A theoretical model of Rab proteins localization on endosomal membranes

Lorenzo Forza

July 2025

Abstract

In this thesis we propose a model for the spatial organization of Rab proteins on cell membranes. We are interested in particular in understanding the process of formation of localized domains of active Rab proteins observed in the experiments. This is a process of symmetry breaking, which is crucial in several vital functions, such as directed migration and cell division. Our starting point is the mean-field kinetic model proposed in Bezeljak et al. (PNAS 117:6540, 2020). After a thorough analysis of their hypotheses based on the literature, we extend their model by explicitly introducing the spatial distribution of molecules and their diffusion on the membrane. Numerical simulations of the resulting model are performed and compared to experimental results by Cezanne et al. (eLife 9:e54434, 2020).

Contents

1 Introduction

This paper is mainly inspired by the article [2], in which an artificial membrane is used to study the behaviour of Rab proteins; behaviour we want to reproduce by using the Gillespie algorithm, together with parameters coming directly from experiment fitting done by the authors in [1]. First, a somehow complete version of the model will be provided, from which the simplified model will be derived, justifying approximations and simplifications using results coming from the literature. Space dependence will then be introduced, allowing for the description and study of clusters of proteins and the inclusion of diffusion. Some attention will be devoted to cytosolic diffusion, usually resorting to the infinite diffusivity picture, without giving any explanation. Another section will be aimed at better understanding what the model is supposed to be and to do. Last before the results will be the computer simulation details that includes model choices and works out parameter conversions from reality into computer simulation. The result section will confront experimental data to the results obtained in the simulations and show that domain formation is something that can indeed be obtained by the model proposed.

Important notes for the reader This thesis tries to start general and work out the simplifications that make most sense by following existing studies and reasoning. Some sections have a general (therefore long, hard and biology-specific) subsection followed by the distilled one where simplifications are made. Analytical or computational studies done will use the simplified model as a baseline, unless otherwise stated; meaning that one trying to reproduce the paper's conclusion or to copy the model as a starting point, does not need to read the general parts. General subsections have the name "general" or "generalized" in them.
Section-specific notes are added at the beginning of each section.

2 Chemicals and processes

2.1 Notes

The expressions "in solution" and "free" refer to the cytosol or whatever solvent is present around the endosome, for instance a laboratory prepared solution. "Free" also means that the molecule/complex being talked about is not in complex with something else.

Some of the descriptions do not represent the detailed nature of the chemicals (missing auxiliary proteins, prenylation of Rab, structure modifications of Rab under GDP or GTP binding). An even more detailed explanation of the molecules at play can be found in [14].

The reactions for the model in section 2.5 are taken almost exactly as they are in [1]. There is a couple of differences between the scheme just mentioned and the on discussed in the reaction section, although some kind of conversion is possible (check Appendix C). This is to have continuity with a model already present in literature for which we are already provided with coefficients from data fitting. The two different approaches are equivalent in the case of diffusion and attachment/detachment reactions much faster than biological reactions"

2.2 Chemical species and their functions

- Rab5 (or Rab): a protein that is able to anchor to the membrane and bind to other chemical species with different effects, specified later. In the proposed model they are present only in complex to something else.
- GDP/GTP (guanosine di/tri-phosphate): inactive and active versions of the nucleotides that can attach to Rab; otherwise they are present in solution.
- Rab:GDP/Rab:GTP: complex made of a Rab and a GDP/GTP molecule. Like Rab, they can bind to the membrane. Rab:GDP is the inactive state and Rab:GTP is the active state. The active state binds to the

membrane in a more stable way making it is less susceptible to extraction than the inactive counterpart. Binding of GDP/GTP to Rab is very stable (halftime of about one hour [**Wanquiong**]).

- GEF (guanine nucleotide exchange factor): a dimer protein composed of a Rabex5 and a Rabaptin5 that is found as a tetramer due to the tendency for Rabaptin5 to be in a homodimer state (cited in [2], figures from [4]and [8]). It has one catalytic site on every Rabex5 and one binding site for Rab5:GTP on every Rabaptin5. GEF is in solution unless it binds a Rab5:GTP on one of the Rabaptin5 sites. GEF assembly/recruitment is a bit more complicated but we will exclude details for simplicity.

- GEF:Rab:GDP/GEF:Rab:GTP **(GEF on the left of the complex)**: the catalytic (Rabex5) site of GEF is bound to a Rab protein; this reduces the stability of GDP/GTP bringing it to detach; another GDP/GTP can attach to the empty site, the complete reaction is a nucleotide exchange (hence the acronym for GEF). In the paper we'll treat GEF as a dimer Rabex5:Rabaptin5, with only one catalysis and one binding site.

- Rab:GTP:GEF **(GEF on the right of the complex)**: the binding site of Rabaptin5 in GEF is occupied by a Rab:GTP. The complex is bound to the membrane and can still catalyze nucleotide exchange; since GEF is bound to the membrane already, this makes catalysis faster by means of faster recruitment. The Rab:GTP bound to the Rabaptin binding site doesn't go thought hydrolysis as long as it is in this state.

- GAP and GAP:Rab:GTP (GTPase activating proteins): GAP can bind to Rab:GTP; when it does, it catalyzes GTP hydrolysis into GDP. Rab already hydrolyzes GTP autonomously but at a much slower rate. Free GAP is only present in solution.

- GDI and GDI:Rab:GDP (Rab GDP-dissociation inhibitor): GDI is a protein that can bind to Rab:GDP; it is present only in solution, as well as its complex with Rab:GDP. GDI can extract Rab:GDP from the membrane and prevent it from binding to the membrane again [**Science**] (until the process is inverted)

- GDF (GDI displacement factor): a membrane localized protein that induces the release of Rab:GDP from the GDI:Rab:GDP complex in the cytosol and binds it to the membrane.

- (Rab:GTP:GEF:Rab:GDP: a GEF that has both catalysis and binding sites occupied)

2.3 General version of the reactions

In this subsection, it is reported a more complete version of the reactions that take place in our system with relatively extensive explanations. In here, the

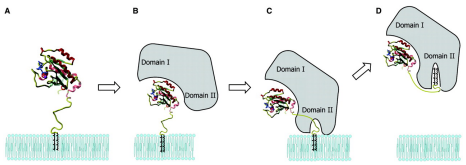

Figure 1: Representation of Rab:GDP extraction performed by GDI. The coloured string is Rab, the grey floatig shape is GDI, the light blue is the membrane.[**Science**]

complete set of reactions is divided into subcategories and simplified, first by condensing into a single reaction processes that either are simultaneous, include proteins that are not considered explicitly in our model, etc. and then by simplifying remaining reactions with approximations from concepts taken from the literature and other coming from knowledge of the system we're using.

2.3.1 GDI and Rab5 membrane binding with GDF

Cytosolic Rab is only present in the complex Rab5:GDP:GDI (see how GDI works in [6] or the scheme from [4]). Membrane localization is aided by other factors; the candidates are GDF (GDI dissociation factor) (see [13], [12], [10] and [11]) and GEF [7]. The thesis will use GDFs. The membrane binding process, as per [5], looks like this:

$$\text{Rab5:GDP:GDI}_{sol}\text{+GDF} \rightarrow \text{Rab5:GDP:GDI:GDF} \rightarrow \text{Rab5:GDP + GDI}_{sol}\text{+GDF} \tag{1}$$

In principle, GDF might have its own membrane binding processes, of the kind (see Yip proteins)

$$\text{GDF}_{sol} \rightleftharpoons \text{GDF} \tag{2}$$

The extraction process, from studies in [6] and [5], doesn't seem to involve anything other than GDI and Rab5:GDP:

$$\text{Rab5:GDP + GDI}_{sol} \rightleftharpoons \text{Rab5:GDP:GDI} \rightarrow \text{Rab5:GDP:GDI}_{sol} \tag{3}$$

to avoid including explicitly GDF into the model, its contribution will be incorporated into the rates by considering it constant and uniform on the membrane. Both reactions can now be written together as:

$$\text{Rab5:GDP:GDI}_{sol} \rightleftharpoons \text{Rab5:GDP + GDI}_{sol} \tag{4}$$

The alternative/additional reaction of Rab membrane binding from papers like [7], that see GEF taking GDF role, can be written as follows:

$$\text{Rab5:GTP:GEF + Rab5:GDP:GDI}_{sol} \rightleftharpoons \text{Rab5:GTP:GEF:Rab5:GDP:GDI}_{sol} \tag{5}$$

$$\text{Rab5:GTP:GEF:Rab5:GDP:GDI} \rightarrow \text{Rab5:GTP:GEF} + \text{Rab5:GDP} + \text{GDI}_{sol} \tag{6}$$

$$\text{or Rab5:GTP:GEF:Rab5:GDP} + \text{GDI}_{sol} \tag{7}$$

Compact version of the reaction looks like:

$$\text{Rab5:GTP:GEF} + \text{Rab5:GDP:GDI}_{sol} \rightarrow \text{Rab5:GTP:GEF} + \text{Rab5:GDP} + \text{GDI}_{sol} \tag{8}$$

This last reaction is not added to the base model.

2.3.2 Rab activation

Reiterating, Rab:GDP is the inactive, weakly bound version of Rab while Rab:GTP is the active and tightly bound one; activation occurs through nucleotide exchange, meaning the extraction of GDP from Rab:GDP and binding of a GTP in its place. The role of GEF in this is to reduce affinity of GDP/GTP to the Rab molecule, catalyzing a detachment of the nucleotide already present, leaving a free site that can bind a free GDP/GTP. The binding of GTP over GDP is favored by the higher concentration of free GTP in the cytosol [9]. Even with this mechanism, Rab is not present on the membrane without GTP/GDP or GEF attached to it [9] and the high speed of the reaction suggests that intermediate states are not of great importance [7]. Other important point to make is that there are two kinds of activation: a slower version that involves membrane bound Rab:GDP to randomly bind to diffusing GEF in solution; and a faster, positive feedback version that sees the GEF anchored to membrane by binding to Rab:GTP.

Non-feedback process Firstly, the non-feedback reaction(s):

$$\text{Rab5:GDP} + \text{GEF}_{sol} \rightleftharpoons \text{GEF:Rab5:GDP} \tag{9}$$

$$\text{GEF:Rab5:GDP} + \text{GTP}_{sol} \rightarrow \text{Rab5:GTP:GEF} + \text{GDP}_{sol} \tag{10}$$

This process is a necessary step towards activation; it is not fast enough to provoke activation by itself but guarantees the presence of some Rab5:GTP on the membrane, essential for the positive feedback. Other details about how this process is actually carried out is not clear so we stick with the model from [1]. Another detail we have evidence for is the distinction between the products of (10) and (11) following [2] we can tell there is immediate handover of the new active Rab5:GTP from the Rabex5 (catalyzer) site to the Rabaptin5 (effector). The version from [1] uses the following instead:

$$\text{GEF:Rab5:GDP} + \text{GTP}_{sol} \rightarrow \text{Rab5:GTP} + \text{GEF}_{sol} + \text{GDP}_{sol} \tag{11}$$

in which GEF is released to the cytosol after the reaction. A simplified version of the complete reaction will then be:

$$\text{Rab5:GDP} + \text{GEF}_{sol} + \text{GTP}_{sol} \rightarrow \text{Rab5:GTP:GEF} + \text{GDP}_{sol} \tag{12}$$

and exclusion of the free guanosines in solution leads to

$$\text{Rab5:GDP} + \text{GEF}_{sol} \rightarrow \text{Rab5:GTP:GEF} \tag{13}$$

Feedback process The feedback reaction consists in the following:

$$\text{Rab5:GTP:GEF} + \text{Rab5:GDP} \rightleftharpoons \text{Rab5:GDP:GEF:Rab5:GTP} \tag{14}$$

$$\text{Rab5:GTP:GEF:Rab5:GDP} \rightarrow \text{Rab5:GTP:GEF} + \text{Rab5:GTP} \tag{15}$$

Again, compacted in a single reaction

$$\text{Rab5:GTP:GEF} + \text{Rab:GDP} \rightarrow \text{Rab5:GTP:GEF} + \text{Rab5:GTP} \tag{16}$$

This reaction has a rate that is much higher than the non-feedback version (as suggested in [1]).
In addition to this, one must remember what specified in reactions (5) and (6), depending on the details of the products.

2.3.3 Rab deactivation (GAP)

With Rab deactivation we refer to the hydrolysis of a phosphate group in GTP into a GDP, deactivating the protein (differently from the GEF case, here the process is not reversible), the reactions (excluding the non important phosphate group) are:

$$\text{Rab5:GTP} \rightarrow \text{Rab5:GDP} \tag{17}$$

and the GAP mediated version [9]

$$\text{Rab5:GTP} + \text{GAP}_{sol} \rightleftharpoons \text{GAP:Rab5:GTP} \tag{18}$$

$$\text{GAP:Rab5:GTP} \rightarrow \text{Rab5:GDP} + \text{GAP}_{sol} \tag{19}$$

The GAP mediated decay is much faster than the spontaneous one. A further simplification of this reaction looks exactly like (17), there will be a difference when looking at the reaction rate that will not be linear in Rab5:GTP concentration but a sigmoid function.

2.3.4 Additional details

In the paper [7], it is claimed that the rate limiting factor for membrane activation comes from the GDI:Rab:GDP dissociation factor (reaction 1 or the left to right reaction in 4). In the same paper is the possibility of GEF to act as a GDF and therefore dissociate the Rab:GDP:GDI complex itself and activate Rab directly. In this case there is also the claim that dissociation becomes essentially irreversible; highlighting how fast GEF mediated activation is when membrane bound, how little affinity there is between Rab5:GTP and GDI and how much time does it take for active Rab to decay into non-active Rab in the absence of GAPs. [4] The structure used for GEF is a simplification. The extended representation of it is Rabex5:Rabaptin5, one has a catalysis site while

the other a binding site for Rab5:GTP. An even more correct representation has to remember that there are 2 Rabex5 and 2 Rabaptin5 in the complex; the number of chemical complexes rises, different stabilities on the membrane, possible directionality (at least for the single GEF) ect. In here the REP molecule is also introduced, it serves as a GDI but only for newly synthesisized Rabs; it functions as a Rab supplier for the membrane but not as an extractor.
In [3], GEF also has the property of preventing hydrolysis of GTP by Rab, excluding the possibility of GEF being membrane bound through GDP at some point. [8] and [4] include structural images regarding the tetramer structure of the GEF complex, plus a justification for the difference in catalysis strength between membrane bound and non-membrane bound GEF: the catalyzer Rabex5 has self inhibiting tails when in solution; membrane binging necessitates a complex formation with the effector Rabaptin5, that creates distance between Rabex5 and the inhibiting tails, boosting catalysis power.

2.4 Legend

To write reaction rates, we make use of alternative names simplify writing. Concentrations will be written by using the alternative names of the chemicals and removing square brackets (for instance [GDI] $\rightarrow$ [I] $\rightarrow$ I) Here are also the names of the reaction coefficients/parameters/concentrations used in the "Reactions" chapter and any other that make use of them.

Rab:GDP $\rightarrow$ D

Rab:GTP $\rightarrow$ T

$\mathrm{GEF}_{sol} \rightarrow \mathrm{E}_{sol}$

Rab:GTP:GEF $\rightarrow$ E

$\mathrm{GDI}_{sol} \rightarrow \mathrm{I}_{sol}$

Rab:GDP:$\mathrm{GDI}_{sol} \rightarrow \mathrm{M}_{sol}$

$\mathrm{GAP}_{sol} \rightarrow \mathrm{A}_{sol}$

R_i: rate of i-th reaction

k_i: multiplicative linear constant of i-th reaction

K_{Hi}, n_i: other coefficients of the i-th reaction, found in the special functions

2.5 Reactions

Reactions will show the reactants and products first, then the reaction rate details. "leftarrow" means the reaction from right to left and "rightarrow" from left to right. For all reactions, first the reaction is provided with "full" names of

the chemicals, then the associated rates will be written using the names written in the legend with the additional convention that concentrations will take only the name of the chemical specie without square brackets ([GDI] $\rightarrow$ [I] $\rightarrow$ I). Alternative versions of rates for the same reaction are explained below.

- GDI_{sol} + Rab:GDP $\rightleftharpoons$ Rab:GDP:GDI_{sol}
  1. $R_1 = k_1 M_{sol} f_1(D,T,E)$ or $\tilde{k}_1 f_1(D,T,E)$ (leftarrow)
  2. $R_2 = k_2 D I_{sol}$ or $\tilde{k}_2 D$ (rightarrow rate)
- GAP_{sol} + Rab:GTP $\rightarrow$ GAP_{sol} + Rab:GDP
  3. $R_3 = k_3 A f_3(T)$ or $\tilde{k}_3 f_3(T)$
- GEF_{sol} + Rab:GDP $\rightarrow$ GEF_{sol} + Rab:GTP
  4. $R_4 = k_4 E_{sol} D$
- GEF_{sol} + Rab:GTP $\rightleftharpoons$ Rab:GTP:GEF
  5. $R_5 = k_5 E_{sol} T$ (rigtharrow)
  6. $R_6 = k_6 E$ (leftarrow)
- Rab:GTP:GEF + Rab:GDP $\rightarrow$ GEF_{sol} + 2Rab:GTP
  7. $R_7 = k_7 E f_7(D)$

Alternative reaction rates The alternative rates (after the "or") mean taking into consideration some aspects that were overlooked to avoid going too much into detail. (The tildes are there to remind that they are not identical one to another, even if clearly related).

R_1 the reaction already factors out GDFs. The complete version would look like this (Michaelis-Menten like; Appendix A): $k_2[GDF]\frac{M_{sol}}{K_{H2}+M_{sol}}$. Assuming [GDF] constant, we find two limit cases: one for very small and one for very big values of M_{sol}, respectively the linear and saturated versions of the reaction, associated then with the first and second versions of the rates. (everything is then incorporated into k_2). Even after considering this, it is possible to linearize the relation at some value of M_{sol}, eg. the initial value.

R_2 if the circumstances allow one to assume I_{sol} doesn't change too much or is constant, we can integrate its contribution into k_1 (useful in high diffusivity, large volume cases or low initial concentrations of M_{sol}).

R_3 In our model we have that the concentration A doesn't change over time, then its contribution is integrated into k_3

Differential equations We can also find the time evolution of the single species; the following equations have to be taken in a coarse grain sense, meaning a portion of space big enough to be able to express concentrations in a somehow continuous way while not big enough to be able to discard localization. Inside one of these regions of space, mean field is (locally) applied

$$\begin{cases} \theta(\frac{\partial I_{sol}}{\partial t} - \Phi_I) = R_1 - R_2 \\ \theta(\frac{\partial M_{sol}}{\partial t} - \Phi_M) = -R_1 + R_2 \\ \theta(\frac{\partial E_{sol}}{\partial t} - \Phi_{E_{sol}}) = -R_5 + R_6 + R_7 \\ \frac{\partial E}{\partial t} = R_5 - R_6 - R_7 + \phi_E \\ \frac{\partial D}{\partial t} = R_1 - R_2 + R_3 - R_4 - R_7 + \phi_D \\ \frac{\partial T}{\partial t} = -R_3 + R_4 - R_5 + R_6 + 2R_7 + \phi_T \end{cases} \tag{20}$$

Making a choice for the model and rewriting explicitly (the tildes are not conserved; what's following will be the first and still pretty general baseline for the thesis), we get

$$\begin{cases} \theta(\frac{\partial I_{sol}}{\partial t} - \Phi_I) = k_1 M_{sol} f_1(D,T,E) - k_2 D I_{sol} \\ \theta(\frac{\partial M_{sol}}{\partial t} - \Phi_M) = -k_1 M_{sol} f_1(D,T,E) + k_2 D I_{sol} \\ \theta(\frac{\partial E_{sol}}{\partial t} - \Phi_{E_{sol}}) = -k_5 E_{sol} D + k_6 E + k_7 E f_7(D) \\ \frac{\partial E}{\partial t} = k_5 E_{sol} D - k_6 E - k_7 E f_7(D) + \phi_E \\ \frac{\partial D}{\partial t} = k_1 M_{sol} f_1(D,T,E) - k_2 D I_{sol} + k_3 f_3(T) - k_4 E_{sol} D - k_7 E f_7(D) + \phi_D \\ \frac{\partial T}{\partial t} = -k_3 f_3(T) - k_5 E_{sol} T + k_6 E + 2k_7 E f_7(D) + \phi_T \end{cases} \tag{21}$$

The factor $\theta = V/S$ has been introduced, explanation regarding why it has to be there is in Appendix B. Essentially it makes it possible to have the correct form of mass conservation while also keeping the reaction dependent solely on the concentrations. That being said, one has to be keep this in mind about the dimensionality and derivation of the coefficients k_i. Other new terms named ϕ `and` Φ have been added. They represent diffusive fluxes of molecules; the capital version Φ is used for when molecules are in the cytosol and the lowercase version ϕ for when they're on the membrane.

Important point now is what functions $f_1,\ f_3$ *and* f_7 are; a more detailed/precise derivation of the model can lead to the following (explained in Appendix A):

$$f_1(D,T,E) = \tilde{f}_1(D+T+E) = 1 - \frac{D+T+E}{K_{H1}} \tag{22}$$

$$f_3(T) = \frac{T^{n_3}}{K_{H3}^{n_3} + T^{n_3}} = \frac{T}{K_{H3} + T} \tag{23}$$

$$f_7(D) = \frac{D^{n_7}}{K_{H7}^{n_7} + D^{n_7}} \tag{24}$$

But a more basic approach for a first study can settle for something simpler:

$$f_1(D, T, E) = 1 \tag{25}$$

$$f_3(T) = T \tag{26}$$

$$f_7(D) = \frac{D^{n_7}}{K_{H7}^{n_7} + D^{n_7}} \tag{27}$$

Parameters would need to be determined experimentally but starting values of $n_3 = 1$ and $n_7 = 2$ can be justified (Appendix A). The 1 value is a safer assumption due to the more straightforward essence of the reaction; for the other value we can only confidently assign $1 \leq n_7 \leq 2$ if we exclude significant changes in GEF depending on the number of binds it makes and avoiding possible bigger compounds that can see an even bigger number of binding sites (oligomerization, beyond the scope of the thesis, would give a $n_7 > 2$). Volume exclusion in the more complex version of f_1 is to prevent impossible concentrations of Rab on the membrane, also eliminating possible blow-up effects that may arise in the study.

2.6 Diffusion

All chemical species are, in principle, capable of moving diffusively following the equation $\frac{\partial}{\partial t}A = D_A \Delta A$ ($\Delta = \nabla^2$, the laplacian). For particles bound to the membrane, the explicit implementation has to take into account the curvature of the membrane.
For the particles that are found in solution, it is customary to disregard the process completely. Since diffusion in solution is faster than all other processes, one usually takes the high diffusion limit, in which all chemical species in solution reach homogeneous concentration before any other reaction takes place (and therefore one needs only one, mean value for all particles in solution). Study and implementation of diffusion will take place in later sections and will make use of the diffusive length $l_D = \sqrt{Dt}$ or $l_D = \sqrt{2Dt}$ coming from the variance of end to end distance in random walks. The later mentions of diffusion will use diffusive length to prove that infinite diffusivity of the cytosol is a valid approximation for the system at hand; and will find the parameters that have to be introduced into the computer simulation as a conversion from real life values.

3 Model variants and objects of interest

While analysis can be carried out in different ways, with different ends and methods, here are some points about legitimacy of the model and region of biological significance.

3.1 Time horizon

Especially when working with time evolving systems, statistical physics is used to look at stationary solutions and draw phase space diagrams about the different kinds of stabilities in relation to some of the parameters of the system. This essentially refers to an infinite time horizon study which is not too suiting for biological phenomena. Experiments regarding Rab5 activation have running times ranging from about ten minutes, up to a couple of hours; early endosomes exist in that state for about 10 minutes and paper [2], which this thesis proposes to reproduce the results of, runs experiments for 15 minutes. Whatever study is being made has to consider that, after these time spans, arising phenomena are not actually of immediate biological interest.

3.2 Space and chemical species representation

Tracking the amount of chemicals present in a given portion of space is a basic component to decide what reactions can take place and at what rate. A first distinction is between cytosol and membrane, the first a 3D, high diffusion region while the second is a 2D, possibly curved, lower diffusion region. Possibilities for space analysis include:

- Mean field approach: assumption used when cytosol diffusion is considered much faster than any other process. Between any two reactions, the chemicals in solution have the time to diffuse to their stationary distribution, an uniform distribution. In this approach, therefore, the cytosol is considered uniform and there is no space dependence.

- Coarse graining: another field approach that keeps space elements to it. There are different ways to implement it but a rule of thumb is taking a limited portion of space and treating that in a mean field way. A key difference with the mean field approach is that different portions of space now communicate in a non trivial way; mechanisms such as diffusion, which was lost in mean field, in now present again. A different way to look at the approach is, for any point in space, define the coarse grained concentration as the average concentration in some volume around the point. The difference between the two approaches is that the first creates a lattice while the second is continuous.

- One-to-one association of space site to chemical: every space site/lattice cell is able to host a single chemical of the ones present in the model. Reactions now have to be computed across different cell sites.

More nuanced versions of cytosolic diffusion without creation of a lattice can be implemented, for instance the definition of a "far" and a "near" region; even if theoretically correct, one should remember that real experiments use a real volume that doesn't have to be exceeded, for example by defining the volume of the simulation as that given by the diffusion length, which grows over time.

4 First analysis

In order to find useful parameters, relations, behaviours etc. we start from an ulterior simplification of the model and try to get to more general cases.

4.1 Cytosolic Diffusion

The way chemical species are treated when in solution will be different for membrane bound due to an important characteristic of chemicals in solution: they move really fast compared to the membrane bound one. In particular if we defined cytosolic diffusion as a reaction, the reaction rate would be much bigger than any biological reaction's. The basic version of the equation related to diffusion is

$$\frac{\partial f}{\partial t} = D\nabla^2 f \tag{28}$$

Other contributions such as convection, spatial dependence of the diffusivity, creation and destruction can be added to get the convection-diffusion equation

$$\frac{\partial f}{\partial t} = \nabla(D\nabla f) - \nabla(\vec{v}f) + R \tag{29}$$

The most general version of the reaction we're interested in is just a diffusion equation with a reaction term, so

$$\frac{\partial f}{\partial t} = D\nabla^2 f - c \tag{30}$$

Stationary concentration profile in spherical symmetry To study the endosome this is a fair starting point: reactions can occur only on its membrane while outside the chemicals just diffuse. The stationary solution tells us what concentrations should we expect near the membrane when activation is going on: locally, near the endosome, we would see a depletion of Rab5:GDP:GDI and a production of free GDI; same concept for GEF. While the stationary solution is not necessarily the one actually found, we can expect that profile to be approached and reached in some characteristic time.
To do spherical symmetry, we write the laplacian explicitly (only the radial component, given the symmetry):

$$(\nabla^2 f(r))_r = \frac{1}{r^{d-1}}\frac{d}{dr}(r^{d-1}f(r)) \tag{31}$$

The stationary profiles are:

1D $f(x) = \frac{c}{2D}x^2 + Ax + B$

2D $f(r) = A\log r + \frac{cr^2}{4D} + B$

3D $f(r) = \frac{A}{r} + \frac{cr^2}{6D} + B$

where A and B have to be determined by boundary conditions. Dimensions smaller than 3 cannot be really said to be stationary without other assumptions (both diverge at infinity). We also specify that reactions occur only on the membrane (now there are two domains). So in the cytosol, the basic diffusion equation is used, while for the membrane we need also the creation/destruction term. We're going to proceed only with the 3D version and pose as boundary conditions:

- The concentrations far form the membrane is some fixed value c_∞
- The first derivative of the concentration near the membrane is dictated by the membrane (either with differentiability or some variant of it)

We can immediately find B: $f(\infty) = c_\infty = B$. To determine A, we first report the values of the first derivative near the membrane (r_0^- and r_0^+ are measures taken on the membrane from inside/outside respectively):

$$f'(r)|_{r=r_0^+} = -\frac{A}{r_0^2} \tag{32}$$

$$f'(r)|_{r=r_0^-} = \frac{r_0 c}{3D} \tag{33}$$

A relation between the two has to be established in order to get a value for A, a simple one is differentiability ($f'(r)|_{r=r_0^+} = f'(r)|_{r=r_0^-}$) that gives $A = -\frac{r_0^3 c}{3D}$, and the complete formula

$$f(r) = c_\infty - \frac{r_0^2 c}{3D}\frac{r_0}{r} \tag{34}$$

with this we can near membrane concentration: $f(r_0) = c_0 = c_\infty - \frac{r_0^2 c}{3D}$.

Diffusion lenght and volume Especially for mean field approaches, where usually one considers the cytosolic concentrations to be the average over the entire volume (well mixed solution), there can be the case in which the volumes are considered infinite (leading to the infinite reservoire case), meaning that nothing that happens on the membrane can modify the reservoire eg. by means of depletion. While this picture may be useful or justified in some cases, some considerations about diffusivity can give a similar approach that is more realistic. Dimensional analysis tells us that, for diffusion

$$< x^2 > \propto Dt \tag{35}$$

we define the standard deviation of the position $l_D = \sqrt{Dt}$, the diffusive length. This, aside from some constant, is the typical distance of a diffusing particle from its starting point. We say that a particle is able to react on the membrane only if it is able to reach the membrane, the diffusive length gives us a criterion for the maximum distance a particle can be at initial time to be able to reach the membrane.

With these considerations, we can study some basics of the factor $\theta = V/S$, the ratio between the volume of the cytosol and the surface area of the membrane.

The physical meaning of θ is found inside the reactions. It is the ratio between changes in membrane and cytosol concentration when the same number of molecules is added. In particular big values of θ mean that cytosol concentrations are almost unaffected by events occurring on the membrane (eg. association or dissociation). Details are in the Appendix B, while here we provide a simple example.

Example For a simple membrane association/dissociation reaction, the count (denoted by the superscript "n") of particles will follow the relation

$$\begin{cases} \frac{\partial A^n}{\partial t} = k_a A^n_{sol} - k_d A^n \\ \frac{\partial A_{sol}}{\partial t} = -k_a A^n_{sol} + k_d A^n \end{cases} \tag{36}$$

in which one can find the particle conservation relation

$$\frac{\partial (A^n + A^n_{sol})}{\partial t} = 0 \tag{37}$$

In order to switch to concentrations, one has to remember that they are written as:

$$A = \frac{A^n}{S}, \quad A_{sol} = \frac{A^n_{sol}}{V} \tag{38}$$

with the notation A = [A] meaning concentrations. By just making a substitution one gets a new writing for the mass action law:

$$\begin{cases} \frac{\partial SA}{\partial t} = k_a V A_{sol} - k_d SA \\ \frac{\partial V A_{sol}}{\partial t} = -k_a V A_{sol} + k_d SA \end{cases} \tag{39}$$

$$\begin{cases} \frac{\partial A}{\partial t} = \theta k_a A_{sol} - k_d A \\ \frac{\partial A_{sol}}{\partial t} = k_a A_{sol} - \theta^{-1} k_d A \end{cases} \tag{40}$$

the relation

$$\frac{\partial (A + A_{sol})}{\partial t} = 0 \tag{41}$$

is not valid, since it became

$$\frac{\partial ([A] + \theta [A_{sol}])}{\partial t} = 0 \tag{42}$$

In which one can see that, for big values of theta, if one changes A by some amount, they will need to change A_{sol} only slightly to satisfy the particle conservation relation.

Caution: by looking at the mass action laws now, it looks like, for big θ,

the A terms should be negligible and we would see an increase of A up to exorbitant values before an equilibrium can be reached. This is an artifact of how the model was introduced: by using the rates k_a, k_d taken constant from the mass action law with particle counts instead of the opposite; those coefficients are case dependent (phenomenological). A correct writing of the mass action law should include time derivatives of the counts of particles and reaction rates that are concentration dependent.
End of example

The calculations are made for spherical geometry; then numerical values are given as examples. The endosome is a sphere of radius r inside the cytosol, another sphere of radius $\sqrt{Dt}+r$. The volume of the sphere is $V = \frac{4}{3}\pi[(\sqrt{Dt}+r)^3 - r^3]$ and the area of the endosome is $S = 4\pi r^2$.

$$\theta = \frac{V}{S} = \frac{r}{3}((\frac{\sqrt{Dt}}{r}+1)^3 - 1) \tag{43}$$

$$\approx \frac{(Dt)^{3/2}}{3r^2} \text{ for } r \ll \sqrt{Dt} \tag{44}$$

Typical diameter of an endosome is between $(0.1-1)\mu m$; the paper [2] uses beads of diameter of $10\mu m$. Typical values of the diffusion in cytosol range from $(10-100)\mu m^2/s$, depending on protein molecular weight; the proteins of the Rab model have molecular weight ranging in$(25-100)kDa$; in principle spanning over all values of D (smaller proteins use the $100\mu m^2/s$ value, bigger ones the $10\mu m^2/s$). Time is the trickier part to correctly implement in the model; in the simple case of mean field approaches we can simply take the time of the experiment; we use the values $t = [15s, 15min, 120min]$.
So for $D = 100\mu m^2/s$ (θ is expressed in μm)

$\theta(\mu m)$	t	15s	15 min	120 min
r(μm)				
0.1		$2 \cdot 10^6$	$9 \cdot 10^8$	$2 \cdot 10^{10}$
1		$2 \cdot 10^4$	$9 \cdot 10^6$	$2 \cdot 10^8$
10		$4 \cdot 10^2$	$1 \cdot 10^6$	$2 \cdot 10^6$

and for $D = 10\mu m^2/s$

$\theta(\mu m)$	t	15s	15 min	120 min
r(μm)				
0.1		$6 \cdot 10^4$	$3 \cdot 10^7$	$6 \cdot 10^8$
1		$6 \cdot 10^2$	$3 \cdot 10^5$	$7 \cdot 10^6$
10		$6 \cdot 10^0$	$4 \cdot 10^4$	$7 \cdot 10^5$

In a problem with diffusion one must take, as the volume of the system, the smaller between cytosolic and diffusive volume; for the obvious fact that a particle cannot come from a place where the cytosol is not present.

Well mixed cytosol To assume well mixed state of the cytosol, the use of diffusive length can be again useful since it defines a volume in which a previously localized particle can be now realistically anywhere inside that volume. If the volume is that of the cytosol and its associated time is smaller to the depletion times of the interested chemical species, then that is a good condition for well mixing. In the realistic biological case, where cells have a diameter of 10-100 μm with an average of about $25\mu m$, the existence of other intracellular objects that reduce the effective volume of the cell (eg. the nucleus). Supply of new proteins and presence of numerous endosomes is also something that may prove of importance. Some values of diffusive length are in the table below, it can be seen that well mixing can be inferred if fast diffusing molecules deplete with characteristic times bigger than 15s, for slow diffusive particles the time scale stands between 15s and 15 min. Most of the molecules in the model diffuse fast with the exception of maybe GEF. From [1], a depletion time can be extracted by observing that the fastest growing activation curves reach half their saturation value in 13-16 minutes. The analysis just made sees mean field cytosolic concentrations a justified approach for Rab5:GDP:GDI but not necessarilly for GEF.

$\sqrt{Dt}(\mu m)$	t	15s	15 min	120 min
D($\mu m^2/s$)				
10(slow)		$1 \cdot 10^1$	$9 \cdot 10^1$	$3 \cdot 10^2$
100(fast)		$4 \cdot 10^1$	$3 \cdot 10^2$	$8 \cdot 10^2$

5 Computer simulation details

The basis for the simulation is the Gillespie algorithm, in which all processes described in the previous chapter are present with their relative rates described in a following section.

5.1 Notes

In this section, over/sub-scripts are used to indicate usually different things, read the legends before/after the equations/definitions to not get lost in possible duplicates for the same notation or different nomenclature for the same object, especially if they're not part of the same discourse.

This section includes also theoric parts for equivalences/transformations/passages from model to model, representation to representation, etc. While this doesn't work towards the end goals of the thesis, it provides a basis from which the reader can easily understand where their own model lies in this description, and therefore how to interpret the results here related to their own.

The part about diffusion implementation uses $l_D = \sqrt{\mathcal{D}t}$ and $l_D = d_c$ while it's arguable that $l_D = \sqrt{2\mathcal{D}t}$ and/or $l_D = 2d_c$ should be used instead. The lack of specification around this point stems from the fact that theoretic values of

biological rates, even in the case of diffusion, are not of easy access and when they are, are usually not very precise.

5.2 Basics of the model

The computer model is based on the Gillespie algorithm: reaction rates are defined, computed for the state of the system at a given time; the first reaction that occurs is decided stochastically and the state is updated accordingly along with the reaction rates. The process is repeated as many times as one desires. Some of the choices around the details of the system are written below.

5.2.1 Molecule and membrane representation

To state the amount of a certain kind molecule in a certain region of space; the way adopted involves associating to a certain chemical specie, an integer number of simulation particles, each representing a number of real molecules, or equivalently, some fraction of a mole. The relation between real and simulation particles is given by a conversion rate discussed more later. Naturally the coefficients of the model will have to be adjusted to take this into account. Given that the membrane will be some lattice, our choice of model will allow every element of the lattice (a cell) to contain as many molecules as it wants; this corresponds to a coarse grained representation for the single cell and justifies defining reaction rates inside the single cells without the need for inter-cell computations.

5.2.2 Membrane

The membrane is modeled as a plain, square-like surface with hexagonal lattice. Since it is a Bravais lattice, processes do not need particular specifications for directionality other than a uniform randomization for diffusion. This model for the membrane will have to represent a real membrane, therefore its total surface will be that of the endosome: S. The total surface will be divided into N_c cells, each with a surface area $S_c = S/N_c$. Each of the hexes can host any number of simulation particles.

5.2.3 Cytosol

In the cytosol, no processes other than diffusion take place. Approaches for including diffusion details can be are numerous and depending on the context of the problem some can be preferred to others; some alternatives:

- 3-Dimensional fine lattice, with cell size comparable to that used for the membrane.
- 3-Dimensional coarse lattice: the use of the high diffusion limit in a less drastic way allows one to define a region of space in the cytosol in which an homogeneous concentration is present; smaller that the whole space, bigger than the membrane lattice elements.

- Cytosol mediated membrane to membrane diffution: a different approach, under some conditions, would try to find where a particle that detaches from the membrane would bind again after cytosolic diffusion.
- Radial distribution of concentration that depends on the rate of absorption of the membrane.
- "far" and "near" regions: depending on the rates of depletion of the chemical species in some regions of the cytosol, especially near the membrane, one can define a "near" region, that has to be simulated, since change happens fast in it; and a "far" region that is unaffected by the rest of the system or alternatively studied in a less precise way

Following what was discussed in the section about cytosolic diffusion, we use a mean field approach for the cytosol since the diffusion lenght in the characteristic time of depletion of the chemical species in solution is comparable and bigger than the cell diameter.

5.3 Reaction rates

The reaction rates will be a conversion of what is used in the theory (i.e. the entries of the Reaction section), parameters that we find a fit for in [1].

5.3.1 Experimental and computer simulation rates

Starting with how conversion should be carried out, a term from the reaction section is taken as an example. The first step is that of passing from an equation that uses concentrations (M = mol/V or mol/S) to one that uses molecule counts (expressed in mol). (Notice that θ gets integrated in the coefficients at a certain point)

$$\theta \frac{\partial M_{sol}}{\partial t} = -k_1 M_{sol} + k_2 D I_{sol} \tag{45}$$

$$\frac{\theta}{V} \frac{\partial M^n_{sol}}{\partial t} = -\frac{k_1}{V}\ M^n_{sol} + \frac{k_2}{SV} D^n I^n_{sol} \tag{46}$$

$$\theta \frac{\partial M^n_{sol}}{\partial t} = -k_1 M^n_{sol} + \frac{k_2}{S} D^n I^n_{sol} \tag{47}$$

$$\frac{\partial M^n_{sol}}{\partial t} = -k^n_1 M^n_{sol} + k^n_2 D^n I^n_{sol} \tag{48}$$

Where we defined M = [M] = concentrations, M^n = counts of M in mol and we will use M^s = counts of M in the computer simulation (and analogous for all other species).

The passage from molecules in mol to simulation particles involves a conversion rate (cr) in this way:

$$cr * M^s = M^n = M * V \tag{49}$$

meaning that for each simulation molecule we have, in reality, cr mols of molecules (in [1], one simulation particle corresponds to 1 nM, here cr = $10^{-9}V$ where V is not specified but simulations suggest a likely value is of 100 μm^3). For $cr = 1/N_A$ the number of simulation particles would be the number of molecules in the system (simply substitute the value and invert the relations to find that this is the case). This parameter has to respect $1/M^n < cr < 1/N_A$; the lower limit corresponds to condensing all the molecules in the system into a singular simulation particle (essentially becomes an ON/OFF state) the higher limit, as anticipated, simulates with a number of particles equal to the number of molecules in the system. The higher limit is not actually necessary but since we would use more particles than there are in reality this will underestimate stochastic effects compared to reality. A smaller value of cr makes the computation faster but less representative of the system.
After this definition, a simple substitution yields:

$$(cr)\frac{\partial M^s_{sol}}{\partial t} = -k^n_1(cr)M^s_{sol} + k^n_2(cr)^2 D^s I^s_{sol} \tag{50}$$

$$\frac{\partial M^s_{sol}}{\partial t} = -k^n_1 M^s_{sol} + k^n_2(cr) D^s I^s_{sol} \tag{51}$$

$$\frac{\partial M^s_{sol}}{\partial t} = -k^s_1 M^s_{sol} + k^s_2 D^s I^s_{sol} \tag{52}$$

Other parameters that vary are the K_{Hi}, with the following passages:

$$f_3(T) = \frac{T}{K_{H3} + T} \tag{53}$$

$$= \frac{ST^n}{K_{H3} + ST^n} = \frac{T^n}{K^n_{H3} + T^n} \tag{54}$$

$$= \frac{(cv)T^s}{K^n_{H3} + (cv)T^s} = \frac{T^s}{K^s_{H3} + T^s} \tag{55}$$

Repeating the same process for all other parameters, we get to Table 1.

5.3.2 Coefficient renormalization under lattice representation

For spatial representation of the membrane, a lattice is added. Periodic boundary conditions are used and we define N_c as the number of cells that make up the lattice.
The aim of this part is to find a way to keep the same time scales for the reactions, but also to consider what parameters would be physically likely in the passage from a mean field picture to a coarse grained one. Here, "i" will be used to identify cells ; it is the cell number. The reaction number is denoted by "α". k_α, k^s_α, $K_{H\alpha}$, $K^s_{H\alpha}$ will keep their meaning of mean-field parameters. Let, then k_{α,N_c}, k^s_{α,N_c}, $K_{H\alpha,N_c}$, K^s_{α,N_c} be their corresponding value when the membrane is divided into N_c coarse grained cells. Isotropy is assumed for the coefficients.

i	k_i^n/k_i	k_i^s/k_i^n	k_i^s/k_i
1	$1/\theta$	1	$1/\theta$
2	1/V	cr	cr/V
3	1	1	1
4	1/V	cr	cr/V
5	1/V	cr	cr/V
6	1	1	1
7	1	1	1

i	K_{Hi}^n/K_{Hi}	K_{Hi}^s/K_{Hi}^n	K_{Hi}^s/K_{Hi}
1	1/S	1/cr	1/S(cr)
3	1/S	1/cr	1/S(cr)
7	1/S	1/cr	1/S(cr)

Table 1: Conversion between values of the linear coefficients of the reaction rates depending on the context. k_i denote the theoretical values (they use molarity for the chemical species), k_i^n assume the geometry of the problem is given and uses molecules counts expressed in mol, k_i^s are the computer simulation coefficients: they assume that every particle in the simulation represents cr (conversion rate) real molecules (expressed in mol). The same thing is applied to the K_{Hi} coefficients

Now the same reactions can happen in any of the cells; but, importantly, the rate at which they do is now local:

$$\begin{aligned} R_\alpha &\to R_{\alpha,i} & R_\alpha^s &\to R_{\alpha,i}^s \\ R_\alpha &= \sum_{i=1}^{N_c} R_{\alpha,i} & R_\alpha^s &= \sum_{i=1}^{N_c} R_{\alpha,i}^s \\ R_\alpha &= k_\alpha * g_\alpha(C) & R_\alpha^s &= k_\alpha^s * g_\alpha^s(C^s) \\ R_{\alpha,i} &= k_{\alpha,i} * g_{\alpha,N_c}(C_i) & R_{\alpha,i}^s &= k_{\alpha,i}^s * g_{\alpha,N_c}^s(C_i^s) \end{aligned}$$

On the left the relation for the concentration representation and on the right, the one for the simulation. $g_\alpha(C)$ identifies the function of the rates that exclude the linear coefficients k_α (and analogous for all the variants); C represents the chemical species.
It's important to notice that the approach on the left uses concentrations while the one on the right uses counts. The concentration approach has now to consider that membrane concentrations are now local and that they use a different surface area from before:

$$\begin{aligned} C &= C^n/S \\ C_i &= C_i^n/S_c = N_c * C_i^n/S \end{aligned} \tag{56}$$

For the simulation approach, the number of particles is the same for both cases, it only changes the fact that now localization is possible:

$$\begin{aligned} (cr)C^s &= C^n \\ (cr)C_i^s &= C_i^n \end{aligned} \tag{57}$$

After this one can also explicitly write the following relations for the sum over the sites of the chemical species with their specific

$$\begin{aligned}
&\sum_{i=1}^{N_c} C_i = 1/S_c \sum_{i=1}^{N_c} C_i^n = C^n/S_c = N_c C^n/S = N_c C \quad \text{(concentrations)} \\
&\sum_{i=1}^{N_c} C_i^s = C^s \quad \text{(simulation particles)}
\end{aligned} \tag{58}$$

The following step is to write the equations for reaction rates and solving them. Taking two as example, for concentrations we get

$$\begin{aligned}
R_1 &= k_1 M_{sol} \\
R_{1,i} &= k_{1,N_c} M_{sol} \\
\sum_{i=1}^{N_c} R_{1,i} &= \sum_{i=1}^{N_c} k_{1,N_c} M_{sol} = N_c k_{1,N_c} M_{sol} = R_1 \\
k_{1,N_c} &= k_1/N_c
\end{aligned} \tag{59}$$

$$\begin{aligned}
R_2 &= k_2 D I_{sol} \\
R_{2,i} &= k_{2,N_c} D_i I_{sol} \\
\sum_{i=1}^{N_c} R_{2,i} &= N_c k_{2,N_c} D I_{sol} = R_2 \\
k_{2,N_c} &= k_2/N_c
\end{aligned} \tag{60}$$

and for simulation particles

$$\begin{aligned}
R_1^s &= k_1^s M_{sol}^s \\
R_{1,i}^s &= k_{1,N_c}^s M_{sol}^s \\
\sum_{i=1}^{N_c} R_{1,i}^s &= \sum_{i=1}^{N_c} k_{1,N_c}^s M_{sol}^s = N_c k_{1,N_c}^s M_{sol}^s = R_1^s \\
k_{1,N_c}^s &= k_1^s/N_c
\end{aligned} \tag{61}$$

$$\begin{aligned}
R_2^s &= k_2^s D^s I_{sol}^s \\
R_{2,i}^s &= k_{2,N_c}^s D_i^s I_{sol}^s \\
\sum_{i=1}^{N_c} R_{2,i}^s &= k_{2,N_c}^s I_{sol}^s \sum_{i=1}^{N_c} D_i^s = k_{2,N_c}^s D^s I_{sol}^s = R_2^s \\
k_{2,N_c}^s &= k_2^s
\end{aligned} \tag{62}$$

Repeating the same steps for all R_αs, we get

α	$k_{\alpha,N_c}/k_\alpha$	$k^s_{\alpha,N_c}/k^s_\alpha$
1	$1/N_c$	$1/N_c$
2	$1/N_c$	1
3*	$1/N_c$	1
4	$1/N_c$	1
5	$1/N_c$	1
6	$1/N_c$	1
7*	$1/N_c$	1

α	$K_{H\alpha,N_c}/K_{H\alpha}$	$K^s_{H\alpha,N_c}/K^s_{H\alpha}$
1	1	$1/N_c$
3	1	$1/N_c$
7	1	$1/N_c$

Table 2: Reaction rate coefficients conversion after changes in membrane representation from mean field to coarse grained with N_c cells. Coefficients for indices 3 and 7 require a more specific study due to the presence of the functions $f_3(T)$ and $f_7(D)$ and the reaction of two membrane species (E_iD_i) for which the sum $\sum_i E_iD_i$ is profile dependent. $K_{H\alpha}$ coefficients should also get better study

5.3.3 Diffusion

As anticipated, diffusion in the cytosol is simply integrated in the model by using a mean field approach for chemical species in solution. About the membrane diffusion, it can only be defined in the simulation after some geometric parameters are added. The diffusivity of every membrane bound molecule is denoted by $\mathcal{D}$, which becomes $\mathcal{D}_D, \mathcal{D}_T, \mathcal{D}_E$ if we specify the chemical. The simulated membrane is supposed to represent the entirety of the real membrane. In Gillespie, diffusion has to be implemented defining a diffusive event. A diffusive event is defined as removing the interested particle from its starting cell and the addition of the same particle in an adjacent cell. To find the reaction rate of such an event, use the following notation ("cell" refers to the lattice element, not the eukaryotic cell): d_c = cell distance (in μm), N_c = cell number, S_c = area of the cell (μm^2, S = total area (μm^2), $\mathcal{D}_{N_c}$ = Gillespie diffusivity (adimensional). Now consider the following points:

- Diffusion length takes the value $l_D = \sqrt{\mathcal{D}t}$ (or the same value multiplied $\sqrt{2}$ with a different approach)
- d_c can be evicted by the formula of the area of an exagon: $A_c = \frac{\sqrt{3}}{2} d_c^2 \rightarrow d_c = \sqrt{\frac{2A}{\sqrt{3}N_c}}$
- a diffusion event will happen then the diffusive length,in a probabilistic sense, reaches the distance of two cells: $l_D = d_c$
- the diffusion rate is $k_D = 1/\tau_D$ where τ_D is the characteristic time of the reaction

By substituting one finds:

$$\begin{aligned}\sqrt{\mathcal{D}\tau_D} &= \sqrt{\frac{2A}{\sqrt{3}N_c}} \\ \tau_D &= \frac{2A}{\sqrt{3}N_C\mathcal{D}} = 1/k_D \\ k_D &= N_c\frac{\mathcal{D}}{A} * \frac{\sqrt{3}}{2}\end{aligned} \tag{63}$$

of course, this rate is for a single particle diffusing; probability of a diffusion event happening will be this value multiplied by the number of particles available for diffusion.

6 Parameters collection

In order to make realistic comparisons to the empirical cases, we have to define more rigorously corresponences between actual experiments and computer parameters. We use this section also to gather all the parameters used.

Endosomes from [2], are spheres (beads) with diameter of $10\mu m$, while real endosomes are much smaller, at about $(0.1-1)\mu m$. $S \approx 3 * 10^{-2 \div 0}$ for real endosomes or $S \approx 3 * 10^2$ for the ones in [2]

Diffusion takes the values $(10-100)\mu m^2/s$ in the cytosol and $(0.01-10)\mu m^2/t$ on the membrane; depending on variables such as molecule size and binding strength.

Human cell size ranges $(10-100)\mu m$ with an average of $25\mu m$; number of endosomes in a human cell seems to be various but mainly ranging in the hundreds or thousands. A really rough approximation for the volume in these cases is $V \approx (1-100)\mu m^3$.

Rate coefficients are taken initially, directly from [1], using the Table 3:

i	k_i^s
1	1.2x10^{-1} s^{-1}
2	2.25x10^5 $M^{-1}s^{-1}$
3	5.25x10^{-4} s^{-1}
4	5x10^4 $M^{-1}s^{-1}$
5	2.875x10^4 $M^{-1}s^{-1}$
6	8.0x10^{-2} s^{-1}
7	10x10^7 $M^{-1}s^{-1}$

Table 3: Values of the linear coefficients of the reaction rates taken from [1]

Although it is not clear what other values of that experiment are, for conversion purposes. K_{H1} and K_{H3} are not used in the simulation and $K_{H7} = 100$ is taken as a initial value. Initial number of simulation particles for the chemical species in mean field are set to have $M^s_{sol} = 500$ and all other species with the same proportions of the particular experiment. Particle number is made to scale with cell number N_c in some way to avoid artifacts such as visualization of domains without the need for feedback mechanisms.

7 Results

The computer simulation uses the reaction scheme in Appendix C with linear reaction coefficients from Table 3 and readjusted those values following the points made in chapter 5.3; the only outlier is k_4 that is take one or two orders of magnitude smaller since not doing this results in homogeneous membrane activation in the characteristic times of the experiment. To keep track of the values used without repeating in the single figures, k_4 will take the values 5.0x10^2 and 5.0x10^3 (low and high k_4 values, both in $\mu M^{-1} \cdot s^{-1}$); $\mathcal{D}$ 0.1 and 1.0 (low and high values; in $\mu m^2/s$; plus E_{sol} concentration will take a high and a low value. The Gillespie algorithm managed to reproduce domain formation in time marks similar to those expected in experiments. A preliminary test to reproduce activation plots from [1] suggests the correct value for the conversion rate (defined in chapter 5.3.1) ranges around 10^7. In experiments from [2] observations about the state of the endosome bead are made at the 15 min mark. The computer simulation is able to reproduce a similar picture to that obtained in this paper with similarly numerous and sized domains at the same time mark. Main figures of [2] are shown in Figure 2 and 3, in both of which domains can be visualized in subfigure C. Results from the computer simulation show in Figures 4 and 5, uses small grids to show that domain formation is indeed possible. A finer grid is used in all other simulation figures; one can highlight how Figures 6 and 7, simulated for low values of GEF concentration give an inactive membrane with some domains in it. All of the simulations mentioned to this point can be seen as recreating what is seen in 3 B) or C). Figure 8 shows an instance in which high GEF concentrations activate the entire membrane in such a way that no domains can be seen, imitating what is shown in Figure 3 D). A lower diffusion version of the same set of parameters of the last mention can be found in 9 in which domains are present despite high Rab activity on the membrane.

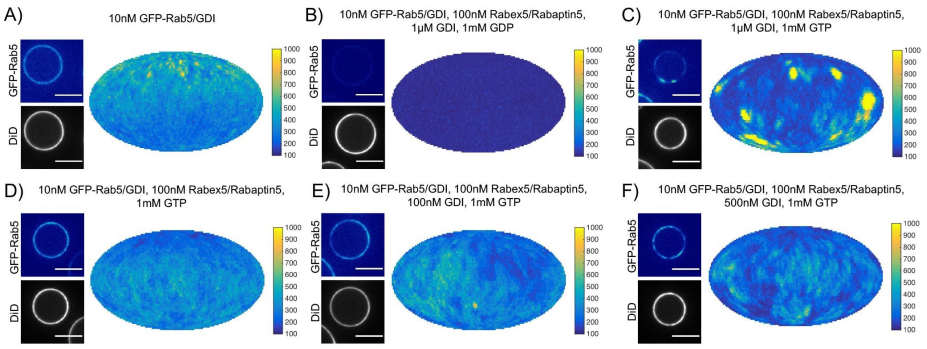

Figure 2: Experiment variants done varying free GDI initial concentration (excluding image B which is a particular case). In subfigure C domains can be observed. Here is shown that low values of I_{sol} prevent domain formation

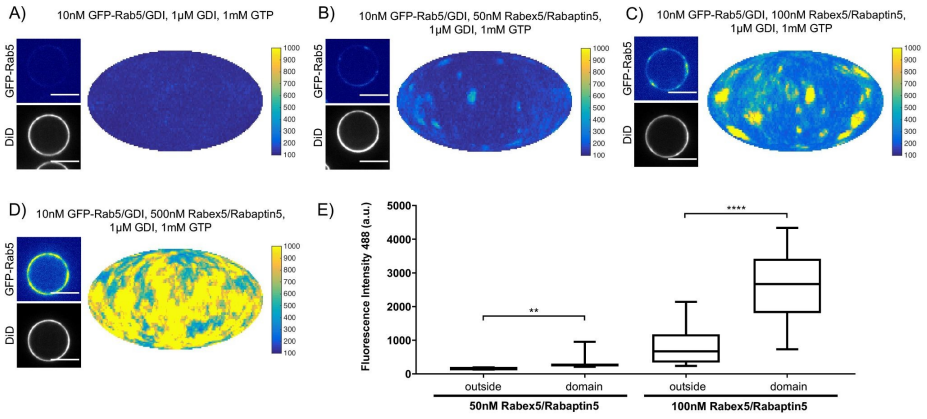

Figure 3: Experiment variants done vaying free GEF initial concentration. In subfigure C domains can be observed. The image shows that controlled values of E_{sol} are needed for domain formation: low values prevent membrane activation while high values activate uniformly the membrane without domains

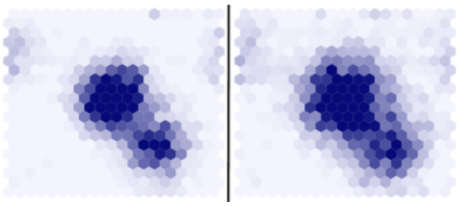

Figure 4: Simulation at 10 (left) and 15 (right) minutes for low k_4, low E_{sol} and high $\mathcal{D}$. One or two domains are present, very little off-domain activity is present

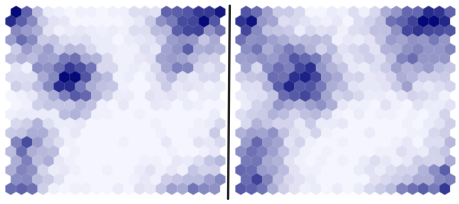

Figure 5: Simulation at 10 (left) and 15 (right) minutes for high k_4, low E_{sol} and high $\mathcal{D}$. Two domains are present (mind the periodic boundary condition); off domain activity is present mainly near the domain sites

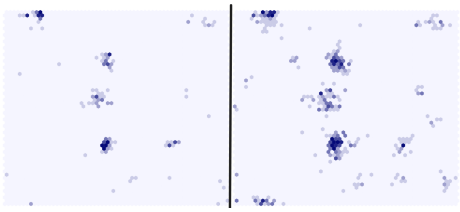

Figure 6: Simulation at 10 (left) and 15 (right) minutes for low k_4, low E_{sol} and low $\mathcal{D}$. There are four-five domains and very little off-domain activity

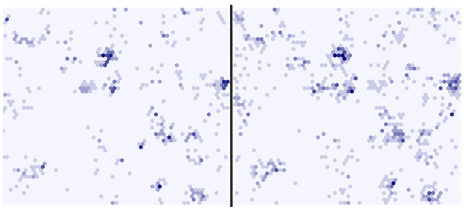

Figure 7: Simulation at 10 (left) and 15 (right) minutes for high k_4, low E_{sol} and low $\mathcal{D}$. There are about seven domain and some off-domain activity

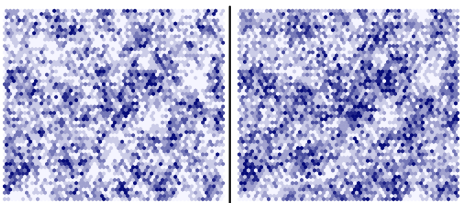

Figure 8: Simulation at 10 (left) and 15 (right) minutes for high k_4, high E_{sol} and high $\mathcal{D}$. Here effects of high amounts of free GEF is felt: all the membrane is active and no domains can be distinguished clearly

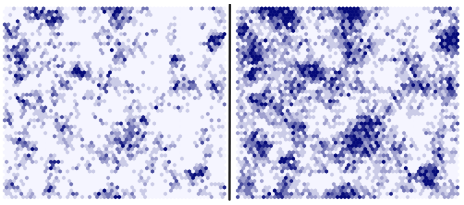

Figure 9: Simulation at 10 (left) and 15 (right) minutes for high k_4, high E_{sol} and low $\mathcal{D}$. About eight large domains can be seen despite high membrane activity.

A Hill function (or generalized Michaelis Menten) and space filling

The Hill function answers the following type of question: "Given a chemical specie A that is able to form complexes with some other specie B; when these are mixed together, what portion of the total binding sites of A are occupied?". Just to give a more concrete example, given hemoglobin has 4 binding sites for oxygen and there are 1000 copies of this protein in a solution with oxygen; what portion of the (4*1000 =) 4000 binding sites is occupied (as a function of concentrations)?

A.1 Michaelis Menten

For our case we can start from f_7; the total reaction reads

$$\text{GAP}_{sol} + \text{Rab5:GTP} \rightleftharpoons \text{GAP:Rab5:GTP} \rightarrow \text{GAP}_{sol} + \text{Rab5:GDP} \tag{64}$$

$$A + B \rightleftharpoons C \rightarrow A + P \tag{65}$$

This represents the case in which the protein that forms complexes is able to bind to a single molecule of the interested type. Defining the reaction rates $k_1,\ k_{-1},\ k_2$ (indexing goes left to right, the $_{-1}$ refers to the inverse process of reaction 1), we get the differential equations

$$\begin{cases} \frac{\partial A}{\partial t} = -k_1 AB + (k_{-1} + k_2)C \\ \frac{\partial B}{\partial t} = -k_1 AB + k_{-1} C \\ \frac{\partial C}{\partial t} = k_1 AB - (k_{-1} + k_2)C \\ \frac{\partial P}{\partial t} = k_2 C \end{cases} \tag{66}$$

by summing the first and third equations we find $\frac{\partial (A+C)}{\partial t} = 0 \rightarrow A + C = A_{tot}$ (a constant), with the last three equations we also get $\frac{\partial (B+C+P)}{\partial t} = 0 \rightarrow B + C + P = B_{tot}$ (a constant). These can be considered conservation relations. At this point we could also rewrite the system of equations by substituting the new relations like this:

$$\begin{cases} A = A_{tot} - C \\ \frac{\partial B}{\partial t} = -k_1 AB + k_{-1} C \\ \frac{\partial C}{\partial t} = k_1 AB - (k_{-1} + k_2)C \\ \frac{\partial P}{\partial t} = k_2 C \end{cases} \tag{67}$$

The simplified reaction to which we want to find a reaction rate reads:

$$\text{GAP}_{sol} + \text{Rab5:GTP} \rightarrow \text{GAP}_{sol} + \text{Rab5:GDP} \tag{68}$$

$$A + B \rightarrow A + C \tag{69}$$

Immediately we find that in the way this reaction is written, the concentration of A is conserved (A = A_{tot}); so if A_{tot} is present in the general reaction rate, it can be substituted with A in the simplified one.
Next step is considering the reaction rate of the simplified reaction to be the same as that of the general one when the compound, C reaches his stationary value. For biological systems this value is found by considering the reaction rates of biological reactions to be much smaller than non-biological ones ($k_2 \ll k_1, k_{-1}$), so we first find the stationary value in the adhesion and dissociation reactions:

$$\begin{cases} k_1 AB = k_{-1} C \\ C = A_{tot} - A \end{cases} \tag{70}$$

gives, after some simple algebra,

$$C = \frac{k_1 A_{tot} B}{k_1 B + k_{-1}} = A_{tot} \frac{B}{K_D + B} \quad (\text{with } K_D = \frac{k_{-1}}{k_1}) \tag{71}$$

now, by how it was defined,

$$\frac{\partial P}{\partial t} = k_2 C = k_2 A_{tot} \frac{B}{K_D + B} \tag{72}$$

This is the Michaelis Menten rate.
the simplified reaction, can be written

$$A + B \to A + P \quad (\text{or } B \to P \ \ (\text{with A-dependent rate})) \tag{73}$$

$$\begin{cases} \frac{\partial A}{\partial t} = 0 \\ \frac{\partial B}{\partial t} = -k_2 A_{tot} \frac{B}{K_D + B} \\ \frac{\partial P}{\partial t} = k_2 A_{tot} \frac{B}{K_D + B} \end{cases} \tag{74}$$

We abstained from using the B_{tot} value, since, in our model, B is Rab5:GTP, which varies through other mechanisms.

A.2 Hill function

The "simple" way to find the Hill function is to use the simplifying assumption that there are no intermediates (i.e. if the chemical specie, A has n binding sites, then A exists with either zero or n molecules attached: empty or saturated) and perfect cooperativity (the way we get to the no intermediates case: a binding makes subsequent bindings to the same protein much more likely, making intermediates less present). In this case we get the relation:

$$A + B^n \rightleftharpoons C \tag{75}$$

and the equations (here already processed to the version we need)

$$\begin{cases} A = A_{tot} - C \\ \frac{\partial C}{\partial t} = k_+ A B^n - k_- C \end{cases} \tag{76}$$

we end up getting the Hill function:

$$C = A_{tot}\frac{B^n}{K_D + B^n} = A_{tot}\frac{B^n}{K_H^n + B^n} \quad (K_H^n = K_D = \frac{k_-}{k_+}) \tag{77}$$

. Usually, if this form is taken, the value of the Hill exponent has to be found empirically (the ideal value is the same as the number of binding sites but the real cases differ from it).

To get a more precise version we could instead follow the same principle of the Michelis Menten part with more binding sites. We go to the case of f_6, in which we have 2 binding sites:

$$A + B \rightleftharpoons C \tag{78}$$
$$C \to A + P \tag{79}$$
$$C + B \rightleftharpoons D \tag{80}$$
$$D \to C + P \tag{81}$$

In these reactions C and D are composed of one A molecule and one or two B molecules respectively. As before, the reaction constants k_i refer to the i-th reaction, and the minus sign in the pedix refers to the inverse reaction (right to left).
The differential equations are now

$$\begin{cases} \frac{\partial A}{\partial t} = -k_1AB + (k_{-1} + k_2)C \\ \frac{\partial B}{\partial t} = -k_1AB + k_{-1}C - k_3BC + k_{-3}D \\ \frac{\partial C}{\partial t} = k_1AB - (k_{-1} + k_2)C - k_3BC + (k_{-3} + k_4)D \\ \frac{\partial D}{\partial t} = k_3BC - (k_{-3} + k_4)D \\ \frac{\partial P}{\partial t} = k_2C + k_4D \end{cases} \tag{82}$$

Again, we can find the "conservation" relations $A + C + D = A_{tot}$ and $B + C + 2D + P = B_{tot}$. Using again the assumption of faster adhesion/dissociation over biological reactions, the stationary values for the compounds are found as such:

$$\begin{cases} 0 = k_1AB - (k_{-1} + k_2)C - k_3BC + (k_{-3} + k_4)D \quad \text{(equation for C)} \\ 0 = k_3BC - (k_{-3} + k_4)D \quad \text{(equation for D)} \\ D = A_{tot} - A + C \quad \text{(conservation law)} \end{cases} \tag{83}$$

solving the algebra, we find (with $K_{Di} = \frac{k_{-i}}{k_i}$):

$$\begin{cases} C = A_{tot}(1 - \frac{B^2}{K_{D1}K_{D3} + K_{D3}B + B^2})\frac{B}{K_{D1} + B} \\ D = A_{tot}\frac{B^2}{K_{D1}K_{D3} + K_{D3}B + B^2} \end{cases} \tag{84}$$

if in the formula for D we exclude the first order term in B in the denominator we get $D \approx A_{tot}\frac{B^2}{K_H^2 + B^2}$ $(K_H, \text{the Hill constant})$; this is the Hill function (with

n = 2). In this case the simplified reaction will take the form

$$A + B \to A + P \quad \text{(or } B \to P \text{ with A-dependent rate)} \tag{85}$$

with

$$\frac{\partial P}{\partial t} = A_{tot}(k_2(1-\frac{B^2}{K_{D1}K_{D3} + K_{D3}B + B^2})\frac{B}{K_{D1} + B} + k_4\frac{B^2}{K_{D1}K_{D3} + K_{D3}B + B^2}) \tag{86}$$

considering $k_2 < k_4$ as the stochastic process now has more ways to happen (in the differential time case, approximately twice $k_4 \approx 2k_2$, given that no significant enzymatic changes occurred)

A.3 Volume exclusion

For the function f_2, the reasoning goes as follows: the probability of the reaction happening is the probability of having the reactants times the probability of finding an empty spot where the reaction can happen. The reaction rate is proportional to this probability. For the reaction

$$\text{GDI:Rab5:GDP}_{sol} \to \text{GDI}_{sol} + \text{Rab5:GDP} \tag{87}$$

$$M_{sol} \to I_{sol} + D \tag{88}$$

$$\begin{cases} P(\text{the reaction occurs}) = \text{coefficient} * P(M) * P(\text{available reaction site}) \\ P(\text{available reaction site}) = \frac{\text{number of available sites}}{\text{total number of sites}} = 1 - \frac{\text{number of occupied sites}}{\text{total number of sites}} \end{cases} \tag{89}$$

the only things that can occupy sites in our model are Rab proteins. The "coefficient" and "P(M)" are treated the same way as in "basic" chemistry. By recasting everything using concentrations, we get

$$R_1 = k_1 M_{sol}(1 - \frac{D + T + E}{K_{H1}}) \tag{90}$$

where K_{H1} is the concentration of (total) sites. Intuitively we could consider the Rab5:GTP:GEF to take more space than the alternatives without the GEF; this is overlooked for simplicity.

B The $\theta = V/S$ factor

The θ factor has the physical meaning of "ratio of change in concentrations between cytosol and membrane after administration of the same number of molecules of some chemical specie". The main idea behind this factor is to take the mass action law, to which mass conservation must be applied. In the case of an homogeneous media (eg. only cytosol or only membrane) the problem

doesn't arise but it does when both membrane and cytosol are present. The concentrations in each media can be written as:

$$\text{concentration}_{sol} = \frac{\#\text{particles in solution}}{\text{(total) volume}}$$
$$\text{concentration}_{membrane} = \frac{\#\text{particles on membrane}}{\text{(total) area}}$$

for the membrane (the "total" in brackets are added because it is right in mean field but not in eg. coarse graining). One immediately sees that, even dimensionally, they differ in a multiplicative length factor; most importantly the total number of particles must stay the same, owing to mass/particle conservation. Such requirement can be written as:

$$\begin{aligned}\#\text{total particles} &= \#\text{particles in solution} + \#\text{particles in membrane} \\ &= V \cdot \text{concentration}_{sol} + S \cdot \text{concentration}_{membrane}\end{aligned}$$

Particle conservation is written as $\frac{d(\#\text{total particles})}{dt} = 0$, so

$$\begin{aligned}\frac{d(\text{concentration}_{membrane})}{dt} &= \frac{V}{S}\frac{d(\text{concentration}_{sol})}{dt} \\ &= \theta\frac{d(\text{concentration}_{sol})}{dt}\end{aligned}$$

So for big values of θ, changes on the membrane do not affect the cytosol, while for small values of θ the opposite is true.

In addition to this, θ was also discussed in a diffusive sense. Importantly, the diffusive lengths are time dependent and grow over time, this should mean that, if the role of theta is taken as just defined, its value should grow over time and allow a flow of new matter from the newly introduced regions over time and adjust concentrations according to some other model that deals with diffusivity.

C Alternative reaction scheme

Using the reactions derived from the general reactions section, one gets to the following scheme, where bold text font is applied to reactions that change from the model used in the rest of the thesis:

- GDI_{sol} + Rab:GDP $\rightleftharpoons$ Rab:GDP:GDI_{sol}
    1. $R_1 = k_1 M_{sol} f_1(D,T,E)$ or $\tilde{k}_1 f_1(D,T,E)$ (leftarrow)
    2. $R_2 = k_2 D I_{sol}$ or $\tilde{k}_2 D$ (rightarrow rate)
- GAP_{sol} + Rab:GTP $\rightarrow$ GAP_{sol} + Rab:GDP
    3. $R_3 = k_3 A f_3(T)$ or $\tilde{k}_3 f_3(T)$

- **GEF$_{sol}$ + Rab:GDP $\rightarrow$ Rab:GTP:GEF**

  4. $R_4 = k_4 E_{sol} D$

- GEF$_{sol}$ + Rab:GTP $\rightleftharpoons$ Rab:GTP:GEF

  5. $R_5 = k_5 E_{sol} T$ (rigtharrow)
  6. $R_6 = k_6 E$ (leftarrow)

- **Rab:GTP:GEF + Rab:GDP $\rightarrow$ Rab:GTP:GEF + Rab:GTP**

  7. $R_7 = k_7 E f_7(D)$

It's easy to see that, in the limit case of very fast attachment/detachment (much faster than biological reactions), the two reaction schemes are equivalent in mean field: at any point in time the concentration of E and E$_{sol}$ in a certain region of space is at equilibrium and this happens between any two biological reactions (removing $E \rightarrow E_{sol}$ has no effect since it is followed immediately by an infinitely fast $E_{sol} \rightarrow E$, if needed).
For finite values of attachment/detachment rates, however, we have to work some more. For sake of simplicity let's only take reactions 5, 6 and 7. Adding the letter "a" to all regarding the literature derived picture (discussed in the "General reaction" section) and with the letter "b" all parameters related to [1]. The formulas for the reaction rates do not change between pictures: only parameters and products. Time evolution of D will be omitted and assumed to be some constant value that is result of the remaining equations.

$$\begin{cases} \frac{\partial E}{\partial t}|_a & = R_5^a - R_6^a \\ \frac{\partial T}{\partial t}|_a & = -R_5^a + R_6^a + R_7^a \\ \frac{\partial (E+T)}{\partial t}|_a & = R_7^a \end{cases} \tag{91}$$

$$\begin{cases} \frac{\partial E}{\partial t}|_b & = R_5^b - R_6^b - R_7^b \\ \frac{\partial T}{\partial t}|_b & = -R_5^b + R_6^b + 2R_7^b \\ \frac{\partial (E+T)}{\partial t}|_b & = R_7^b \end{cases} \tag{92}$$

In both cases, the third equation is a linear composition of the first two, therefore not an independent relation; representing the total number of Rab molecules on the membrane (this can be the activation parameter). A possible requirement for the two schemes to be equivalent, is to ask that the total number of bound Rab molecules is the same at any point in time. The requirement just prescribed amounts to saying $R_7^a = R_7^b \rightarrow k_7^a E^a(f_7(D)) = k_7^b E^b(f_7(D))$ (f_7 in brackets since we assumed D to be equal in both cases; f_7 itself could change but we assume it not to). To proceed in the calculations we ask the reactions to occur at the same starting values of concentrations (banally (same initial condition $\rightarrow$ same result) requires same initial conditions), and that the conservation laws are satisfied ($E^a + T^a = T^a_{tot}$ and $E^a + \theta E^a_{sol} = \theta E^{tot}_{sol}$ (and the same for variables denominated "b").

By just making the substitution, of the conservation laws into, eg. the equation for $\frac{\partial E}{\partial t}$, and finding the equilibrium point we get to the relations

$$E^{a,*} = \frac{1}{2}(T_t^a + \theta E_s^t + \theta \frac{k_6^a}{k_5^a})(1 - \sqrt{1 - \frac{4\theta^{-1} E_s^t T_t^a}{(\theta^{-1} T_t^a + E_s^t + \frac{k_6^a}{k_5^a})^2}}) \tag{93}$$

$$E^{b,*} = \frac{1}{2}(T_t^b + \theta E_s^t + \theta \frac{k_6^b + k_7^b}{k_5^b})(1 - \sqrt{1 - \frac{4\theta^{-1} E_s^t T_t^b}{(\theta^{-1} T_t^b + E_s^t + \frac{k_6^b + k_7^b}{k_5^b})^2}}) \tag{94}$$

multiplying these by their respective k_7 and expanding the square root for big values of θ, the final relation (one equation with three unknowns: k_5, k_6, k_7 in one picture and all else given) consists in equating a truncation of the following two series:

$$\frac{k_7^a E_s^t T_t^a}{\theta^{-1} T_t^a + E_s^t + \frac{k_6^a}{k_5^a}} * (1 + \frac{E_s^t T_t^a}{(\theta^{-1} T_t^a + E_s^t + \frac{k_6^a}{k_5^a})^2} + 2\theta^{-1} (\frac{E_s^t T_t^a}{(\theta^{-1} T_t^a + E_s^t + \frac{k_6^a}{k_5^a})^2})^2 + \cdots) \tag{95}$$

$$\frac{k_7^b E_s^t T_t^b}{\theta^{-1} T_t^b + E_s^t + \frac{k_6^b + k_7^b}{k_5^b}} * (1 + \frac{E_s^t T_t^b}{(\theta^{-1} T_t^b + E_s^t + \frac{k_6^b + k_7^b}{k_5^b})^2} + 2\theta^{-1} (\frac{E_s^t T_t^b}{(\theta^{-1} T_t^b + E_s^t + \frac{k_6^b + k_7^b}{k_5^b})^2})^2 + \cdots) \tag{96}$$

A similar approach has to be reiterated for the rate k_4 and additional equations have to be introduced to match the degrees of freedom; the thesis won't go deeper into the derivation of the new coefficients.

Bibliography

[1] Bezeljak et al. "Stochastic activation and bistability in a Rab GTPase regolatory network". In: *PNAS* (2020). DOI: /10.1073/pnas.1921027117.

[2] Cezanne et al. "A non-linear system patterns Rab5 GTPase on the membrane". In: *eLife* (2020). DOI: https://doi.org/10.7554/eLife.54434.

[3] Hart et al. "A GDP Dissociation Inhibitor That Serves as aGTPase Inhibitor for the Ras-Like ProteinCDC42Hs". In: *Science* (1992). DOI: 10.1126/science.1439791.

[4] Pylypenko et al. "Rab GTPases and their interacting protein partners: Structural insights into Rab functional diversity." In: *Pub Med* (2017). DOI: doi:10.1080/21541248.2017.1336191.

[5] Pylypenko et al. "Structure of doubly prenylated Ypt1:GDI complex and the mechanism of GDI-mediated Rab recycling". In: *The EMBO Journal* (2006). DOI: https://doi.org/10.1038/sj.emboj.7600921.

[6] Rak et al. "Structure of Rab GDP-Dissociation Inhibitor in Complex with Prenylated YPT1 GTPase." In: *Science* (2003). DOI: https://doi.org/10.1126/science.1087761.

[7] Wu et al. "Membrane targeting mechanism of Rab GTPases elucidated by semisynthetic protein probes". In: *Nature Chemical Biology* (2010). DOI: https://doi.org/10.1038/nchembio.386.

[8] Zhang et al. "Molecular mechanism for Rabex-5 GEF activation by Rabaptin-5". In: *eLife* (2014). DOI: https://doi.org/10.7554/eLife.02687.

[9] Rehmann H Wittinghofer A Bos JL. "GEFs and GAPs: Critical Elements in the Control of Small G Proteins". In: *Cell* 129.5 (June 2007), pp. 865–877. ISSN: 0092-8674. DOI: 10.1016/j.cell.2007.05.018.

[10] S. Pfeffer. "A model for Rab GTPase localization". In: *Biochemical Society Transaction* (2005). DOI: https://doi.org/10.1042/BST0330627.

[11] Dikran Aivazian Suzanne Pfeffer. "Targeting Rab GTPases to distinct membrane compartments". In: *Nat Rev Mol Cell Biol.* (2004). DOI: 10.1038/nrm1500.

[12] Suzanne R. Pfeffer Ulf Sivars Dikran Aivazian. "Purification and Properties of Yip3/PRA1 as a Rab GDI Displacement Factor". In: *Science Direct* (2006). DOI: https://doi.org/10.1016/S0076-6879(05)03030-2.

[13] Suzanne R. Pfeffer Ulf Sivars Dikran Aivazian. "Yip3 catalyses the dissociation of endosomal Rab–GDI complexes". In: *Nature* (2003). DOI: https://doi.org/10.1038/nature02057.

[14] Song C Yuan W. "The Emerging Role of Rab5 in Membrane Receptor Trafficking and Signaling Pathways". In: *Biochemistry Research International* 2020 (Feb. 2020), pp. 1–10. ISSN: 2090-2255. DOI: 10.1155/2020/4186308.