strain design is more complex:

- Strain 1: $\Delta bem1$ $pCDC42-CDC42^{SW}$ -mNeogreen, $pCDC42$ -mCherry @HO locus
- Strain 2: $\Delta bem1$ $pCDC42-CDC42$ $pCDC42$ -mNeogreen $pCDC42$ -mCherry, @HO locus
- Strain 3: $\Delta bem1$ $\Delta bem3$ $pCDC42-CDC42^{SW}$ -mNeogreen $pCDC42$ -mCherry, @HO locus
- Strain 4: $\Delta bem1$ $\Delta bem3$ $pCDC42-CDC42$ $pCDC42$ -mNeogreen $pCDC42$ -mCherry, @HO locus

Strain 1 is used to correlate colors from a distribution under selection and not under selection. To know what a normal correlation is (due to effects as from volume correlations), this will have to be statistically different from the correlation found when comparing two

proteins not under selection, which is the case in strain 2. In both cases, any correlation is expected to vanish when the strains become fit (transgenerational feedback works best when the strain is ill), which is the case in strains 3 and 4.

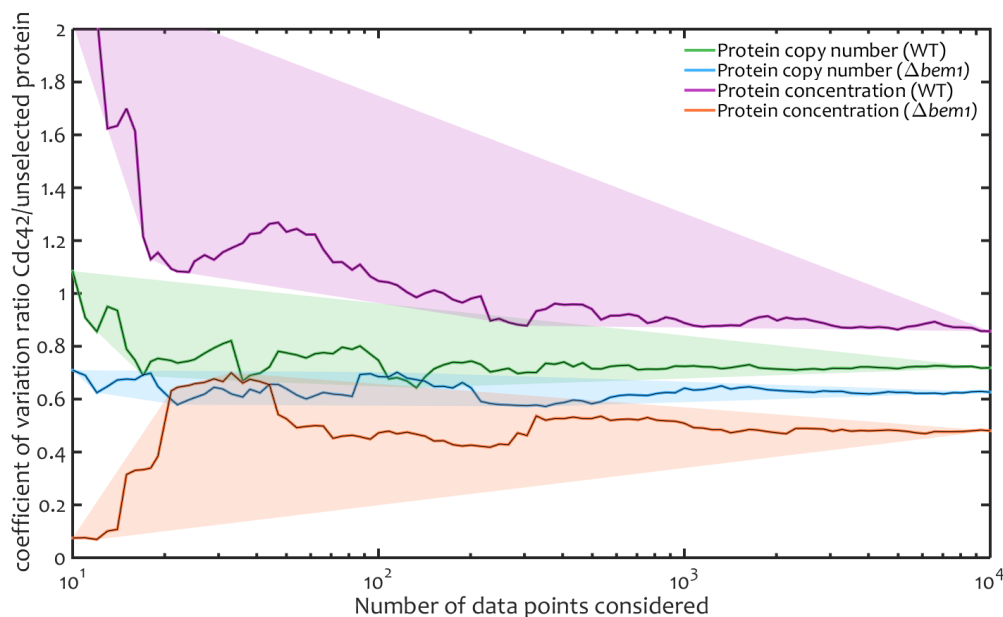


Figure 53 Feasibility of discriminating transgenerational feedback with single cell microscopy using 100-1000 cells. Plotted is the expected ratio of coefficients of variation at the end of G1, as a function of number of analyzed cells, with bootstrapped (10000 samples) confidence intervals, simulated from the model from chapter 3. Green and blue are for Cdc42 and an unselected protein copy number, purple and orange for concentrations.

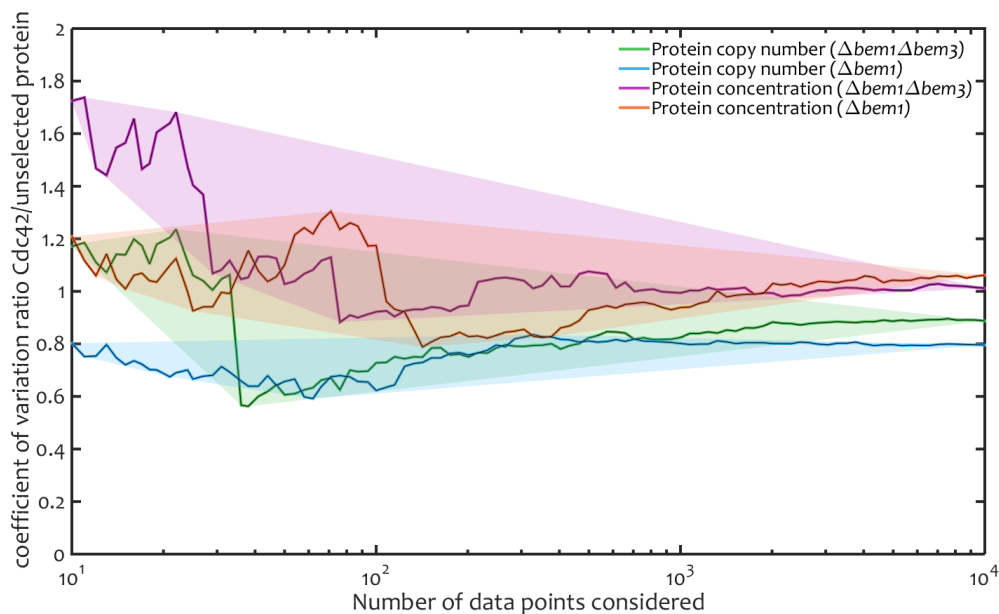


Figure 54 Feasibility of discriminating transgenerational feedback with flow cytometry using 10000 cells. Plotted is the expected ratio of coefficients of variation of asynchronous cells, as a function of number of analyzed cells, with bootstrapped (10000 samples) confidence intervals, simulated from the model from chapter 3. Green and blue are for Cdc42 and an unselected protein copy number, purple and orange for concentrations.

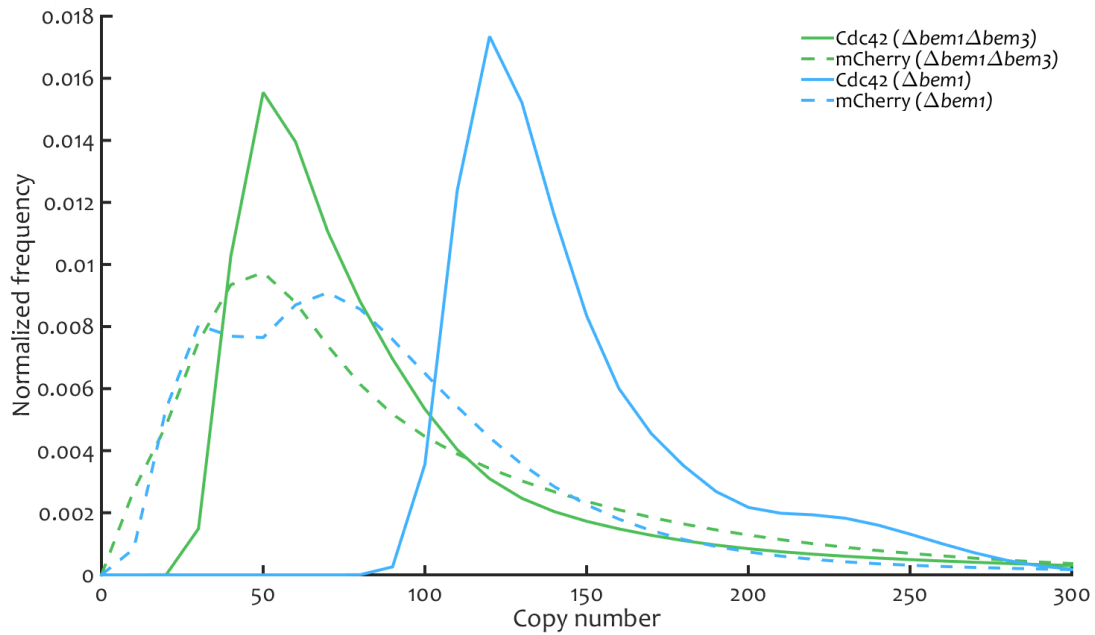


Figure 55 Example of simulated concentration distributions, processed by Matlab's kernel smoothing density estimator *ksdensity*, at the end of G1, for 100 cells.

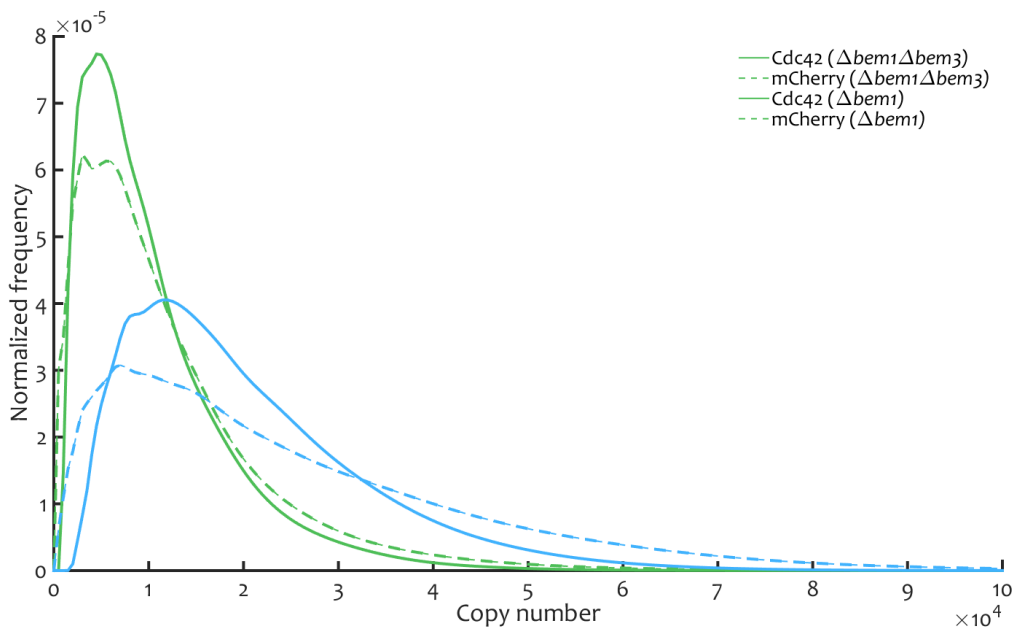


Figure 56 Example of simulated copy number distributions, processed by Matlab's kernel smoothing density estimator *ksdensity*, of 10000 asynchronous cells.

From Figure 52, we may think that correlating the left tail of the distribution magnifies the differences in correlations. However, with the typical flow cytometry data size of 10^4 , excluding too much data means losing too much statistical power. We would not be able to prove anymore that due to a fitness barrier, Cdc42 in the $\Delta bem1$ background is correlated less to a protein not under selection, than to be expected when comparing two proteins that are not under selection. This is what would verify the claim that fitness leaves a detectable

memory effect in the population protein distribution, which is reduced when the fitness landscape flattens, as in the $\Delta bem1 \Delta bem3$, demonstrating transgenerational feedback.

As seen from Figure 57, including at least the lower half of the data may generate a statistically significant difference in correlation between Cdc42(selected)/unselected and unselected/unselected proteins. Cdc42 and the unselected protein correlate more when including high copy numbers, as these typically correspond to high volumes. Still, a small difference is likely to be measured. For cells where the volume distribution is much smaller (fit cells), there should be no detectable correlations, showing that the feedback vanishes for a flat mesotype, and that noise alone does not generate correlations.

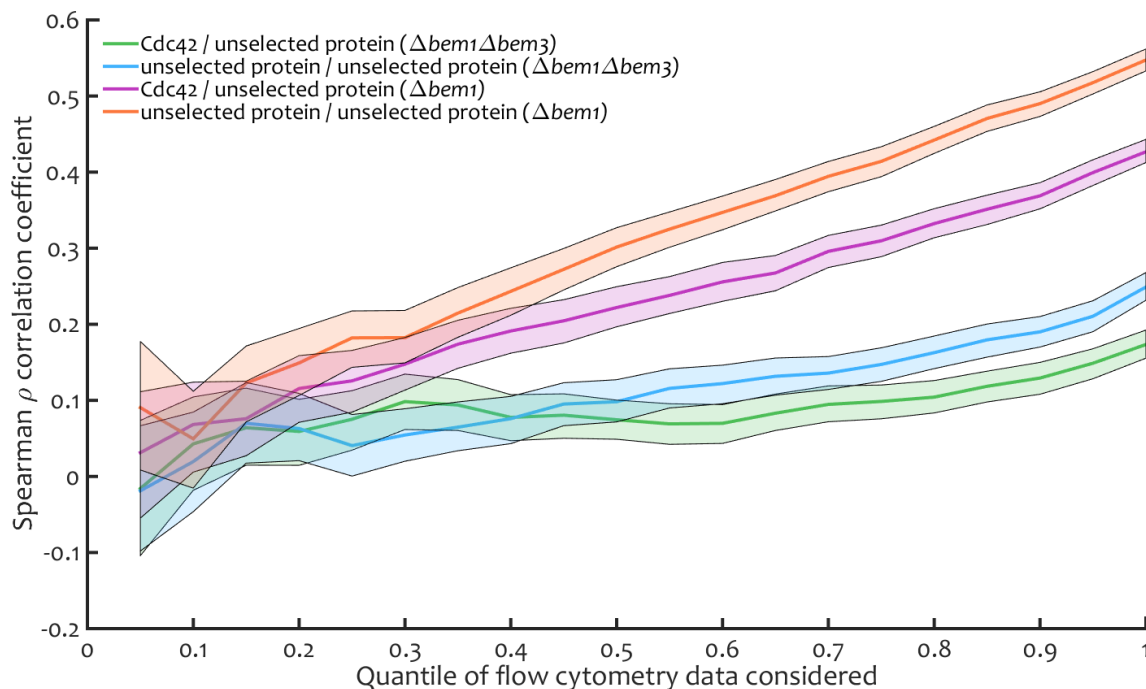


Figure 57 Flow cytometry data does not benefit from discarding quantiles for demonstrating transgenerational feedback. Spearman rank cross correlation of protein copy number as simulated from the growth model of Chapter three, for $\Delta bem1 \Delta bem3$ and $\Delta bem1$ between Cdc42 (under selection) and an unselected protein or between two unselected proteins. 95% confidence intervals are retrieved through bootstrapping (1000 samples). This is plotted as a function of quantile of the flow cytometry data (10000 points in total) used.

4.5 Discussion

In the previous chapter, we have seen that even simple mesotypes can amount to elaborate phenotypes, as exemplified by the noisy protein Cdc42 for polarizing budding yeast. However, the model used to demonstrate this still contained various details that are not crucial for the description of e.g., the GAP epistasis in the $\Delta bem1$. Therefore, the goal for this chapter has been to dive into the fundamental reasons for non-trivial phenotype generation from relatively simple geno-/mesotypes. Ultimately, this is crucial to gain an understanding of the general genotype-to-phenotype map, and the rules of evolution, which couple phenotype back to genotype (within a certain environment).

A minimalistic, two-state model has been constructed, without any cell cycle and cell growth details. The only features are stochastic production of protein per generation, and a simple sigmoidal mesotype curve determining progeny after every generation depending on the current protein copy number. Only two protein number bins exist; low and high. While seemingly simplistic, this model captures experimental observations such as the GAP epistasis in the $\Delta bem1$ and observed fitness landscape shapes. Importantly, it provided the analytical tractability to illustrate the important concepts from cellular to population dynamics. Concretely, the role of noise and its intrinsically active mechanism surfaced, namely protein expression noise generating a process called *transgenerational feedback*.

Noise is typically considered deleterious (e.g., [254]) or something that may be selected against (e.g., [255]). This model allowed the quantification of its effects, which were far from predominantly negative. The following quantifiable consequences were observed (conclusions in **bold**, hypotheses or claims in *italic*):

- **Noise limits the sustainably viable protein expression space if the mesotype has a sharp lower boundary (e.g., Cdc42 in a $\Delta bem1$), which is typically associated with an essential gene (although much more pronounced in absence of the transgenerational feedback discussed further on). Conversely, it expands the sustainably viable protein space if the mesotype has a sharp upper boundary (e.g., any GAP in a $\Delta bem1$), which is typically associated with a non-essential gene.**

Interestingly, it means that high noise, even at steady state growth, can be selected for and forms a good short-term evolutionary solution. While high noise and narrow landscapes are not encountered in the sample of [268], trajectories are imaginable where increasing the noise of a protein (which would face a sharp upper boundary) has a larger target size (a simpler solution) than finding the mutation to precisely get the right expression level. On the long-term the latter mutation will be found, ultimately reverting the noise increase thereafter. This trajectory still accommodates the observation of [268], which states that fitness costs in sharp fitness landscapes cause selection for lower noise, motivating the claim:

Increasing expression noise can provide a convenient short-term solution for adaptation.

- Noise allows the temporal survival of a colony outside the protein expression range where the colony is sustainably viable (the more noise the better). This can be sizeable, for example doubling the temporary survivable region during 15 generations for Cdc42 in a *Δbem1* colony of size 10^6 .

The notion of an optimal genotype is very relative. Other circumstances, such as an initial condition (e.g., does the cell start as a spore, from quiescence or from a log-phase ancestor) and the environment determine how well the cell is equipped for fast growth. Noise is important to buffer against invariably changing circumstances, which can alter at a faster pace than that genomic adaptation can occur. Noise acts a temporary buffer, for a number of generations that can be high enough to hold on until more permanent genomic changes arise.

This buffering shares some similarities to bet-hedging [269], where a population sacrifices fitness under physiological conditions to improve its odds during stress. However, the previous discussion point showed such a trade-off does not have to exist. Moreover, as explained further on in this discussion, the added role of memory is essential, such as in bet hedging in *Bacillus subtilis* where the variable but heritable phosphorylation state of Spo0A diversifies sporulation fates in the population [270].

With the previous paragraph in mind and because noise is both the buffer and the cause of phenotypical variation, noise works instead in a manner reminiscent of evolutionary capacitance, where genotypical variation is initially nullified in terms of selection. When e.g., an environmental change switches the nullification off, selection is allowed on a more diverse population than normally possible. A notable example is the heat-stress protein Hsp90 [271], [272] in fruit fly and plant model systems respectively. As a chaperone, it assists folding of unstable proteins, until stress causes an overload for its function. At this point, many unstable folds become phenotypically apparent, allowing selection to take place. In the case of yeast Sup35 prion formation, impairing the function (translation termination) that normally suppressed genetic variation is also widespread and hereditary [273].

A difference emerges when regarding noise as an evolutionary capacitor in the traditional sense. Even after a perturbation primes selection on the diverse population, noise continues to facilitate evolution outside the sustainably viable genetic zone (the genotype space with appropriate protein expression for durable survival). While not completely nullifying certain genetic variation, it can stall the purging effect of selection for relevant time scales. From that perspective, noise most prominently buffers after the perturbation.

It may therefore be more appropriate to treat noise as an inductor-capacitor (LC-) circuit, familiar in electronics. There (simply put), a closed electronic circuit comprising in the minimal form of a coil (the inductor) connected to two conducting plates (the capacitor) allows storage of energy. The capacitor has electrical energy stored in the space between the plates, which upon discharge pushes a current through the coil, which absorbs the energy into a

magnetic field. Conversely, the latter field may induce a current reverting the energy back to the capacitor. As a consequence, energy is moved back-and-forth.

In this analogy, noise oscillates between inducing the variation, and uses the capacitance to release the evolutionary potential to the population. Continuing with this electronic analogy, the ‘resonance frequency’ of the LC-circuit is then set by the time between perturbations and the time to complete genetic selection. This leads to the formulation of the claim:

Expression noise can act as an evolutionary capacitor, or even a LC-circuit, allowing survival of deleterious genotypes within a certain environment until a rescuing mutation occurs or until a change of environment ‘reshuffles the deck’.

- **Two mutations that conserve the shape of a non-flat mesotype landscape containing an inviable region, exhibit negative (positive) epistasis when from the same (opposite) sign. This also holds true for a generic mutation (noted only through changing cell cycle time) combined with a mutation preserving a lower bound mesotype, while the opposite holds for the less common upper bound mesotype. However, when the worst-case fitness is still viable, all mesotype-preserving mutations (shifting only the mesotypical inflection point) can exhibit both positive and negative epistasis.**

When combined with observations on mutations in literature, this conclusion has important implications. For example, since there are much more deleterious than beneficial mutations [258], the amount of combinations between same sign mutations outnumber the opposite sign combinations. The subsequently predicted epistatic encounters from this model are then consistent with the observation that negative epistasis is more abundant than positive.

Additionally, it has been observed in e.g., [260] and [274], that (beneficial) adaptive mutations, even while originating from unrelated modules, tend to exhibit negative epistasis. From [20], we could deduce (see also next point) that the epistasis generating adaptive mutations will typically have lower bounds in their mesotypes. Together, this leads to the prediction that adaptive mutations will usually interact negatively with each other, as they are the same sign in a typically lower bound mesotype, providing evidence to the claim:

Noise contributes to the biological basis for the observed global negative epistatic coupling of adaptive mutations.

Noise then forms part of a broader documented range of explanations for the diminishing returns of adaptive mutations. For example, in [275] some alleles found in experimental evolution of *Methylobacterium extorquens* exhibited negative epistasis due to the global cost of protein overexpression. In another study (in yeast) [276], the “modular life model” is defined. In this paper authors show that diminishing returns occur when genes (in a data set of >28000 SNPs) are categorized into modules, and assuming each module has a maximum functionality (how fast this maximum is reached depends on the environment). In a more

mechanistically detailed approach, the study on directed protein evolution in [277] showed how for a single protein (an arylesterase in *E. coli*) mutations become less beneficial. The described multiple molecular principles might apply to other enzymes as well.

- **Taking into account how expression noise affects observed fitness landscapes, a simple sigmoidal mesotype landscape, interpretable as a Hill curve, is retrieved in 88% of tested fitness landscapes from [20].**

The two-state model from this chapter contained only a minimal level of detail: a simple mesotypical landscape and expression noise. These two features combined were enough to improve upon the fits from [20] 88% of the time (metric adjusted R-squared). This suggests that realized fitness curves are often a result of simple mesotypes with noisy protein copy numbers. The fact that the mesotypes used here were a simple Hill curve is even more remarkable, since it is typically explained along the lines of linear, cooperative chemical reactions. While this simple interpretation has some caveats (see e.g, [278]), for example for ultrasensitive processes with very high cooperativity, it forms a good starting point to find a molecular justification for every reverse-engineered mesotype. Even when more sophisticated mesotypes are chosen (for an overview see [279]), including noise is the only way for observed fitness landscapes to give clues on what happens at the molecular level in a top-down fashion. This is very important, particularly in the context of building “levels” in the genotype-phenotype map, as explained in Chapter 1. Thus:

Noise is necessary to translate observed fitness landscapes to relatively simple lower-level processes, forming a critical part of the bridge from molecular processes to phenotype.

- **The aforementioned conclusions hold for noise coming from any sensible, unimodal protein expression distribution, including common choices as the gamma and log-normal distribution.**

While translating real protein expression distributions to simple, parametric forms can be subject to multiple acceptable formulations, general rules can be postulated. In yeast, most genes lack dosage regulation [11], and the long half-lives [235] also suggests little cell cycle regulation, favouring unimodal distributions. Molecular burst models [189], [237] suggest skewed distributions as a negative binomial or an even simpler continuous gamma variant are appropriate at the population level. Down to the single-cell level, this still suggests gamma-like distributions [280], like a log-normal (which also has positive skew and fat tails). Across tested distributions, what matters most are the first two moments; the mean level compared to the location of the fitness feature, and the noise level (using the coefficient of variation as a metric). These heavily influence the doubling times, the epistatic regions, etc. Seemingly:

Given a mean expression level, the noise level is crucial for fitness, while the exact family (log-normal, gamma, etc.) of the expression noise distribution is of marginal importance.

While noise is important, the question was also why noise can be utilized in this way. What was found is that an epigenetic mechanism called transgenerational feedback (akin to [41]) is very likely responsible, as a Lamarckian survival-of-the-fittest effect based on protein copy numbers. Provided it is experimentally shown (which was the topic of 4.4), its existence now seems so self-evident that it is worth shining the light on some important, quantifiable conclusions that have been reached.

- **The differential inheritance of protein copy numbers based on fitness serves as a beneficial, impromptu mnemonic tool for a non-flat mesotype landscape; it always expands the sustainably viable protein expression space and never decreases fitness.**

If there is sufficient noise (this depends also on population size, so say $V > 0.05$ for $N_0 = 10^6$), there is always a part of the population that is better equipped in terms of protein copy number for the fitness landscape determined by the genotype and environment. As their offspring inherit proteins from the ancestors, the new starting point is better than those with ancestors having suboptimal copy numbers. Those with by chance better copy numbers will become more abundant, as they can pass down their advantage. In this way, the fitness landscape imprints itself in the population, helping it overcome its suboptimal genotype. This lets itself be felt in terms of e.g., viable protein expressions (at most expanding it by a sizeable 40% (factor $\sqrt{2} - 1$) for sharp boundaries), which results in an expansion of the evolutionary accessible mutational space and hence the evolvability. Incidentally the transgenerational feedback works best at the typical noise levels found in yeast genes [229] and for sharper boundaries. Evidently, for flat landscapes there can be no effect.

Pending final experimental validation, simple epigenetic inheritance of noisy protein copy numbers provides a transgenerational feedback to serve as an exclusively beneficial, fast and reversible adaptation mechanism.

- **Transgenerational feedback expands the region in expression space where epistatic mutants can be encountered when the mesotype is sharp. The region can shrink for shallow landscapes.**

It is an important question where epistasis comes from in the mechanistic sense. As aforementioned, inferred mesotypes typically have lower bounds. In this case, particularly for low noise, the driving force for noise to cause epistasis (which occurs as explained earlier in this discussion) is transgenerational feedback. In the extreme case of maximum sharpness, absence of transgenerational feedback could mean a 65% reduction in epistatic expression space. This prompts the hypothesis:

Transgenerational feedback is the likely driving mechanism for the epistasis caused by noise, including the global negative epistasis coupling for which noise is deemed partially responsible.

- When applying the two-state model to observed fitness landscapes from [20], the transgenerational feedback emerged as an evolutionary factor to consider for essential genes. Concretely, for 11 out of 25 of the essential genes fitness is increased >5% for >5% of the commonly explored evolutionary space between gene deletion and duplication, which allows the typical beneficial mutation [259] in that space to become accessible (for 8 of those 11 even in >10% of the space). The same holds for a small minority of the non-essential genes.

While non-essential genes are relatively free to evolve, the evolution of essential genes provides a certain bottle-neck in terms of adaptability. To obtain predictability of possible evolutionary trajectories, all the possible accessible mutations need to be considered. Transgenerational feedback is responsible for a significant addition to these paths. The question is the degree to which we can extend this finding.

Very little had to be assumed for the feedback to work. Preferably, $V < 1$ for the strongest effect, which comes naturally in the models used in this dissertation as long as protein life time is much longer than mRNA lifetime. This still holds true to some extent for higher organisms as mouse fibroblasts and humans, see [281]. The protein lifetimes from that paper seem long enough to allow some memory across generations, another important factor. Moreover, turnover characteristics seem well conserved across mammals [282], which suggested to authors a new turnover model. This was needed to avoid an in their eyes unlikely conserved elaborate regulatory mechanism to adequately maintain the right copy number in every situation. However, the basic and automatic transgenerational feedback will contribute to this, reducing the need for extensive regulation. Therefore, extrapolating the results from this chapter to the yeast genome and beyond leads to the following hypothesis:

Evolvability of 25% to 60% of the essential genes in yeast are significantly affected by transgenerational feedback. The mRNA and protein lifetimes, together with the simple nature of the mechanism, suggest similar numbers for other organisms, including mammals.

Returning to the principal goal in this dissertation, adding to rules of life (defined by evolution and self-organization) in the context of genotype-phenotype map constructions, transgenerational feedback provides an interesting new epigenetic mechanism. It exclusively acts as a promoter of evolvability, a rare property for an evolutionary mechanism. The feedback circumvents the supposed trade-off between robustness and adaptability. In the worst case it is neutral, while it is strongest when it is most needed, at the edge of viability. In combination with noise, realized fitness landscapes are non-trivially different from the mesotypes one would predict on the basis of protein function. The feedback and noise therefore work in tandem to alter the genotype-phenotype map, and form part of the rules of evolution.

4.6 Supplementary information on methods

4.6.1 Epistasis fraction explained and as function of noise level

To determine the contributions to the observed epistasis for the GAP deletions in a $\Delta bem1$ background due to Cdc42 copy number noise, the epistasis is calculated using 4.4, with the experimental data from [24], see also Table 5. The calculated observed value was -0.4943.

The fitted fitness values follow from substituting different values for c_i for different strain backgrounds in 4.3, with a cumulative distribution function (CDF) of choice that has coefficient of variation V and mean 1 (the ratio c_i is normalized with respect to endogenous expression). The CDFs are specified as in Table 10 (the three parameter one allows some degeneracy), to become a function of V alone:

Table 10 Distribution types and parameter choice for CDFs with mean 1 and coefficient of variation V .

CDF	Parameters			
Gamma	$k = 1/V^2, \theta = V^2$ (shape, scale)			
Triangular	a (lower limit)	b (mode location)	c (upper limit)	Condition
	$1 - V\sqrt{6}$	1	$1 + \sqrt{6}$	$V \leq 1/\sqrt{6}$
	0	$\frac{3}{2} - \sqrt{6V^2 - 3/4}$	$\frac{3}{2} + \sqrt{6V^2 - 3/4}$	$\frac{1}{\sqrt{6}} < V \leq 1/\sqrt{2}$
	not defined	not defined	not defined	$V > 1/\sqrt{2}$
Uniform	a (lower limit)	c (upper limit)	Condition	
	$1 - V\sqrt{3}$	$1 + \sqrt{3}$	$V \leq 1/\sqrt{3}$	
	not defined	not defined	$V > 1/\sqrt{3}$	
Pareto	$\alpha = 1 + \sqrt{1 + 1/V^2}, x_m = (\alpha - 1)/\alpha$ (shape, scale)			
Log-normal	$\mu = -\sigma^2/2, \sigma = \sqrt{\ln(1 + V^2)}$			

For Cdc42, analysis on the measured flow cytometry data had shown $V = 0.832$. Fixing V to this value (or the maximum allowed within a CDF to be defined) then resulted in a three-parameter fit to find the appropriate c_i 's that minimized the sum of absolute standard scores (time deviations from observed doubling times, divided by the observed standard error). The c_i 's of $\Delta bem1, \Delta bem1 \Delta bem3$ and $\Delta bem1 \Delta bem2$ are the parameters, as the c_i of $\Delta bem1 \Delta bem3 \Delta bem2$ follows from Figure 32, since $c_{i,\Delta bem1 \Delta bem3 \Delta bem2} = c_{i,\Delta bem1 \Delta bem3} + c_{i,\Delta bem1 \Delta bem2} - c_{i,\Delta bem1}$, with the restriction that it is positive. The objective minimization was conducted in Matlab R2014b using successively *fminsearch*, *fminsearch* and *fminunc*, all with maximum function evaluations and iterations set to 10000. Each use the previous output as its starting point. The first starting vector was $(c_{i,\Delta bem1}, c_{i,\Delta bem1 \Delta bem2}, c_{i,\Delta bem1 \Delta bem3}) = (0.975, 0.675, 0.45)$. Relative GAP fractions then follow from $(c_{i,\Delta bem1} - c_{i,\Delta bem1 \Delta GAP})/c_{i,\Delta bem1}$ and are plotted in Figure 39. Then, fixing these c_i 's, coefficient of variation V is varied to see how epistasis varies along this dimension. This changes CDFs through Table 10 and fitness through 4.3 resulting in Figure 78 (see Appendix K).

4.6.2 Fitness/survival contours as function of noise level and MI-point location

A grid of c_i from 0.001 to 2.25 with step 0.00075 and V from 0.001 to 1 with step 0.001 was generated, which defined the gamma CDF output values using Table 10 for the contour plots in Figure 41, Figure 80 and Figure 45. Fitness on this grid then followed from 4.19 and 4.20. For consistency with results without transgenerational feedback (explained further on), fitness values are rounded to 6 decimals. The lower/upper boundary example was implemented using $k = +/ -10^5$, $g_0 = 0$, while the shallow example had $k = -10^5$, $g_0 = 1.2$ to insert in 4.15. In case $0 < \lambda_{max} < 1$, this does not result in structural growth and a defined fitness. Instead, it provides the basis for the number of survivable generations n_{gen} until the population size reaches 1, starting with a population size N_0 of 10^6 , through:

$$N_0 \lambda_{max}^{n_{gen}} = 1 \Rightarrow n_{gen} = -\log(N_0) / \log \lambda_{max} \quad (4.6)$$

Analogously, fitness and n_{gen} grids were calculated without transgenerational feedback replacing 4.19 with 4.36. However, as explained in the text, the finite population size is relevant here for the growth factors / doubling times, which modifies λ_{max} slightly to:

$$\left\{ \begin{array}{l} \frac{g(4c_i/3)[1 - F(4c_i/3)]^* + g(2c_i/3)[F(2c_i/3)]^*}{[1 - F(4c_i/3)]^* + [F(2c_i/3)]^*}, [1 - F(4c_i/3)]^* = 0 \wedge [F(2c_i/3)]^* > 0 \\ g(4c_i/3), [F(2c_i/3)]^* = 0 \\ g(2c_i/3), [1 - F(4c_i/3)]^* = 0 \wedge [F(2c_i/3)]^* > 0 \end{array} \right.$$

where $[\]^*$ denotes rounding to $[\log_{10} N_0]$ number of digits, which is 6 here.

Contours were generated for these grids of fitness and number of survivable generations through the *contour* function in Matlab 2014b. Because of the rounding of the CDF, this also sets the accuracy of the calculated fitnesses, which is rounded to the same number of digits (here, 6). Derivatives needed for the epistasis generating function (4.6) plots are approximated differences along the line $V = 0.5$, using Matlab's *diff*. As the rounding without feedback induced some numerical discontinuities, the derivatives are smoothed by a 25-point robust locally estimated scatter plot smoothing. These calculations yield Figure 48, Figure 81 (Appendix L) and Figure 50. Incidentally, Figure 38 contains a contour plot example regarding epistasis. The functional form of the fitness is given as $\omega_r = \exp(-F_V(1.5)c_i^{1-(V-0.5)/1.3})$.

4.6.3 Epistasis generic mutations figure

For the epistasis between a mesotype-preserving and generic mutation, the sharp lower boundary case is considered, with an average V of 0.22 [229]. The relative fitness point of origin (to which fitness differences are measured) is 0.96 ($c_i = 0.9375$), with 1 as the optimum without generic mutations. The sweep along the horizontal axis is performed in analogy with 4.6.2, with feedback, keeping V fixed and with a grid of c_i from 0.6675 to 1.14 (steps of 0.0037). As fitness are inversely proportional to cycle time T_0 , generic mutations are absorbed into the model by changing cycle time to θT_0 and shifting mean expression, so that:

$$\omega_{r,mut\ generic}(c_i) = \theta / \omega_r(c_i / \theta)$$

with θ ranging from 0.9 to 1.15 (step 0.0005) in Figure 42. Fitness values are then converted to epistasis values ϵ through 4.4, for the deviation in expected effect of combining the mesotype-preserving shift in c_i mutation (relative to the origin, $c_i = 0.9375$) and the shift in θ mutations (relative to the origin, $\theta = 1$), so:

$$\epsilon = \log_{10} \left(\frac{\omega_{generic} \omega_{mesotype}}{\omega_{generic+mesotype} \omega_{origin}} \right)$$

Sign epistasis occurs when the generic mutation is beneficial relative to the origin but is deleterious when the mesotype-preserving mutation is already in place, or vice versa.

4.6.4 Fitting empirical fitness landscapes

Empirical fitness landscapes (glucose data) from [20] were re-analyzed. There, fluorescence intensity values at various expression levels were given, which need to be translated to c_i 's from the two-state model. This implies dividing by WT intensities, but it is not so robust to do this directly from the reported fluorescence measurements of WT expression, since these contained sub-zero and missing values. Instead, converting intensities to c_i 's is done indirectly by first converting to protein copy numbers.

Estimates for the average copy number of these gene products are obtained from [234]. As this data set did not contain all of the original 81 genes, 8 were discarded. The fluorescence/protein conversion factor was calculated robustly by finding the median of the measured WT expression in fluorescence and the 73 remaining WT copy number estimates, and applied to the reported relative (to WT) promoter strengths. Then, the copy numbers for the 73 genes across all expression levels are normalized using their respective WT copy number values. Since c_i is the ratio of the concentration location of the mesotypical inflection point, divided by the mean expressed concentration, different c_i per protein are retrieved by divided to these normalized expression values.

The two-state model is then fitted with four free parameters; c_i , k , g_0 and V , where V is actually an effective average of the noise of all used promoters. The fitness values in [20] were defined as average number of doublings in a certain time frame, divided by that of WT, which in the two-state model translates as $\lambda_{max}/2$. This is because the growth factors of a general background and that of WT are λ_{max} (4.20) and 2 (perfection) respectively. The four-parameters set yielding the lowest sum of squared residuals of the measured fitness values and two-state model predictions were then found using Matlab's *fminsearch* (maximum function evaluations and iterations 10^4 , function and parameter change tolerance 10^{-18}).

Appropriate starting values for the search were found by finding the lowest sum of squared residuals in a grid of parameter values, with ranges of -2 to 2 (20 logarithmically spaced

points) for c_i , 0.001 to 1.999 (20 linearly spaced points) for g_0 , $-10^{1.999}$ to $-10^{-1.999}$ (20 logarithmically spaced points), 0 and $10^{-1.999}$ to $10^{1.999}$ for k and finally, 0.001 to 0.999 (20 linearly spaced points) for V . For the search is conducted, parameters are transformed using the log and atanh functions and their inverses to impose the bounds $c_0 > 0$, $0 \leq g_0 \leq 2$, $-200 \leq k \leq 200$ and $0 \leq V \leq 1$. The first restriction is for sensibility, the restriction on fitness landscape depth ensures the number of progeny is between 0 and 2 per division, k is only between these values as large values of $|k|$ are essentially indiscernible for interpretation as a cooperativity coefficient, and values of V below 1 are typical for yeast [229].

Then, the adjusted R-squared of the original impulse fits were calculated for all genes, and only those genes for which the two-state model fit provided better adjusted R-squared were kept for further analysis (64 out of 73). To predict the effect of transgenerational feedback, the actual coefficients of variation are needed (not the effective fitted V from the collection of promoters). Linking the 64 remaining genes to observed V 's from [229], left only 50 genes for further analysis, depicted in Figure 44.

For these genes, relative fitness was calculated (as defined in 4.19, 4.36 of this dissertation, not as defined in [20]) with and without feedback. The relative percentual benefit of feedback was calculated along a grid of 0 to 2 times WT expression (0.001 step size). The percentage of points on the grid that were viable with feedback and whose feedback benefit exceeded 5% were plotted in Figure 51. The essential genes were gathered from <https://www.yeastgenome.org/observable/APO:0000112> (date of access 01-08-2019), selecting for inviable null mutants in S288c, and marked in the plot.

For the predicted range of (essential) genes that the feedback affects, we note that 11 out of the tested 25 essential genes are affected by at least 5% in their evolutionary relevant expression space (and 14 less than 5%). If we consider the likelihood of encountering an affected essential gene out of the pool of all essential genes binomial and assuming no prior knowledge about the fraction affected (uniform prior, beta(1,1) distribution), the posterior density of the affected fraction follows a beta(11, 14) (representing 11 affected, 14 not affected) distribution. Consequently, the highest posterior density interval (as the credible interval) for the fraction affected is between 0.26 and 0.63, so the fraction affected essential genes should roughly be between 25% and 60%.

Similarly, for the non-essential genes, the posterior density of the affected fraction follows a beta(1 affected, 24 not affected) distribution. The highest posterior density interval (as the credible interval) for the fraction affected is between 0.002 and 0.17, so the fraction affected non-essential genes should roughly be below 17%. By Monte Carlo simulation of a 20%/80% mixture of beta(11,14) and beta(1,24) distributions, we get an estimate of the distribution of affected genes, assuming 20% is essential and 80% is non-essential [59]. Taking the 2.5% and 97.5% quantiles of the resulting distribution, the bounds on the total of affected genes can be generated, which are 8% and 25% respectively.

4.6.5 Simulated noise levels and correlations in feedback experimental design

To get an idea of the influence of $[Cdc42]_{min}$ on concentration of Cdc42, a sweep was conducted of normalized values $[Cdc42]_{min}/(P_{WT}/4/3\pi r_{min}^3)$ from 0.03 to 0.18 (step 0.03), 0.2 to 0.28 (step 0.02) and 0.3 to 0.45 (step 0.01) in the model from Chapter 3 with standard parameters (see Table 4) in the $\Delta bem1$ background. This yielded simulated volume and copy number count (a simulated “flow cytometry measurement”) for Cdc42, measured at a single time point where the colony size is around $5 \cdot 10^6$. Dividing copy numbers by the volumes yielded concentrations, whose distributions were approximated by a kernel smoothing function (Matlab R2014b *ksdensity*) at points 0 to $500 \text{ \#}/\mu^3$, with step size $5 \text{ \#}/\mu^3$. These were plotted together with the calculated respective doubling times in Figure 52.

Then, simulations continued for two values of the Cdc42 threshold, as fitted in 3.3.1. Volume and copy number counts for Cdc42 were calculated in the simulations, and this time also for another protein with identical turnover characteristics but without function in polarity. The samples were taken at either the last simulated end of G1 (containing $\sim 1/2$ the cells) or at a single time point (for Figure 53 and Figure 54 respectively), both where the colony size is around $5 \cdot 10^6$. Again, only live cells were considered for the simulated volume and flow cytometer data. The data was then reduced to 10000 cells, a typical experimental size.

Subsequently, the data is further divided, such that the effect of sample size is visible. The 100 sizes considered are logarithmically spaced between 10 and the maximum of 10000, and the data is such that e.g., all data included in the sample with size 10 is also included in the sample of size 11, etc. For every sample size, we generate 10000 bootstrapped samples to compute the coefficient of variations (squared) for the copy numbers and concentration in $\Delta bem1 \Delta bem3$ and the $\Delta bem1$ for Cdc42 and the Cdc42-like proteins (mCherry in the experimental proposal). The latter protein is assumed to be long lived [266], [267], and for the simulation a factor three for the protein lifetime above that of Cdc42 is chosen. The 95% confidence intervals of the ratio of V 's of Cdc42/Cdc42-like followed from the 0.025 and 0.975 quantiles of the bootstrapped results. The median is chosen as the highlighted values between these bounds, shown in Figure 53 and Figure 54. As an visual example of feasibility of the experiments, Figure 55 and Figure 56 show the protein (copy number / concentration) distributions at 100 (end of G1) and 10000 (arbitrary cell cycle stage) cells respectively.

Then, repeating simulations only having another Cdc42-like protein added to be traced, it is possible to calculate correlations between Cdc42 and the Cdc42-like protein, as well as between Cdc42-like proteins. Under optimal growth conditions (not the case in Appendix N), the copy numbers correlate differently across backgrounds, which depends also on the health of the subsets of cells considered. To envision this, data is sorted and correlations are plotted (from 1000 bootstrap samples, 95% confidence intervals from 0.025 and 0.975 quantiles as before) for a grid of left quantiles (0.05 to 1 with step size 0.05) of copy number data. As described in 4.4.3, the Spearman ρ rank correlation coefficient is used, to be robust against experimental deviations of matched mean expression of Cdc42 and Cdc42-like proteins.

5 Conclusions and discussion

This century holds great promise for synthetic biology and medicine. In the far future, it will be possible to repair any genetic defect, find weak spots for every pathogen, regenerate organs, and device organisms to manufacture any organic material we want. Key in these advances is complete understanding of how genes cause traits (phenotypes), i.e., the genotype-phenotype (GP)-map. For example, the powers of Crispr-Cas already allow vast possibilities in genetic manipulations, and the future of prime editing [283] looks even more promising. But what is currently most limiting is not the question *how* to make desired changes, but rather *what* these are. Therefore, increased efforts in deciphering the rules for the built-up of biological systems, which *self-organize* and *evolve*, are needed.

Important open questions include how the structure of the GP-map is composed. What are possible stepping stones or more metaphorically, islands in the ocean between genotype and phenotype? Is there at least one model system where the complete voyage across water can be validated and what boat (experimental design) would be needed? Also, what hidden treasures can be found on these islands, perhaps a spring of epistasis, or a chest containing an evolutionary mechanism?

My vision on how to accomplish understanding of the GP-map is working our way from genotype to (population) phenotype through levels, where each contains relatively simple, coarse-grained rules of the level below. As chapter one and three explained in more detail, these level definitions are far from trivial. Actually, it is not even certain that the most suitable level choice to understand a self-organizing system is the same as for an evolving one. Several attempts have been made in literature, each with their own purpose, but not oriented on gaining general and fundamentally justifiable insights in the process of phenotype generation that can be translated to new, understandable rules of life.

Throughout this dissertation, I have proposed and tested a new definition to bridge genotype and phenotype, called **mesotype**. This is a biophysically justified simplification of the effect of varying protein concentrations, which is derived from molecular mechanisms. For example, the mesotype can be as simple as “divide when a protein concentration exceeds a certain threshold”. If needed, the description can be made more detailed, but in any case, the mesotype relies on deep knowledge of the molecular mechanisms emerging from the proteins involved in the phenotype of interest. Consequently, this currently imposes strong requirements on model systems where this level definition can be tested.

As demonstrated in chapter one (1.3), polarization in budding yeast fulfills all requirements. This relatively modular process involves about ten core proteins to *self-organize* a protein (Cdc42) pattern that is used to mark the next site of division and conveniently, it has precedent in evolutionary studies as well. Decades of research allow the formation of an

interpretable physical interaction map, which must precede experimental efforts to reduce the system to a tractable form. Only then it is feasible to distinguish so-called functional subunits [51] that make up molecular mechanisms, which are ultimately responsible for the generation of the mesotype that acts as a summary of protein fitness landscapes.

However, the concept of functional subunits is new itself and required validation. To this end, its relatively extreme prediction for the yeast polarity mutant $\Delta bem1$ (lacking an important Cdc42 recruitment player) was validated, which was the focus of attention in Chapter two. In short, the prediction was that on the cellular level, a sharp minimum Cdc42 concentration boundary existed for polarization (and hence viability). The validation required variably inducing Cdc42 expression (through the galactose-sensitive *GAL1* promoter), and a fluorescent tag to cross-validate expression. Despite the collection of data on the population level, where such strong dependencies are diluted (through the mesotype to phenotype transition), we measured the expected fitness effects of variable Cdc42 expression and polarity genotype (such as GAP deletions, the deactivating agent for Cdc42) in section 2.3. En passant, the efforts to make this experiment possible also yielded the first fully functional fluorescently labeled Cdc42 in budding yeast, mimicking the strategy employed in [185] for fission yeast.

With the functional subunits validated, the focus was shifted to validating the utility of the mesotype concept, which allows translation of molecular mechanisms to population phenotypes. This entails for example, quantitative understanding of how the aforementioned sharp Cdc42 boundary is diluted in population fitness curves, or the emergence of complex phenotypes as measured in [24]. To this end, a cell growth model was constructed in chapter three, incorporating protein turnover, simple cell expansion dynamics and the mesotype. As simulations of this model successfully replicated the observed, non-standard phenotypes (section 3.3), this inspired enough confidence in the mesotype-including growth model to make new general predictions that stem from deeper understanding of the GP-map.

Interestingly from an evolutionary perspective, small changes in the environment can have large impacts on competitive fitness. For example, the model from Chapter 3 predicts that while the $\Delta bem1$ grows 5 times slower than WT under lab conditions, the difference is only 20% when suboptimal food slows WT down by 30% (see Figure 34 in 3.4.1). This equalizing behavior is also consistent with observations in growth assays done in Chapter 2. This supports the general suggestion that lab conditions are ideal for (almost) perfectly-tuned strains as WT. Yet, out in the wild where nutrition is rarely optimal, genetic backgrounds that are severely hampered have ample time to catch-up while the best genotypes are stalled. This makes an important difference for experimental evolution (in the lab) and historical evolution (as occurred in the wild). The explorable evolutionary space may in the latter case be much larger, as are the consequences of drift, as the forces of selection are either weak or short-lived (the brief moment food is abundant).

A second hypothesis also emerged, namely the evolutionary importance of expression noise, which was further considered in chapter four. As results from chapter three did not reveal that details in cell growth or degradation were of much importance, a model containing only noisy protein production and the mesotype was considered. Cells were divided into two bins; having a high or low protein number. Several interesting results follow from this minimal model.

Firstly, an important and likely abundant evolutionary mechanism is found that underlies expression noise, called transgenerational feedback. For suboptimal genotypes, it can utilize the noise to create a sub-population of more favorable protein copy numbers, which causes a bias for the next generation to inherit better protein copy numbers than their ancestors had. This forms a simple epigenetic adaptation mechanism, with which the population ‘memorizes’ the fitness landscape.

Particularly when the mesotypical landscape is rather sharp compared to the broadness of the protein production distribution, this confers an advantage; steady state growth is improved and population survival times for sustainably non-viable genotypes are increased. Based on observed landscapes in [20], I predict 25-60% of the essential genes to have expanded their permitted expression range by >5%, a significant contribution to their evolvability (see Figure 51 in 4.3.5). Moreover, this mechanism is never deleterious and does not sacrifice robustness for adaptability, a notable feature.

The transgenerational feedback, while seemingly inevitable, will have to be validated in at least one model system. Again, polarization in budding yeast, and in particular the $\Delta bem1$ situation seems suited for the job. Section 4.4 describes an experimental design appropriate for testing the presence of the feedback. When proven, and given that this epigenetic mechanism does not require any additional functions in the cell (e.g., no formation of histones that curl DNA) nor does it come at any energetic cost, I expect this mechanism to be a factor in many organisms. Moreover, the balance between mRNA, protein and cell life times of mammals [281] and yeast does not seem to differ much, and turnover seems conserved [282] (which could erase the protein copy number memory), further suggesting that the feedback exists as a general phenomenon across species.

Secondly, expression noise emerges as a source of epistasis, mainly the negative variant. The implications may be profound. Suppose we consider two unrelated mutations, which both individually affect fitness, but one through affecting the respective mean protein copy number and another generically through a change in average cell cycle time. These fall into the category of mutations investigated in chapter 4. As the cell moves away from the optimal protein concentration, negative epistasis fueled by noise is shown to be more likely than positive epistasis. This can provide the “biological basis for global coupling”, leading to diminishing returns in evolutionary adaptation (see sections 4.2.1.7 and 4.2.2.3), an open question put forward in [260].

Ultimately, the question remains whether the bottom-up approach from genotype through protein, functional subunits, molecular mechanisms, mesotype, phenotype and ultimately, population phenotype will always be so tractable and fruitful. Fortunately, we see that most observed fitness landscapes in yeast are readily decomposed into expression noise and a tractable mesotype (a simple Hill curve with documented interpretations). This looks promising for future built-ups of the GP-map in other model systems and organisms (see 4th discussion point in 4.5).

In summary, important rules of life, when defined as self-organizing, evolving systems, are found: functional subunits are an experimentally verified way of coarse-graining genotypes. The mesotype in turn provides a suitable level above this, for both self-organization and evolution, to bridge the GP-transition. From this we learned that the likely abundant transgenerational feedback is a prime example that evolutionary mechanisms do not always trade-off adaptability with robustness. Moreover, pervasive negative epistasis can have its source in inevitable expression noise. How much further these level definitions will bring us will follow from applying them to more situations in more organisms, but I hope to have contributed to a promising start.

Bibliography

- [1] P. L. Luisi, "About various definitions of life," *Orig. Life Evol. Biosph.*, vol. 28, no. 4–6, pp. 613–622, 1998.
- [2] E. Karsenti, "Self-organization in cell biology: a brief history," *Nat. Rev. Mol. Cell Biol.*, vol. 9, no. 3, p. 255, 2008.
- [3] C. Ofria and C. O. Wilke, "Avida: A Software Platform for Research in Computational Evolutionary Biology," *Artif. Life*, vol. 10, no. 2, pp. 191–229, Mar. 2004.
- [4] G. M. Hodgson and T. Knudsen, *Darwin's conjecture: The search for general principles of social and economic evolution*. University of Chicago Press, 2010.
- [5] L. Smolin, "Time, laws, and the future of cosmology," *Phys. Today*, vol. 67, no. 3, pp. 38–43, Mar. 2014.
- [6] J. Näsvall, L. Sun, J. R. Roth, and D. I. Andersson, "Real-time evolution of new genes by innovation, amplification, and divergence," *Science*, vol. 338, no. 6105, pp. 384–387, 2012.
- [7] Z. D. Blount, J. E. Barrick, C. J. Davidson, and R. E. Lenski, "Genomic analysis of a key innovation in an experimental *Escherichia coli* population," *Nature*, vol. 489, no. 7417, pp. 513–518, Sep. 2012.
- [8] M. A. Fisher, K. L. McKinley, L. H. Bradley, S. R. Viola, and M. H. Hecht, "De Novo Designed Proteins from a Library of Artificial Sequences Function in *Escherichia Coli* and Enable Cell Growth," *PLoS ONE*, vol. 6, no. 1, p. e15364, Jan. 2011.
- [9] S. C. Patel and M. H. Hecht, "Directed evolution of the peroxidase activity of a de novo-designed protein," *Protein Eng. Des. Sel.*, vol. 25, no. 9, pp. 445–452, Sep. 2012.
- [10] A. M. Selmecki *et al.*, "Polyploidy can drive rapid adaptation in yeast," *Nature*, vol. 519, no. 7543, pp. 349–352, Mar. 2015.
- [11] M. Springer, J. S. Weissman, and M. W. Kirschner, "A general lack of compensation for gene dosage in yeast," *Mol. Syst. Biol.*, vol. 6, May 2010.
- [12] A. C. Gerstein and S. P. Otto, "Ploidy and the Causes of Genomic Evolution," *J. Hered.*, vol. 100, no. 5, pp. 571–581, Sep. 2009.
- [13] R. B. Langkjaer, P. F. Cliften, M. Johnston, and J. Piškur, "Yeast genome duplication was followed by asynchronous differentiation of duplicated genes," *Nature*, vol. 421, no. 6925, p. 848, 2003.
- [14] L. Cui, "Widespread genome duplications throughout the history of flowering plants," *Genome Res.*, vol. 16, no. 6, pp. 738–749, Jun. 2006.
- [15] T. D. Cuyper and P. Hogeweg, "Virtual Genomes in Flux: An Interplay of Neutrality and Adaptability Explains Genome Expansion and Streamlining," *Genome Biol. Evol.*, vol. 4, no. 3, pp. 212–229, Mar. 2012.
- [16] M. Lynch and G. K. Marinov, "The bioenergetic costs of a gene," *Proc. Natl. Acad. Sci.*, p. 201514974, 2015.
- [17] S. F. Greenbury, S. Schaper, S. E. Ahnert, and A. A. Louis, "Genetic Correlations Greatly Increase Mutational Robustness and Can Both Reduce and Enhance Evolvability," *PLoS Comput. Biol.*, vol. 12, no. 3, p. e1004773, Mar. 2016.
- [18] A. Crombach and P. Hogeweg, "Evolution of evolvability in gene regulatory networks," *PLoS Comput. Biol.*, vol. 4, no. 7, p. e1000112, 2008.
- [19] M. Kimura, *The neutral theory of molecular evolution*. Cambridge University Press, 1983.

- [20]L. Keren *et al.*, “Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness,” *Cell*, vol. 166, no. 5, p. 1282–1294.e18, Aug. 2016.
- [21]J. van Leeuwen *et al.*, “Exploring genetic suppression interactions on a global scale,” *Science*, vol. 354, no. 6312, pp. aag0839–aag0839, Nov. 2016.
- [22]A. B. Gjuvsland, B. J. Hayes, S. W. Omholt, and Ö. Carlborg, “Statistical epistasis is a generic feature of gene regulatory networks,” *Genetics*, vol. 175, no. 1, pp. 411–420, 2007.
- [23]D. J. Kvittek and G. Sherlock, “Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape,” *PLoS Genet.*, vol. 7, no. 4, p. e1002056, Apr. 2011.
- [24]L. Laan, J. H. Koschwanez, and A. W. Murray, “Evolutionary adaptation after crippling cell polarization follows reproducible trajectories,” *eLife*, vol. 4, p. e09638, 2015.
- [25]S. Wright, *The roles of mutation, inbreeding, crossbreeding, and selection in evolution*, vol. 1. na, 1932.
- [26]R. J. Woods, J. E. Barrick, T. F. Cooper, U. Shrestha, M. R. Kauth, and R. E. Lenski, “Second-order selection for evolvability in a large *Escherichia coli* population,” *Science*, vol. 331, no. 6023, pp. 1433–1436, 2011.
- [27]F. J. Poelwijk, S. Tănase-Nicola, D. J. Kiviet, and S. J. Tans, “Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes,” *J. Theor. Biol.*, vol. 272, no. 1, pp. 141–144, Mar. 2011.
- [28]J. A. Draghi, T. L. Parsons, G. P. Wagner, and J. B. Plotkin, “Mutational robustness can facilitate adaptation,” *Nature*, vol. 463, no. 7279, pp. 353–355, Jan. 2010.
- [29]A. J. Stewart, T. L. Parsons, and J. B. Plotkin, “Environmental robustness and the adaptability of populations,” *Evolution*, vol. 66, no. 5, pp. 1598–1612, May 2012.
- [30]G. P. Wagner and J. Zhang, “The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms,” *Nat. Rev. Genet.*, vol. 12, no. 3, pp. 204–213, Mar. 2011.
- [31]M. Costanzo *et al.*, “The genetic landscape of a cell,” *science*, vol. 327, no. 5964, pp. 425–431, 2010.
- [32]J. T. Gao *et al.*, “Modular coherence of protein dynamics in yeast cell polarity system,” *Proc. Natl. Acad. Sci.*, vol. 108, no. 18, pp. 7647–7652, May 2011.
- [33]L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, “From molecular to modular cell biology,” *Nature*, vol. 402, pp. C47–C52, 1999.
- [34]S. Johnson and N. S. Jones, “Looplessness in networks is linked to trophic coherence,” *Proc. Natl. Acad. Sci.*, vol. 114, no. 22, pp. 5618–5623, May 2017.
- [35]C. Adami, “What is complexity?,” *BioEssays*, vol. 24, no. 12, pp. 1085–1094, Dec. 2002.
- [36]M. Tessera and G. A. Hoelzer, “On the thermodynamics of multilevel evolution,” *Biosystems*, vol. 113, no. 3, pp. 140–143, Sep. 2013.
- [37]M. A. Maslin, C. M. Brierley, A. M. Milner, S. Shultz, M. H. Trauth, and K. E. Wilson, “East African climate pulses and early human evolution,” *Quat. Sci. Rev.*, vol. 101, pp. 1–17, Oct. 2014.
- [38]K. Hammerschmidt, C. J. Rose, B. Kerr, and P. B. Rainey, “Life cycles, fitness decoupling and the evolution of multicellularity,” *Nature*, vol. 515, no. 7525, pp. 75–79, Nov. 2014.
- [39]A. H. Yona, I. Frumkin, and Y. Pilpel, “A Relay Race on the Evolutionary Adaptation Spectrum,” *Cell*, vol. 163, no. 3, pp. 549–559, Oct. 2015.
- [40]E. Heard and R. A. Martienssen, “Transgenerational Epigenetic Inheritance: Myths and Mechanisms,” *Cell*, vol. 157, no. 1, pp. 95–109, Mar. 2014.

- [41] B. Xue and S. Leibler, "Evolutionary learning of adaptation to varying environments through a transgenerational feedback," *Proc. Natl. Acad. Sci.*, vol. 113, no. 40, pp. 11266–11271, Oct. 2016.
- [42] M. G. J. de Vos, A. Dawid, V. Sunderlikova, and S. J. Tans, "Breaking evolutionary constraint with a tradeoff ratchet," *Proc. Natl. Acad. Sci.*, Nov. 2015.
- [43] R. Maharjan and T. Ferenci, "Mutational Signatures Indicative of Environmental Stress in Bacteria," *Mol. Biol. Evol.*, vol. 32, no. 2, pp. 380–391, Feb. 2015.
- [44] R. P. Maharjan and T. Ferenci, "A shifting mutational landscape in 6 nutritional states: stress-induced mutagenesis as a series of distinct stress input–mutation output relationships," *PLoS Biol.*, vol. 15, no. 6, p. e2001477, 2017.
- [45] C. V. Jack, C. Cruz, R. M. Hull, M. A. Keller, M. Ralsler, and J. Houseley, "Regulation of ribosomal DNA amplification by the TOR pathway," *Proc. Natl. Acad. Sci.*, vol. 112, no. 31, pp. 9674–9679, 2015.
- [46] S. Camazine, J.-L. Deneubourg, N. R. Franks, J. Sneyd, E. Bonabeau, and G. Theraula, *Self-organization in biological systems*, vol. 7. Princeton university press, 2003.
- [47] L. Laan *et al.*, "Cortical Dynein Controls Microtubule Dynamics to Generate Pulling Forces that Position Microtubule Asters," *Cell*, vol. 148, no. 3, pp. 502–514, Feb. 2012.
- [48] T. Freisinger *et al.*, "Establishment of a robust single axis of cell polarity by coupling multiple positive feedback loops," *Nat. Commun.*, vol. 4, p. 1807, May 2013.
- [49] Y. K. Takahashi, H. Kori, and N. Masuda, "Self-organization of feed-forward structure and entrainment in excitatory neural networks with spike-timing-dependent plasticity," *Phys. Rev. E*, vol. 79, no. 5, May 2009.
- [50] A. H. Chau, J. M. Walter, J. Gerardin, C. Tang, and W. A. Lim, "Designing Synthetic Regulatory Networks Capable of Self-Organizing Cell Polarization," *Cell*, vol. 151, no. 2, pp. 320–332, Oct. 2012.
- [51] F. Brauns, L. Iñigo de la Cruz, W. Daalman, I. de Bruin, E. Frey, and L. Laan, "Multiple distinct pattern forming mechanisms coexist in the polarization machinery of budding yeast," *to be submitted*.
- [52] A. B. Chitnis, D. Dalle Nogare, and M. Matsuda, "Building the posterior lateral line system in zebrafish," *Dev. Neurobiol.*, vol. 72, no. 3, pp. 234–255, Mar. 2012.
- [53] D. Dalle Nogare *et al.*, "Leading and trailing cells cooperate in collective migration of the zebrafish posterior lateral line primordium," *Development*, vol. 141, no. 16, pp. 3188–3196, Aug. 2014.
- [54] F. Caudron and Y. Barral, "A Super-Assembly of Whi3 Encodes Memory of Deceptive Encounters by Single Cells during Yeast Courtship," *Cell*, vol. 155, no. 6, pp. 1244–1257, Dec. 2013.
- [55] S. L. Rea, B. H. Graham, E. Nakamaru-Ogiso, A. Kar, and M. J. Falk, "Bacteria, yeast, worms, and flies: Exploiting simple model organisms to investigate human mitochondrial diseases," *Dev. Disabil. Res. Rev.*, vol. 16, no. 2, pp. 200–218, Aug. 2010.
- [56] H. Mewes *et al.*, "Overview of the yeast genome," *Nature*, vol. 387, no. 6632, pp. 7–8, 1997.
- [57] J. M. Cherry *et al.*, "Saccharomyces Genome Database: the genomics resource of budding yeast," *Nucleic Acids Res.*, vol. 40, no. D1, pp. D700–D705, Jan. 2012.
- [58] C. Stark, "BioGRID: a general repository for interaction datasets," *Nucleic Acids Res.*, vol. 34, no. 90001, pp. D535–D539, Jan. 2006.
- [59] G. Giaever and C. Nislow, "The Yeast Deletion Collection: A Decade of Functional Genomics," *Genetics*, vol. 197, no. 2, pp. 451–465, Jun. 2014.

- [60]L. W. Bergman, "Growth and maintenance of yeast," in *Two-hybrid systems*, Springer, 2001, pp. 9–14.
- [61]J. S. Smith and D. Burke, Eds., *Yeast genetics: methods and protocols*. New York: Humana Press, 2014.
- [62]R. Mans, M. Wijsman, P. Daran-Lapujade, and J.-M. Daran, "A protocol for introduction of multiple genetic modifications in *Saccharomyces cerevisiae* using CRISPR/Cas9," *FEMS Yeast Res.*, vol. 18, no. 7, Nov. 2018.
- [63]R. Dhar, R. SäGesser, C. Weikert, J. Yuan, and A. Wagner, "Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution: Adaptation of *S. cerevisiae* to saline stress," *J. Evol. Biol.*, vol. 24, no. 5, pp. 1135–1153, May 2011.
- [64]Z. Cakar, U. Seker, C. Tamerler, M. Sonderegger, and U. Sauer, "Evolutionary engineering of multiple-stress resistant," *FEMS Yeast Res.*, vol. 5, no. 6–7, pp. 569–578, Apr. 2005.
- [65]E. T. Diepeveen, V. Pourquie, T. Gehrmann, T. Abeel, and L. Laan, "Patterns of conservation and diversification in the fungal polarization network," *bioRxiv*, p. 154641, 2017.
- [66]S. G. Martin and R. A. Arkowitz, "Cell polarization in budding and fission yeasts," *FEMS Microbiol. Rev.*, vol. 38, no. 2, pp. 228–253, Mar. 2014.
- [67]S. G. Martin, "Spontaneous cell polarization: Feedback control of Cdc42 GTPase breaks cellular symmetry," *BioEssays*, vol. 37, no. 11, pp. 1193–1201, Nov. 2015.
- [68]B. Klünder, T. Freisinger, R. Wedlich-Söldner, and E. Frey, "GDI-Mediated Cell Polarization in Yeast Provides Precise Spatial and Temporal Control of Cdc42 Signaling," *PLoS Comput. Biol.*, vol. 9, no. 12, p. e1003396, Dec. 2013.
- [69]K. J. A. Vendel, S. Tschirpke, F. Shamsi, M. Dogterom, and L. Laan, "Minimal *in vitro* systems shed light on cell polarity," *J. Cell Sci.*, vol. 132, no. 4, p. jcs217554, Feb. 2019.
- [70]M. Pelillo, I. Poli, A. Roli, R. Serra, D. Slanzi, and M. Villani, Eds., *Artificial Life and Evolutionary Computation*, vol. 830. Cham: Springer International Publishing, 2018.
- [71]S. Etienne-Manneville, "Cdc42 - the centre of polarity," *J. Cell Sci.*, vol. 117, no. 8, pp. 1291–1300, Mar. 2004.
- [72]A. Hall, "The cellular functions of small GTP-binding proteins," *Science*, vol. 249, no. 4969, pp. 635–640, 1990.
- [73]B. E. Caplin, L. A. Hettich, and M. S. Marshall, "Substrate characterization of the *Saccharomyces cerevisiae* protein farnesyltransferase and type-I protein geranylgeranyltransferase," *Biochim. Biophys. Acta BBA-Protein Struct. Mol. Enzymol.*, vol. 1205, no. 1, pp. 39–48, 1994.
- [74]G. Koch, K. Tanaka, T. Masuda, W. Yamochi, H. Nonaka, and Y. Takai, "Association of the Rho family small GTP-binding proteins with Rho GDP dissociation inhibitor (Rho GDI) in *Saccharomyces cerevisiae*," *Oncogene*, vol. 15, no. 4, p. 417, 1997.
- [75]K. Shinjo *et al.*, "Molecular cloning of the gene for the human placental GTP-binding protein Gp (G25K): identification of this GTP-binding protein as the human homolog of the yeast cell-division-cycle protein CDC42," *Proc. Natl. Acad. Sci.*, vol. 87, no. 24, pp. 9853–9857, 1990.
- [76]J. L. Johnson, J. W. Erickson, and R. A. Cerione, "New Insights into How the Rho Guanine Nucleotide Dissociation Inhibitor Regulates the Interaction of Cdc42 with Membranes," *J. Biol. Chem.*, vol. 284, no. 35, pp. 23860–23871, Aug. 2009.
- [77]E. Boulter *et al.*, "Regulation of Rho GTPase crosstalk, degradation and activity by RhoGDI1," *Nat. Cell Biol.*, vol. 12, no. 5, pp. 477–483, May 2010.

- [78]J. Peterson, "Interactions between the bud emergence proteins Bem1p and Bem2p and Rho- type GTPases in yeast," *J. Cell Biol.*, vol. 127, no. 5, pp. 1395–1406, Dec. 1994.
- [79]I. Bose, J. E. Irazoqui, J. J. Moskow, E. S. G. Bardes, T. R. Zyla, and D. J. Lew, "Assembly of Scaffold-mediated Complexes Containing Cdc42p, the Exchange Factor Cdc24p, and the Effector Cla4p Required for Cell Cycle-regulated Phosphorylation of Cdc24p," *J. Biol. Chem.*, vol. 276, no. 10, pp. 7176–7186, Mar. 2001.
- [80]L. Kozubowski, K. Saito, J. M. Johnson, A. S. Howell, T. R. Zyla, and D. J. Lew, "Symmetry-Breaking Polarization Driven by a Cdc42p GEF-PAK Complex," *Curr. Biol.*, vol. 18, no. 22, pp. 1719–1726, Nov. 2008.
- [81]C. Tiedje, I. Sakwa, U. Just, and T. Höfken, "The rho gdi rdi1 regulates rho gtpases by distinct mechanisms," *Mol. Biol. Cell*, vol. 19, no. 7, pp. 2885–2896, 2008.
- [82]Y. Zheng, R. Cerione, and A. Bender, "Control of the yeast bud-site assembly GTPase Cdc42. Catalysis of guanine nucleotide exchange by Cdc24 and stimulation of GTPase activity by Bem3," *J. Biol. Chem.*, vol. 269, no. 4, pp. 2369–2372, 1994.
- [83]J. L. Bos, H. Rehmann, and A. Wittinghofer, "GEFs and GAPs: Critical Elements in the Control of Small G Proteins," *Cell*, vol. 129, no. 5, pp. 865–877, Jun. 2007.
- [84]S. E. Smith, B. Rubinstein, I. Mendes Pinto, B. D. Slaughter, J. R. Unruh, and R. Li, "Independence of symmetry breaking on Bem1-mediated autocatalytic activation of Cdc42," *J. Cell Biol.*, vol. 202, no. 7, pp. 1091–1106, Sep. 2013.
- [85]M.-P. Gulli, M. Jaquenoud, Y. Shimada, G. Niederhäuser, P. Wiget, and M. Peter, "Phosphorylation of the Cdc42 exchange factor Cdc24 by the PAK-like kinase Cla4 may regulate polarized growth in yeast," *Mol. Cell*, vol. 6, no. 5, pp. 1155–1167, 2000.
- [86]C.-C. Kuo, N. S. Savage, H. Chen, C.-F. Wu, T. R. Zyla, and D. J. Lew, "Inhibitory GEF Phosphorylation Provides Negative Feedback in the Yeast Polarity Circuit," *Curr. Biol.*, vol. 24, no. 7, pp. 753–759, Mar. 2014.
- [87]S. Aronov, R. Gelin-Licht, G. Zipor, L. Haim, E. Safran, and J. E. Gerst, "mRNAs Encoding Polarity and Exocytosis Factors Are Cotransported with the Cortical Endoplasmic Reticulum to the Incipient Bud in *Saccharomyces cerevisiae*," *Mol. Cell. Biol.*, vol. 27, no. 9, pp. 3441–3455, May 2007.
- [88]A. R. Marquitz, J. C. Harrison, I. Bose, T. R. Zyla, J. N. McMillan, and D. J. Lew, "The Rho-GAP Bem2p plays a GAP-independent role in the morphogenesis checkpoint," *EMBO J.*, vol. 21, no. 15, pp. 4012–4025, 2002.
- [89]Y. Zheng, R. Cerione, and A. Bender, "Control of the yeast bud-site assembly GTPase Cdc42. Catalysis of guanine nucleotide exchange by Cdc24 and stimulation of GTPase activity by Bem3," *J. Biol. Chem.*, vol. 269, no. 4, pp. 2369–2372, 1994.
- [90]G. R. Smith, S. A. Givan, P. Cullen, and G. F. Sprague, "GTPase-Activating Proteins for Cdc42," *Eukaryot. Cell*, vol. 1, no. 3, pp. 469–480, Jun. 2002.
- [91]D. McCusker *et al.*, "Cdk1 coordinates cell-surface growth with the cell cycle," *Nat. Cell Biol.*, vol. 9, no. 5, pp. 506–515, May 2007.
- [92]M. Knaus, M.-P. Pelli-Gulli, F. Van Drogen, S. Springer, M. Jaquenoud, and M. Peter, "Phosphorylation of Bem2p and Bem3p may contribute to local activation of Cdc42p at bud emergence," *EMBO J.*, vol. 26, no. 21, pp. 4501–4513, 2007.
- [93]Z. Tong, X.-D. Gao, A. S. Howell, I. Bose, D. J. Lew, and E. Bi, "Adjacent positioning of cellular structures enabled by a Cdc42 GTPase-activating protein mediated zone of inhibition," *J. Cell Biol.*, vol. 179, no. 7, pp. 1375–1384, Dec. 2007.

- [94] T. H. Chuang, X. Xu, U. G. Knaus, M. J. Hart, and G. M. Bokoch, "GDP dissociation inhibitor prevents intrinsic and GTPase activating protein-stimulated GTP hydrolysis by the Rac GTP-binding protein.," *J. Biol. Chem.*, vol. 268, no. 2, pp. 775–778, 1993.
- [95] A. C. Wild, J. W. Yu, M. A. Lemmon, and K. J. Blumer, "The p21-activated Protein Kinase-related Kinase Cla4 Is a Coincidence Detector of Signaling by Cdc42 and Phosphatidylinositol 4-Phosphate," *J. Biol. Chem.*, vol. 279, no. 17, pp. 17101–17110, Apr. 2004.
- [96] L. J. Watson, G. Rossi, and P. Brennwald, "Quantitative Analysis of Membrane Trafficking in Regulation of Cdc42 Polarity," *Traffic*, vol. 15, no. 12, pp. 1330–1343, Dec. 2014.
- [97] B. D. Slaughter, J. R. Unruh, A. Das, S. E. Smith, B. Rubinstein, and R. Li, "Non-uniform membrane diffusion enables steady-state cell polarization via vesicular trafficking," *Nat. Commun.*, vol. 4, p. 1380, Jan. 2013.
- [98] T. Leeuw *et al.*, "Pheromone Response in Yeast: Association of Bem1p with Proteins of the MAP Kinase Cascade and Actin," *Science*, pp. 1210–1213, 1995.
- [99] A. E. Adams, "Relationship of actin and tubulin distribution to bud growth in wild-type and morphogenetic-mutant *Saccharomyces cerevisiae*," *J. Cell Biol.*, vol. 98, no. 3, pp. 934–945, Mar. 1984.
- [100] K. R. Ayscough, J. Stryker, N. Pokala, M. Sanders, P. Crews, and D. G. Drubin, "High rates of actin filament turnover in budding yeast and roles for actin in establishment and maintenance of cell polarity revealed using the actin inhibitor latrunculin-A," *J. Cell Biol.*, vol. 137, no. 2, pp. 399–416, 1997.
- [101] M. Evangelista *et al.*, "Bni1p, a yeast formin linking cdc42p and the actin cytoskeleton during polarized morphogenesis," *Science*, vol. 276, no. 5309, pp. 118–122, 1997.
- [102] I. Sagot, S. K. Klee, and D. Pellman, "Yeast formins regulate cell polarity by controlling the assembly of actin cables," *Nat. Cell Biol.*, vol. 4, no. 1, p. 42, 2002.
- [103] D. C. Prosser, T. G. Drivas, L. Maldonado-Báez, and B. Wendland, "Existence of a novel clathrin-independent endocytic pathway in yeast that depends on Rho1 and formin," *J. Cell Biol.*, vol. 195, no. 4, pp. 657–671, Nov. 2011.
- [104] R. A. Arkowitz and N. Lowe, "A small conserved domain in the yeast Spa2p is necessary and sufficient for its polarized localization," *J. Cell Biol.*, vol. 138, no. 1, pp. 17–36, 1997.
- [105] R. Wedlich-Soldner, S. Altschuler, L. Wu, and R. Li, "Spontaneous cell polarization through actomyosin-based delivery of the Cdc42 GTPase," *Science*, vol. 299, no. 5610, pp. 1231–1235, 2003.
- [106] R. C. Aguilar *et al.*, "Epsin N-terminal homology domains perform an essential function regulating Cdc42 through binding Cdc42 GTPase-activating proteins," *PNAS*, vol. 103, no. 11, pp. 4116–4121, 2006.
- [107] B. L. Goode, J. A. Eskin, and B. Wendland, "Actin and Endocytosis in Budding Yeast," *Genetics*, vol. 199, no. 2, pp. 315–358, Feb. 2015.
- [108] E. Garí Marsol, T. Volpe, H. Wang, C. Gallego, B. Futcher, and M. Aldea, "Whi3 binds the mRNA of the G (1) cyclin CLN3 to modulate cell fate in budding yeast," *Genes Dev.* 2001 Vol 15 Núm 21 P 2803-8, 2001.
- [109] H. Wang, E. Garí, E. Vergés, C. Gallego, and M. Aldea, "Recruitment of Cdc28 by Whi3 restricts nuclear accumulation of the G1 cyclin-Cdk complex to late G1," *EMBO J.*, vol. 23, pp. 180–190, 2004.
- [110] M. Tyers, G. Tokiwa, R. Nash, and B. Futcher, "The Cln3-Cdc28 kinase complex of *S. cerevisiae* is regulated by proteolysis and phosphorylation.," *EMBO J.*, vol. 11, no. 5, pp. 1773–1784, 1992.

- [111] Y. Cai and B. Futcher, "Effects of the Yeast RNA-Binding Protein Whi3 on the Half-Life and Abundance of CLN3 mRNA and Other Targets," *PLoS ONE*, vol. 8, no. 12, p. e84630, Dec. 2013.
- [112] E. Vergés, N. Colomina, E. Garí, C. Gallego, and M. Aldea, "Cyclin Cln3 Is Retained at the ER and Released by the J Chaperone Ydj1 in Late G1 to Trigger Cell Cycle Entry," *Mol. Cell*, vol. 26, no. 5, pp. 649–662, Jun. 2007.
- [113] C. Gallego, E. Garí, N. Colomina, E. Herrero, and M. Aldea, "The Cln3 cyclin is down-regulated by translational repression and degradation during the G1 arrest caused by nitrogen deprivation in budding yeast," *EMBO J.*, vol. 16, no. 23, pp. 7196–7206, 1997.
- [114] B. Wendland, "Epsins: adaptors in endocytosis?," *Nat. Rev. Mol. Cell Biol.*, vol. 3, no. 12, p. 971, 2002.
- [115] D. Mukherjee *et al.*, "The yeast endocytic protein Epsin 2 functions in a cell-division signaling pathway," *J. Cell Sci.*, vol. 122, no. 15, pp. 2782–2782, Aug. 2009.
- [116] T. Lechler, A. Shevchenko, A. Shevchenko, and R. Li, "Direct Involvement of Yeast Type I Myosins in Cdc42-Dependent Actin Polymerization," *J. Cell Biol.*, vol. 148, no. 2, pp. 363–374, Jan. 2000.
- [117] A. B. Goryachev and M. Leda, "Many roads to symmetry breaking: molecular mechanisms and theoretical models of yeast cell polarity," *Mol. Biol. Cell*, vol. 28, no. 3, pp. 370–380, Feb. 2017.
- [118] R. Kawasaki, K. Fujimura-Kamada, H. Toi, H. Kato, and K. Tanaka, "The upstream regulator, Rsr1p, and downstream effectors, Gic1p and Gic2p, of the Cdc42p small GTPase coordinately regulate initiation of budding in *Saccharomyces cerevisiae*," *Genes Cells*, vol. 8, no. 3, pp. 235–250, 2003.
- [119] D. Stuart and C. Wittenberg, "CLN3, not positive feedback, determines the timing of CLN2 transcription in cycling cells.," *Genes Dev.*, vol. 9, no. 22, pp. 2780–2794, Nov. 1995.
- [120] M. Costanzo *et al.*, "CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast," *Cell*, vol. 117, no. 7, pp. 899–913, 2004.
- [121] L. R. Schenkman, C. Caruso, N. Pagé, and J. R. Pringle, "The role of cell cycle-regulated expression in the localization of spatial landmark proteins in yeast," *J. Cell Biol.*, vol. 156, no. 5, pp. 829–841, Mar. 2002.
- [122] C. Koch, A. Schleiffer, G. Ammerer, and K. Nasmyth, "Switching transcription on and off during the yeast cell cycle: Cln/Cdc28 kinases activate bound transcription factor SBF (Swi4/Swi6) at start, whereas Clb/Cdc28 kinases displace it from the promoter in G2.," *Genes Dev.*, vol. 10, no. 2, pp. 129–141, 1996.
- [123] C. Wittenberg, K. Sugimoto, and S. I. Reed, "G1-specific cyclins of *S. cerevisiae*: cell cycle periodicity, regulation by mating pheromone, and association with the p34CDC28 protein kinase," *Cell*, vol. 62, no. 2, pp. 225–237, 1990.
- [124] A. Doncic, M. Falleur-Fettig, and J. M. Skotheim, "Distinct Interactions Select and Maintain a Specific Cell Fate," *Mol. Cell*, vol. 43, no. 4, pp. 528–539, Aug. 2011.
- [125] R. Sopko, D. Huang, J. C. Smith, D. Figeys, and B. J. Andrews, "Activation of the Cdc42p GTPase by cyclin-dependent protein kinases in budding yeast," *EMBO J.*, vol. 26, no. 21, pp. 4487–4500, 2007.
- [126] V. Archambault, E. J. Chang, B. J. Drapkin, F. R. Cross, B. T. Chait, and M. P. Rout, "Targeted proteomic study of the cyclin-Cdk module," *Mol. Cell*, vol. 14, no. 6, pp. 699–711, 2004.

- [127] Y. Shimada, M.-P. Gulli, and M. Peter, "Nuclear sequestration of the exchange factor Cdc24 by Far1 regulates cell polarity during yeast mating," *Nat. Cell Biol.*, vol. 2, no. 2, pp. 117–124, 2000.
- [128] L. S. Garrenton, A. Braunwarth, S. Irniger, E. Hurt, M. Kunzler, and J. Thorner, "Nucleus-Specific and Cell Cycle-Regulated Degradation of Mitogen-Activated Protein Kinase Scaffold Protein Ste5 Contributes to the Control of Signaling Competence," *Mol. Cell Biol.*, vol. 29, no. 2, pp. 582–601, Jan. 2009.
- [129] A. Doncic and J. M. Skotheim, "Feedforward Regulation Ensures Stability and Rapid Reversibility of a Cellular State," *Mol. Cell*, vol. 50, no. 6, pp. 856–868, Jun. 2013.
- [130] J. R. Buchan, D. Muhrad, and R. Parker, "P bodies promote stress granule assembly in *Saccharomyces cerevisiae*," *J. Cell Biol.*, vol. 183, no. 3, pp. 441–455, Nov. 2008.
- [131] S. Kroschwald *et al.*, "Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules," *Elife*, vol. 4, p. e06807, 2015.
- [132] F. E. Loughlin *et al.*, "The zinc fingers of the SR-like protein ZRANB2 are single-stranded RNA-binding domains that recognize 5' splice site-like sequences," *Proc. Natl. Acad. Sci.*, vol. 106, no. 14, pp. 5581–5586, 2009.
- [133] D. J. Hogan, D. P. Riordan, A. P. Gerber, D. Herschlag, and P. O. Brown, "Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System," *PLoS Biol.*, vol. 6, no. 10, p. e255, Oct. 2008.
- [134] C. Lee, P. Occhipinti, and A. S. Gladfelter, "PolyQ-dependent RNA–protein assemblies control symmetry breaking," *J. Cell Biol.*, vol. 208, no. 5, pp. 533–544, Mar. 2015.
- [135] D. F. Porter, Y. Y. Koh, B. VanVeller, R. T. Raines, and M. Wickens, "Target selection by natural and redesigned PUF proteins," *Proc. Natl. Acad. Sci.*, vol. 112, no. 52, pp. 15868–15873, Dec. 2015.
- [136] A.-C. Gavin *et al.*, "Functional organization of the yeast proteome by systematic analysis of protein complexes," *Nature*, vol. 415, no. 6868, p. 141, 2002.
- [137] K. Tarassov *et al.*, "An in vivo map of the yeast protein interactome," *Science*, vol. 320, no. 5882, pp. 1465–1470, 2008.
- [138] N. Colomina, F. Ferrezuelo, H. Wang, M. Aldea, and E. Garí, "Whi3, a Developmental Regulator of Budding Yeast, Binds a Large Set of mRNAs Functionally Related to the Endoplasmic Reticulum," *J. Biol. Chem.*, vol. 283, no. 42, pp. 28670–28679, Oct. 2008.
- [139] S. Sharifpoor *et al.*, "Functional wiring of the yeast kinome revealed by global analysis of genetic network motifs," *Genome Res.*, vol. 22, no. 4, pp. 791–801, Apr. 2012.
- [140] K. J. Blumer, J. E. Reneke, and J. Thorner, "The STE2 gene product is the ligand-binding component of the alpha-factor receptor of *Saccharomyces cerevisiae*," *J. Biol. Chem.*, vol. 263, no. 22, pp. 10836–10842, 1988.
- [141] D. C. Hagen, G. McCaffrey, and G. F. Sprague, "Evidence the yeast STE3 gene encodes a receptor for the peptide pheromone a factor: gene sequence and implications for the structure of the presumed receptor," *Proc. Natl. Acad. Sci.*, vol. 83, no. 5, pp. 1418–1422, 1986.
- [142] M. Whiteway *et al.*, "The STE4 and STE18 genes of yeast encode potential β and γ subunits of the mating factor receptor-coupled G protein," *Cell*, vol. 56, no. 3, pp. 467–477, 1989.
- [143] J. E. Hirschman, G. S. De Zutter, W. F. Simonds, and D. D. Jenness, "The G $\beta\gamma$ Complex of the Yeast Pheromone Response Pathway SUBCELLULAR FRACTIONATION AND PROTEIN-PROTEIN INTERACTIONS," *J. Biol. Chem.*, vol. 272, no. 1, pp. 240–248, 1997.

- [144] S. Klein, H. Reuveni, and A. Levitzki, "Signal transduction by a nondissociable heterotrimeric yeast G protein," *Proc. Natl. Acad. Sci.*, vol. 97, no. 7, pp. 3219–3223, 2000.
- [145] T. Leeuw, C. Wu, J. D. Schrag, M. Whiteway, D. Y. Thomas, and E. Leberer, "Interaction of a G-protein β -subunit with a conserved sequence in Ste20/PAK family protein kinases," *Nature*, vol. 391, no. 6663, p. 191, 1998.
- [146] Y. Feng, L. Y. Song, E. Kincaid, S. K. Mahanty, and E. A. Elion, "Functional binding between G β and the LIM domain of Ste5 is required to activate the MEKK Ste11," *Curr. Biol.*, vol. 8, no. 5, pp. 267–282, 1998.
- [147] K.-Y. Chol, B. Satterberg, D. M. Lyons, and E. A. Elion, "Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in *S. cerevisiae*," *Cell*, vol. 78, no. 3, pp. 499–512, 1994.
- [148] R. E. Lamson, M. J. Winters, and P. M. Pryciak, "Cdc42 Regulation of Kinase Activity and Signaling by the Yeast p21-Activated Kinase Ste20," *Mol. Cell. Biol.*, vol. 22, no. 9, pp. 2939–2951, May 2002.
- [149] M. Peter, A. M. Neiman, H. O. Park, M. Van Lohuizen, and I. Herskowitz, "Functional analysis of the interaction between the small GTP binding protein Cdc42 and the Ste20 protein kinase in yeast," *EMBO J.*, vol. 15, no. 24, pp. 7046–7059, 1996.
- [150] Y. Wang, W. Chen, D. M. Simpson, and E. A. Elion, "Cdc24 Regulates Nuclear Shuttling and Recruitment of the Ste5 Scaffold to a Heterotrimeric G Protein in *Saccharomyces cerevisiae*," *J. Biol. Chem.*, vol. 280, no. 13, pp. 13084–13096, Apr. 2005.
- [151] A.-C. Butty, P. M. Pryciak, L. S. Huang, I. Herskowitz, and M. Peter, "The role of Far1p in linking the heterotrimeric G protein to polarity establishment proteins during yeast mating," *Science*, vol. 282, no. 5393, pp. 1511–1516, 1998.
- [152] M. Qi, "Formin-induced actin cables are required for polarized recruitment of the Ste5 scaffold and high level activation of MAPK Fus3," *J. Cell Sci.*, vol. 118, no. 13, pp. 2837–2848, Jul. 2005.
- [153] D. M. Lyons, S. K. Mahanty, K.-Y. Choi, M. Manandhar, and E. A. Elion, "The SH3-domain protein Bem1 coordinates mitogen-activated protein kinase cascade activation with cell cycle control in *Saccharomyces cerevisiae*," *Mol. Cell. Biol.*, vol. 16, no. 8, pp. 4095–4106, 1996.
- [154] P. Wiget, Y. Shimada, A.-C. Butty, E. Bi, and M. Peter, "Site-specific regulation of the GEF Cdc24p by the scaffold protein Far1p during yeast mating," *EMBO J.*, vol. 23, no. 5, pp. 1063–1074, 2004.
- [155] H. G. Dohlman and J. Thorner, "Regulation of G protein-initiated signal transduction in yeast: paradigms and principles," *Annu. Rev. Biochem.*, vol. 70, no. 1, pp. 703–754, 2001.
- [156] E. A. Elion, "The Ste5p scaffold," *J. Cell Sci.*, vol. 114, no. 22, pp. 3967–3978, 2001.
- [157] M. Good, G. Tang, J. Singleton, A. Reményi, and W. A. Lim, "The Ste5 Scaffold Directs Mating Signaling by Catalytically Unlocking the Fus3 MAP Kinase for Activation," *Cell*, vol. 136, no. 6, pp. 1085–1097, Mar. 2009.
- [158] F. van Drogen, V. M. Stucke, G. Jorritsma, and M. Peter, "MAP kinase dynamics in response to pheromones in budding yeast," *Nat. Cell Biol.*, vol. 3, no. 12, p. 1051, 2001.
- [159] D. Matheos, M. Metodiev, E. Muller, D. Stone, and M. D. Rose, "Pheromone-induced polarization is dependent on the Fus3p MAPK acting through the formin Bni1p," *J. Cell Biol.*, vol. 165, no. 1, pp. 99–109, Apr. 2004.

- [160] H. B. Kim, "Cellular morphogenesis in the *Saccharomyces cerevisiae* cell cycle: localization of the CDC3 gene product and the timing of events at the budding site," *J. Cell Biol.*, vol. 112, no. 4, pp. 535–544, Feb. 1991.
- [161] E. G. Flescher, "Components required for cytokinesis are important for bud site selection in yeast," *J. Cell Biol.*, vol. 122, no. 2, pp. 373–386, Jul. 1993.
- [162] D. Freifelder, "Bud position in *Saccharomyces cerevisiae*," *J. Bacteriol.*, vol. 80, no. 4, p. 567, 1960.
- [163] A. Fujita *et al.*, "A yeast gene necessary for bud-site selection encodes a protein similar to insulin-degrading enzymes," *Nature*, vol. 372, no. 6506, p. 567, 1994.
- [164] P. J. Kang, E. Angerman, C.-H. Jung, and H.-O. Park, "Bud4 mediates the cell-type-specific assembly of the axial landmark in budding yeast," *J. Cell Sci.*, vol. 125, no. 16, pp. 3840–3849, Aug. 2012.
- [165] P. J. Kang, J. K. Hood-DeGrenier, and H.-O. Park, "Coupling of septins to the axial landmark by Bud4 in budding yeast," *J. Cell Sci.*, vol. 126, no. 5, pp. 1218–1226, Mar. 2013.
- [166] P. J. Kang, M. E. Lee, and H.-O. Park, "Bud3 activates Cdc42 to establish a proper growth site in budding yeast," *J. Cell Biol.*, vol. 206, no. 1, pp. 19–28, Jul. 2014.
- [167] A. Bender, "Genetic evidence for the roles of the bud-site-selection genes BUD5 and BUD2 in control of the Rsr1p (Bud1p) GTPase in yeast," *Proc. Natl. Acad. Sci.*, vol. 90, no. 21, pp. 9926–9929, 1993.
- [168] C. Camus, E. Boy-Marcotte, and M. Jacquet, "Two subclasses of guanine exchange factor (GEF) domains revealed by comparison of activities of chimeric genes constructed from CDC25, SDC25 and BUD5 in *Saccharomyces cerevisiae*," *Mol. Gen. Genet. MGG*, vol. 245, no. 2, pp. 167–176, 1994.
- [169] A. L. Marston, T. Chen, M. C. Yang, P. Belhumeur, and J. Chant, "A localized GTPase exchange factor, Bud5, determines the orientation of division axes in yeast," *Curr. Biol.*, vol. 11, no. 10, pp. 803–807, 2001.
- [170] P. J. Kang, L. Béven, S. Hariharan, and H.-O. Park, "The Rsr1/Bud1 GTPase interacts with itself and the Cdc42 GTPase during bud-site selection and polarity establishment in budding yeast," *Mol. Biol. Cell*, vol. 21, no. 17, pp. 3007–3016, 2010.
- [171] Y. Zheng, A. Bender, and R. A. Cerione, "Interactions among proteins involved in bud-site selection and bud-site assembly in *Saccharomyces cerevisiae*," *J. Biol. Chem.*, vol. 270, no. 2, pp. 626–630, 1995.
- [172] M. E. Lee, W.-C. Lo, K. E. Miller, C.-S. Chou, and H.-O. Park, "Regulation of Cdc42 polarization by the Rsr1 GTPase and Rga1, a Cdc42 GTPase-activating protein, in budding yeast," *J. Cell Sci.*, vol. 128, no. 11, pp. 2106–2117, Jun. 2015.
- [173] A. Fujita *et al.*, "Rax1, a protein required for the establishment of the bipolar budding pattern in yeast," *Gene*, vol. 327, no. 2, pp. 161–169, Mar. 2004.
- [174] T. Chen *et al.*, "Multigenerational cortical inheritance of the Rax2 protein in orienting polarity and division in yeast," *Science*, vol. 290, no. 5498, pp. 1975–1978, 2000.
- [175] E. Bi and H.-O. Park, "Cell Polarization and Cytokinesis in Budding Yeast," *Genetics*, vol. 191, no. 2, pp. 347–387, Jun. 2012.
- [176] P. J. Kang, E. Angerman, K. Nakashima, J. R. Pringle, and H.-O. Park, "Interactions among Rax1p, Rax2p, Bud8p, and Bud9p in marking cortical sites for bipolar bud-site selection in yeast," *Mol. Biol. Cell*, vol. 15, no. 11, pp. 5145–5157, 2004.

- [177] A.-B. Krappmann, N. Taheri, M. Heinrich, and H.-U. Mösch, "Distinct domains of yeast cortical tag proteins Bud8p and Bud9p confer polar localization and functionality," *Mol. Biol. Cell*, vol. 18, no. 9, pp. 3323–3339, 2007.
- [178] J. E. Irazoqui, A. S. Gladfelter, and D. J. Lew, "Scaffold-mediated symmetry breaking by Cdc42p," *Nat. Cell Biol.*, vol. 5, no. 12, pp. 1062–1070, Dec. 2003.
- [179] B. Woods, H. Lai, C.-F. Wu, T. R. Zyla, N. S. Savage, and D. J. Lew, "Parallel Actin-Independent Recycling Pathways Polarize Cdc42 in Budding Yeast," *Curr. Biol.*, vol. 26, no. 16, pp. 2114–2126, Aug. 2016.
- [180] J. C. Harrison, E. S. Bardes, Y. Ohya, and D. J. Lew, "A role for the Pkc1p/Mpk1p kinase cascade in the morphogenesis checkpoint," *Nat. Cell Biol.*, vol. 3, no. 4, p. 417, 2001.
- [181] J. E. Irazoqui, A. S. Howell, C. L. Theesfeld, and D. J. Lew, "Opposing roles for actin in Cdc42p polarization," *Mol. Biol. Cell*, vol. 16, no. 3, pp. 1296–1304, 2005.
- [182] I. T. Kleijn, L. H. J. Krah, and R. Hermsen, "Noise propagation in an integrated model of bacterial gene expression and growth," *PLOS Comput. Biol.*, vol. 14, no. 10, p. e1006386, Oct. 2018.
- [183] J. Liu, H. Martin-Yken, F. Bigey, S. Dequin, J.-M. François, and J.-P. Capp, "Natural Yeast Promoter Variants Reveal Epistasis in the Generation of Transcriptional-Mediated Noise and Its Potential Benefit in Stressful Conditions," *Genome Biol. Evol.*, vol. 7, no. 4, pp. 969–984, Apr. 2015.
- [184] K. Witte, D. Strickland, and M. Glotzer, "Cell cycle entry triggers a switch between two modes of Cdc42 activation during yeast polarization," *Elife*, vol. 6, 2017.
- [185] F. O. Bendezú, V. Vincenzetti, D. Vavylonis, R. Wyss, H. Vogel, and S. G. Martin, "Spontaneous Cdc42 Polarization Independent of GDI-Mediated Extraction and Actin-Based Trafficking," *PLOS Biol.*, vol. 13, no. 4, p. e1002097, Apr. 2015.
- [186] K. Stengel and Y. Zheng, "Cdc42 in oncogenic transformation, invasion, and tumorigenesis," *Cell. Signal.*, vol. 23, no. 9, pp. 1415–1423, Sep. 2011.
- [187] L. E. Arias-Romero and J. Chernoff, "Targeting Cdc42 in cancer," *Expert Opin. Ther. Targets*, vol. 17, no. 11, pp. 1263–1273, Nov. 2013.
- [188] A. Khmelinskii *et al.*, "Incomplete proteasomal degradation of green fluorescent proteins in the context of tandem fluorescent protein timers," *Mol. Biol. Cell*, vol. 27, no. 2, pp. 360–370, 2016.
- [189] N. Friedman, L. Cai, and X. S. Xie, "Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression," *Phys. Rev. Lett.*, vol. 97, no. 16, Oct. 2006.
- [190] O. Bodnar and C. Elster, "On the adjustment of inconsistent data using the Birge ratio," *Metrologia*, vol. 51, no. 5, pp. 516–521, Oct. 2014.
- [191] J. K. Kruschke, "Bayesian estimation supersedes the t test.," *J. Exp. Psychol. Gen.*, vol. 142, no. 2, p. 573, 2013.
- [192] R. E. Kass and A. E. Raftery, "Bayes Factors," *J. Am. Stat. Assoc.*, vol. 90, no. 430, pp. 773–795, Jun. 1995.
- [193] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," 1970.
- [194] R. Wedlich-Soldner, "Robust cell polarity is a dynamic state established by coupling transport and GTPase signaling," *J. Cell Biol.*, vol. 166, no. 6, pp. 889–900, Sep. 2004.
- [195] S. Carson, H. B. Miller, M. C. Srougi, and D. S. Witherow, *Molecular biology techniques: a classroom laboratory manual*. Academic Press, 2019.

- [196] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nat. Methods*, vol. 6, no. 5, pp. 343–345, May 2009.
- [197] R. D. Gietz and R. H. Schiestl, "High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method," *Nat. Protoc.*, vol. 2, no. 1, pp. 31–34, Jan. 2007.
- [198] V. Lundblad and K. Struhl, "Yeast," in *Current Protocols in Molecular Biology*, F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl, Eds. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2008.
- [199] R. Escalante-Chong *et al.*, "Galactose metabolic genes in yeast respond to a ratio of galactose and glucose," *Proc. Natl. Acad. Sci.*, vol. 112, no. 5, pp. 1636–1641, Feb. 2015.
- [200] C. Heij, P. de Boer, P. H. Franses, T. Kloek, and H. K. van Dijk, *Econometric methods with applications in business and economics*. Oxford University Press, 2004.
- [201] S. M. Stigler, "Studies in the History of Probability and Statistics. XXXII: Laplace, Fisher, and the discovery of the concept of sufficiency," *Biometrika*, vol. 60, no. 3, pp. 439–445, 1973.
- [202] J. Mandel, *The statistical analysis of experimental data*. Courier Corporation, 2012.
- [203] D. M. Roden, "Personalized medicine and the genotype–phenotype dilemma," *J. Interv. Card. Electrophysiol.*, vol. 31, no. 1, pp. 17–23, Jun. 2011.
- [204] G. Zhang and D. W. Nebert, "Personalized medicine: Genetic risk prediction of drug response," *Pharmacol. Ther.*, vol. 175, pp. 75–90, Jul. 2017.
- [205] S. Kosuri *et al.*, "Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*," *Proc. Natl. Acad. Sci.*, vol. 110, no. 34, pp. 14024–14029, Aug. 2013.
- [206] M. Kayser and P. M. Schneider, "DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations," *Forensic Sci. Int. Genet.*, vol. 3, no. 3, pp. 154–161, Jun. 2009.
- [207] T. A. Manolio *et al.*, "Finding the missing heritability of complex diseases," *Nature*, vol. 461, no. 7265, pp. 747–753, Oct. 2009.
- [208] A. J. Marian, "Molecular genetic studies of complex phenotypes," *Transl. Res.*, vol. 159, no. 2, pp. 64–79, Feb. 2012.
- [209] O. Zuk, E. Hechter, S. R. Sunyaev, and E. S. Lander, "The mystery of missing heritability: Genetic interactions create phantom heritability," *Proc. Natl. Acad. Sci.*, vol. 109, no. 4, pp. 1193–1198, 2012.
- [210] M. D. Ritchie, E. R. Holzinger, R. Li, S. A. Pendergrass, and D. Kim, "Methods of integrating data to uncover genotype–phenotype interactions," *Nat. Rev. Genet.*, vol. 16, no. 2, pp. 85–97, Feb. 2015.
- [211] F. Corson and E. D. Siggia, "Geometry, epistasis, and developmental patterning," *Proc. Natl. Acad. Sci.*, vol. 109, no. 15, pp. 5568–5575, 2012.
- [212] H. M. Salis, E. A. Mirsky, and C. A. Voigt, "Automated design of synthetic ribosome binding sites to control protein expression," *Nat. Biotechnol.*, vol. 27, no. 10, pp. 946–950, Oct. 2009.
- [213] E. M. Ramos *et al.*, "Phenotype–Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources," *Eur. J. Hum. Genet.*, vol. 22, no. 1, p. 144, 2014.
- [214] C. J. Mungall *et al.*, "The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D712–D722, Jan. 2017.

- [215] T. D. Cuypers and P. Hogeweg, "Virtual Genomes in Flux: An Interplay of Neutrality and Adaptability Explains Genome Expansion and Streamlining," *Genome Biol. Evol.*, vol. 4, no. 3, pp. 212–229, Mar. 2012.
- [216] A. Nakauma and G. S. van Doorn, "Reconstructing the genotype-to-fitness map for the bacterial chemotaxis network and its emergent behavioural phenotypes," *J. Theor. Biol.*, vol. 420, pp. 200–212, May 2017.
- [217] A. Blanco-Gómez *et al.*, "Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes," *BioEssays*, vol. 38, no. 7, pp. 664–673, Jul. 2016.
- [218] B. John and K. R. Lewis, "Chromosome Variability and Geographic Distribution in Insects," *Science*, vol. 152, no. 3723, p. 711, May 1966.
- [219] M. K. Yu *et al.*, "Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems," *Cell Syst.*, vol. 2, no. 2, pp. 77–88, Feb. 2016.
- [220] I. I. Gottesman and T. Gould, "The endophenotype concept in psychiatry," *Res. Adv. Genet. Genomics Implic. Psychiatry*, pp. 63–84, 2005.
- [221] E. H. During, R. S. Osorio, F. M. Elahi, L. Mosconi, and M. J. de Leon, "The concept of FDG-PET endophenotype in Alzheimer's disease," *Neurol. Sci.*, vol. 32, no. 4, pp. 559–569, Aug. 2011.
- [222] S. Schnell and T. E. Turner, "Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws," *Prog. Biophys. Mol. Biol.*, vol. 85, no. 2–3, pp. 235–260, Jun. 2004.
- [223] F. Brauns, J. Halatek, and E. Frey, "Phase-space geometry of reaction–diffusion dynamics," *ArXiv Prepr. ArXiv181208684*, 2018.
- [224] J. Halatek and E. Frey, "Rethinking pattern formation in reaction–diffusion systems," *Nat. Phys.*, Feb. 2018.
- [225] J. Halatek, F. Brauns, and E. Frey, "Self-organization principles of intracellular pattern formation," *Phil Trans R Soc B*, vol. 373, no. 1747, p. 20170107, 2018.
- [226] J. J. Turner, J. C. Ewald, and J. M. Skotheim, "Cell Size Control in Yeast," *Curr. Biol.*, vol. 22, no. 9, pp. R350–R359, May 2012.
- [227] A. I. Goranov *et al.*, "The rate of cell growth is governed by cell cycle stage," *Genes Dev.*, vol. 23, no. 12, pp. 1408–1422, Jun. 2009.
- [228] F. Ferrezuelo *et al.*, "The critical size is set at a single-cell level by growth rate to attain homeostasis and adaptation," *Nat. Commun.*, vol. 3, p. 1012, Aug. 2012.
- [229] Y. T. Chong *et al.*, "Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis," *Cell*, vol. 161, no. 6, pp. 1413–1424, Jun. 2015.
- [230] A. S. Howell, M. Jin, C.-F. Wu, T. R. Zyla, T. C. Elston, and D. J. Lew, "Negative Feedback Enhances Robustness in the Yeast Polarity Establishment Circuit," *Cell*, vol. 149, no. 2, pp. 322–333, Apr. 2012.
- [231] X. Darzacq *et al.*, "In vivo dynamics of RNA polymerase II transcription," *Nat. Struct. Mol. Biol.*, vol. 14, no. 9, pp. 796–806, Sep. 2007.
- [232] E. Gur and R. T. Sauer, "Degrons in protein substrates program the speed and operating efficiency of the AAA+ Lon proteolytic machine," *Proc. Natl. Acad. Sci.*, vol. 106, no. 44, pp. 18503–18508, 2009.
- [233] A. Raj and A. van Oudenaarden, "Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences," *Cell*, vol. 135, no. 2, pp. 216–226, Oct. 2008.

- [234] N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, and M. Mann, "Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells," *Nat. Methods*, vol. 11, no. 3, pp. 319–324, Mar. 2014.
- [235] R. Christiano, N. Nagaraj, F. Fröhlich, and T. C. Walther, "Global Proteome Turnover Analyses of the Yeasts *S. cerevisiae* and *S. pombe*," *Cell Rep.*, vol. 9, no. 5, pp. 1959–1965, Dec. 2014.
- [236] J. Grigull, S. Mnaimneh, J. Pootoolal, M. D. Robinson, and T. R. Hughes, "Genome-Wide Analysis of mRNA Stability Using Transcription Inhibitors and Microarrays Reveals Posttranscriptional Control of Ribosome Biogenesis Factors," *Mol. Cell. Biol.*, vol. 24, no. 12, pp. 5534–5547, Jun. 2004.
- [237] V. Shahrezaei and P. S. Swain, "Analytical distributions for stochastic gene expression," *Proc. Natl. Acad. Sci.*, vol. 105, no. 45, pp. 17256–17261, 2008.
- [238] B. Futcher, G. I. Latter, P. Monardo, C. S. McLaughlin, and J. I. Garrels, "A sampling of the yeast proteome," *Mol. Cell. Biol.*, vol. 19, no. 11, pp. 7357–7368, 1999.
- [239] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *J. Comput. Phys.*, vol. 22, no. 4, pp. 403–434, 1976.
- [240] M. K. Yu *et al.*, "Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems," *Cell Syst.*, vol. 2, no. 2, pp. 77–88, Feb. 2016.
- [241] A. H. Y. Tong *et al.*, "Global mapping of the yeast genetic interaction network," *science*, vol. 303, no. 5659, pp. 808–813, 2004.
- [242] E. N. Smith and L. Kruglyak, "Gene–Environment Interaction in Yeast Gene Expression," *PLoS Biol.*, vol. 6, no. 4, p. e83, Apr. 2008.
- [243] A. T. Layton, N. S. Savage, A. S. Howell, S. Y. Carroll, D. G. Drubin, and D. J. Lew, "Modeling Vesicle Traffic Reveals Unexpected Consequences for Cdc42p-Mediated Polarity Establishment," *Curr. Biol.*, vol. 21, no. 3, pp. 184–194, Feb. 2011.
- [244] M. Johnston and R. W. Davis, "Sequences that regulate the divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*," *Mol. Cell. Biol.*, vol. 4, no. 8, pp. 1440–1448, 1984.
- [245] A. Kolkman *et al.*, "Proteome analysis of yeast response to various nutrient limitations," *Mol. Syst. Biol.*, vol. 2, May 2006.
- [246] B. Lehner, "Molecular mechanisms of epistasis within and between genes," *Trends Genet.*, vol. 27, no. 8, pp. 323–331, Aug. 2011.
- [247] Z. Zhang, W. Qian, and J. Zhang, "Positive selection for elevated gene expression noise in yeast," *Mol. Syst. Biol.*, vol. 5, no. 1, p. 299, Jan. 2009.
- [248] B. Lehner, "Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast," *PloS One*, vol. 5, no. 2, p. e9035, 2010.
- [249] C. H. Waddington, "Canalization of development and the inheritance of acquired characters," *Nature*, vol. 150, no. 3811, p. 563, 1942.
- [250] S. M. Ross, *Introduction to probability models*. Academic press, 2014.
- [251] K. S. Trivedi and A. Bobbio, *Reliability and availability engineering: modeling, analysis, and applications*. Cambridge University Press, 2017.
- [252] J. da Silva, M. Coetzer, R. Nedellec, C. Pastore, and D. E. Mosier, "Fitness Epistasis and Constraints on Adaptation in a Human Immunodeficiency Virus Type 1 Protein Region," *Genetics*, vol. 185, no. 1, pp. 293–303, May 2010.
- [253] S. Basu and A. DasGupta, "The mean, median, and mode of unimodal distributions: a characterization," *Theory Probab. Its Appl.*, vol. 41, no. 2, pp. 210–223, 1997.

- [254] Z. Wang and J. Zhang, "Impact of gene expression noise on organismal fitness and the efficacy of natural selection," *Proc. Natl. Acad. Sci.*, vol. 108, no. 16, pp. E67–E76, Apr. 2011.
- [255] H. B. Fraser, A. E. Hirsh, G. Giaever, J. Kumm, and M. B. Eisen, "Noise Minimization in Eukaryotic Gene Expression," *PLoS Biol.*, vol. 2, no. 6, p. e137, Apr. 2004.
- [256] B. A. Bassett, "The Error Function and The Kink Soliton," *ArXiv Prepr. Cond-Mat9604120*, 1996.
- [257] Y. Kamada, T. Sekito, and Y. Ohsumi, "Autophagy in yeast: ATOR-mediated response to nutrient starvation," in *TOR*, Springer, 2004, pp. 73–84.
- [258] A. Eyre-Walker and P. D. Keightley, "The distribution of fitness effects of new mutations," *Nat. Rev. Genet.*, vol. 8, no. 8, pp. 610–618, Aug. 2007.
- [259] S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, and G. Sherlock, "Quantitative evolutionary dynamics using high-resolution lineage tracking," *Nature*, vol. 519, no. 7542, pp. 181–186, Mar. 2015.
- [260] S. Kryazhimskiy, D. P. Rice, E. R. Jerison, and M. M. Desai, "Global epistasis makes adaptation predictable despite sequence-level stochasticity," *Science*, vol. 344, no. 6191, pp. 1519–1522, 2014.
- [261] S. Venkataram *et al.*, "Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast," *Cell*, vol. 166, no. 6, p. 1585–1596.e22, Sep. 2016.
- [262] J. Uličný, "Lorenz-Mie light scattering in cellular biology," *Gen Physiol Biophys*, vol. 11, pp. 133–151, 1992.
- [263] A. L. Koch, B. R. Robertson, and D. K. Button, "Deduction of the cell volume and mass from forward scatter intensity of bacteria analyzed by flow cytometry," *J. Microbiol. Methods*, vol. 27, no. 1, pp. 49–61, 1996.
- [264] F. Baganz, A. Hayes, D. Marren, D. C. Gardner, and S. G. Oliver, "Suitability of replacement markers for functional analysis studies in *Saccharomyces cerevisiae*," *Yeast*, vol. 13, no. 16, pp. 1563–1573, 1997.
- [265] V. K. Mutalik *et al.*, "Precise and reliable gene expression via standard transcription and translation initiation elements," *Nat. Methods*, vol. 10, no. 4, pp. 354–360, Apr. 2013.
- [266] X. Li *et al.*, "Generation of destabilized green fluorescent protein as a transcription reporter," *J. Biol. Chem.*, vol. 273, no. 52, pp. 34970–34975, 1998.
- [267] E. L. Snapp, "Fluorescent proteins: a cell biologist's user guide," *Trends Cell Biol.*, vol. 19, no. 11, pp. 649–655, Nov. 2009.
- [268] L. Keren *et al.*, "Noise in gene expression is coupled to growth rate," *Genome Res.*, vol. 25, no. 12, pp. 1893–1902, Dec. 2015.
- [269] T. Philippi and J. Seger, "Hedging one's evolutionary bets, revisited," *Trends Ecol. Evol.*, vol. 4, no. 2, pp. 41–44, 1989.
- [270] J.-W. Veening, E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen, "Bet-hedging and epigenetic inheritance in bacterial cell development," *Proc. Natl. Acad. Sci.*, vol. 105, no. 11, pp. 4393–4398, 2008.
- [271] S. L. Rutherford and S. Lindquist, "Hsp90 as a capacitor for morphological evolution," *Nature*, vol. 396, no. 6709, pp. 336–342, 1998.
- [272] C. Queitsch, T. A. Sangster, and S. Lindquist, "Hsp90 as a capacitor of phenotypic variation," *Nature*, vol. 417, no. 6889, pp. 618–624, 2002.

- [273] R. Halfmann, D. F. Jarosz, S. K. Jones, A. Chang, A. K. Lancaster, and S. Lindquist, "Prions are a common mechanism for phenotypic inheritance in wild yeasts," *Nature*, vol. 482, no. 7385, pp. 363–368, 2012.
- [274] A. Wünsche, D. M. Dinh, R. S. Satterwhite, C. D. Arenas, D. M. Stoebel, and T. F. Cooper, "Diminishing-returns epistasis decreases adaptability along an evolutionary trajectory," *Nat. Ecol. Evol.*, vol. 1, no. 4, pp. 1–6, 2017.
- [275] H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, and C. J. Marx, "Diminishing returns epistasis among beneficial mutations decelerates adaptation," *Science*, vol. 332, no. 6034, pp. 1190–1192, 2011.
- [276] X. Wei and J. Zhang, "Patterns and Mechanisms of Diminishing Returns from Beneficial Mutations," *Mol. Biol. Evol.*, vol. 36, no. 5, pp. 1008–1021, May 2019.
- [277] N. Tokuriki, C. J. Jackson, L. Afriat-Jurnou, K. T. Wyganowski, R. Tang, and D. S. Tawfik, "Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme," *Nat. Commun.*, vol. 3, no. 1, Jan. 2012.
- [278] H. Prinz, "Hill coefficients, dose–response curves and allosteric mechanisms," *J. Chem. Biol.*, vol. 3, no. 1, pp. 37–44, Mar. 2010.
- [279] M. I. Stefan and N. Le Novère, "Cooperative Binding," *PLoS Comput. Biol.*, vol. 9, no. 6, p. e1003106, Jun. 2013.
- [280] S. Covo and A. Elalouf, "A novel single-gamma approximation to the sum of independent gamma variables, and a generalization to infinitely divisible distributions," *Electron. J. Stat.*, vol. 8, no. 1, pp. 894–926, 2014.
- [281] B. Schwanhäusser *et al.*, "Global quantification of mammalian gene expression control," *Nature*, vol. 473, no. 7347, pp. 337–342, May 2011.
- [282] S. B. Cambridge, F. Gnad, C. Nguyen, J. L. Bermejo, M. Krüger, and M. Mann, "Systems-wide Proteomic Analysis in Mammalian Cells Reveals Conserved, Functional Protein Turnover," *J. Proteome Res.*, vol. 10, no. 12, pp. 5275–5284, Dec. 2011.
- [283] A. V. Anzalone *et al.*, "Search-and-replace genome editing without double-strand breaks or donor DNA," *Nature*, vol. 576, no. 7785, pp. 149–157, Dec. 2019.
- [284] C.-F. Wu *et al.*, "Role of competition between polarity sites in establishing a unique front," *Elife*, vol. 4, p. e11611, 2015.
- [285] D. I. Johnson, "Cdc42: an essential Rho-type GTPase controlling eukaryotic cell polarity," *Microbiol Mol Biol Rev*, vol. 63, no. 1, pp. 54–105, 1999.
- [286] K. G. Kozminski, A. J. Chen, A. A. Rodal, and D. G. Drubin, "Functions and functional domains of the GTPase Cdc42p," *Mol. Biol. Cell*, vol. 11, no. 1, pp. 339–354, 2000.
- [287] T. Stewart, L. W. G. Strijbosch, H. Moors, and P. van Batenburg, "A simple approximation to the convolution of gamma distributions," 2007.
- [288] M. Kumar, M. S. Mommer, and V. Sourjik, "Mobility of Cytoplasmic, Membrane, and DNA-Binding Proteins in Escherichia coli," *Biophys. J.*, vol. 98, no. 4, pp. 552–559, Feb. 2010.
- [289] M. H. Neumann and O. Hössjer, "On the effect of estimating the error density in nonparametric deconvolution," *J. Nonparametric Stat.*, vol. 7, no. 4, pp. 307–330, Jan. 1997.
- [290] C. A. Coelho and J. T. Mexia, "On the distribution of the product and ratio of independent generalized gamma-ratio random variables," *Sankhyā Indian J. Stat.*, pp. 221–255, 2007.

Appendix

A Material list

Table 11 List of used materials, kits, equipment and corresponding software packages

Item	Company
Media	
Agar (plates)	VWR
CSM (including single and double dropouts)	Formedium
Dextrose	Sigma-Aldrich
Galactose	Sigma-Aldrich
Peptone	VWR
Raffinose pentahydrate	Sigma-Aldrich
Sorbitol	Sigma-Aldrich
Yeast extract	VWR
Yeast nitrogen base	Sigma-Aldrich
Yeast nitrogen base, low fluorescence	Formedium
Water / MilliQ	Merck Millipore
Antibiotics	
Carbenicillin	Formedium
G418 Disuphate	Formedium
Nourseothricin sulfate (clonNAT)	Melford Biolaboratories
Hygromycin B	Formedium
Enzymes	
BamHI	New England Biolabs
DpnI	New England Biolabs
EcoRI-HF	New England Biolabs
PsalI	New England Biolabs
XhoI	New England Biolabs
Zymolyase	Zymo research
PCR related	
Agarose (DNA gels)	Sigma-Aldrich
CutSmart Buffer	New England Biolabs
DNA ladders	Eurogentec & Melford
dNTP mix	Thermo Fisher
GoTaq buffer	Promega
GoTaq polymerase	Promega
Loading Dye	Melford Biolaboratories
Phusion buffer	Thermo Fisher
Phusion polymerase	Thermo Fisher
Primers	IDT
Sequencing	Macrogen
SYBR Safe DNA Gel stain	Thermo Fisher / Invitrogen

TAE	Formedium
Container-related materials	
96 well plates for growth assays (Cat. no. 267427)	Thermo Scientific
96 well plates for microscopy Cat. no. 5241-20)	Zell-Kontakt GmbH
Accuvettes	Coulter
Concanavalin A	Sigma-Aldrich
Cover slips	VWR / Thermo Scientific
Glass slides	VWR / Thermo Scientific
Other reagents	
Coulter Clenz Cleaning Agent	Coulter
Isoton II diluent	Coulter
Lithium Acetate	Sigma-Aldrich
PBS	VWR
Potassium acetate	Sigma-Aldrich
SDS	Biorad
ssDNA (for transformation)	Sigma-Aldrich
TE	VWR
Tris (pH 7.5)	VWR
Triton	Sigma-Aldrich
Kits	
Plasmid Miniprep (PureYield) Kit	Promega
Wizard SV Gel and PCR clean up system	Promega
YeaStar Genomic DNA extraction kit	Zymo research
Hardware and software	
Brightfield light source pE-100	CoolLED
C1000 Touch Thermal Cycler (PCR machine)	Biorad
Camera (Orca Flash v4)	Hamamatsu
FACScan (flow cytometer for GFP strains)	BD
FACSCelesta Flow cytometer (sfGFP strains)	BD
Fluorescence excitation source Spectra X	Lumencor
Gel Doc EZ (Gel imager)	Biorad
Matlab R2014b	Mathworks
Microscope Ti Eclipse (60x Plan Apo objective, x1.5	Nikon
NIS Elements version 4.3-4.5	Nikon
Plate reader	TECAN Infinite 200 pro
Software flow cytometry GFP strains	FlowJo CE
Software flow cytometry sfGFP strains	BD FACSDiva
Sonicator	Qsonica
Spectrophotometer (11 series) for OD and DNA	Denovix
Temperature regulation microscopy	Okolab incubator
Z2 Coulter Counter Analyzer	Beckman Coulter

B Strain and plasmid list

Table 12 List of strains used in this dissertation (continues on next page). Green stands for dominant allele, red for recessive and orange for a diploid mix of dominant and recessive alleles at that locus (white is not determined).

Name	Former	Source	Mating Type	Background	BEM1	BEM2	BEM3	NRP1	SPA2	CDC42	can1
YWKD001											
/ YidB001	YLL3a	Laan Lab collection	α	W303	KanMx		clonNAT	Hygro	GFP URA		1-100
YWKD005	YLL131a	Laan Lab collection	a	W303					GFP URA		MFAprHIS3
YWKD006	YLL132a	Laan Lab collection	a	W303					GFP URA		MFAprHIS3
YWKD007	YLL137a	Laan Lab collection	a	W303					GFP URA		MFAprHIS3
YWKD023		Wedlich-Söldner Lab									
/ YMS003	RWS116	collection	a	W303				Hygro			1-100
YWKD024		Wedlich-Söldner Lab									
/ YMS004	RWS119	collection	a	W303							1-100
YWKD038		Wedlich-Söldner Lab									
/ YMS002	RWS1421	collection	a	W303							1-100
YWKD043	YMP203	M. Peter Lab	a	W303							
YWKD045	YLL112	Laan Lab collection	Diploid	W303	KanMx		clonNAT				MFAprHIS3
YWKD047	YMS006a	Marit Smeets	Diploid	W303	KanMx		clonNAT			Cdc42-URA-Gal1pr-GFP-CDC42	MFAprHIS3
YWKD053d	New	YWKD045	Diploid	W303	KanMx		clonNAT			Cdc42-URA-Gal1pr-CDC42	MFAprHIS3
YWKD054abc	New	YWKD045	Diploid	W303	KanMx		clonNAT			URA-Gal1pr-CDC42	MFAprHIS3
YWKD055c	New	YWKD045	Diploid	W303	KanMx		clonNAT			URA-Gal1pr-sfGFP-CDC42	MFAprHIS3
YWKD061a,b	New	YWKD055c	Diploid	W303	KanMx		clonNAT			URA-Gal1pr-sfGFP-CDC42	MFAprHIS3
YWKD062a,b,c	New	YWKD055c	a	W303							
YWKD065a			a (62a/c),								
/ YidB003	New	YWKD055c	a or α (62b)	W303						URA-Gal1pr-sfGFP-CDC42	MFAprHIS3
YWKD068a,b	New	YWKD055c	a or α	W303	KanMx						?
YWKD069a,b,c	New	YWKD055c	a	W303	KanMx						MFAprHIS3
YWKD070a,b,c	New	YWKD055c	a	W303	KanMx					URA-Gal1pr-sfGFP-CDC42	MFAprHIS3
YWKD071a,b,c	New	YWKD054b	a	W303						URA-Gal1pr-sfGFP-CDC42	MFAprHIS3
YWKD072a,b,c	New	YWKD054b	a	W303						URA-Gal1pr-CDC42	MFAprHIS3
YWKD073a,b,c	New	YWKD054b	a	W303						URA-Gal1pr-CDC42	MFAprHIS3
YWKD073a,b,c	New	YWKD054b	a	W303	KanMx		clonNAT			URA-Gal1pr-sfGFP-CDC42	MFAprHIS3
YLIC008	New	YWKD043xYWKD001	Diploid	W303							1-100
YLIC009a2	New	YLIC008	α	W303							
YLIC1051	New	YLIC009xYWKD005	Diploid	W303	KanMx		clonNAT	Hygro	GFP URA		MFAprHIS3
YFM007a	New	YWKD055c	a	W303						URA-Gal1pr-sfGFP-CDC42	?

Name	LEU2	HIS3	URA3	BUD4	ADE2	TRP1	Other
YWKD001							
/YtdB001	3,112	11,15		S288C			
YWKD005	3,112	11,15		S288C			
YWKD006	3,112	11,15		S288C			
YWKD007	3,112	11,15		S288C			
YWKD023			<i>ura3</i>				
/YMS003		11,15			1	YipLac204-MET-CLN2	<i>cln1Δ::HisG, cln2Δ, cln3Δ::HisG</i>
YWKD024							
/YMS004		11,15	<i>Gal1pr-myc-GFP-CDC42</i>		1	YipLac204-MET-CLN2	<i>cln1Δ::HisG, cln2Δ, cln3Δ::HisG</i>
YWKD038							
/YMS002		11,15	<i>CDC42pr-myc-GFP-CDC42</i>		1	YipLac204-MET-CLN2	<i>cln1Δ::HisG, cln2Δ, cln3Δ::HisG</i>
YWKD043	3,112	11,15			<i>ade2</i>	<i>trp1</i>	<i>cln1Δ::HisG, cln2Δ, cln3Δ::HisG</i>
YWKD045	3,112	11,15		S288C			
YWKD047	3,112	11,15		S288C			
YWKD053d	3,112	11,15		S288C			
YWKD054abc	3,112	11,15		S288C			
YWKD055c	3,112	11,15		S288C			
YWKD061a,b	3,112	11,15		S288C			
YWKD062a,b,c	3,112	11,15		S288C			
YWKD065a							
/YtdB003	3,112	11,15		S288C			
YWKD068a,b	3,112	11,15		S288C			
YWKD069a,b,c	3,112	11,15		S288C			
YWKD070a,b,c	3,112	11,15		S288C			
YWKD071a,b,c	3,112	11,15		S288C			
YWKD072a,b,c	3,112	11,15		S288C			
YWKD073a,b,c	3,112	11,15		S288C			
YLIC008	3,112	11,15			<i>ade2</i>	<i>trp1</i>	<i>CLA4/cln4::LEU2</i>
YLIC009a2	3,112	11,15			<i>ade2</i>	<i>trp1</i>	<i>cln4::LEU2</i>
YLIC1051	3,112	11,15		S288C/W303	<i>ade2</i>	<i>trp1</i>	<i>CLA4/cln4::LEU2</i>
YFM007a	3,112	11,15		S288C			

Table 13 Used plasmids throughout this dissertation, with their origin and most important features.

Name	Former name	Source	Contains	Selectable markers
pWKD001	pRL368	Rong Li Lab	Gal1pr-GFP-Cdc42, URA3	Ampicillin
pWKD006	New	pWKD001	Gal1pr-Cdc42, URA3	Ampicillin
pWKD008	sfGFP pSB1C3	See 2.5	sfGFP	Chloramphenicol
pWKD009	New	pWKD006	Gal1pr-Cdc42, URA3, Post-Cdc42 homology regions	Ampicillin, URA3
pWKD010	New	pWKD006	Gal1pr-Cdc42, URA3, Pre/Post-Cdc42 homology regions	Ampicillin, URA3
pWKD011	New	pWKD010	Gal1pr-sfGFP-Cdc42 (sandwich), URA3, Pre/Post-Cdc42 homology regions	Ampicillin, URA3

C Preparatory experiments with a Cdc42-probe

C.1 Visible, inducible and functional Cdc42-probe

C.1.1 Cdc42 copy number control through the Gal-promoter

In order to proof correct modelling of the details of GAP action, precise control of Cdc42 is required. To control the Cdc42 copy numbers in our experiments, we place the Gal1/10 promoter ahead of our Cdc42-fluorophore construct. It is known to have a large dynamical range (about 1000-fold [244]), making it a well-suitable promoter to toggle between viable and inviable Cdc42 expression levels. This is particularly true since Cdc42 is quite an abundant protein [234], so strong maximum expression is needed. Care must be taken for the choice of medium. Galactose is necessary for modulating the induction for variable expression, but is also a sugar that yeast can metabolize. As is explained further in the supplement of Chapter 2, 2.6, medium is not supplemented with glucose as this modifies expression, but raffinose is used instead [199].

Ideally, we would like to characterize the expression levels of Gal promoter in this context (strain background, media, etc.) relative to the endogenous expression. This is possible by measuring fluorescent intensities in bulk of the strains with the traditional N-term fusions *CDC42pr-GFP-CDC42* and *GAL1pr-GFP-CDC42* from [48]. We assume results reasonably transfer to the expression of the new fluorophore fusion described in the next section, although being situated at a different locus, a different start after the start codon (gene start of *CDC42* instead of *GFP*) and a different 3' UTR.

While at first sight the use of the N-term GFP fusion seems inappropriate, since it was argued in 2.1 that this construct has functionality deficiencies, this will not impede with the expression count. In fact, measuring instead strains with the fluorophore-construct whose construction we describe in the next subsection can fully restore Cdc42 function when measured through fitness, can have a surprising downside. As explained in Chapter 4, the fitness associated with each expression level couples back to the *observed* expression levels through a process called transgenerational feedback. Of course, the improved construct functionality is necessary in all other contexts, particularly those relating Cdc42 levels to fitness as in 2.3.

C.1.2 Fluorophore fusion

The previous section mentioned the problem with previous versions of fluorescent Cdc42 under the Gal-promoter, as in [105]. Additionally, the importance must be noted of the 3' UTR region. It has been shown that for Cdc42, this region is of particular importance for correct localization of mRNAs [87], which is potentially compromised when using exogenous fluorescent Cdc42 versions if the terminator region is not included.

Despite these deficiencies, this N-term fusion is still useful for characterization of the Gal-promoter. To ensure correct mRNA localization, sfGFP is inserted into an endogenous version of Cdc42 of a diploid *S. cerevisiae* (background W303, from [24]), yielding the *GAL1pr-sfGFP-CDC42^{SW}* construct. This leaves the terminator region unchanged. Importantly, the choice of sfGFP should yield a faster maturation time than with mCherry [188], which has been used in a similar sandwich construct in [284], and was found to be only partially functional. Authors in that paper conducted a growth assay on a plate, showing that this fusion as the only source of Cdc42 allowed similar growth to WT at 24 °C, mildly decreased growth at 30 °C, and significantly slower growth at higher temperatures. The goal is to eliminate the 30 °C growth defect with the use of sfGFP.

Furthermore, the transformed diploid is heterozygous for *BEM1* and *BEM3*. This allows generation of the required haploid polarity mutants containing the fluorescent Cdc42 version following from sporulation. Schematically, this insertion (without promoter) is depicted in Figure 58, where it can be seen that sfGFP is quite a heavy addition to Cdc42. While the logic is the same as in [185], the amino acids differ at the insertion location authors used in *S. pombe*. Still, given the location of the domains there is no obvious reason to reconsider the location. Placing the construct between L134 and R135 leaves the most important domains uninterrupted, except arguably a GEF interacting site, which should have other dominant sites.

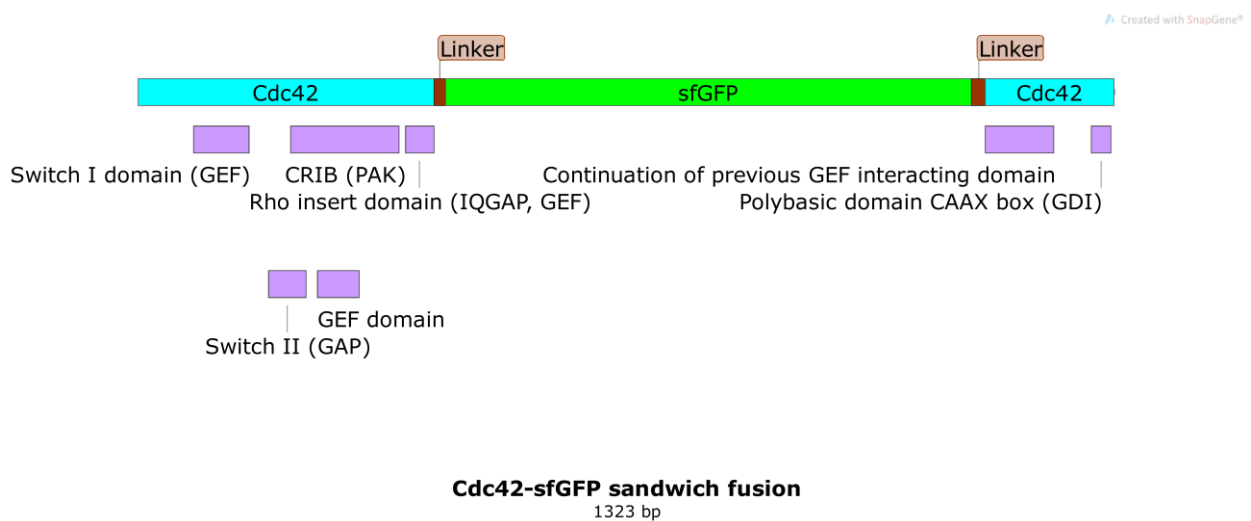


Figure 58 Domain-wise (purple) overview of the Cdc42(blue)-sfGFP(green) sandwich construct made in *S. cerevisiae*. Domain information from [285] and for the switch II region from [286].

The goal is now to validate this probe. Subsequently, a series of tests, each shining light on a different manifestation of fitness effects of Cdc42 concentrations, is used to cross-validate the prediction concerning strong Cdc42 threshold to polarize, dependent on background. Ultimately, these provide supporting evidence for the quantitative assay in Chapter 2 (2.3).

C.2 Probe validation

C.2.1 Fluorescence assays

C.2.1.1 Gal promoter varies in target range

Ideally, the Gal promoter in our constructs is able to vary expression between an order of magnitude below and above endogenous expression of Cdc42. This way strong under- and overexpression can be established using a single construct. To determine the dynamical range relative to endogenous Cdc42 expression, we turn to flow cytometry, which allows measurement of the Cdc42 fluorescent protein distribution across the population.

As previously explained, we acquired the Cdc42 distributions using the poorly functional N-term GFP fusion. The strains containing this fusion (RWS119 and RWS1421) have GFP-Cdc42 under both the *GAL1* and the endogenous promoter respectively [48]. This distribution of GFP-Cdc42 under various expression levels (through different galactose concentrations) can then be normalized to represent actual protein numbers by comparing to the reference distribution of GFP-Cdc42 under the endogenous promoter.

As explained in more detail in 3.2.1.3 and Appendix E, the copy number distributions follow after some data processing. First, gamma distributions are fitted on the data (as expected from [189]). Removal of the background by deconvolution is approximated by analytical means [287]. This yielded Cdc42 copy number distributions under 0.05%, 0.2% and 2% (galactose) inducer levels (see Figure 59). What can be seen is that in practice we can approximately get an order of magnitude below and above endogenous expression. Also, the distribution is quite wide, e.g., at maximum induction part of the population has 3 times the average endogenous expression, and part has up to 12. The notion that expression is very noisy, particularly for Cdc42 has important implications for quantitative predictions on fitness, as done in Chapter 3.

C.2.1.2 Cdc42-probe suitable for abundance, not localization measurements

With the promoter part verified, we can address the fluorescent behavior of the new construct, the Cdc42-sfGFP sandwich fusion. As a qualitatively diagnostic, time lapse movies are made of several strains containing the *GAL1pr-sfGFP-CDC42^{SW}* construct, with different genetic background and ploidy. Sometimes for the *GAL1pr-sfGFP-CDC42^{SW}* strains, strong fluorescent punctae were visible, particularly in aging cells. These differ distinctly from the abnormal internal (somewhat ring-like) structures that can sometimes be observed in the old fusion (see e.g., in [48]). The strong spots accumulated during the life time of a cell, but were not transferred to new daughter cells, which always started 'clean'. These were not related to polarization, and are usually mobile. In extremely affected cells, also stationary spots were present (see Figure 60).

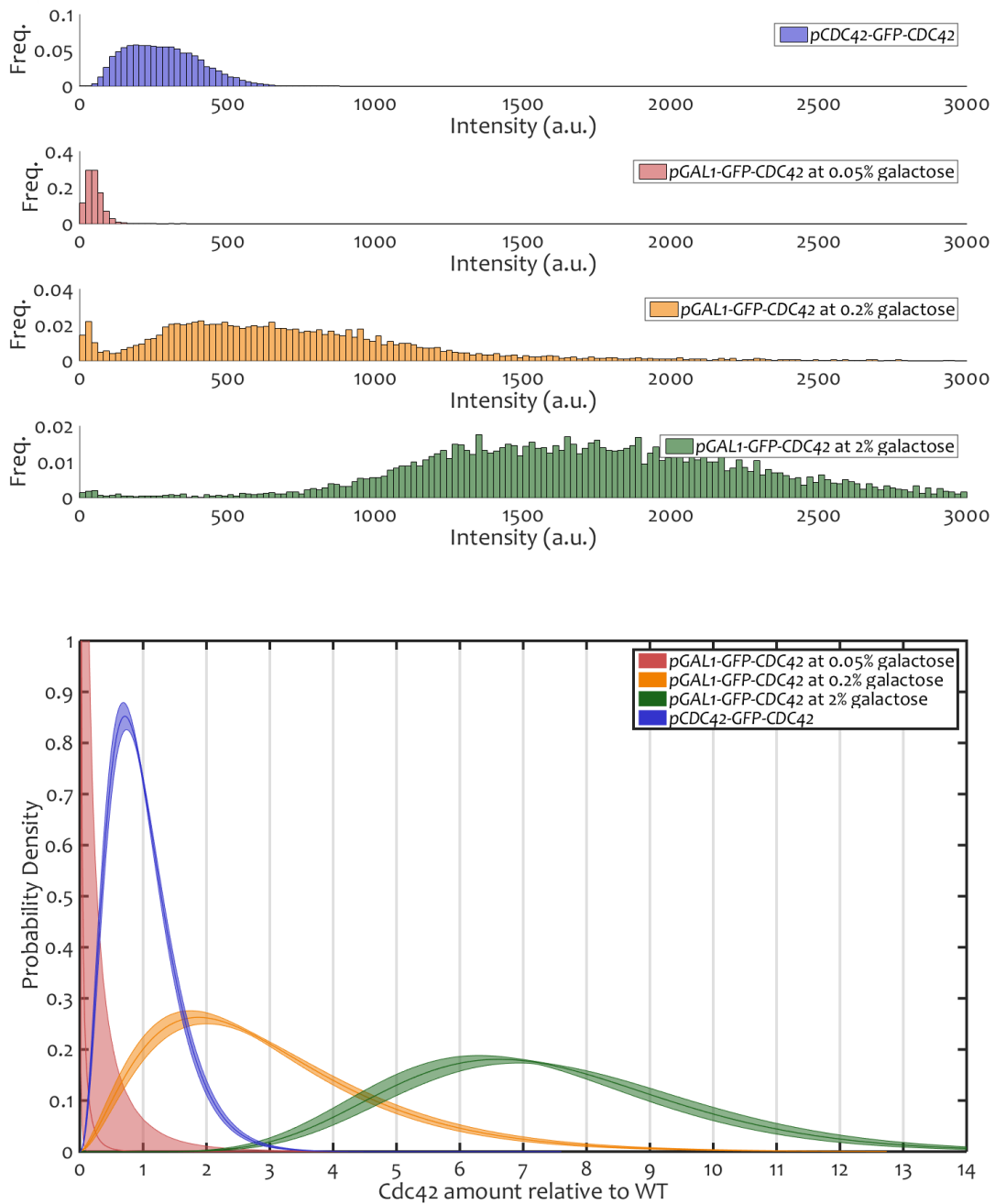


Figure 59 Top, four normalized histograms with gated data from flow cytometry experiments of GAL1pr-GFP-CDC42 (0.05% galactose in red, 0.2% in orange and 2% in green) and CDC42pr-GFP-CDC42 (blue), bottom: expression levels normalized to WT endogenous expression from flow cytometry measurements of GAL1pr-GFP-CDC42 normalized to CDC42pr-GFP-CDC42 with distribution fitting errors at different induction levels. Horizontal axis showing expression is in units of endogenous Cdc42 expression.

Despite heavy presence of these spots, cells were still able to divide properly (see again Figure 60), suggesting that it is not Cdc42 that is particularly strongly concentrated at these points. Otherwise, these would strongly interfere with the polarity cue, whose strength is on the verge of being visualized (the signal-to-noise is narrowly too low). This implies that

fluorescent remnants are not properly degraded, while the *Cdc42* is. This behavior is documented in literature [188], where authors show that sfGFP suffers from incomplete degradation, which influences the perceived protein turnover.

While unlikely to affect the *Cdc42* concentrations in the cell, the artefact impedes good single cell fluorescent measurements of the polarity spot. However, noting that mainly old cells are affected, the influence on the population level will not be very strong. Therefore, we turn to bulk measurements of protein distribution through flow cytometry to gain a better insight in the magnitude of this effect.

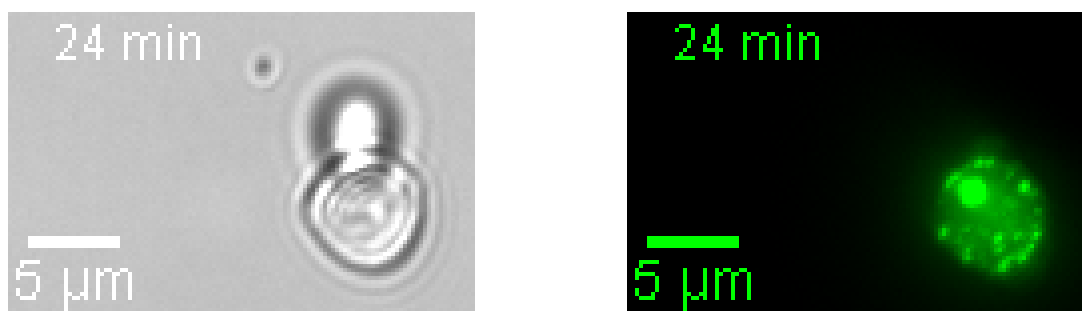


Figure 60 Example of a YWKO065a cell, which has *GAL1pr-sfGFP-CDC42^{SW}* as its sole *Cdc42* source and otherwise a WT polarity background at 0.1% galactose, which has many fluorescent protein foci, as seen in the green fluorescent channel frame on the right. However, the brightfield image reveals this cell is still budding.

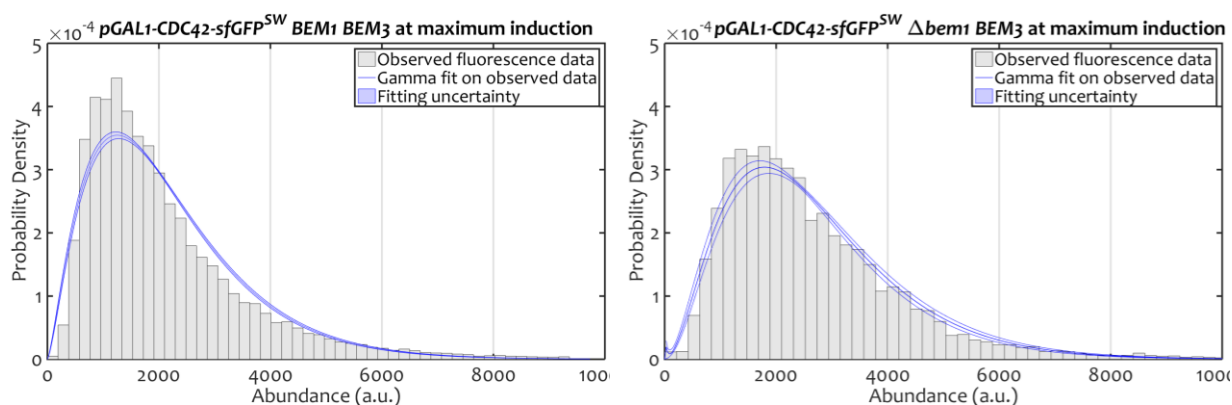


Figure 61 Protein abundance distribution examples (grey with double gamma fit in blue) from flow cytometry of *GAL1pr-sfGFP-CDC42^{SW}* as the only source of *Cdc42* in an otherwise WT (left, YWKO065a) or Δ *bem1* (right, YWKO069a) background at maximal induction (at least 2% galactose) concentration.

Using flow cytometry, we this time measure *GAL1pr-sfGFP-CDC42^{SW}* strains to assess the influence of artefacts. Two examples of fluorescent protein distributions are given in Figure 61. As hoped, the gamma shapes of the distribution are conserved in this construct. Only a mild excess around the peak in observed for WT, which is likely due to the degradation artefacts also exhibited in fluorescence microscopy. While the artefacts are frustrating single cell microscopy measurements for *Cdc42* localization, considering the reasonably small size of

this effect across the whole population, the sandwich fusion construct is still useful for determining Cdc42 content.

C.2.1.3 Cdc42-needy cells have slow starts

After having assessed the minor influence of artefacts across the population, we can expand on the influence that different Cdc42 concentrations have in different backgrounds, and how this manifests itself in Cdc42 distributions using the flow cytometry assay. However, when comparing the strain background at various inducer concentrations, it should be noted that in this flow cytometry (with the sfGFP construct) evaporation was not very well countered. Subsequently, effective galactose concentration could be up to most twice as large as intended. Accounting for this, and taking into account the lack of technical replicates for this assay, not much is gained in terms of relative expression levels at different inducer concentration. Still, we can determine clear differential expression.

One of the distinct differences across media/strain background is the presence of two peaks in the Cdc42 distribution, that relate to an 'on' and 'off' state of the promoter. The setup of the experiment was such that media for cells was refreshed hours before measurement, at which point cells had been in saturation. Typically, cells degrade their proteins [257] and will 'reboot' in G1. We expect this to form a bigger challenge for $\Delta b e m 1$ backgrounds than for WT, considering they need to muster more Cdc42 in order to polarize. Indeed, it was quite apparent that fitting the data with two gamma distributions attributed a sizeable fraction of $\Delta b e m 1$ backgrounds (see Figure 62) in the lower state of production, presumably corresponding to problems with 'rebooting'. The pure $\Delta b e m 1$ was even worse off than the $\Delta b e m 1 \Delta b e m 3$, which is consistent with the notion that the former requires a higher abundance of Cdc42 (see Figure 16). Problems quickly vanished as expression became stronger, showing a first indication of Cdc42 dependence on fitness.

These slow starts let themselves be further felt in the typical strength of the on-state. Figure 63 compares expression across backgrounds. While the former allows a qualitative comparison of the whole distribution, the latter condenses the information to two summary statistics to visualize more quantitatively what the effect of higher induction is: median and width (standard deviation). For convenience, all summary statistics have been background corrected (intensity counts of samples without sfGFP) and a reference point is chosen (WT at maximum expression), to which all values are normalized.

The following trends are then visible: firstly, expression increases with higher induction (progressively higher medians for 0.05%, 0.15%, 0.2% and 2% for all colors/backgrounds), as it should. Secondly, the width of the distribution increases with higher induction (progressively higher distribution widths for 0.05%, 0.15%, 0.2% and 2% for all colors/backgrounds). Finally, both previous trends are more pronounced in the $\Delta b e m 1$ background, particularly when Bem3 is still present (blue). The low expression of Cdc42 seems to stop this background in its tracks early on. However, once expression is suitable also for this background, its expression is

the one that increases the most. Conversely, WT performance appears mostly insensitive to expression, in line with the view that the viable WT expression range is very broad.

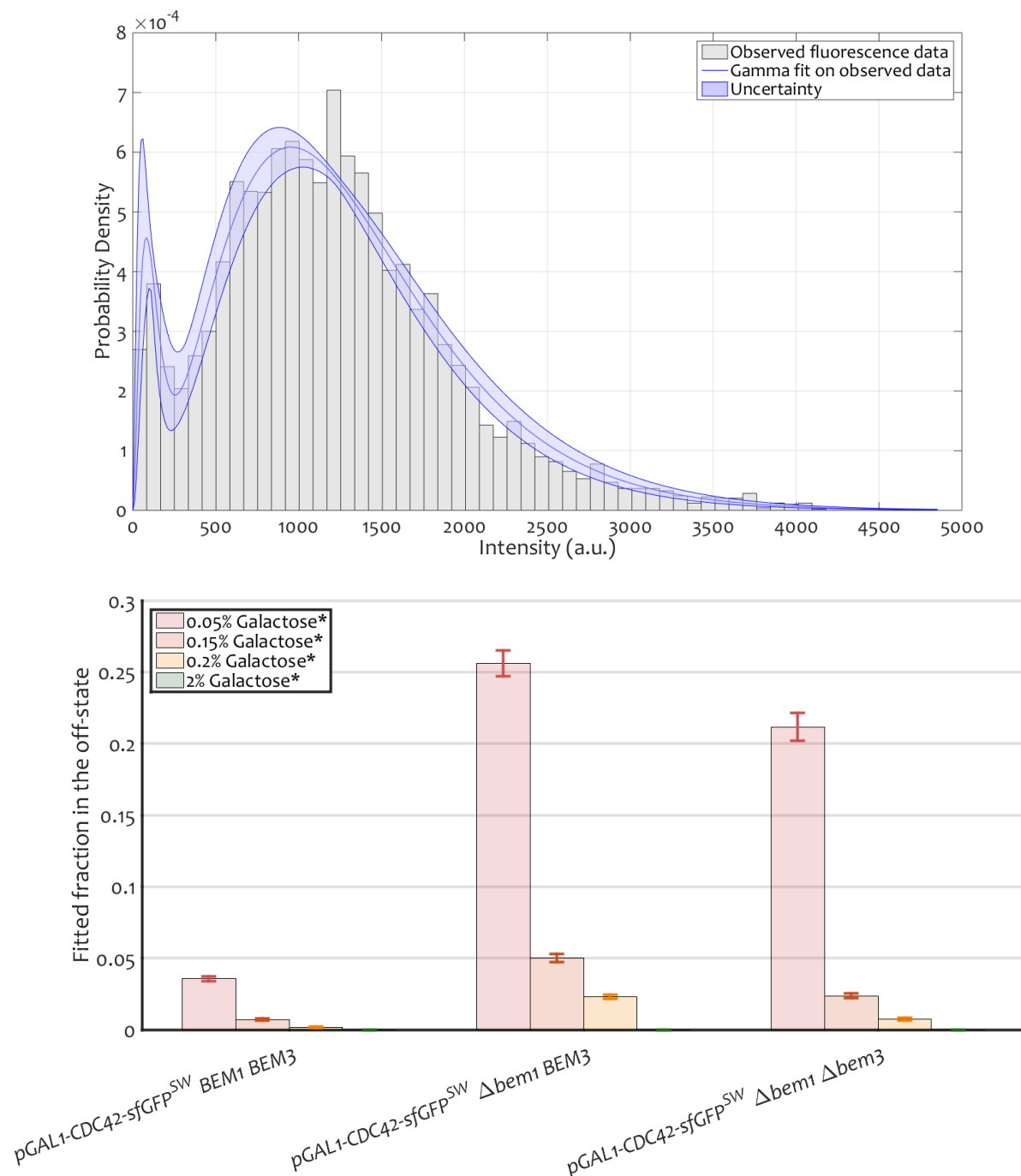


Figure 62 Top: Example of a Δbem1 replicate with *GAL1pr-sfGFP-CDC42^{SW}* intensity histogram from flow cytometry. Two distinct peaks are visible, the 'off' (lower) and 'on' (higher) state, which have been fitted by a mixture of two gamma distributions with the blue patch showing the uncertainty due to the fitting error on the parameters. Bottom: Fraction of the population found in the lower, 'off'-state of protein production when the *sfGFP-CDC42^{SW}* distribution is fitted with two gamma distributions. Colors denote at which induction concentration was measured, which due to evaporation uncertainty is at least 0.05% galactose (red), 0.15% (orange), 0.2% (yellow) and 2% (green). Error bars represent the standard fitting errors on these fractions, for three polarity backgrounds: WT (YWKD065a), Δbem1 (YWKD069a, b, c combined) and $\Delta\text{bem1} \Delta\text{bem3}$ (YWKD070a, b, c combined).

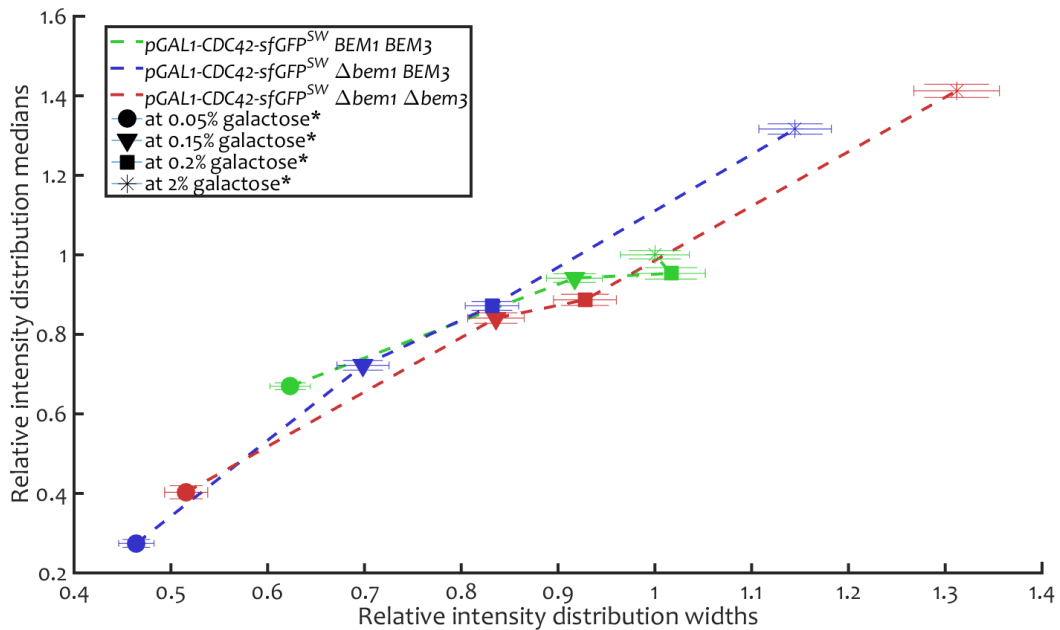


Figure 63 Summary statistics on the *sfGFP* intensity distributions as measured in flow cytometry, corrected for background. All values are relative, i.e., they are normalized to WT values at maximum expression for convenience. The median (vertical axis) is plotted against the standard deviation (horizontal axis) for strains with *GAL1pr-sfGFP-CDC42^{SW}* as the only source of *Cdc42* in an otherwise WT (green line, YWKD065a), Δ *bem1* (blue line, YWKD069a/b/c) and Δ *bem1* Δ *bem3* (red line, YWKD070a) background. Values are the weighted average of mentioned biological replicates that yielded at least 1000 points. Markers denote at which induction concentration was measured, which due to evaporation uncertainty is at least 0.05% galactose (square), 0.15% (triangle), 0.2% (square) and 2% (asterisk). Error bars construct the approximate 95% confidence interval.

C.2.2 Volumetric assay

C.2.2.1 Slow medium great equalizer

After the fluorescence tests, we assessed as a final check the effect of the *GAL1pr-sfGFP-CDC42^{SW}* construct on volumes. Disentangling possible unwanted side-effects from the expected influence that variable expression of *Cdc42* has on volumes is not trivial. Notably, within the Δ *bem1* there is a quite strong documented corollary of fitness on size [24], which we suppress to double-check the correct functionality of the construct.

Therefore, we decide to conduct this experiment in an environment such that fitness differences are minimal, which implies relatively poor medium. Concretely, low amino acid concentrations (and autoclaving for sterilization) impede prolonged growth under maximum speed. In 3.4.1, we model the effect of slow growth conditions, to more quantitatively understand why this equalizes fitness differences so well. Experimentally, a rough growth assay in this medium (shown in 2.3.2 and Appendix C.3) shows that for example the Δ *bem1* is still well alive at 0.04% galactose, despite that endogenous expression is estimated around 0.1% (see Figure 59), which is narrowly enough in optimal growth conditions (YPD). So, when

maximum growth conditions given a media type are not optimal because e.g., amino acids are not abundant, this benefits mutants short of *Cdc42*, allowing us here to focus on the effect of sugar concentration and the sfGFP construct itself.

C.2.2.2 Minor influence sfGFP sandwich on volumes

To condense the amount of information, we reduce the distribution data to two summary statistics; the median and the standard deviation. An example for WT is found in Figure 64. This allows visualization of the size trends dependent on induced expression and polarity background.

Firstly, we see that simply adding galactose in the medium affects volumes Figure 65, even for pure WT, which present another obstacle in assessing the construct. Volumes increase with sugar concentration, even beyond values for growth in dextrose. As the typical size (as expressed in the median) increases, the width of the distribution scales with it. This does not seem a background dependent effect, as this also holds for the $\Delta bem1 \Delta bem3$. As shown in Appendix C.3, this is likely an osmolarity effect due to high sugar concentration. While it is possible to mitigate this when buffering with sorbitol, lower osmolarity values are more conventional. Therefore, we decided to simply de-trend the data of the strains without endogenous *CDC42*, looking instead at relative sizes compared to WT under the same conditions (Figure 66).

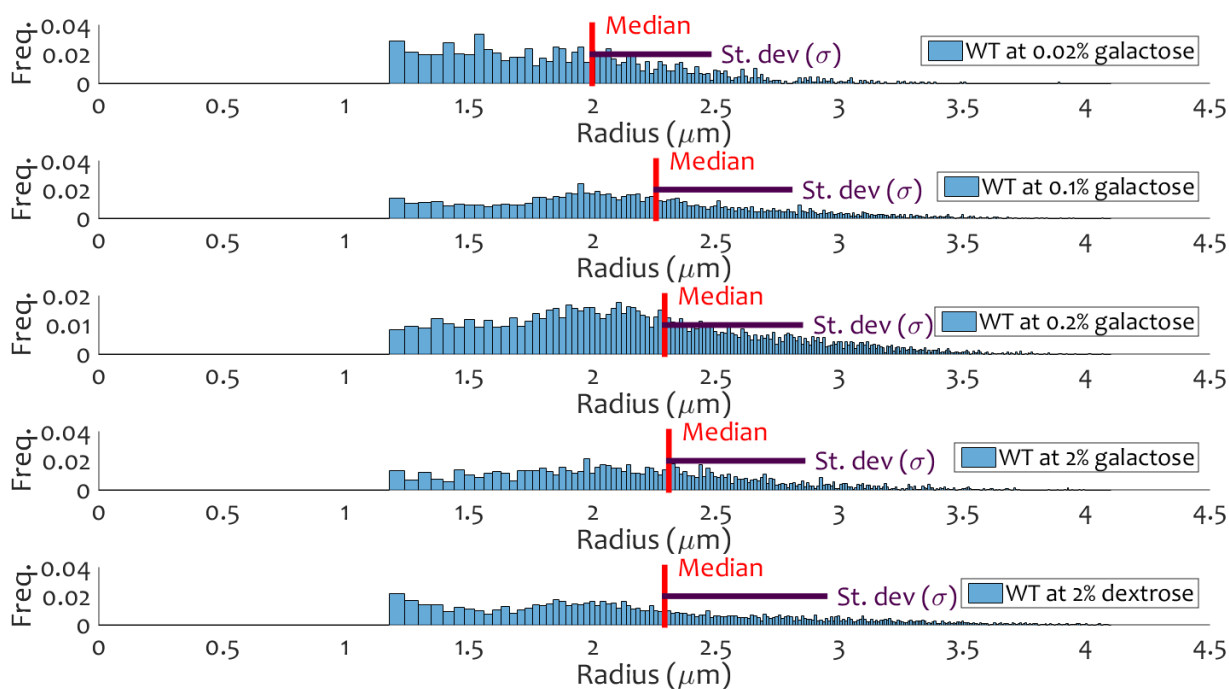


Figure 64 Histogram of sizes (radii) of WT in various media, together with the median (red line) and standard deviation (purple), for simplicity assuming all sizes are at the center of the bin in which they belong.

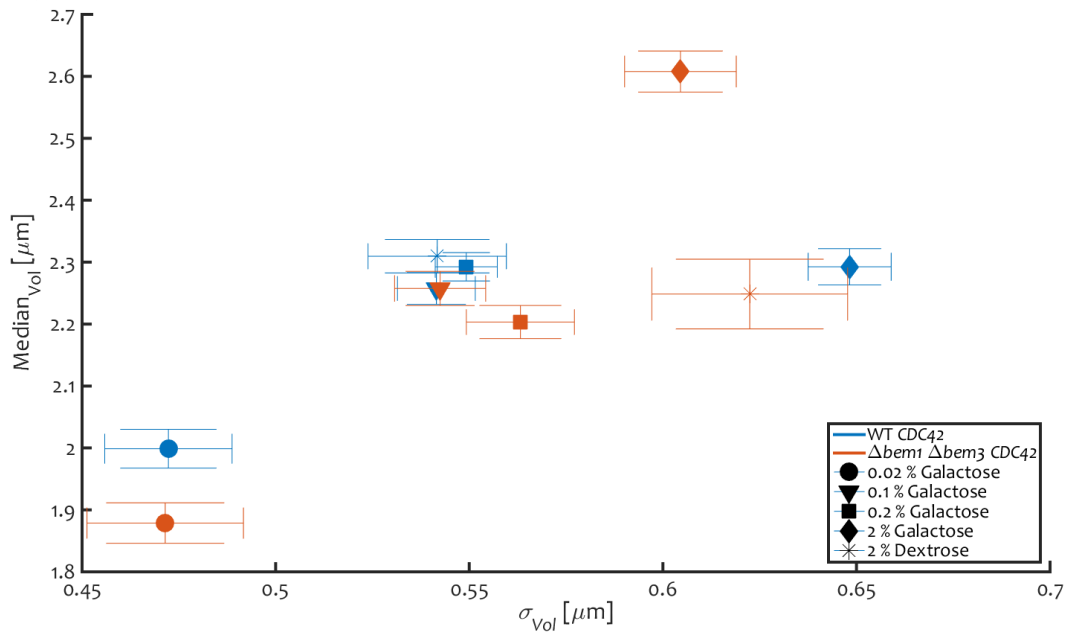


Figure 65 Median and standard deviation of volumes as measured with a Coulter Counter for strains with endogenous CDC42 in a WT background (YKWD001, blue) and $\Delta\text{bem1 } \Delta\text{bem3}$ (YKWD007, orange) background, for five galactose concentrations; 0.02% (circular markers), 0.1% (triangular markers), 0.2% (square markers), 2% (diamond markers) and 0% (also no raffinose, instead 2% dextrose, asterisks as markers). Error bars show the 95% confidence interval.

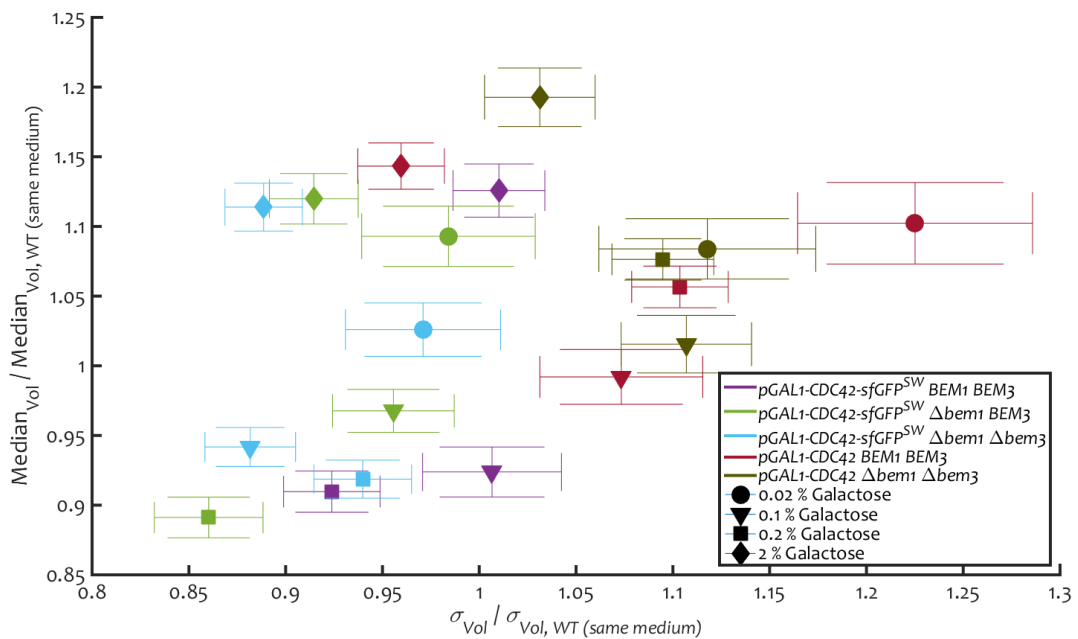


Figure 66 Median and standard deviation of volumes as measured with a Coulter Counter for strains with GAL1pr-sfGFP-CDC42^{SW} as the only source of Cdc42 in an otherwise WT (purple, YKWD065a), Δbem1 (green, YKWD069a/b) and $\Delta\text{bem1 } \Delta\text{bem3}$ (light blue, YKWD070a/b/c) background and with GAL1pr-CDC42 as the only source of Cdc42 in an otherwise WT (burgundy, YKWD071a) and $\Delta\text{bem1 } \Delta\text{bem3}$ (olive, YKWD073a/b) background, for four galactose concentrations; 0.02% (circular markers), 0.1% (triangular markers), 0.2% (square markers) and 2% (diamond markers). When multiple biological replicates were present, values and corresponding error bars result from weighted averaging. Error bars show the 95% confidence interval.

Now, the non-endogenous Cdc42 strains (either with sfGFP or not, but always under the Gal1-promoter) exhibit small trends upwards when supplied with more galactose, resulting in higher induction. These are presumably some residual fitness effects, with background benefitting from more Cdc42. Still, overall the media has largely equalized fitness, as under this concentration the $\Delta bem1$ under fast growth conditions was expected to be unviable and is at the edge of viability 30% larger than WT [24]. It is an interesting find that the influence of the environment on successfully crossing the presumed Cdc42 barrier resulting from its genotype can be well controlled in the experiment.

The most important conclusion follows when comparing the strains with/without sfGFP sandwiched in the Cdc42. We see in Figure 66 only a +/- 10% difference in volumes compared to WT without and with sfGFP respectively. Remarkably, cells with sfGFP in the construct become slightly smaller, so there is no sign of deleterious effects of this construct on fitness, which scales with volumes [24] (at constant osmolarity).

The causes of the small volume deviation might be a change in diffusive constant important for the reaction-diffusion pathway, as sfGFP is more than twice the size of Cdc42 (Figure 58). Cytoplasmic mobility is shown in *E. coli* to depend sharply on the molecular weight around the weight of Cdc42 [288]. Additionally, the fast maturation of sfGFP and how it evades degradation [188] may contribute to a more stable and faster fold of Cdc42, which might weaken the role of Rdi1 to protect Cdc42 from degradation [77]. This may result in slightly higher overall Cdc42 number, which is beneficial.

Ultimately, the small volume deviations and the on population-scale small fluorescent artefacts inspire sufficient trust to proceed with strains containing the *GAL1pr-sfGFP-CDC42^{SW}* construct for the conclusive growth assay, which are conducted in 2.3.

C.3 Rough growth assay in suboptimal growth conditions

This assay (Figure 67) has comparable medium to the volumetric assay in Chapter 2. However, as evaporation was poorly controlled, actual galactose concentration may be up to twice as high, not the case in the volumetric assay). Curves are intended to reflect growth as during the volumetric assay, which does not imply the fastest growth at very low OD, but somewhere higher up the curve (OD>0.1, as OD base starting level is typically already around 0.08). Also, when supplemented with sorbitol such that all strains are measured in the same high osmolarity media, trends as function of galactose concentration vanish (see Figure 68).

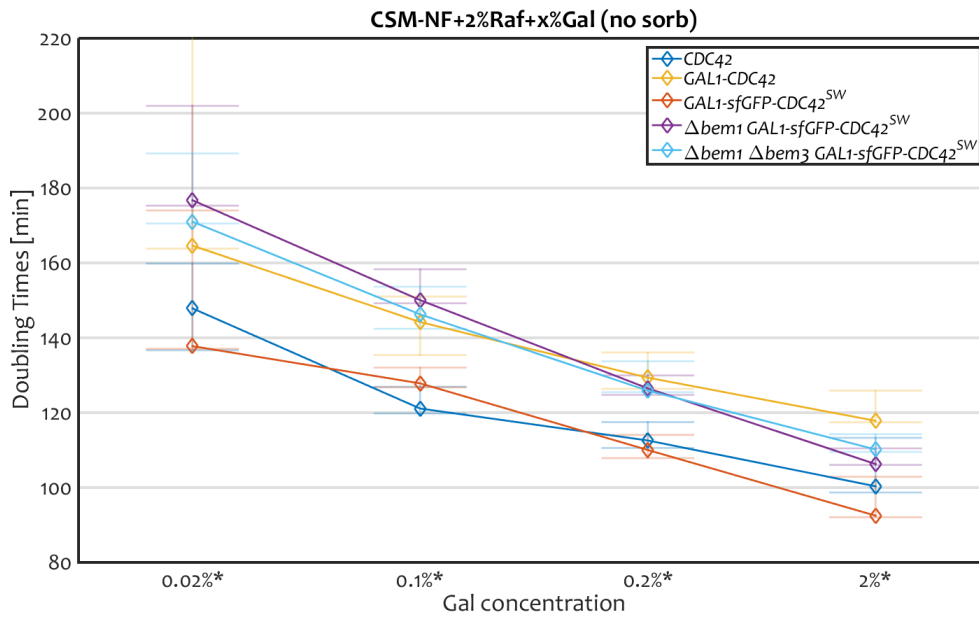


Figure 67 Doubling times as measured from a TECAN plate reader for $GAL1pr$ -sfGFP-CDC42^{SW} as the only source of Cdc42 in an otherwise WT (orange line, YWKD065a), $\Delta bem1$ (purple line, YWKD069a/b/c) and $\Delta bem1 \Delta bem3$ (light blue, YWKD070a/b/c) background, for $GAL1pr$ -CDC42 as the only source of Cdc42 in an otherwise WT (yellow, YWKD071a/b/c) background and pure WT (dark blue line, YWKD062/b/c). Values are linear fits using weighted least squares (inverse of approximated measurements errors as weights) intended on the regime where most of the log-phase growth takes place, defined as $OD > 0.1$ (so not necessarily the 'fastest part'). The induction concentrations at the horizontal axis represent due to evaporation uncertainty at least 0.02%, 0.1%, 0.2% and 2% galactose. Error bars represent lowest and highest replicate value.

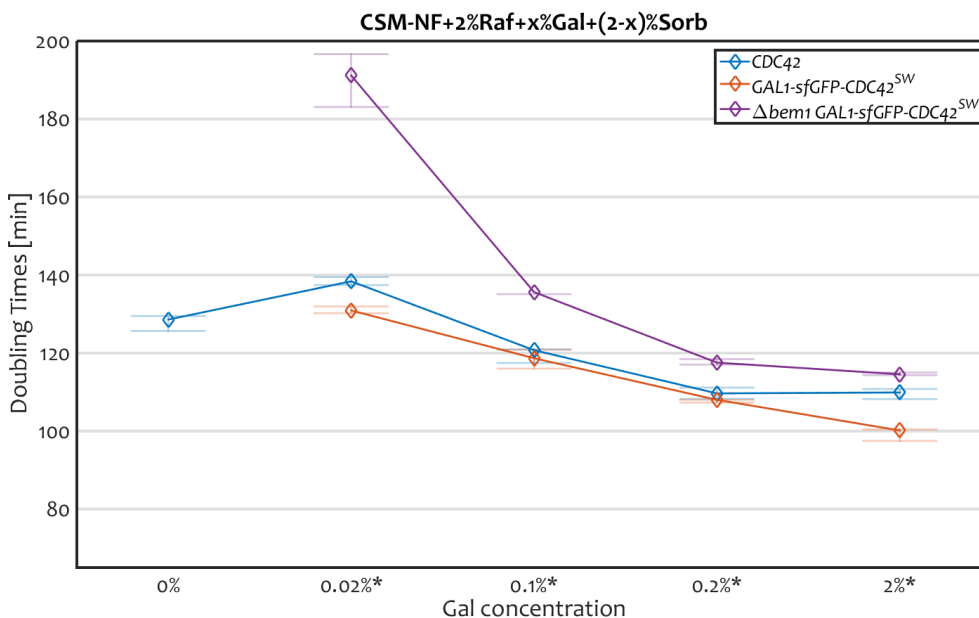


Figure 68 Doubling times as measured from a TECAN plate reader for $GAL1pr$ -sfGFP-CDC42^{SW} as the only source of Cdc42 in an otherwise WT (orange line, YWKD065a), $\Delta bem1$ (purple line, YWKD069a) background and pure WT (dark blue line, YWKD062c). Values are linear fits using weighted least squares (inverse of approximated measurements errors as weights) intended on the regime where most of the log-phase growth takes place, defined as $OD > 0.1$ (so not necessarily the 'fastest part'). The induction concentrations at the horizontal axis represent due to evaporation uncertainty at least 0.02%, 0.1%, 0.2% and 2% galactose. Error bars represent lowest and highest replicate value. Medium was supplemented with sorbitol such that the osmolarity of all media are the same.

D Membrane growth rate considerations for calibration

When calibrating the growth parameters to match the WT experimental doubling time of 83 min., we first consider the rough dependency of the doubling time on the growth parameters. We assume that WT is relatively optimized such that polarization time is always $t_{pol,min}$, and that switching from isotropic to polarized growth is limited by the minimum G1 time ($t_{G1,min}$) requirement. So, changing membrane growth rate C_1 does not strongly alter fitness. In that case, the smallest size after G1 r_1 follows from growing for $t_{G1,min}$ to reach r_{min} and for $t_{pol,min}$ afterwards (t_{P1} total in G1):

$$4\pi r_1^2 = 4\pi r_{min}^2 + C_1 t_{pol,min} \Rightarrow r_1 = (r_{min}^2 + C_1 t_{pol,min}/(4\pi))^{1/2}$$

The radius of the bud, r_{bud} , then becomes:

$$r_{bud} = 0.7^{1/3} r_1 = 0.7^{1/3} (r_{min}^2 + C_1 t_{pol,min}/(4\pi))^{1/2} \quad (D.1)$$

which takes amount of time t_{P2} to grow, governed by:

$$\begin{aligned} 4\pi r_{bud}^2 &= C_2 t_{P2} \Rightarrow 4\pi 0.7^{2/3} (r_{min}^2 + C_1 t_{pol,min}) = c_P C_1 t_{P2} \\ \Rightarrow t_{P2} &= \frac{4\pi 0.7^{2/3}}{c_P C_1} r_{min}^2 + \frac{0.7^{2/3}}{c_P} t_{pol,min} \end{aligned} \quad (D.2)$$

For self-consistency, this will only hold when the daughter cell can also obtain the same size, otherwise G1 growth is too fast:

$$4\pi r_{min}^2 \geq 4\pi r_{bud}^2 + C_1 t_{G1,min} \Rightarrow C_1 \leq 4\pi (r_{min}^2 - r_{bud}^2)/t_{G1,min}$$

Now suppose G1 growth is maximal, so we set $C_1 = 4\pi (r_{min}^2 - r_{bud}^2)/t_{G1,min}$ in D.1 and D.2:

$$t_{P2} = \frac{t_{G1,min} 0.7^{2/3}}{c_P (r_{min}^2 - r_{bud}^2)} r_{min}^2 + \frac{0.7^{2/3}}{c_P} t_{pol,min} \quad (D.3)$$

$$\begin{aligned} r_{bud}^2 &= 0.7^{2/3} (r_{min}^2 + (r_{min}^2 - r_{bud}^2) t_{pol,min}/t_{G1,min}) \\ \Rightarrow r_{bud}^2 &= \frac{0.7^{2/3} (1 + t_{pol,min}/t_{G1,min})}{(1 + 0.7^{2/3} t_{pol,min}/t_{G1,min})} r_{min}^2 \end{aligned} \quad (D.4)$$

Plugging these values into D.3 yields

$$t_{P2} = \frac{0.7^{2/3}}{c_P (1 - 0.7^{2/3})} t_{G1,min} + \left(\frac{0.7^{4/3}}{c_P (1 - 0.7^{2/3})} + \frac{0.7^{2/3}}{c_P} \right) t_{pol,min} \quad (D.5)$$

Then for the doubling time t_{tot} of that mother cell:

$$\begin{aligned}
t_{tot} &= \left(1 + \frac{0.7^{2/3}}{c_p(1 - 0.7^{2/3})}\right)t_{G1,min} + \left(1 + \frac{0.7^{4/3}}{c_p(1 - 0.7^{2/3})} + \frac{0.7^{2/3}}{c_p}\right)t_{pol,min} \\
&= \left(1 + \frac{0.7^{2/3}}{c_p(1 - 0.7^{2/3})}\right)(t_{G1,min} + t_{pol,min}) \tag{D.6}
\end{aligned}$$

The value of C_1 is then:

$$C_1 = \frac{4\pi \left(r_{min}^2 - 0.7^{2/3} \left(r_{min}^2 + \frac{C_1 t_{pol,min}}{4\pi} \right) \right)}{t_{G1,min}} = \frac{4\pi \left(1 - 0.7^{2/3} \right) r_{min}^2}{t_{G1,min} \left(1 + 0.7^{2/3} \frac{t_{pol,min}}{t_{G1,min}} \right)} \tag{D.7}$$

Substituting the time values assumed in this model (see Table 4, section 3.2), we get:

$$C_1 = 0.136 r_{min}^2 \tag{D.8}$$

The population doubling time is now to some extent robust against changes in C_1 , as one may expect in WT. For example, doubling C_1 only changes up polarized growth; more yet faster growth is required. Supposing a 20/60 min. balance between G1/G2 times for calibrated WT, the speed gain in polarized growth would be 30 min. (60/2), while the loss due to more G1 growth is 10 min., so 20 min. time gain overall for the first generation. For the cells that have already had one daughter, this gain would be 10 min. instead (loss has doubled), and for two daughter mothers this would be about 0. But because these generations are smaller fractions of the population, the total population doubling time would likely only have about 10 min. gain from doubling C_1 . This shows a rather broad optimum.

However, for the smallest daughter size will increase as well. Doubling C_1 would imply twice the added membrane growth for about a quarter of the cell cycle. For self-consistency, $r_{min} = r_{bud}$ in D.1 gives:

$$\left(1 - 0.7^{2/3}\right)r_{bud}^2 = 0.7^{2/3} \frac{C_1 t_{F1}}{4\pi} \Rightarrow r_{bud} = \sqrt{\frac{0.7^{2/3} C_1 t_{F1}}{4\pi \left(1 - 0.7^{2/3}\right)}} \tag{D.9}$$

For the usually assumed parameters, this is about 1.82 μm . While an increase of a factor of 2 in C_1 would gain at most 15% in doubling time, it will also increase the smallest daughter size by more than 40%. So an increase in C_1 quickly becomes incompatible with [228].

Decreasing C_1 has a much larger effect on doubling times. Decreasing a factor of two means for cells (that do not yet have a daughter) almost a doubling of G1 phase (20 to 35 min., as only time before polarization gets doubled), while polarized growth phase will double as well from 60 to 115 min. Subsequent generations suffer less, but overall having $C_1 = 0.136 r_{min}^2$ is hence relatively optimized.

E Reconstruction of burst parameters from FACS data

In order to characterize expression of the endogenous *CDC42* and analogously, expression under the inducible promoter *GAL1*, a fluorescent protein, in this case GFP, must be attached to Cdc42. While this notably leads to loss-of-function when N-terminally fused, an additional, endogenous version of Cdc42 must remain present. However, this is not a big disadvantage: overexpression of Cdc42 may have otherwise led to a fitness effect, which in turn may influence the shape of the population protein distribution.

Presumably, measuring population protein distributions through fluorescent intensity by flow cytometry should yield a gamma distribution, given the small mRNA/protein life time ratios $\tau_{1/2,Cdc42}$ [189], [235], [236]. That this seems true is shown in Figure 69 to Figure 72. This data was acquired from strains YMS02-04, replicates of RWS116, 119, and RWS1421 from the Wedlich-Söldner lab, measured by Marit Smeets (see also 2.5) and gated to remove artefacts.

There is evidence for a mixture distribution, mainly for the lowest galactose concentration. This suggests a significant fraction of the population seems to have the promoter in the “off” state. Yet, another paper [199] did not see this, so presumably these are just cells that had not yet left lag phase, so we will fit using mixture distributions, but take the one with the highest mean as the *GAL1* promoter distribution. The mixture distribution also gives a slightly better fit for the endogenous expression (see Figure 72, suggesting minor regulation (possibly a transcription factor causing a two-level expression system) which will be neglected.

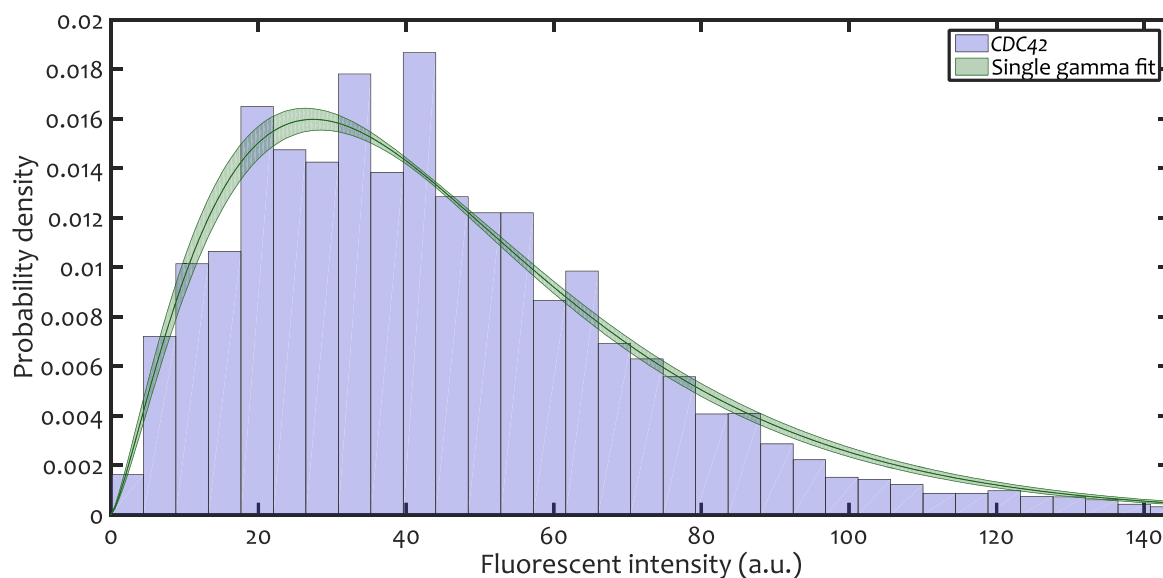


Figure 69 Flow cytometry data of YMS02, which has non-fluorescent, endogenous *CDC42* to obtain a measure for the autofluorescence. In green is the gamma distribution fit on the data, with the corresponding 95% confidence interval (CI) given the parameter uncertainty after estimation. The CI resulted from Monte Carlo simulation of parameter values assuming normal errors and zero cross correlation in uncertainty estimates for k and θ .

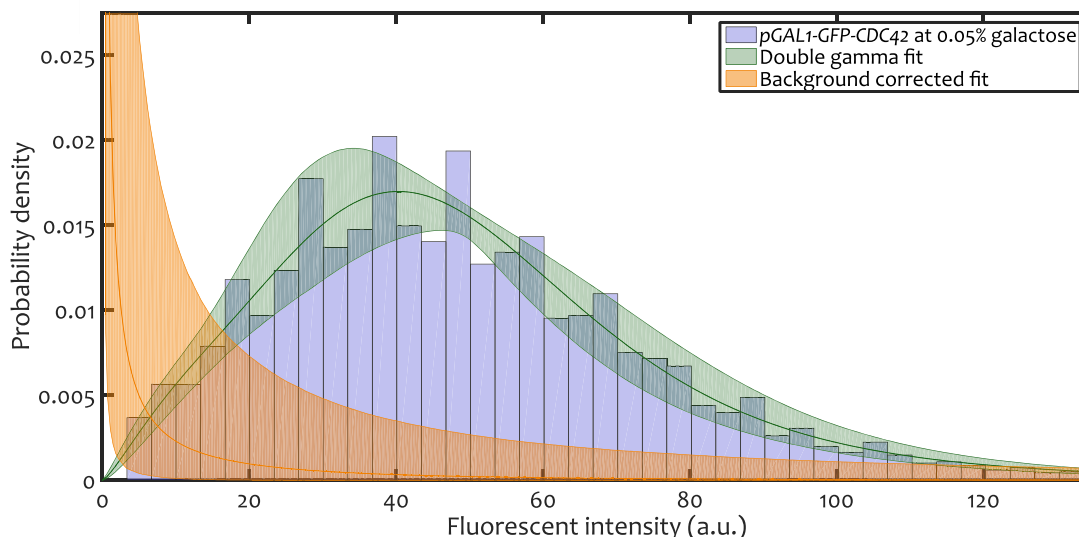


Figure 70 Flow cytometry data of YMS03 with 0.05% galactose, which has GAL1-GFP-CDC42. In green is the gamma distribution fit on the data, in orange is the background corrected gamma fit, with the corresponding 95% confidence interval (CI) given the parameter uncertainty after estimation. The CI resulted from Monte Carlo simulation of parameter values assuming normal errors and zero cross correlation in uncertainty estimates for k and θ .

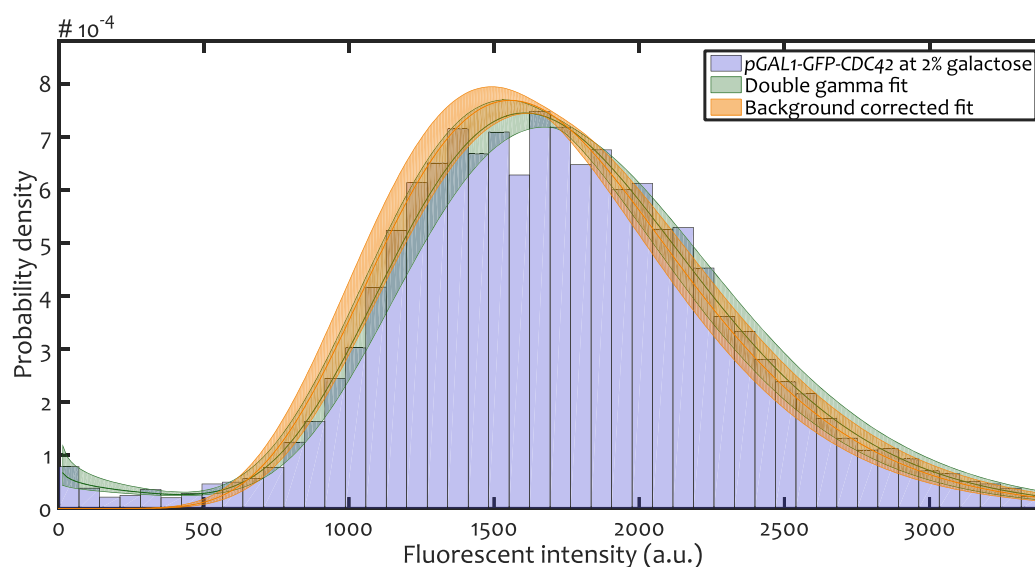


Figure 71 Flow cytometry data of YMS03 with 2% galactose, which has GAL1-GFP-CDC42. In green is the gamma distribution fit on the data, in orange is the background corrected gamma fit, with the corresponding 95% confidence interval (CI) given the parameter uncertainty after estimation. The CI resulted from Monte Carlo simulation of parameter values assuming normal errors and zero cross correlation in uncertainty estimates for k and θ .

There is significant autofluorescence given the histogram of the strain without GFP, particularly for the *pGAL1-GFP-CDC42* strain in medium with 0.05% galactose. This would have to be “subtracted” from the observations, yielding a distribution for the fluorescent Cdc42 copy number. In statistical terms, the observations should be deconvolved (of the background). Conceptually, one could try to remove the autofluorescence from the observed distributions in two ways: either non-parametrically or assuming certain properties for the underlying distributions (modelling).

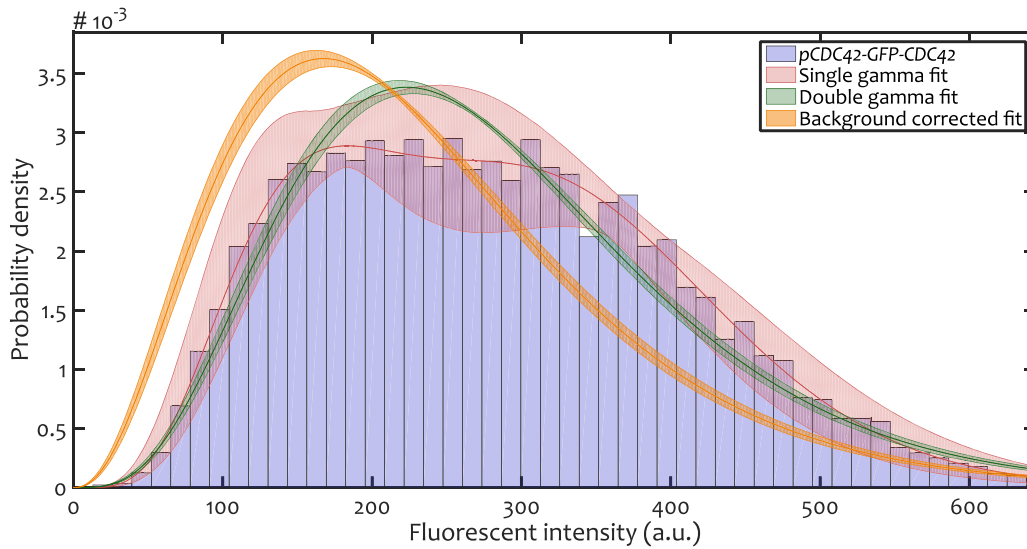


Figure 72 Flow cytometry data of YMS04 with 0.2% galactose, which has pCDC42-GFP-CDC42 to obtain a measure for the endogenous expression. In green is the gamma distribution fit on the data, in orange is the background corrected gamma fit and in red a double gamma mixture distribution fit, with the corresponding 95% confidence interval (CI) given the parameter uncertainty after estimation. The CI resulted from Monte Carlo simulation of parameter values assuming normal errors and zero cross correlation in uncertainty estimates for k and θ .

The non-parametric approach is not straightforward. Convolution in the Fourier domain means multiplication, so in very basic terms, one should take the Fourier transform, divide out the noise (the autofluorescent signal) and do an inverse Fourier transform. This last step is difficult; the necessary integral is hard to compute or does not exist because the empirically derived probability distributions are typically not so well behaved. Researchers usually smoothen the discrete observations data with kernels. Without assuming any theoretical knowledge on the noise distribution, this problem gets difficult, but is solved, e.g. here [289].

The parametric approach seems more feasible. The non-parametric approach ignores how much is already known about the underlying processes, knowledge that plays to our advantage. From a theoretical point of view, assuming exponential waiting times between bursts of expression, one would expect a gamma distribution (see [189], [237]), which fits the noise well. That this also holds for the convolved observations can be explained by [287], where it is shown that one can approximate the convolution of two gamma probability density functions by another gamma function. This is a very useful property of the Gamma distribution, because it becomes possible to assume a Gamma distribution for the unobserved Cdc42 distribution, with parameters inferred from the observed distributions.

Concretely, the unobserved, real Cdc42 distribution is reconstructed as follows: The observed, convoluted Cdc42 distribution is modelled as a mixture distribution subject to a random variable having probability p to be “off”, following Γ_{off} , and $(1 - p)$ to be “on” following Γ_{on} . These are decomposed separately. The actual Cdc42 signal S then follows the distribution:

$$S \sim f(x) = \begin{cases} \Gamma(k_{s,off}, \theta_{s,off}), & \text{probability } p \\ \Gamma(k_{s,on}, \theta_{s,on}), & \text{probability } (1 - p) \end{cases} \quad (\text{E.1})$$

where parameters follow from:

$$k_{t,off} = \frac{(k_b\theta_b + k_{s,off}\theta_{s,off})^2}{k_b\theta_b^2 + k_{s,off}\theta_{s,off}^2} \quad (\text{E.2})$$

$$\theta_{t,off} = \frac{k_b\theta_b^2 + k_{s,off}\theta_{s,off}^2}{k_b\theta_b + k_{s,off}\theta_{s,off}} \quad (\text{E.3})$$

and similarly for “on”, with $k_{t,off/on}$ and $\theta_{t,off/on}$ resulting from maximum likelihood estimation (MLE) on the (total, subscript t) observed signal, and k_b and θ_b resulting from fitting the background (subscript b) signal. The MLE also gives 95% confidence intervals.

Similarly, (but now neglect the “off” state) this is done for the endogenous promoter case, yielding a gamma distribution with a mean given by $k_{endo}\theta_{endo}$, and a confidence interval following propagation of error of the ML estimates. We define this mean as “the WT amount Cdc42”. This can be used to scale the x-axis of the distribution Γ_s to have Cdc42 concentration (in WT units) instead of arbitrary fluorescence units. Gamma distributions are easily scaled:

$$\Gamma_{scaled} = \Gamma_s(k_s, \theta_s / (k_{endo}\theta_{endo})) \quad (\text{E.4})$$

Again, using Monte Carlo simulations gives the error estimates of θ . The result of scaling should be only a slightly wider distribution. The resulting confidence interval will be asymmetric with respect to the mean value, for the fold Cdc42 increase when adding various amounts of galactose. The resulting, fitted flow cytometry data, normalized by the endogenous expression are then given in Figure 59 (bottom) in Appendix C.2.1.1.

F Effect of stochastic GAP production on Cdc42/GAP ratio

If GAPs are explicitly included into the growth model, polarization will not depend on Cdc42 concentration relative to a threshold, but on the ratio of Cdc42 and GAP concentrations relative to a threshold. In order to understand the distribution of the protein ratio, we use that within every generation, the production of new protein follows approximately a gamma distribution, by the following reasoning. The distribution of proteins Z by the model assumptions is a convolution of exponentially distributed burst sizes (as the random variable is a sum of exponentially distributed random variables):

$$Z = \sum_{i=1}^m X_i$$

where m itself is Poisson distributed (number of bursts within a deterministic amount of time when burst waiting times are exponentially distributed).

This sum of exponentials follows a Gamma distribution with scale parameter as the inverse of the individual exponential parameter and the shape parameter is Poisson distributed. This is because an exponential distribution is just a gamma distribution (with shape parameter 1 and scale as inverse exponential parameter), and the sum of gamma's can be approximated by another gamma [287]. The latter is exact when scale parameters are identical, which is true if all exponentials are independent and identically distributed (not the case if degradation is an important factor, as then the first burst is effectively weaker than the last). So:

$$[Cdc42] \sim \Gamma(x; k_1, \theta_1) \quad , \quad [GAP] \sim \Gamma(x; k_2, \theta_2)$$

Hence, assuming that both Cdc42 and the GAPs approximately follow a gamma distribution within a cell cycle, the ratio of the two will follow a beta prime or generalized gamma ratio distribution [290]:

$$\frac{[Cdc42]}{[GAP]} \sim \frac{(x/q)^{k_1-1} (1 + (x/q))^{-k_1-k_2}}{qB(k_1, k_2)}$$

with $q = \theta_1/\theta_2$. Since from [229] we know typically $k_1 \ll k_2$ and $k_2 \gg 1$, the m -th moment as given in [290] reduces to:

$$E \left[\left(\frac{[Cdc42]}{[GAP]} \right)^m \right] = \left(\frac{1}{q} \right)^{-m} \frac{\Gamma(k_1 + m)}{\Gamma(k_1)} \frac{\Gamma(k_2 - m)}{\Gamma(k_2)} \approx \left(\frac{q}{k_2} \right)^m \frac{\Gamma(k_1 + m)}{\Gamma(k_1)}$$

which are the moments of a gamma distribution with shape k_1 , as for Cdc42, but rescaled with scale parameter q/k_2 . This makes intuitive sense, because in the limit that GAPs are completely constitutively expressed, its gamma distribution just turns into a delta peak around one value. In other words, the burst dynamics of Cdc42 are dominant for the protein noise relevant for reaching the polarity threshold.

G Convergence and error estimation

The following plot shows the convergence to an equilibrium state, for two opposite ends of the fitness spectrum; approximately WT and $\Delta bem1$, see Figure 73 and Figure 74. Simulations are run with $[Cdc42]_{min,rel} = \frac{[Cdc42]_{min}}{P_{WT} / \frac{4}{3}\pi r_{min,WT}^3}$ of 0.03 and 0.45 respectively, and otherwise standard parameters (see Table 4). The choice of initial conditions is by default 1000 cells without protein at $1.1r_{min,WT}$, and is shown to be inconsequential for convergence.

Additionally, to demonstrate the reproducibility of the equilibrium results, an estimate of the numerical error is obtained that inevitably occurs when estimating doubling time. This is shown in Figure 75. It depicts two main error sources: one is due to truncation of the simulation at a colony size of 5 million to maintain reasonable computation times (presented on the x-axis), and the other is due to the intrinsic randomness of protein bursts that the simulation take into account, leading to slightly different results from different random seeds (the multiple points per genotype). The chosen genotypes reflect fast, medium and slow growth (approximately WT, $\Delta bem1\Delta bem2$ ($c_{min,rel} = 0.27$) and $\Delta bem1$ respectively).

It can be seen that cutting off the simulations at the point of 5 million cells seems appropriate, with deviations of typically 0.1% for adding 100 minutes to the colony simulation time. The doubling time estimates are instead dominated by the error from the “technical replicates” (different seeds). However, it is typically confined to 0.5%, up to typically 1% for slow growing cells. This is smaller than the typical experimental error of roughly 5%, showing that this level of accuracy for the simulations is sufficient.

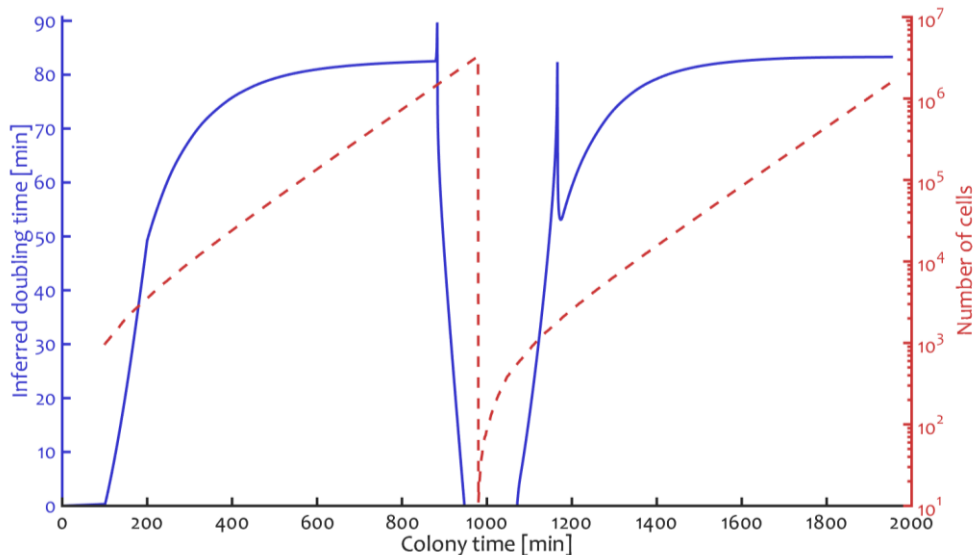


Figure 73 Simulated doubling times (blue filled line) and colony size (red dashed line) for (approximately) WT as a function of time from ancestor. Colony size / number of cells are in this case a subset of the actual colony size, namely those whose time stamps fall into the appropriate one-minute bins to be considered for the doubling times. With the dilution step half-way (leaving only 1000 cells in reality), this renders doubling time estimates (average of previous 100 200-point moving windows) around this time very unreliable, but convergence follows above a colony size of 100000.

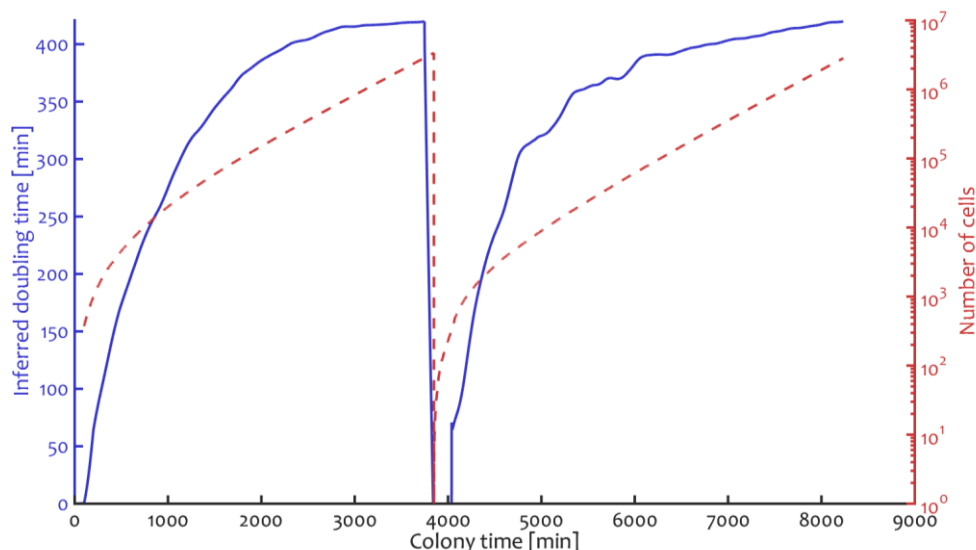


Figure 74 Simulated doubling times (blue filled line) and colony size (red dashed line) for (approximately) Δbem1 as a function of time from ancestor. Colony size / number of cells are in this case a subset of the actual colony size, namely those whose time stamps fall into the appropriate one-minute bins to be considered for the doubling times. With the dilution step half-way (leaving only 1000 cells in reality), this renders doubling time estimates (average of previous 100 200-point moving windows) around this time very unreliable, but convergence follows above a colony size of 100000.

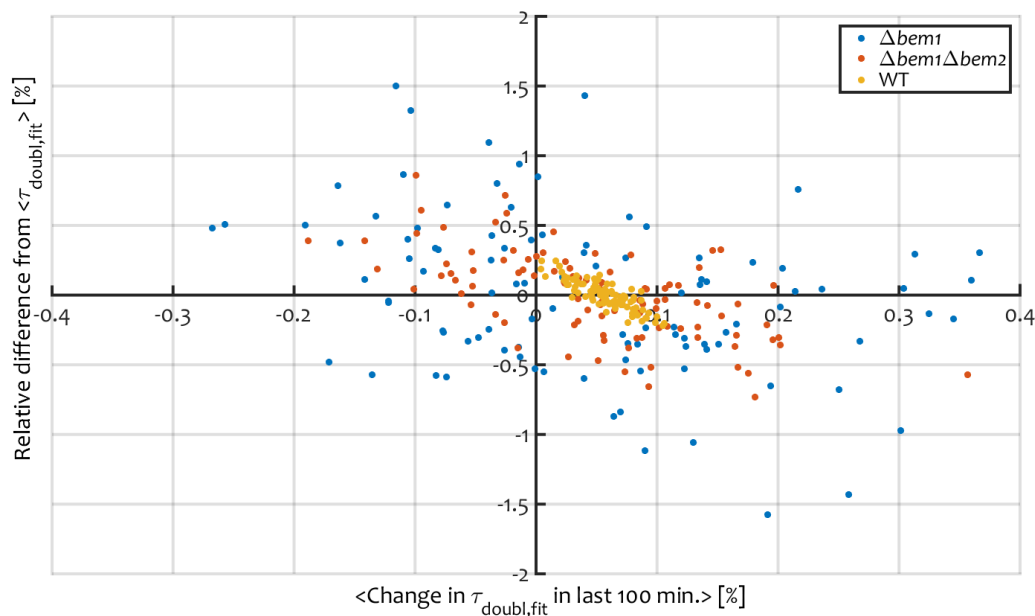


Figure 75 Scatter plot of relative error in doubling times (in %, centered around 0) and average relative change for in doubling time (in %, defined as the 100 times the fitted slope on fitted doubling times over the last 100 time windows of 200 min. in the simulation that reaches a colony size of at least 5 million cells) converted to a percentage for 100 different random seeds at the start of the simulations (mimicking “technical replicates”) for three genotypes, which approximately correspond to WT (fast growing), $\Delta\text{bem1}\Delta\text{bem2}$ (medium growing), Δbem1 (slow growing).

H Single cell test case

In order to more easily assess the correctness of the numerical implementation of the growth rate model, the results from the perspective of a single cell are considered. What can be seen in Figure 76 is that for this example cell (from a $\Delta bem1$ background simulation, with Cdc42 threshold approximately as $\Delta bem1\Delta bem2$), there is continuous degradation of Cdc42, but an even steeper decline in concentration due to dilutions. The discontinuous increases show protein expression bursts. Indeed, the cell can only divide (discontinuous drops in protein number absent in protein concentration) when the concentration of Cdc42 exceeds the red zone at an earlier time before budding (there is polarization time and polarized growth in between these events). Otherwise, the cell will continue to grow until it exceeds the maximum cell size, after which it is pronounced dead. This is all according to the way the model is designed, validating the numerical implementation.

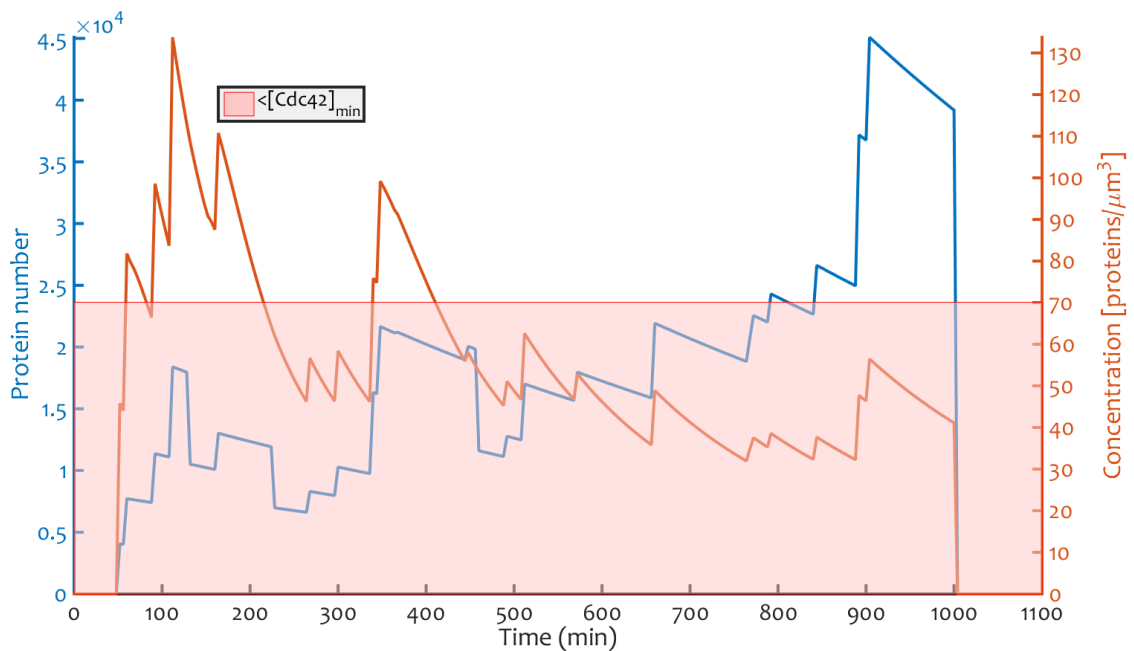


Figure 76 Time trace of the Cdc42 copy number (blue, left y-axis) and corresponding concentrations (orange, right y-axis) in an approximate $\Delta bem1\Delta bem2$. The red patch denotes the region where the concentration is below the minimum concentration to polarize. Where protein number and concentration hit zero again, the cell is dead.

I Relaxation of model assumptions

To test the robustness of the growth model results, the potential effect of relaxing certain model assumptions and varying parameter values is investigated. For otherwise WT parameters, the effect of minimum Cdc42 concentrations on doubling times is depicted in Figure 77. The curves for the $\Delta bem1$ backgrounds show the same trends. Curves are created by first adjusting C_1 and C_2 by the same factor such that WT doubling time comes as close to 83 minutes as possible. This is accomplished by modifying this factor in the model generating doubling times in a parameter sweep, and then linear interpolating to get an 83-minute doubling time. The sweeps for this pre-factor of C_1 and C_2 are 0.3 to 0.5 (step 0.02), 0.65 to 0.75 (step 0.02), 0.5 to 0.7 (step 0.02) and 0.4 to 0.5 (step 0.01) for assumption sets with twice the normal C_2/C_1 , 85% the normal r_{min} , fixed r_{bud} and constant volume expansion respectively.

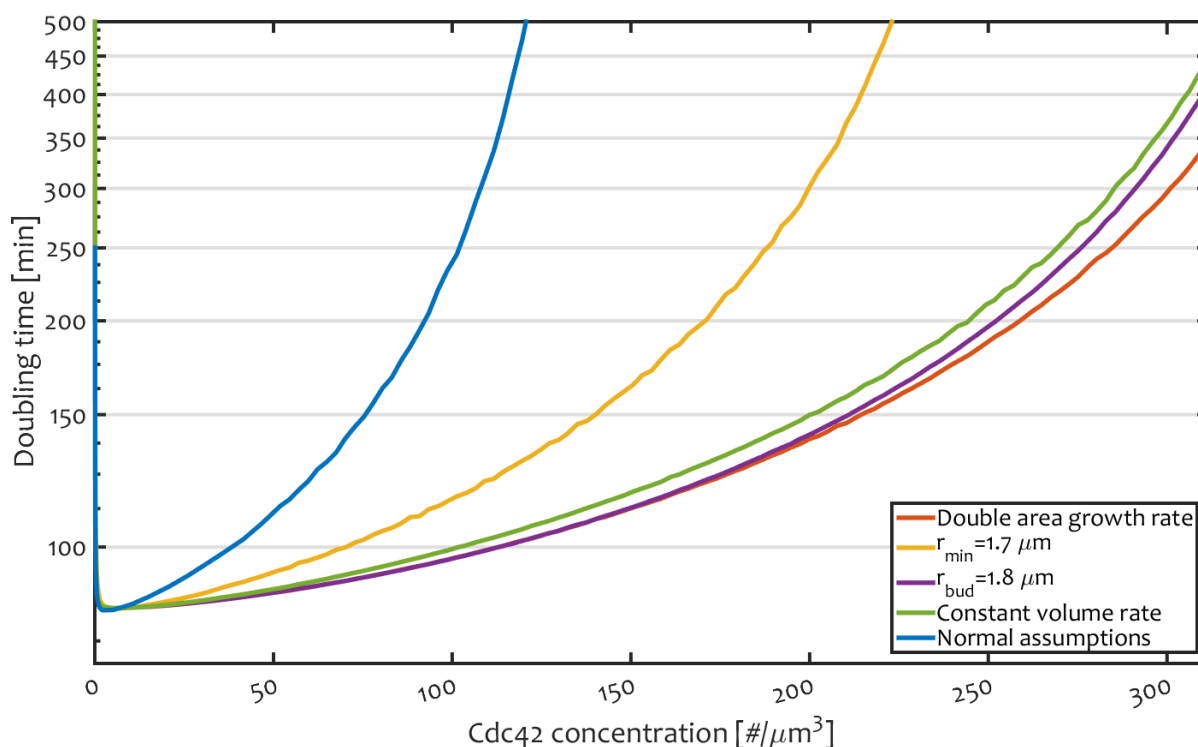


Figure 77 Simulated doubling times as a function of minimum Cdc42 concentration to polarize for WT parameters under various assumptions: in blue the typical assumed settings, in red with twice the normal membrane area growth rate during polarized growth, in yellow with a smaller minimum radius to polarize, in purple with a constant small daughter size and in green with constant volume instead of area growth rate. Area growth rates are recalibrated for every assumption set to match the WT doubling time of 83 minutes.

While the exact values of the fitted Cdc42 thresholds shift for the GAP deletion genotypes, the model fit remains accurate in all cases (see Table 14), with the mean absolute error staying within 40% of the original settings, and always less than 1 standard error from the experimental values. The relative effect of the GAP deletion is essentially constant.

*Table 14 Growth model fitting parameters represented as $[Cdc42]_{min,\Delta bem1}$, $[Cdc42]_{min,WT}$, relative effective GAP concentration effect of the *bem3* deletion and *bem2* deletion respectively for the model assumption set 1 to 5 (standard settings, twice the normal membrane area growth rate during polarized growth, smaller minimum radius to polarize, constant small daughter size and constant volume instead of area growth rate respectively), when fitted to the experimental data set for the $\Delta bem1$, $\Delta bem1\Delta bem3$, $\Delta bem1\Delta bem2$, $\Delta bem1\Delta bem3\Delta bem2$, WT, $\Delta bem3$, $\Delta bem2$ and $\Delta bem3\Delta bem2$ mutants [24]. The last row shows the mean absolute error (MAE), defined as the average of all deviations of the fitted simulated doubling time from the experimental doubling time divided by their respective experimental standard error for these mutants.*

	Set 1 (Default)	Set 2 ($2C_2$)	Set 3 ($r_{min} =$ $1.7 \mu m$)	Set 4 ($r_{bud} =$ $1.8 \mu m$)	Set 5 ($dV/dt =$ c)
$[Cdc42]_{min,\Delta bem1}$	116	327	214	314	308
$[Cdc42]_{min,WT}$	6	15	11	14	15
$[BEM3]/[GAP]_{tot,eff}$	0.62	0.63	0.62	0.64	0.63
$[BEM2]/[GAP]_{tot,eff}$	0.34	0.35	0.35	0.35	0.35
MAE	0.88	0.57	0.90	0.43	0.68

J Epistasis prediction from growth model

The growth model of Chapter 3 can predict epistasis between mutations that affect the model parameters. As an indication of this, in service of 3.4.2, the epistatic interactions between WT and $\Delta bem1$ is considered, when varying five model parameters. This leads to the epistatic predictions described in the aforementioned subsection, where parameter changes are coupled to the mutants related in literature, see Table 15.

Table 15 WT and $\Delta bem1$ simulated doubling times for various parameter sets of the growth model of Chapter 3. The standard parameters set refers to Table 4 with the fits of 3.3.1, for the other cases indicated parameters are changed relative to that set. The exception is the small/large bud size set, where changes are relative to the standard set with fixed bud radius of 2.5 μm , as normally daughters scale in size with mothers. Epistasis reported is defined as in equation 4.4. Literature categories as described in 3.4.2 are connected to the corresponding cases.

Case description	Relates to literature category	Parameters changed relative to standard	Parameters change [%]	Doubling time WT [min.]	Doubling time $\Delta bem1$ [min.]	Epistasis
Standard				82	396	
Fast membrane growth	Beneficial	C_1, C_2	5	82	624	-0.20
Slow membrane growth	Deleterious	C_1, C_2	-5	82	290	0.14
Slow G1	Slow G1 phase	$t_{G1,min,WT}$	20	94	903	-0.30
Fast G1	Fast G1 phase	$t_{G1,min,WT}$	-20	71	232	0.17
Slow degradation	Proteasomal	$t_{1/2,CDC42}$	40	82	314	0.10
Fast degradation		$t_{1/2,CDC42}$	-40	83	962	-0.38
Small bursts	Ribosomal	$p_{b,WT}$	-5	82	527	-0.12
Large bursts		$p_{b,WT}$	5	82	319	0.09
Standard with bud radius = 2.5 μm				89	234	
Small bud size	Small size at Start	r_{bud}	-20	64	113	0.18
Large bud size	Large size at Start	r_{bud}	20	119	2633	-0.92

K Effect of noise level in the two-state model

Figure 78 depicts the relative multiplicative epistasis as in equation 4.4 for the GAP deletions in the Δbem1 background. The two-state model is assumed with the Cdc42 case (sharp lower boundary mesotype). Fits are first performed on experimental doubling times from [24], then the coefficient of variation is varied (see also 4.6.1).

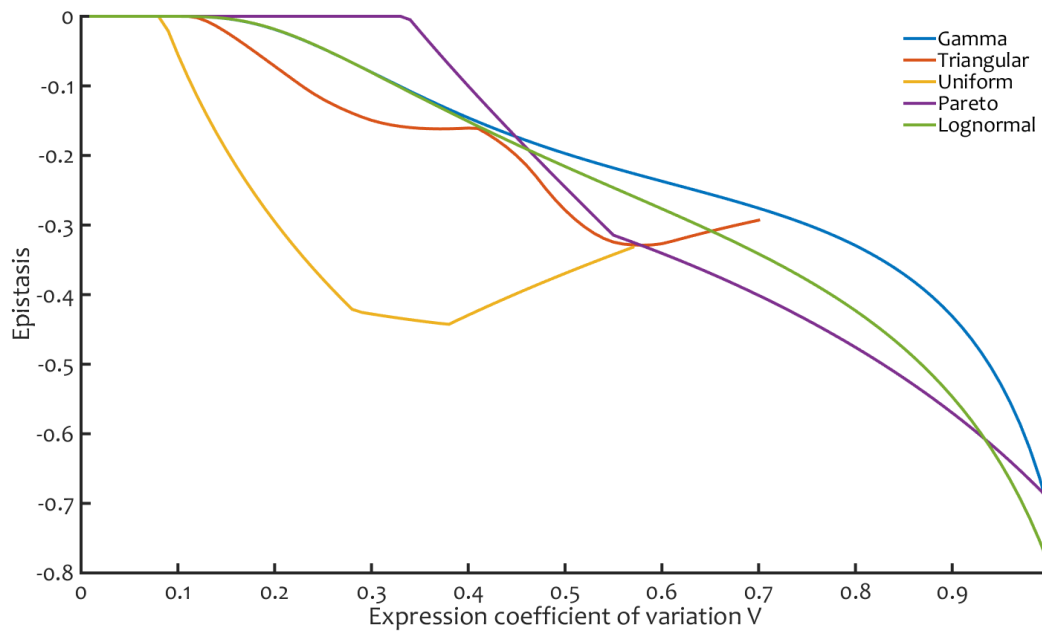


Figure 78 Epistasis as function of coefficient of variation for the volume-free two state model in the Cdc42 case.

L Two-state model fitness considering noisy GAPs in a $\Delta bem1$ background

Figure 80 shows the two state model (see graphical representation for the GAPs in Figure 79) contours for the fitness and survivable number of generations with transgenerational feedback, assuming a sharp upper mesotype boundary (GAP case). Figure 81 has the case without feedback placed on top of this.

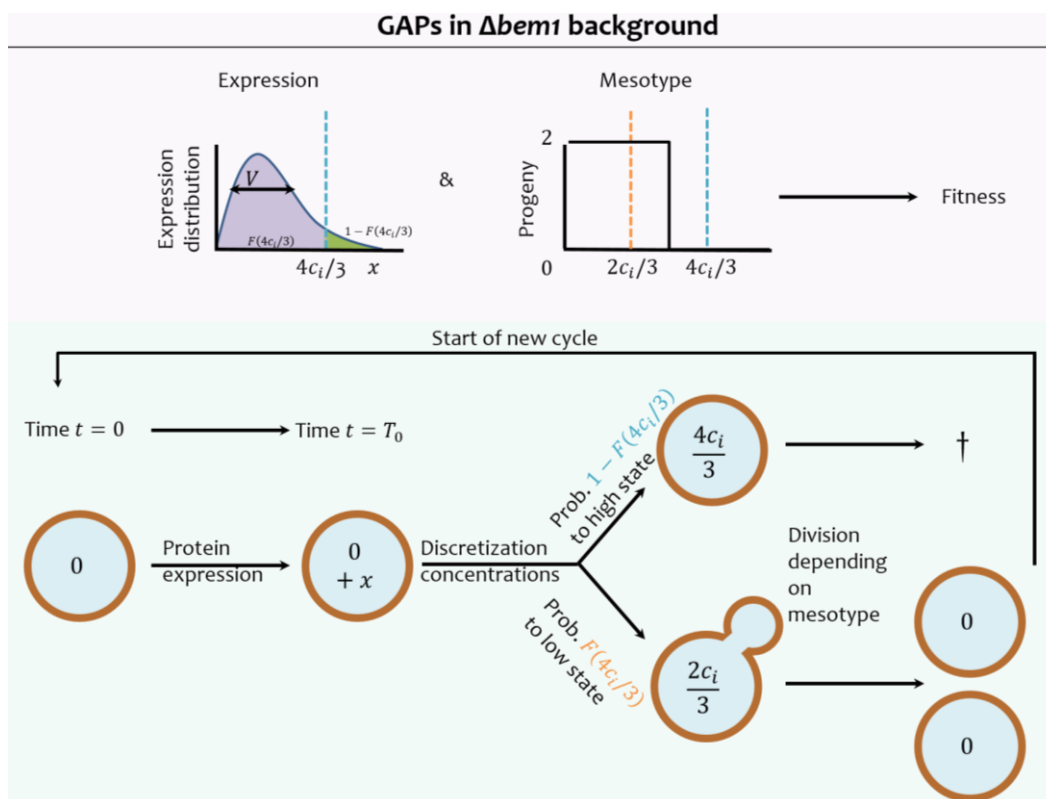


Figure 79 Two state model for fitness reflecting yeast cells polarizing with success depending solely on GAP concentration. Noisy expression during cell cycle time T_0 , combined with the simplified mesotypical rule for progeny depending on GAP concentration yields fitness, as stochastic switching occurs between the low and high GAP concentration state, with probabilities following from cumulative distribution function F , which is a function of noise level/ coefficient of variation V .

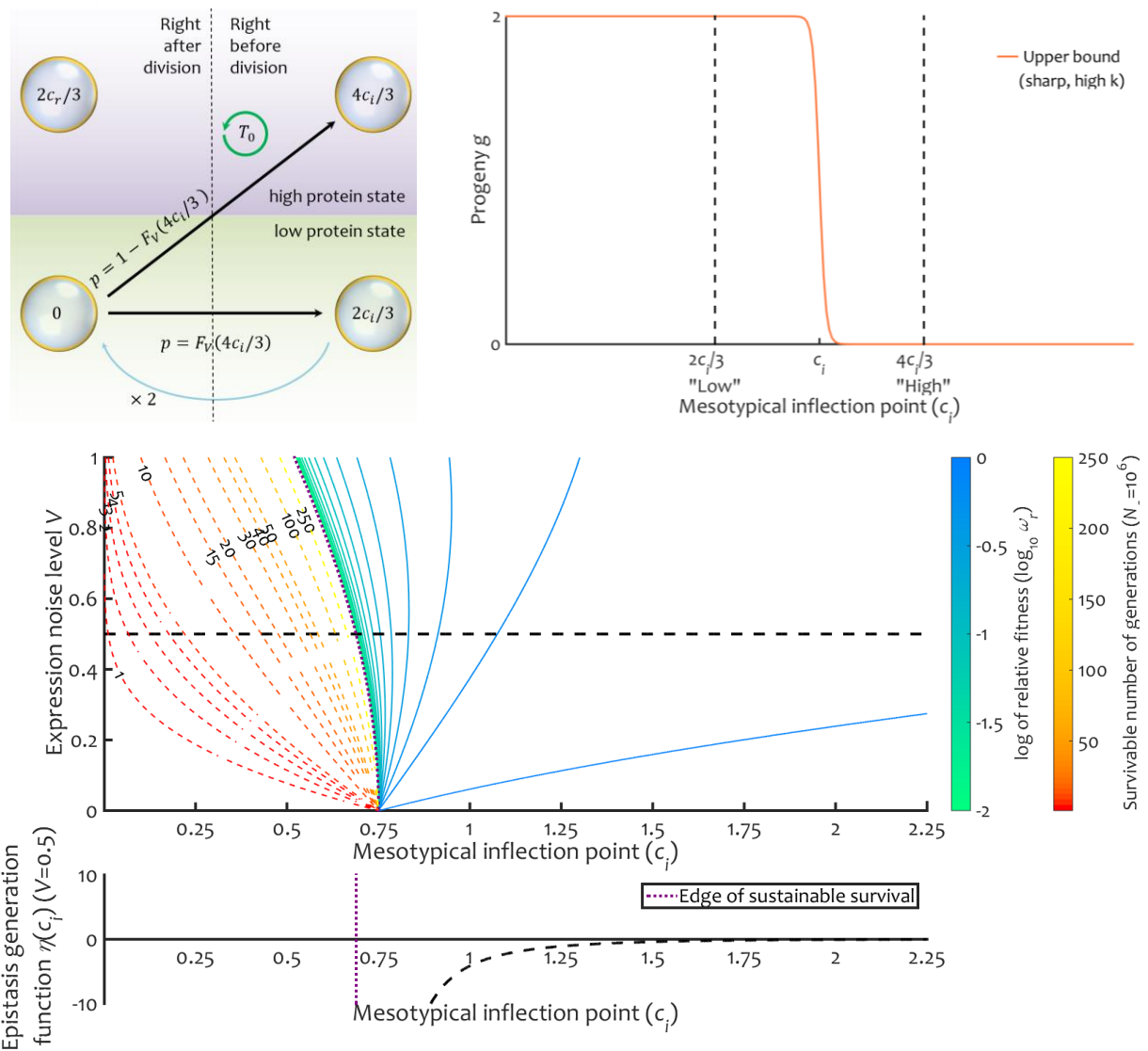


Figure 80 Noise in sharp upper boundary mesotypes (GAP case) extends both the sustainable and unsustainable growth regions, as well as the region that harbours epistasis. Top: Two state model schematic summary (left) and mesotype (right) of the sharp upper boundary (GAP) case. Middle: Contour plot of log relative fitness (blue to green) and number of generations a population of a million cells survives (red to yellow) as function of mesotypical inflection point of the sharp upper boundary in the fitness landscape, and the noise level (coefficient of variation), assuming a gamma protein expression distribution. The purple dotted line indicates the survival edge. The dashed black line denotes the line $V = 0.5$, along which the plot in the bottom is valid. Bottom: Second derivative of log relative fitness with respect to mesotypical inflection point (along the dashed line of the top axis). Here, it is always negative for values of c_i that support growth, allowing negative epistasis for mutations of the same sign.

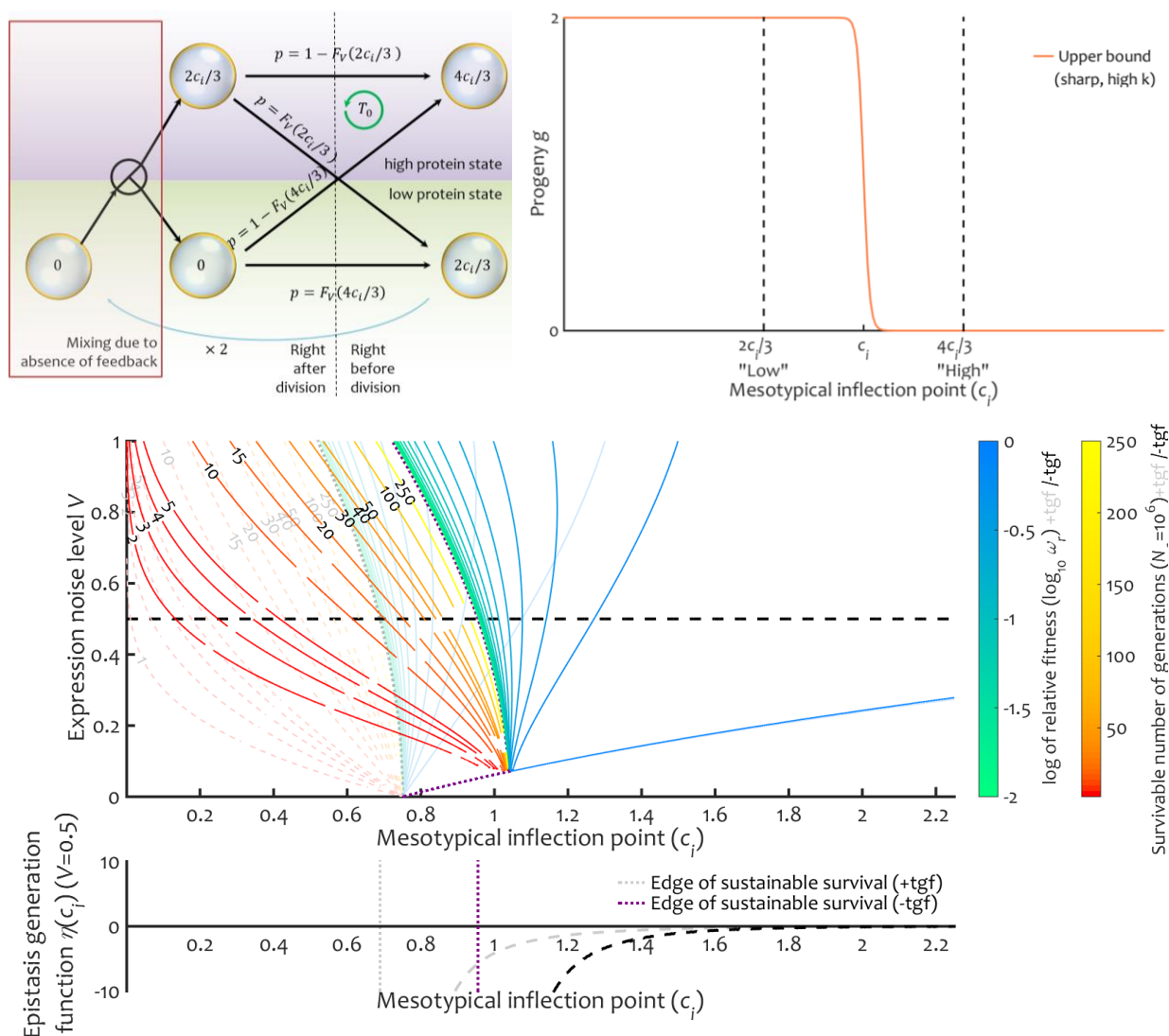


Figure 81 Absence of transgenerational feedback in sharp upper boundary mesotypes (GAP case) limits both the sustainable and unsustainable growth regions. Top: Two state model schematic summary (left) and mesotype (right) of the sharp upper boundary (GAP) case without transgenerational feedback. Middle: Contour plot of log relative fitness (blue to green) and number of generations a population of a million survives (red to yellow) as function of mesotypical inflection point of the sharp upper boundary in the fitness landscape, and the noise level (coefficient of variation), assuming a gamma protein expression distribution, with transgenerational feedback (transparent) and without (opaque). The purple dotted line indicates the survival edge. The dashed black line denotes the line $V = 0.5$, along which the plot in the bottom is valid. Bottom: Second derivative of log relative fitness with respect to mesotypical inflection point (along the dashed line of the top axis). Here, it is always negative for values of c_i that support growth, allowing negative epistasis for mutations of the same sign.

M Derivations volume-free two-state model

M.1 Concentration thresholds

M.1.1 Sharp boundaries, with feedback, log-normal approximation

For sharp lower boundaries, equation 4.3 showed the limiting concentration c_i equals 1.5 times the median of the protein production distribution. Assuming a log-normal distribution with parameters μ and σ , the mean (normalized to 1 here) is $e^{\mu+\sigma^2/2}$, so that $\mu = -\sigma^2/2$. Rewriting the variance in terms of the coefficient of variation gives:

$$V^2 * (\text{mean})^2 = V^2 = \text{variance} = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2} = e^{\sigma^2} - 1 \Rightarrow \sigma = \sqrt{\ln(1 + V^2)}$$

Subsequently, the limiting concentration is given by:

$$c_{i,\text{lim},\text{Cdc42}} = \frac{3}{2}e^{\mu} = \frac{3}{2}e^{-\ln(1+V^2)/2} = \frac{3}{2}e^{\ln(1+V^2)^{-1/2}} = \frac{3/2}{\sqrt{1+V^2}} \quad (\text{M.10})$$

which for small V reduces to:

$$c_{i,\text{lim},\text{Cdc42}} \approx \frac{3}{2} - \frac{3}{4}V^2 \quad (\text{M.11})$$

Conversely, for sharp upper boundaries (e.g., for the GAPs), we get:

$$\begin{aligned} \begin{bmatrix} f(2c_i/3) \\ f(0) \end{bmatrix} &= \begin{bmatrix} 0 & 0 \\ 2F_V(2c_i/3) & 2F_V(4c_i/3) \end{bmatrix} \begin{bmatrix} f(2c_i/3) \\ f(0) \end{bmatrix} = M_{\text{GAP},s} \vec{f} \Rightarrow \lambda_{\text{max}} = 2F(4c_i/3) \\ &\Rightarrow \omega_r = \log_2 \lambda_{\text{max}} = \log_2 2F(4c_i/3) \end{aligned} \quad (\text{M.12})$$

with the limit of c_i such that $2F(4c_i/3) > 1$, implying:

$$\Rightarrow c_{i,\text{lim},\text{GAP}} = \frac{3}{4} \cdot \text{median} = \frac{3}{4} \frac{1}{\sqrt{1+V^2}} \approx \frac{3}{4} - \frac{3V^2}{8} \quad (\text{M.13})$$

M.1.2 Sharp boundaries, no feedback, log-normal approximation

The growth factor $\gamma_{\text{Cdc42},-tgf}$ from generation to generation for the Cdc42 case, reverting to state vector $\vec{f}_{eq,ns} = [1 - F(4c_i/3), F(2c_i/3)]$ after every generation, is given by (with ι as a vector of ones):

$$\begin{aligned} \frac{\iota' M_{\text{Cdc42},s} \vec{f}_{eq,ns}}{\iota' \vec{f}_{eq,ns}} &= 2 \frac{(1 - F_V(4c_i/3))(1 - F_V(2c_i/3)) + F_V(2c_i/3)(1 - F_V(4c_i/3))}{1 + F_V(2c_i/3) - F_V(4c_i/3)} \\ &\Rightarrow \gamma_{\text{Cdc42},-tgf} = 2 \frac{1 - F_V(4c_i/3)}{1 + F_V(2c_i/3) - F_V(4c_i/3)} \end{aligned} \quad (\text{M.14})$$

So, to maintain $g_{s-tgf} \geq 1$, then $F(2c_{i,lim}/3) = 1 - F(4c_{i,lim}/3)$. Assuming a log-normal distribution and using the **tanh** approximation for erf [256], we obtain:

In the log-normal case, we have:

$$\begin{aligned} F(c) &= \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\ln c - \mu}{\sqrt{2}\sigma}\right) \approx \frac{1}{2} + \frac{1}{2} \tanh\left(\sqrt{2/\pi\sigma^2} (\ln c - \mu)\right) \quad (\text{M.15}) \\ &= \frac{1}{2} + \frac{1}{2} \tanh\left(\sqrt{2/\pi\sigma^2} \ln(ce^{\sigma^2/2})\right) \end{aligned}$$

For the typically small V , we can simply say $\sigma = V$ (up to third order correct), and defining:

$$A = \sqrt{2/\pi V^2} \quad (\text{M.16})$$

then we obtain:

$$F(c) \approx \frac{1}{2} + \frac{1}{2} \tanh\left(\ln\left((ce^{\sigma^2/2})^A\right)\right) = \frac{(ce^{V^2/2})^{2A}}{(ce^{V^2/2})^{2A} + 1} \quad (\text{M.17})$$

Substituting this into the condition $F(2c_{i,lim}/3) = 1 - F(4c_{i,lim}/3)$ without feedback gives:

$$\begin{aligned} \frac{(2c_{i,lim,-tgf} e^{V^2/2}/3)^{2A}}{(2c_{i,lim,-tgf} e^{V^2/2}/3)^{2A} + 1} &= \frac{1}{(4c_{i,lim,-tgf} e^{V^2/2}/3)^{2A} + 1} \Rightarrow c_{i,lim,-tgf} = \frac{3e^{-V^2/2}}{2\sqrt{2}} \Rightarrow \\ c_{i,lim,-tgf} &\approx \frac{1}{\sqrt{2}} \left(\frac{3}{2} - \frac{3}{4} V^2\right) = \frac{c_{lim,+tgf}}{\sqrt{2}} \quad (\text{M.18}) \end{aligned}$$

Note that this is the same limit as with the GAP case (sharp upper boundary). Analogously:

$$\begin{aligned} \gamma_{GAP,-tgf} &= \frac{l' M_{GAP,s} \vec{f}_{eq,ns}}{l' \vec{f}_{eq,ns}} = 2 \frac{F_V(2c_{i,lim}/3)(1 - F_V(4c_i/3)) + F_V(4c_i/3)F_V(2c_{i,lim}/3)}{1 + F_V(2c_{i,lim}/3) - F_V(4c_i/3)} \\ &\Rightarrow \gamma_{GAP,-tgf} = 2 \frac{F_V(2c_{i,lim}/3)}{1 + F_V(2c_{i,lim}/3) - F_V(4c_i/3)} \quad (\text{M.19}) \end{aligned}$$

This limits $c_{i,lim}$ again at $F(2c_{i,lim}/3) = 1 - F(4c_i/3)$, the same as for Cdc42 (only now c_i must exceed this).

M.2 Epistatic region width approximations

M.2.1 Sharp boundaries, with feedback, log-normal approximation

The epistatic width region w_e has been defined in equation 4.9. Using equation 4.3 to fill in the relative fitness for the last term, we get:

$$1 - F_V(2c_{i,\omega_r=1-\Delta\omega_r}/3) = 2^{-\Delta\omega_r}$$

Assuming log-normal protein production distribution, and recalling M.17, this yields:

$$\begin{aligned} \frac{1}{(2c_{i,\omega_r=1-\Delta\omega_r}e^{V^2/2}/3)^{2A} + 1} &= 2^{-\Delta\omega_r} \Rightarrow (2c_{i,\omega_r=1-\Delta\omega_r}e^{V^2/2}/3)^{2A} = 2^{\Delta\omega_r} - 1 \\ c_{i,\omega_r=1-\Delta\omega_r} &= \frac{3}{2}e^{-\frac{V^2}{2}}(2^{\Delta\omega_r} - 1)^{1/(2A)} \approx \frac{3}{2}e^{-\frac{V^2}{2}}(\Delta\omega_r \ln 2)^{(1/2A)} \\ &\approx \left(\frac{3}{2} - \frac{3}{4}V^2\right) \left(\frac{1}{2A} \ln(\Delta\omega_r \ln 2) + 1\right) \approx \frac{3}{2} - \frac{3}{4}V^2 + \frac{3}{4A} \ln(\Delta\omega_r \ln 2) - \frac{3}{4} \frac{V^2}{2A} \ln(\Delta\omega_r \ln 2) \end{aligned}$$

Combining with the concentration threshold M.10 derived earlier, the epistatic width becomes:

$$\begin{aligned} \Rightarrow w_e &= c_{i,F_V(2c_i/3)=1/2} - c_{i,\omega_r=1-\Delta\omega_r} = -\frac{3}{4A} \ln(\Delta\omega_r \ln 2) + \frac{3V^2}{8A} \ln(\Delta\omega_r \ln 2) \\ &= \frac{3V}{4} \sqrt{\frac{\pi}{2}} \ln\left(\frac{1}{\Delta\omega_r \ln 2}\right) - \frac{3V^3}{8} \sqrt{\frac{\pi}{2}} \ln\left(\frac{1}{\Delta\omega_r \ln 2}\right) \end{aligned}$$

Analogously, we define $w_e = c_{i,\omega_r=1-\Delta\omega_r} - c_{F_V(4c_i/3)=1/2}$ for the GAPS., the first term follows from M.12 and the log-normal approximation:

$$\begin{aligned} F_V(4c_{i,\omega_r=1-\Delta\omega_r}/3) &= 2^{-\Delta\omega_r} \Rightarrow \frac{(4c_{i,\omega_r=1-\Delta\omega_r}e^{V^2/2}/3)^{2A}}{(4c_{i,\omega_r=1-\Delta\omega_r}e^{V^2/2}/3)^{2A} + 1} \approx 2^{-\Delta\omega_r} \approx 1 - \Delta\omega_r \ln 2 \\ &\Rightarrow (\Delta\omega_r \ln 2)(4c_{i,\omega_r=1-\Delta\omega_r}e^{V^2/2}/3)^{2A} = 1 - \Delta\omega_r \ln 2 \\ \Rightarrow c_{i,\omega_r=1-\Delta\omega_r} &= \frac{3}{4}e^{-\frac{V^2}{2}} \left(\frac{1 - \Delta\omega_r \ln 2}{\Delta\omega_r \ln 2}\right)^{\frac{1}{2A}} \approx \frac{3}{4} \left(1 - \frac{V^2}{2}\right) \left(\frac{1}{2A} \ln\left(\frac{1 - \Delta\omega_r \ln 2}{\Delta\omega_r \ln 2}\right) + 1\right) \\ &= \frac{3}{4} - \frac{3V^2}{8} + \frac{3}{8A} \ln\left(\frac{1 - \Delta\omega_r \ln 2}{\Delta\omega_r \ln 2}\right) \end{aligned}$$

Combining with the concentration threshold M.13 derived earlier, we get

$$w_e = \frac{3}{8A} \ln\left(\frac{1 - \Delta\omega_r \ln 2}{\Delta\omega_r \ln 2}\right) \approx \frac{3V}{8} \sqrt{\frac{\pi}{2}} \ln\left(\frac{1 - \Delta\omega_r \ln 2}{\Delta\omega_r \ln 2}\right)$$

M.2.2 Sharp lower boundary, no feedback, log-normal approximation

Similarly, the epistatic width without transgenerational feedback is defined as:

$$w_{e,-tgf} \stackrel{\text{def}}{=} c_{i,F_V(2c_i/3)=1/2,-tgf} - c_{i,\omega_r=1-\Delta\omega_r,-tgf}$$

For low noise (see derivation at relative fitness section, M.21 and threshold M.18):

$$\frac{\omega_{r,-tgf}}{\omega_{r,+tgf}} \approx 1 - \frac{1}{\omega_{r,+tgf}} \left(\frac{4c_i}{3\sqrt{2}} \right)^{3/V} = 1 - \frac{1}{\omega_{r,+tgf}} (c_i/c_{i,lim,-tgf})^{3/V}$$

At low noise, $\omega_{r,+tgf} = 1$ at $c_{i,lim,-tgf}$, so:

$$\begin{aligned} \omega_{r,-tgf} &\approx 1 - (c_i/c_{i,lim,-tgf})^{3/V} \Rightarrow \Delta\omega_r = (c_{i,\omega_r=1-\Delta\omega_r,-tgf}/c_{i,lim,-tgf})^{3/V} \\ &\Rightarrow c_{i,\omega_r=1-\Delta\omega_r,-tgf} = c_{i,lim,-tgf} (\Delta\omega_r)^{V/3} \end{aligned}$$

Combining with the concentration threshold derived earlier, we get

$$w_{e,-tgf} = c_{i,lim,-tgf} \left(1 - (\Delta\omega_r)^{V/3} \right) \approx \frac{V}{3} c_{i,lim,-tgf} \ln \left(\frac{1}{\Delta\omega_r} \right) \approx \frac{V}{4} \sqrt{2} \ln \left(\frac{1}{\Delta\omega_r} \right)$$

Compared to the situation with transgenerational feedback, we get:

$$\frac{w_e}{w_{e,-tgf}} = \frac{\frac{3}{4} V \sqrt{\frac{\pi}{2}} \ln \left(\frac{1}{\Delta\omega_r \ln 2} \right)}{\frac{V}{4} \sqrt{2} \ln \left(\frac{1}{\Delta\omega_r} \right)} = \frac{3}{2} \sqrt{\pi} \left(1 + \frac{\ln(\ln 2)}{\ln(\Delta\omega_r)} \right) \approx 3 \text{ fold}$$

M.3 Number of survivable generations approximations

M.3.1 Sharp lower boundary, with feedback, log-normal approximation

In the main text, we had equation 4.13 for the survivable number of generations for Cdc42 the case. With the log-normal distribution approximation M.17, this yields:

$$n_{gen} = \frac{\log N_0 - \log \left(\frac{(4c_i e^{V^2/2}/3)^{2A} + 1}{(2c_i e^{V^2/2}/3)^{2A} + 1} \right)}{\log \left(\frac{(2c_i e^{V^2/2}/3)^{2A} + 1}{2} \right)}$$

Note that $2c_i e^{V^2/2}/3$ has the same Taylor expansion around small V up to third order as $c_i/c_{i,lim}$. Therefore, we rewrite this term as $1 + (c_i/c_{i,lim} - 1)$ and expand near $c/c_{lim} \approx 1$:

$$\begin{aligned} n_{gen} &\approx \frac{\log N_0 - \log \left(\frac{1 + 2^{2A} + 2A (c_i/c_{i,lim} - 1) 2^{2A}}{2 + 2A (c_i/c_{i,lim} - 1)} \right)}{\log \left(1 + A (c_i/c_{i,lim} - 1) \right)} \\ &\approx \frac{\log N_0 - \log \left(\frac{1 + 2A \left(\frac{c_i}{c_{i,lim}} - 1 \right)}{2 + 2A \left(\frac{c_i}{c_{i,lim}} - 1 \right)} \right)}{A \left(\frac{c_i}{c_{i,lim}} - 1 \right)} \approx \frac{\log N_0 - \log \left(2^{2A} \left(\frac{1}{2} + A \left(\frac{c_i}{c_{i,lim}} - 1 \right) / 2 \right) \right)}{A \left(\frac{c_i}{c_{i,lim}} - 1 \right)} \end{aligned}$$

Also, assuming small V (so large A), then:

$$n_{gen} \approx \frac{\log N_0 - (2A - 1) \log 2 - A \left(\frac{c_i}{c_{i,lim}} - 1 \right)}{A \left(\frac{c_i}{c_{i,lim}} - 1 \right)} \approx \frac{\frac{1}{A} \log N_0 - \log 4}{\frac{c_i}{c_{i,lim}} - 1} - 1 = \frac{V \sqrt{\frac{\pi}{8}} \log N_0 - \log 4}{\frac{c_i}{c_{i,lim}} - 1} - 1$$

M.3.2 Sharp lower boundary, no feedback, log-normal approximation

With feedback, one could derive the number of survivable generations:

$$2(1 - F(4c_i/3))(2 - 2F(2c_i/3))^{n_{gen} - 1} = 2(1 - F(4c_i/3))(\gamma_{Cdc42,+tgf})^{n_{gen} - 1} \approx 1/N_0$$

So now, without feedback, we would have:

$$2(1 - F(4c_i/3))(\gamma_{Cdc42,-tgf})^{n_{gen} - 1} \approx 1/N_0$$

Recall from M.14 we had:

$$\gamma_{Cdc42,-tgf} = 2 \frac{1 - F_V(4c_i/3)}{1 + F_V(2c_i/3) - F_V(4c_i/3)} \Rightarrow$$

$$\gamma_{Cdc42,-tgf} = \frac{\left(1 - F_V\left(\frac{4c_i}{3}\right)\right) \gamma_{Cdc42,+tgf}}{1 - F_V\left(\frac{4c_i}{3}\right) + F_V\left(\frac{2c_i}{3}\right) \left(F_V\left(\frac{4c_i}{3}\right) - F_V\left(\frac{2c_i}{3}\right)\right)} = B \gamma_{Cdc42,+tgf} \quad (\text{M.20})$$

This means that in analogy with the case with feedback:

$$n_{gen} = \frac{\log N_0 - \log \left(B \frac{1 - F(2c_i/3)}{1 - F(4c_i/3)} \right)}{-\log \left(2B(1 - F(2c_i/3)) \right)}$$

$$= \frac{\log N_0 - \log \left(\frac{1 - F(2c_i/3)}{1 - F(4c_i/3)} \right) + \log \left(\frac{1 - F(4c_i/3) + F(2c_i/3)(F(4c_i/3) - F(2c_i/3))}{1 - F(4c_i/3)} \right)}{\log \left(\frac{1}{1 - F(2c_i/3)} \right) - \log 2 + \log \left(\frac{1 - F(4c_i/3) + F(2c_i/3)(F(4c_i/3) - F(2c_i/3))}{1 - F(4c_i/3)} \right)}$$

In this expression the original expression with feedback can be recognized, except with the same added term in the numerator and denominator. Performing the log-normal substitution M.17 on argument B (and for simplification, define $z = \frac{4}{3}c_i e^{V^2/2}$ and assume small V):

$$\frac{1}{B} \approx \frac{1}{\left(\frac{z}{2}\right)^{2A} + 1} + \frac{\left(\frac{z}{2}\right)^{2A}}{\left(\frac{z}{2}\right)^{2A} + 1} \left(\frac{\left(\frac{z}{2}\right)^{2A}}{\left(\frac{z}{2}\right)^{2A} + 1} - \frac{\left(\frac{z}{2}\right)^{2A}}{\left(\frac{z}{2}\right)^{2A} + 1} \right) \approx 1 + \frac{\left(\frac{z}{2}\right)^{4A} (2^{2A} - 1)}{\left(\left(\frac{z}{2}\right)^{2A} + 1\right)^2}$$

$$= 1 + (2^{2A} - 1)(F(2c_i/3))^2$$

This yields:

$$n_{gen,-tgf} = \frac{\log N_0 - \log\left(\frac{1 - F(2c_i/3)}{1 - F(4c_i/3)}\right) + \log\left(1 + \left(2^{\sqrt{8/\pi V^2}} - 1\right)(F(2c_i/3))^2\right)}{\log\left(\frac{1}{1 - F(2c_i/3)}\right) - \log 2 + \log\left(1 + \left(2^{\sqrt{8/\pi V^2}} - 1\right)(F(2c_i/3))^2\right)}$$

M.4 Relative fitness approximations

M.4.1 Sharp lower boundary, no feedback, log-normal approximation

In the main text we had (using the identity in M.20):

$$\frac{\omega_{r,-tgf}}{\omega_{r,+tgf}} = 1 - \frac{1}{\omega_{r,+tgf}} \log\left(\frac{1 - F(4c_i/3) + F(2c_i/3)(F(4c_i/3) - F(2c_i/3))}{1 - F(4c_i/3)}\right)$$

Again, using the log-normal distribution:

$$\begin{aligned} \Rightarrow \frac{\omega_{r,-tgf}}{\omega_{r,+tgf}} &= 1 - \frac{1}{\omega_{r,+tgf}} \log\left(\frac{\frac{1}{\left(\frac{4}{3}c_i e^{\frac{V^2}{2}}\right)^{2A} + 1} + \frac{\left(\frac{2}{3}c_i e^{\frac{V^2}{2}}\right)^{2A}}{\left(\frac{2}{3}c_i e^{\frac{V^2}{2}}\right)^{2A} + 1} \left(\frac{\left(\frac{4}{3}c_i e^{\frac{V^2}{2}}\right)^{2A}}{\left(\frac{4}{3}c_i e^{\frac{V^2}{2}}\right)^{2A} + 1} - \frac{\left(\frac{2}{3}c_i e^{\frac{V^2}{2}}\right)^{2A}}{\left(\frac{2}{3}c_i e^{\frac{V^2}{2}}\right)^{2A} + 1}}{\frac{1}{\left(\frac{4}{3}c_i e^{\frac{V^2}{2}}\right)^{2A} + 1}}\right)} \\ &= 1 - \frac{1}{\omega_{r,+tgf}} \log\left(\frac{1 + 2\left(\frac{2}{3}c_i e^{V^2/2}\right)^{2A} + \left(\frac{2}{3}c_i e^{V^2/2}\right)^{2A} \left(\frac{4}{3}c_i e^{V^2/2}\right)^{2A}}{\left(\left(\frac{2}{3}c_i e^{V^2/2}\right)^{2A} + 1\right)^2}\right) \end{aligned}$$

Given that $c_i < \frac{3}{4}\sqrt{2}$ and $V \ll 1$:

$$\begin{aligned} \frac{\omega_{r,-tgf}}{\omega_{r,+tgf}} &\approx 1 - \frac{1}{\omega_{r,+tgf}} \log\left(1 + \left(\frac{2}{3}c_i\right)^{2A} \left(\frac{4}{3}c_i\right)^{2A}\right) = 1 - \frac{1}{\omega_{r,+tgf}} \log\left(1 + (4c_i/(3\sqrt{2}))^{4A}\right) \\ &\approx 1 - \frac{1}{\omega_{r,+tgf}} (4c_i/(3\sqrt{2}))^{4A} = 1 - \frac{1}{\omega_{r,+tgf}} (4c_i/(3\sqrt{2}))^{\frac{1}{V}\sqrt{\frac{32}{\pi}}} \Rightarrow \\ \frac{\omega_{r,-tgf}}{\omega_{r,+tgf}} &\approx 1 - \frac{1}{\omega_{r,+tgf}} \left(\frac{4c_i}{3\sqrt{2}}\right)^{3/V} = 1 - \frac{1}{\omega_{r,+tgf}} \left(\frac{c_i}{c_{i,lim,-tgf}}\right)^{3/V} \quad (\text{M.21}) \end{aligned}$$

M.4.2 General case, with feedback

From the main text, we had:

$$\lambda_{max} = \frac{tr(M) + \sqrt{tr(M)^2 - 4 \det(M)}}{2}$$

$$\det(M) = g(4c_i/3)(1 - F(2c_i/3))g(2c_i/3)F(4c_i/3) - g(2c_i/3)F(2c_i/3)g(4c_i/3)(1 - F(4c_i/3))$$

This can be approximated by $\det(M) \approx g(4c_i/3)(1 - F(2c_i/3))g(2c_i/3)F(4c_i/3)$ when noise is low. To illustrate this, divide the first term by the second term of the determinant in the log-normal case example (using the aforementioned tanh approximation M.17):

$$\frac{F\left(\frac{2c_i}{3}\right)\left(1 - F\left(\frac{4c_i}{3}\right)\right)}{\left(1 - F\left(\frac{2c_i}{3}\right)\right)F\left(\frac{4c_i}{3}\right)} = \frac{\left(\frac{2}{3}c_i e^{\frac{V^2}{2}}\right)^{2A}}{\left(\frac{2}{3}c_i e^{\frac{V^2}{2}}\right)^{2A} + 1} \frac{1}{\left(\frac{4}{3}c_i e^{\frac{V^2}{2}}\right)^{2A} + 1} = \frac{\left(\frac{2}{3}c_i e^{\frac{V^2}{2}}\right)^{2A}}{\left(\frac{4}{3}c_i e^{\frac{V^2}{2}}\right)^{2A}} = \left(\frac{1}{2}\right)^{2A} \approx 2^{-\frac{1.6}{V}}$$

So, for small V the second term of determinant is negligible.

$$\begin{aligned} \Rightarrow \lambda_{max} &= \frac{tr(M) + \sqrt{tr(M)^2 - 4 \det(M)}}{2} \\ &\approx \frac{g(4c_i/3)(1 - F(2c_i/3)) + g(2c_i/3)F(4c_i/3) + |g(4c_i/3)(1 - F(2c_i/3)) - g(2c_i/3)F(4c_i/3)|}{2} \\ &= \max\left(g(4c_i/3)(1 - F(2c_i/3)), \quad g(2c_i/3)F(4c_i/3)\right) \end{aligned}$$

M.4.3 General case, no feedback

The growth factor from generations to generations for the general case, reverting to state vector $\vec{f}_{eq,ns} = [1 - F(4c_i/3), F(2c_i/3)]$ after every generation, is given by (with ι as a vector of ones):

$$\begin{aligned} \gamma_{-tgf} &= \frac{\iota' M_{general,s} \vec{f}_{eq,ns}}{\iota' \vec{f}_{eq,ns}} = \\ &= \frac{g\left(\frac{4c_i}{3}\right)\left(1 - F\left(\frac{2c_i}{3}\right)\right)\left(1 - F\left(\frac{4c_i}{3}\right)\right) + g\left(\frac{4c_i}{3}\right)\left(1 - F\left(\frac{4c_i}{3}\right)\right)F\left(\frac{2c_i}{3}\right) + g\left(\frac{2c_i}{3}\right)F\left(\frac{2c_i}{3}\right)\left(1 - F\left(\frac{4c_i}{3}\right)\right) + g\left(\frac{2c_i}{3}\right)F\left(\frac{4c_i}{3}\right)F\left(\frac{2c_i}{3}\right)}{1 - F\left(\frac{4c_i}{3}\right) + F\left(\frac{2c_i}{3}\right)} \end{aligned}$$

Simplifying yields:

$$\gamma_{-tgf} = \frac{g(4c_i/3)(1 - F(4c_i/3)) + g(2c_i/3)F(2c_i/3)}{1 - F(4c_i/3) + F(2c_i/3)} \quad (\text{M.22})$$

Then:

$$\omega_{r,-tgf} = \log_2(\gamma_{-tgf}) = \log_2\left(\frac{g(4c_i/3)(1 - F(4c_i/3)) + g(2c_i/3)F(2c_i/3)}{1 - F(4c_i/3) + F(2c_i/3)}\right)$$

M.4.4 General case, feedback never decreases fitness

To prove the transgenerational feedback is not deleterious, we must prove that the generation-by-generation growth factors with feedback is never lower than without. Using the expressions in equations 4.20 and M.22, we must prove:

$$\begin{aligned} \Delta\gamma &\equiv \lambda_{max} - \gamma_{-tgf} \geq 0 \\ \Rightarrow \frac{\text{tr}(M) + \sqrt{\text{tr}(M)^2 - 4 \det(M)}}{2} &\geq \frac{g(4c_i/3)(1 - F(4c_i/3)) + g(2c_i/3)F(2c_i/3)}{1 - F(4c_i/3) + F(2c_i/3)} \\ \Rightarrow \text{tr}(M)^2 - 4 \det(M) - \left(2 \frac{g(4c_i/3)(1 - F(4c_i/3)) + g(2c_i/3)F(2c_i/3)}{1 - F(4c_i/3) + F(2c_i/3)} - \text{tr}(M)\right)^2 &\geq 0 \Rightarrow \\ 0 \leq \left(g\left(\frac{4c_i}{3}\right)\right)^2 \left(1 - F\left(\frac{2c_i}{3}\right)\right)^2 + \left(g\left(\frac{2c_i}{3}\right)\right)^2 \left(F\left(\frac{4c_i}{3}\right)\right)^2 + 2g\left(\frac{4c_i}{3}\right)g\left(\frac{2c_i}{3}\right) \left(2F\left(\frac{2c_i}{3}\right) - F\left(\frac{2c_i}{3}\right)F\left(\frac{4c_i}{3}\right) - F\left(\frac{4c_i}{3}\right)\right) \\ - \frac{\left(2g\left(\frac{4c_i}{3}\right)\left(1 - F\left(\frac{4c_i}{3}\right)\right) + 2g\left(\frac{2c_i}{3}\right)F\left(\frac{2c_i}{3}\right) - \left(1 - F\left(\frac{4c_i}{3}\right) + F\left(\frac{2c_i}{3}\right)\right)\left(g\left(\frac{4c_i}{3}\right) - g\left(\frac{4c_i}{3}\right)F\left(\frac{2c_i}{3}\right) + g\left(\frac{2c_i}{3}\right)F\left(\frac{4c_i}{3}\right)\right)\right)^2}{\left(1 - F\left(\frac{4c_i}{3}\right) + F\left(\frac{2c_i}{3}\right)\right)^2} \end{aligned}$$

Expanding and simplifying leads to proving:

$$\frac{F(2c_i/3)(F(4c_i/3) - F(2c_i/3))(g(4c_i/3) - g(2c_i/3))^2(1 - F(4c_i/3))}{(1 - F(4c_i/3) + F(2c_i/3))^2} \geq 0$$

This is indeed always true, as $F(4c_i/3) > F(2c_i/3)$. As can be seen, the feedback has no effect in flat landscapes, when $g(2c_i/3) = g(4c_i/3)$. When the growth factor is larger than one, this means fitness with feedback is never worse than without feedback. When the growth factor is below one, this reflects that the number of generations that can be survived is never less with feedback than without feedback, as then the growth factor implies the decay of the population per generation.

N Preliminary experiment on Cdc42 distribution statistics

In section 4.4, an experimental design is discussed where from the population distribution of Cdc42, the effects of selection on Cdc42 become apparent. When fitness relies directly on the copy number of the protein, the fitness (and volume) will invariably affect the copy number again as a feedback mechanism. We called this transgenerational feedback, after [41]. However, the flow cytometer experiments discussed in 4.4.1 (see Methods in 2.6.3.2.2) give some preliminary insights in the feasibility of visualizing the transgenerational feedback of Cdc42 selection. The general idea is to compare the Cdc42 distribution to a color channel where there is/are protein(s) not under selection. Here, in red there is only background, i.e. autofluorescent proteins that should not be under selection.

Generally, the correlation between sfGFP and the autofluorescence increases (see Figure 82). As higher sugar content increases volumes (Figure 65) which scales with protein number, correlations increase. Of interest are the relative differences (Table 9, 4.4.1).

Based on the volume effects (measured with the pure WT with non-inducible Cdc42), there should be correlations between the red and green channel for all backgrounds. However, relative to the pure WT, there is less correlation than expected. Particularly the $\Delta bem1$ background is relatively anti-correlated. A survival bias for small cells with higher than expected amount of Cdc42 is a plausible explanation, considering this effect is also stronger in the less healthy $\Delta bem1$ than for the $\Delta bem1 \Delta bem3$. Conversely, this correlation difference should be smaller for fitter cells, having e.g., high Cdc42 levels or the WT background. Indeed, correlation between colors at high galactose are largely identical (within 0.07) and the WT background generally has the smallest differences with the control. This provides some indication that fitness couples back to observed Cdc42 copy numbers. Further implications of the feedback effect on fitness are described in Chapter 4.

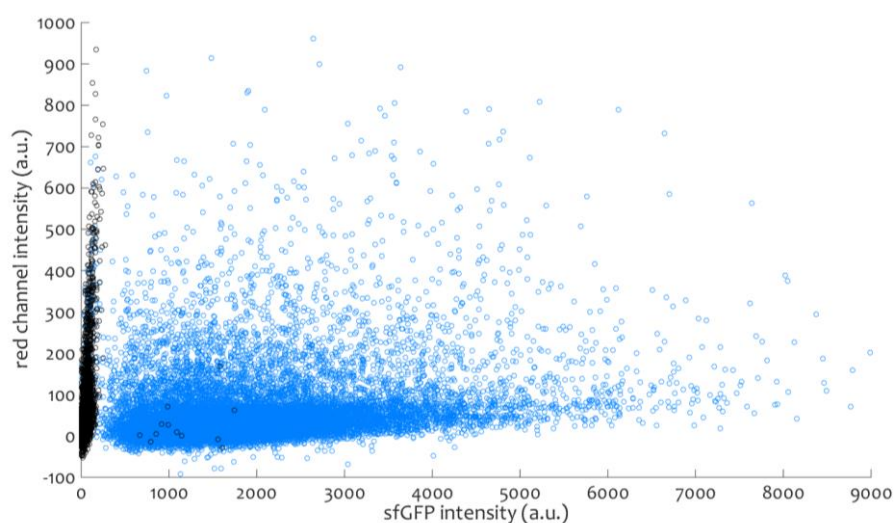


Figure 82 Intensity scatter plot example of the green (sfGFP) and red channel for a GAL1pr-sfGFP-CDC42^{SW} replicate (blue) and pure WT (black). The colors for the black dots are more correlated, while the blue dots are more scattered.

Acknowledgements

With achieving this milestone in my academic education, I would like to highlight the people that have made important contributions. This research project has been an enriching process, that one does not complete alone. Firstly, I thank the members of my committee for their thorough reading and evaluation of my work, in particular my co-promoter Liedewij who in addition gave me the opportunity to conduct this fascinating research and for her involvement in the process. I am grateful for my paranymphs, who have always followed my trajectory with sincere interest. Moreover, my strong appreciation goes out to my colleagues, whose intellectual stimuli as well as their warmth have accompanied me for the past years. Similarly, I owe gratitude to all the students I have supervised over the years, with whom it has been a pleasure to work. Furthermore, I would also like to thank my friends, with whom I happily share my spare time, and my family whose presence is always felt even with a far physical distance. A special mention is for Ms. van Dijk and prof. de Hoop, which have stood at the fundamentals of my academic career. Finally, I would like to strongly express my gratitude to my mother, who has always given me her unconditional love and support, which was unmissable for me and without which I could not have achieved my academic accomplishments.

Curriculum Vitae

Personal information

Family name: Daalman
 Given names: Werner Karl-Gustav
 Date of birth: December 5th, 1988
 Place of birth: Rotterdam, the Netherlands
 Nationality: Dutch
 Gender: Male

Working experience

Dec 2014 – May 2020 PhD student at the Laan Lab, department of Bionanoscience, Delft University of Technology, Delft, the Netherlands

Education

2010-2014 Master's degree Applied Physics at Delft University of Technology, Delft, the Netherlands (grade point average 8.33), thesis "Noise and loss in superconducting resonators", grade 8.5.

2010-2012 Master's degree Quantitative Finance at Erasmus University Rotterdam, Rotterdam, the Netherlands (**cum laude**), thesis "On unspanned factors and estimation of Wishart term structure models", grade 9.

2007-2014 Enrolled as a student in Astronomy at Leiden University, Leiden, the Netherlands

2006-2010 Bachelor's degree Applied Physics at Delft University of Technology, Delft, the Netherlands (**cum laude**), double minor Astronomy and Econometrics

2000-2006 Secondary school (VWO, all four profiles (grade point average 8.53) at Erasmiaans Gymnasium, Rotterdam, the Netherlands



9 789085 934370 >