PHD THESIS

SPATIAL STRUCTURES AND REGULATION IN BIOLOGICAL SYSTEMS

PERNILLE YDE

SUPERVISORS: MOGENS HØGH JENSEN,
JOACHIM MATHIESEN
AND ALA TRUSINA

FACULTY OF SCIENCE
THE NIELS BOHR INSTITUTE
UNIVERSITY OF COPENHAGEN
DENMARK

JANUARY 2014

CONTENTS

ABSTRACT

Spatial arrangements are very important for many biological systems. This thesis presents several different studies of biological systems, which are related to spatial arrangements at two different levels: one is the growth of biological macromolecules, here related to protein aggregation, and the other is the spatial regulation of biological systems, here related to different aspects of the inflammatory response. All systems are studied using computational modelling and mathematical analysis.

The first part of the thesis explores different protein aggregation scenarios. In Chapter 1, we consider a previously studied and very general aggregation model describing frangible linear filaments. This model is especially relevant for the growth of amyloid fibres, that have been related to a number of serious human diseases, and which are known to grow in an accelerated self-enhanced manner.We derive an approximate analytical mathematical expression for the time evolution of the length distribution of the aggregate population, and we discuss the accuracy of the analytical expression. We also compare the model of frangible linear aggregation to experimentally obtained length distributions of growing insulin filaments.

In Chapter 2, we consider the aggregation of the protein p25α, influenced by the presence of the secondary chemical species heparin. Different concentrations of heparin present different environmental conditions, which cause the protein aggregates to form different structural shapes. We construct a mathematical model, which is fitted to experimental data for p25α aggregation at different heparin levels. The model incorporates a logistic-like growth assumption, which is motivated in the beginning of the chapter, and which represents an alternative model for accelerated growth of amyloid fibres.

In Chapter 3, we consider the complex aggregation patterns of the whey proteins β-lactoglobulin (bLG) and α-lactalbumin (aLA), influenced by several different environmental conditions, which cause the aggregates to form the different structural shapes - here the varying environmental conditions are different pH and calcium concentrations. We construct a mathematical model for the aggregation process, and fit the model to an array of experimental data. The model reproduces the dynamics of the aggregation process and predicts final size distributions of the aggregates, which agree well with the expectation based on experimental measurements.

The second part of the thesis explores different spatial aspects of inflammatory response. In Chapter 4 we address the problem of cytokine signal transmission and the subsequent white blood cell recruitment during inflammatory response. We construct a simple model of the inflammatory response in a tissue cell, based on the regulatory network of the transcription factor NF-κB. We show that

the simple model is able to produce either transient or continuous amplification of the cytokine signal depending on the external and internal conditions of the cell. We then construct a multicellular model of the tissue and show how coupled cells are able to function as an excitable medium and propagate waves of high cytokine concentration through the tissue. If the internal regulation in the cells is over-productive, the model predicts a continuous amplification of cytokines, which spans the entire system and resembles a situation of chronic inflammation in the tissue.

In Chapter 5 we consider inflammatory response in the islets of Langerhans, which are responsible for regulating the levels of blood sugar (by releasing insulin and glucagon) and which are located in the pancreas. Low-grade chronic inflammation and over-production of the cytokine IL-1β are characteristic features of islets in patients with type II diabetes. We expand the model of Chapter 4 in order to study the inflammatory response in islets of Langerhans, with a special focus on the influence imposed by the spatial conditions - namely the sizes and different possible shapes of the islets of Langerhans. In agreement with experimental observations, we find that large islets are especially prone to transition into a state of chronic low-grade inflammation. Additionally, we find that different islet shapes may influence the risk of developing chronic inflammation - an observation, which implicates a connection between the distribution of different islet shapes and a protective function.

SUMMARY IN DANISH - DANSK RESUMÉ

Rumlige aspekter er meget vigtige i mange biologiske systemer. Denne afhandling præsenterer en række forskellige studier af biologiske systemer, hvor rumlige aspekter spiller ind på to forskellige niveauer: Det ene er væksten og dermed opnåelsen af forskellige strukturer af biologiske aggregater og det andet er rumlig regulering i biologiske systemer, her relateret til forskellige rumlige aspekter af inflammatorisk respons.

Første del af denne afhandling udforsker forskellige protein-aggregerings scenarier. I kapitel 1 betragter vi en allerede eksisterende, og meget generel aggregerings model, der beskriver fragmenterende (dvs. knækkende) lineære filamenter. Denne model er især relevant for væksten af "amyloid" fibriller, der er forbundet med en række alvorlige sygdomme, og som vokser på en accelereret og selv-inducerende facon, hvilket har vist sig at være velbeskrevet af aggregerings modellen, der inkluderer fragmentering. Vi udleder et approximativt anaytisk matematisk udtryk for tidsudviklingen af længdefordelingen af aggregat-populationen, og vi diskuterer nøjagtigheden af denne analytiske løsning. Vi sammenligner også modellen for lineær aggregering og fragmentering med eksperimentelle længdefordelinger af voksende insulinfilamenter. I kapitel 2, betragter vi aggregering af proteinet p25α, influeret af tilstedeværelsen af det sekundære kemiske stof heparin. Forskellige koncentrationer af heparin udgør en miljømæssig omstændighed, der bevirker at aggregeringsprocessen fører til dannelse af forskellige strukturelle former, og vi konstruerer en model, som er baseret på eksperimentelle data for forskellige heparin niveauer. Modellen indeholder en antagelse af logistisk-lignende vækst, som er motiveret i begyndelsen af kaptitlet, og som udgør en alternativ model for accelereret vækst af "amyloid" fibriller. I kapitel 3, betragter vi komplekse aggregeringsmønstre af valleproteinerne β-lactoglobulin (bLG) og α-lactalbumin (aLA), påvirket af flere miljømæssige forhold, som fører til aggregater af forskellige strukturelle former - de varierende miljømæssige forhold er her forskellig pH og calcium koncentration. Vi opsætter en matematisk model for aggregeringsprocessen, og fitter modellen til eksperimentelle data. Modellen reproducerer dynamikken i aggregeringsprocessen og fører til størrelsesfordelinger, som tilsvarer det forventede udfra eksperimentelle målinger.

Den anden del af afhandlingen udforsker forskellige rumlige aspekter af inflammatorisk respons. I kapitel 4 betragter vi udfordingen i cytokin signal transmission og den efterfølgende rekruttering af hvide blodlegemer under inflammatorisk respons. Vi opsætter en simpel model for inflammatorisk respons i en celle, der baseres på det regulatoriske netværk af transkriptionsfaktoren NF-κB. Vi viser, at den simple model, afhængigt af de eksterne og interne forhold

i cellen, er i stand til at frembringe enten forbigående eller kontinuerlig forstærkning af cytokin-signalet. Vi opsætter desuden en model for vævet (dvs. mange celler ved siden af hianden) og viser, hvordan rumligt koblede celler er i stand til at fungere som et eksiterbart medium og dermed udbrede bølger af høj cytokin koncentration gennem vævet. Hvis den interne regulering i cellerne er overproduktiv, forudsiger vævsmodellen en kontinuerlig forstærkning af cytokiner, der breder sig over hele systemet, og dermed gengiver en situation med kronisk inflammation i vævet. I kapitel 5 betragter vi inflammatorisk respons i de Langerhanske øer, der befinder sig i bugspytkirtlen og er ansvarlige for at regulere blodsukkerniveauet (ved at frigive insulin og glucagon). Svag kronisk inflammation og overproduktion af cytokinet IL-1β er karakteristisk for Langerhanske øer i patienter med type II diabetes. Vi udvider modellen i kapitel 4 for at studere inflammatorisk respons i Langerhanske øer, med særligt fokus på indflydelsen fra diverse rumlige forhold - nemlig størrelsen og forskellige rumlige udformninger af de Langerhanske øer. I overensstemmelse med eksperimentelle observationer, finder vi, at store øer er særligt tilbøjelige til at indgå i en tilstand af kronisk inflammation. Derudover finder vi, at forskellige udformninger af øerne, kan påvirke risikoen for at udvikle kronisk inflammation - en observation, der indikerer en forbindelse mellem rumlig udformning og funktion.

ACKNOWLEDGEMENTS

I am very grateful to have been given the opportunity to do a PhD in the CMOL-group, where the atmosphere is always friendly and warm, and where there are many interesting people who are always keen to discuss their scientific projects. There are many people, whom I wish to thank and to whom I am very grateful.

Mogens Høgh Jensen has been my main supervisor throughout many years. Mogens has been involved in all of my projects, has always been ready to help, and we have had many fruitful discussions. Also, Mogens makes a true effort to ensure, that all of his students are happy with their work, and he has always encouraged me to pursue the projects that I find interesting, and which will help me achieve my future career goals. Last, but not least, Mogens has always kept his door open for any kind of problem, and I am very grateful to have a friend that I can always go to.

Ala Trusina has been my main supervisor for the work done on inflammation and the NF-κB system. Ala has helped me with the project, provided many good ideas and her ability to bridge biology and physics has been very inspiring. Also Thomas Holst-Hansen, who has joined the project, is always enthusiastic and wonderful to work with.

Joachim Matthiesen has also supervised me, has provided excellent ideas for my theoretical work on protein aggregation, and has strengthened my mathematical understanding and competences. Tuomas Knowles has been my supervisor during my stay abroad at the University of Cambridge (UK), and his sharp mind and very positive attitude has been very inspiring and have taught me a lot. I was very happy to visit the University of Cambridge, and I would like to thank the entire group for a great stay. For my work on protein aggregation I have also collaborated with Daniel Otzen and Søren Bang Nielsen, who are always hard working and have provided more experimental data, than my modelling could keep up with.

I would like to thank the entire CMOL-group, for providing an excellent workspace, where ideas are plentiful and help is always near. And last but not least, I would like to thank my family, friends and especially my dear boyfriend Martin for all the chocolate, love and support.

LIST OF PUBLICATIONS

1. **Pernille Yde**, Thomas C. T. Michaels, Julian Willis, Alexander K. Buell, Daniel Otzen, Mogens H. Jensen and Tuomas P. J. Knowles. The length distribution of frangible filaments: an analytical study in the continuum approximation. *Manuscript in progress, aimed at The Journal of Chemical Physics* (2013)

2. Søren Bang Nielsen, **Pernille Yde**, Lise Giehm, Sabrina Sundbye, Gunna Christiansen, Joachim Mathiesen, Mogens Høgh Jensen, Poul Henning Jensen and Daniel E. Otzen. Multiple Roles of Heparin in the Aggregation of p25α. *Journal of Molecular Biology*, 421, 601-615 (2012)

3. **Pernille Yde**, Benedicte Mengel, Mogens H Jensen, Sandeep Krishna and Ala Trusina. Modeling the NF-κB mediated inflammatory response predicts cytokine waves in tissue. *BMC Systems Biology* 5:115 (2011)

4. **Pernille Yde**, Mogens Høgh Jensen, and Ala Trusina Analyzing inflammatory response as excitable media *Physical Review E* 84:051913 (2011)

Publications not included in this thesis:

5. Joachim Mathiesen, **Pernille Yde** and Mogens H. Jensen. Modular networks of word correlations on Twitter. *Scientific Reports*, 2, 814 (2012)

Part I

GROWTH OF BIOLOGICAL AGGREGATES

INTRODUCTION

In this part of the thesis I shall consider different aspects of the growth, and hence the achievement of different spatial structures of large protein aggregates. This part of the thesis was originally intended to investigate the highly ordered amyloid fibres, but we shall also consider less ordered aggregates and different environmental influences of increasing complexity.

Structure and Function

Spatial and structural arrangements are very important in biology. The molecular building blocks of biological systems include many large and complex molecules, which have unique spatial structures, that have evolved to perform specific functions. In many cases structure and function are inextricably linked. For example, the reactive sites of enzymes must be exposed correctly in order to function correctly, and in many cases the enzymes are able to change their function by changing their structure. Also, the functions of macromolecules such as microtubules, biological membranes and DNA are directly linked to their spatial structure. The assembly of large biomolecules is therefore a process of vital importance as even small deviations from the desired structure can have huge consequences for functionality.

Amyloid Fibres

Amyloid fibres are highly ordered protein aggregates of inappropriately folded proteins, which polymerize to form cross-beta structures resulting in long linear filaments. Amyloid fibres have been related to a number of serious human diseases - examples include type II diabetes and and a number of neurodegenerative disorders, such as Parkinson's and Alzheimer's disease. Numerous experimental and theoretical studies have focussed on the growth kinetics of amyloid fibres, and it is a well established fact that the fibres tend to grow in an accelerated fashion, where the initial growth is very slow (this part of the aggregation process is referred to as the *lag-time*), before the growth accelerates to become increasingly fast. In experiments where the monomer-pool is limited, the growth eventually decelerates and comes to an end, as the number of monomers available for aggregation decreases.

Accelerated Growth

The accelerated fashion in which the amyloid fibres grow could, in theory, be explained by different mechanisms. In Chapter 1 we shall consider the growth of frangible linear aggregates, where the overall growth is accelerated by fragmentation, which functions as a so-called *secondary nucleation event* (i.e. the number of growth-sites increases as the number of filament ends increases). The theoretical results of Chapter 1 are fitted to experimental data of growing insulin filaments.

Accelerated growth could also occur as a result of logistic-like growth, where the already aggregated mass induces further growth in a self-enhancing fashion. This kind of growth could be relevant for different spatial aggregate structures, and in Chapter 2 we briefly discuss the growth of spherical versus linear aggregates. In the same chapter we develop a model, that assumes a logistic-like growth, and apply this model to an aggregation process, which is also influenced by external environmental conditions.

Environmental Influence

The aggregation of biomolecules may be influenced by many environmental conditions, such as pH, temperature and the presence of catalysts or other interacting molecules. In Chapter 2 we consider the aggregation of tubulin polymerization-promoting protein (p25α), influenced by a secondary chemical species (Heparin). The presence of different concentrations of Heparin, causes the aggregation process to result in two structurally different aggregate forms, and we develop in Chapter 2 a model, which is fitted to experimental data.

In Chapter 3 we consider an aggregation process of even higher complexity. Here we develop a model for the aggregation of mixed whey proteins, influenced by two varying environmental conditions (pH and calcium concentration) and the model is fitted to an array of experimental data. Again we observe that the differing environmental conditions lead to aggregates of different spatial structures.

1 FRANGIBLE LINEAR FILAMENTS

In this chapter I shall present a mathematical model for the growth kinetics of frangible linear filaments. The linear self-assembly of filamentous structures is a process of fundamental importance to the normal functioning of nature, as well as the formation of amyloid fibres observed in relation to diseases such as type II diabetes, Parkinson's and Alzheimer's diseases [13, 18, 27, 58, 66, 71].

The model presented in this chapter has been studied before [14, 15, 16, 38], and we shall here apply a continuum approximation in order to develop approximate analytical expressions, that are compared with numerical solutions (all numerical solutions are obtained using a fourth order Runge Kutta integration, implemented in c++). Finally, we compare the model with experimental data for growing insulin filaments.

RELEVANT PUBLICATION:
Pernille Yde, Thomas C. T. Michaels, Julian Willis, Alexander K. Buell, Daniel Otzen, Mogens H. Jensen and Tuomas P. J. Knowles. The length distribution of frangible filaments: an analytical study in the continuum approximation. *Manuscript in progress, aimed at The Journal of Chemical Physics* (2013)

1.1 SECONDARY NUCLEATION EVENTS

Early investigations of filamentous growth [53, 54] focused on simple nucleation events followed by linear polymerization. In 1962 Oosawa presented solutions to the kinetic equations for irreversible growth, however, Oosawa's theory was not able to fully explain the experimental observations of amyloid growth.

Experimental observations often report a characteristic *time-lag* followed by a rapid acceleration, in the growth of the total mass of aggregates $M(t)$. In experiments where the monomer-pool is limited, the growth eventually decelerates and comes to an end, and as a result the growth curve for $M(t)$ typically has a sigmoid shape, implying a somewhat self-enhanced growth process. In order to explain this behaviour *secondary nucleation events* were proposed to take place in cohesion with the simple nucleation (sometimes referred to as homogeneous nucleation) originally proposed by Oosawa [8, 24, 25, 62]. Fragmentation, lateral growth or nucleation on the surface of existing polymers were all proposed as possible secondary nucleation events, and in this chapter, we shall focus on fragmentation.

The Length Distribution

The length distribution of an ensemble of growing linear aggregates, represents a simple structural feature, which is often of great interest - for example, the lengths of growing amyloid fibres is a key parameter for determining the progression of disease. Understanding the mechanisms that lead to different aggregate sizes (different filament lengths) is therefore an important aspect of understanding the aggregation process.

Oosawa also studied the development of the length distribution of filaments [53, 54] and showed that the length distribution initially develops into a Poisson distribution, before relaxing over a longer time scale, into an exponential distribution. Oosawa's equilibrium length distribution does not have a "peak" - or in other words the equilibrium length distribution does not lead to a characteristic length, and Oosawa comments in his textbook[53]: "*in polymers growing one-dimensionally, length distributions having a sharp maximum cannot be realized as a true equilibrium*". This is, however, not the case for true biological filaments, whose length distributions have been observed to have a peak around a characteristic length[6, 73]

Fragmentation

It has previously been demonstrated[16, 38], that the growth kinetics of amyloid fibres, can successfully be explained by a model including fragmentation as a secondary nucleation event. Fragmentation is characterised as a nucleation event because it increases the total number of growth-sites, by increasing the total number of filament ends.

As we shall see below, the presence of fragmentation also gives rise to a possible "peak" in the length distribution. This maximum appears due to the emergence of a characteristic length, in the dynamic steady state, where growth is balanced by fragmentation.

1.2 MASTER EQUATION

The length distribution of the filaments is given by the concentration $f(j,t)$ of filaments of aggregation number j at time t, where we define the aggregation number j as the size of the filament measured in number of monomers. We consider a system of completely well-mixed monomers and frangible growing filaments, which undergo the growth processes sketched in Figure 1. Filaments can increase in size through elongation processes, i.e. the addition of single monomers onto the ends of the filament, or shrink through dissociation from either ends and through filament fragmentation. Here, we consider only the formation of *linear* filaments i.e. no branching or clumping processes are allowed. We also include a minimal aggregation number n_c, which defines the smallest stable filament

size. Filaments of aggregation number smaller than n_c will disintegrate into monomers and thereby re-enter the monomer pool. We assume that the formation of these aggregates takes place through a primary nucleation process in which n_c monomers spontaneously bind together to form a nucleus.

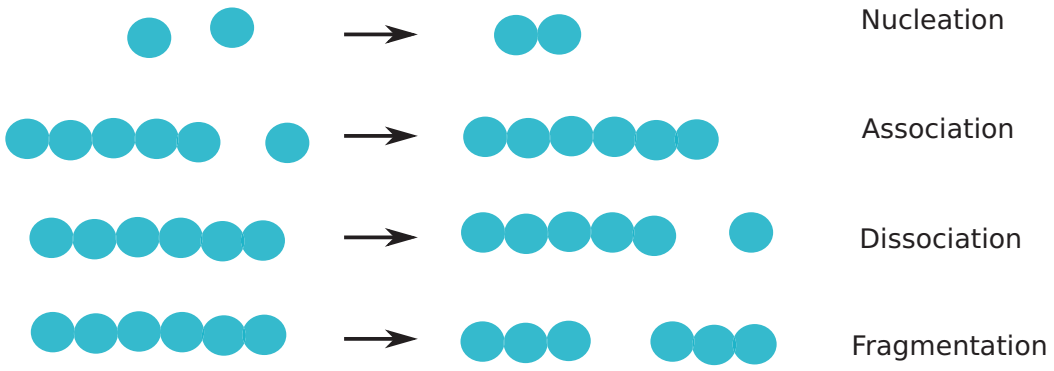


Figure 1: Sketch of linear aggregates that undergo the processes of nucleation, association, dissociation and fragmentation. Here we sketch a nucleation process as the spontaneous binding of *two* monomers - i.e. here $n_c = 2$.

The behaviour of the ensemble of monomers and aggregates can be formulated within the formalism of kinetic differential equations. On accounting for the elementary mechanisms of Figure 1, the temporal evolution of the concentration $f(j,t)$ is described by the following master equation [16, 38]:

$$\begin{aligned}\frac{\partial f(j,t)}{\partial t} &= 2k_+ m(t)\,[f(j-1,t) - f(j,t)] \\ &\quad + 2k_{off}\,[f(j+1,t) - f(j,t)] \\ &\quad - k_-(j-1)f(j,t) \\ &\quad + 2k_- \sum_{i=j+1}^{\infty} f(i,t) \\ &\quad + k_n m(t)^{n_c}\delta_{j,n_c} \end{aligned} \tag{1}$$

where $m(t)$ denotes the concentration of monomers at time t. The terms in equation (1), that are proportional to k_+, describe growth through association of monomers and the factor two accounts for the fact that association can happen at either end of the filament. Similarly, terms proportional to k_{off} accounts for dissociation of monomers at the ends of the filaments. Terms proportional to k_- account for fragmentation processes: a filament of lenght j can break at $(j-1)$ sites and the rate-constant k_- equals the breakage-probability per time and per binding site. The term $-k_-(j-1)f(j,t)$ thus accounts for the loss due to fragmentation of filaments of length j. The term $+2k_-\sum_{i=j+1}^{\infty} f(i,t)$ accounts for the creation of filaments of length j due to fragmentation of longer filaments. The probability that a filament of length i breaks is given by $P(break|i) = k_-(i-1)f(i,t)$. The probability that one of the fragments has length

j *given* that the filament breaks is given by $P(j|\mathrm{break}\, i) = \frac{2}{i-1}$. Hence the total probability that a filament of length j is created is given by $P(j|\mathrm{break}\, i)P(\mathrm{break}|i) = 2k_- f(i,t)$. In the special case where $i = 2j$ we have $P(j|\mathrm{break}\, i) = \frac{1}{i-1}$, but then *two* filaments of length j are created and the factor 2 re-enters. Finally, the term proportional to k_n describes nucleation of new filaments of length $j = n_c$.

The master equation defined by equation (1) resembles a very simple model of frangible linear aggregates. The model can easily be evaluated numerically and our main focus, through the rest of this chapter, will be to derive approximate analytical solutions to the length distribution given by $f(j,t)$. By the end of the chapter we return to discussing the validity of this simple model, and we compare the results to experimental data.

Master Equation in Continuous Scheme

In the limit of large j, the discrete system (1) can be approximated by its continuum limit. The discrete field $f(j,t)$ is replaced by at continuous field $\mathbf{f}(x,t)$ such that:

$$\mathbf{f}(x,t) = f(j,t)$$

where x denotes the length of the filaments. If the monomers are associated with a length a, the master equation can be rewritten as a function of the continuous variable $x = ja$. Finite differences are replaced by partial derivatives (Taylor expansion up to first order):

$$\mathbf{f}(x \pm a, t) - \mathbf{f}(x,t) \quad \approx \quad \pm \frac{\partial \mathbf{f}(x,t)}{\partial x} a \tag{2}$$

The sum is replaced by an integral and because we assume $x >> a$ we approximate the integration limit, $(x+a) \approx x$:

$$2k_- \sum_{z=x+a}^{\infty} \mathbf{f}(z,t) \approx \frac{2k_-}{a} \int_x^{\infty} \mathbf{f}(z,t)\,dz \tag{3}$$

Similarly we approximate $(x-a) \approx x$, in the second term:

$$\frac{k_-}{a}(x-a)\mathbf{f}(x,t) \approx \frac{k_-}{a} x\mathbf{f}(x,t) \tag{4}$$

In the continuum limit, it makes good sense, that a filament of length x is able to break anywhere along the continuous length. Therefore rewriting $(x+a) = x$ and $(x-a) = x$ is not only justified because $x >> a$, but also because it corresponds to a scenario, in which the filaments can break anywhere along the length x (also infinitely close to the ends).

Finally, we define $\nu(t) = 2a(k_+ m(t) - k_{off})$, and the continuous master equation describing the time evolution of the length distribution of frangible growing filaments reads:

$$\begin{aligned}\frac{\partial \mathbf{f}(x,t)}{\partial t} &= -v(t)\frac{\partial \mathbf{f}(x,t)}{\partial x} - \frac{k_-}{a}x\mathbf{f}(x,t) \\ &\quad + 2\frac{k_-}{a}\int_x^\infty \mathbf{f}(z,t)\,dz \\ &\quad + k_n m(t)^{n_c}\delta(x - n_c a)\end{aligned} \tag{5}$$

Boundary Conditions

We shall impose the boundary condition $\mathbf{f}(x,t) = 0$ for $x < n_c a$, as filaments of lengths shorter than the nucleus size will become unstable and dissolve into monomers. In order to obtain a continuous distribution at the boundary, we shall also impose the boundary condition $\mathbf{f}(x,t) = 0$ for $x = n_c a$ [26].

We may also write $\mathbf{f}(x,t) = 0$ for $x \to \infty$, as the filaments will not be able to grow indefinitely.

Initial Conditions

We consider scenarios in which the growth process is initialized by seed material - i.e. some relatively small "seed-filaments" are initially present in the system. The initial solution also contains an initial number of monomers. We now define a unitless mass which is measured in terms of monomers, hence the mass of a monomer is 1 and the mass of a filament of length x is equal to x (or in the discrete case equal to j). Initially, the total mass in the solution is given by $m_{tot} = M(0) + m(0)$ where $M(0)$ describes the mass of the seeds and $m(0)$ desccribes the initial mass of the monomers.

Two Different Growth Conditions

We shall consider two different scenarios: open and closed systems.

The open system is a fictitious situation in which the monomer pool is infinitely large, and the monomer concentration is kept constant: $m(t) = m(0)$.

In the closed system we consider a finite volume in which the total mass is conserved. We may therefore write $m_{tot} = M(t) + m(t)$, where $M(t)$ is the total mass of all filaments in the solution at time t. The closed system is the most realistic system which would apply to experiments, whereas the open system is unrealistic but useful to consider in order to gain information about the initial behavior of the system.

1.3 TIME EVOLUTION OF THE MOMENTS

From previous studies [14, 15, 16, 38] we know that it is possible to obtain analytical expressions for the first two principal moments of the length distribution. Here we shall derive the corresponding continuous equations for the moments. We define the number-concentration of filaments (zero'th moment):

$$P(t) = \int_{n_c a}^{\infty} \mathbf{f}(x,t)\,dx$$

and the mass-concentration of filaments (first moment):

$$M(t) = \int_{n_c a}^{\infty} x\mathbf{f}(x,t)\,dx$$

Differential equations for $P(t)$ and $M(t)$ are obtained by carrying out the calculations:

$$\begin{aligned}\frac{dP(t)}{dt} &= \frac{d}{dt}\int_{n_c a}^{\infty} \mathbf{f}(t,x)\,dx \\ \frac{dM(t)}{dt} &= \frac{d}{dt}\int_{n_c a}^{\infty} x\mathbf{f}(t,x)\,dx \end{aligned} \tag{6}$$

The equation for $P(t)$ is obtained as follows:

$$\begin{aligned}\frac{dP(t)}{dt} &= -\nu(t)\underbrace{\int_{n_c a}^{\infty} \frac{\partial \mathbf{f}(t,x)}{\partial x}dx}_{=\mathbf{f}(t,\infty)-\mathbf{f}(t,n_c a))=0} - \frac{k_-}{a}\int_{n_c a}^{\infty} x\mathbf{f}(t,x)\,dx \\ &\quad + \frac{2k_-}{a}\int_{n_c a}^{\infty} dx\int_{x}^{\infty} \mathbf{f}(t,z)\,dz \\ &\quad + k_n m(t)^{n_c}\int_{n_c a}^{\infty} \delta(x-n_c a)\,dx \\ &= 0 - \frac{k_-}{a}M(t) \\ &\quad + \frac{2k_-}{a}\int_{n_c a}^{\infty} dx\int_{x}^{\infty} \mathbf{f}(t,z)\,dz + k_n m(t)^{n_c}\end{aligned}$$

The order of integration is swapped in the following way:

$$\int_{n_c a}^{\infty} dx\int_{x}^{\infty} dz = \int_{n_c a}^{\infty} dz\int_{n_c a}^{z} dx$$

and finally we obtain:

$$\begin{aligned}\frac{dP(t)}{dt} &= -\frac{k_-}{a}M(t) + \frac{2k_-}{a}\int_{n_c a}^{\infty} dz \int_{n_c a}^{z} \mathbf{f}(t,z)dx \\ &\quad + k_n m(t)^{n_c} \\ &= -\frac{k_-}{a}M(t) + \frac{2k_-}{a}\int_{n_c a}^{\infty} (z - n_c a)\mathbf{f}(t,z)dz \\ &\quad + k_n m(t)^{n_c} \\ &= k_-\left(\frac{1}{a}M(t) - 2n_c P(t)\right) + k_n m(t)^{n_c}\end{aligned} \tag{7}$$

Similarly we obtain the following equation for $M(t)$:

$$\begin{aligned}\frac{dM(t)}{dt} &= -\nu(t)\int_{n_c a}^{\infty} x\frac{\partial \mathbf{f}(t,x)}{\partial x}dx - \frac{k_-}{a}\int_{n_c a}^{\infty} x^2\mathbf{f}(t,x)dx \\ &\quad + \frac{2k_-}{a}\int_{n_c a}^{\infty} dx \int_{x}^{\infty} x\mathbf{f}(t,z)dz \\ &\quad + k_n m(t)^{n_c}\int_{n_c a}^{\infty} x\delta(x - n_c a)dx \\ &= -\nu(t)\underbrace{[x\mathbf{f}(t,x)]_{n_c a}^{\infty}}_{=0} + \nu(t)\int_{n_c a}^{\infty}\mathbf{f}(t,x)dx \\ &\quad - \frac{k_-}{a}\int_{n_c a}^{\infty} x^2\mathbf{f}(t,x)dx \\ &\quad + \frac{2k_-}{a}\int_{n_c a}^{\infty} dz\int_{n_c a}^{z} x\mathbf{f}(t,z)dx \\ &\quad + n_c a k_n m(t)^{n_c} \\ &= \nu(t)P(t) - \frac{k_-}{a}\int_{n_c a}^{\infty} x^2\mathbf{f}(t,x)dx \\ &\quad + \frac{2k_-}{a}\int_{n_c a}^{\infty}\frac{1}{2}\left(z^2 - n_c^2 a^2\right)\mathbf{f}(t,z)dz \\ &\quad + n_c a k_n m(t)^{n_c} \\ &= \nu(t)P(t) - k_- n_c^2 a P(t) \\ &\quad + n_c a k_n m(t)^{n_c} \\ &= \left(\nu(t) - k_- n_c^2 a\right)P(t) + n_c a k_n m(t)^{n_c}\end{aligned} \tag{8}$$

In many cases the terms regarding the spontaneous nucleation events (terms proportional to k_n) are relatively small compared to the other

terms in the equations, as long as $P(t)$ and $M(t)$ are not very small. As we shall consider seeded systems $P(t)$ and $M(t)$ are never very small, and we therefore disregard the terms regarding spontaneous nucleation. We now have the following system of equations describing the time evolution of the two first principal moments:

$$\begin{aligned} \frac{dP(t)}{dt} &= k_- \left(\frac{1}{a} M(t) - 2 n_c P(t) \right) \\ \frac{dM(t)}{dt} &= \left(\nu(t) - k_- n_c^2 a \right) P(t) \end{aligned} \tag{9}$$

where $\nu(t) = 2a(k_+ m(t) - k_{off}) = 2a(k_+(m_{tot} - M(t)) - k_{off})$. Especially the entry of $M(t)$ through $\nu(t)$, makes the equations complicated to solve analytically.

1.4 OPEN SYSTEM

At the early stages of the aggregation process the monomer concentration is close to the initial concentration $m(t) \approx m(0)$. In the "Open System Approximation" the monomer concentration is constant: $m(t) = m(0) = m_0$ and the system is therefore expected to describe the early stages of the aggregation process very well. In this scenario we denote the length distribution $\mathbf{f}_0(x,t)$ and the principal moments $P_0(t)$ and $M_0(t)$. The variable $\nu(t)$ also becomes a constant $\nu(t) = 2a(k_+ m_0 - k_{off}) = \nu_0$.

Disregarding spontaneous nucleation events (terms proportional to k_n) the master equation now has the form:

$$\frac{\partial \mathbf{f}_0(x,t)}{\partial t} = -\nu_0 \frac{\partial \mathbf{f}_0(x,t)}{\partial x} - \frac{k_-}{a} x \mathbf{f}_0(x,t) + 2\frac{k_-}{a} \int_x^\infty \mathbf{f}_0(z,t) dz \tag{10}$$

Moments - Open System

The simplification $\nu(t) = \nu_0$ makes it easy to solve the principal moments (given by equations (9)) analytically, and we obtain the following solutions for $P_0(t)$ and $M_0(t)$:

$$\begin{aligned} P_0(t) &= A_1 e^{\kappa_+ t} + A_2 e^{\kappa_- t} \\ M_0(t) &= A_3 e^{\kappa_+ t} + A_4 e^{\kappa_- t} \end{aligned} \tag{11}$$

where

$$\kappa_\pm = -k_- n_c \pm \sqrt{\frac{k_- \nu_0}{a}}$$

and

$$\begin{aligned} A_{1,2} &= \frac{1}{2}\left(P(0) \pm \sqrt{\frac{k_-}{a\nu_0}}\left(M(0) - n_c a P(0)\right)\right) \\ A_{3,4} &= \frac{1}{2}\left(M(0) \pm \sqrt{\frac{ak_-}{\nu_0}} n_c \left(M(0) + n_c a P(0)\right) \pm \sqrt{\frac{\nu_0 a}{k_-}} P(0)\right) \end{aligned}$$

An analytical solution for $\mathbf{f}_0(x,t)$ demands a little more work, and we shall start by considering the limiting behavior before we derive a full solution.

Limiting Behavior of the Open System

In the limit $t \to \infty$ the term with the positive exponent κ_+ will dominate $P_0(t)$:

$$P_0(t) = \int_{n_c a}^{\infty} \mathbf{f}_0(x,t)dx \to A_1 e^{\kappa_+ t} \qquad \text{for} \quad t \to \infty$$

and therefore we deduce that $\mathbf{f}_0(x,t)$ must have the limiting form:

$$\mathbf{f}_0(x,t) \to \mathbf{f}_0^{limit}(x,t) = X(x)e^{\kappa_+ t} \tag{12}$$

where

$$\int_{n_c a}^{\infty} X(x)dx = A_1 \tag{13}$$

Differentiating the master equation (10) with respect to x we obtain the following equation:

$$\begin{aligned} \frac{\partial^2 \mathbf{f}_0^{limit}(x,t)}{\partial x \partial t} &= -\nu_0 \frac{\partial^2 \mathbf{f}_0^{limit}(x,t)}{\partial x^2} \\ &\quad -\frac{k_-}{a}\left(\mathbf{f}_0^{limit}(x,t) + x\frac{\partial \mathbf{f}_0^{limit}(x,t)}{\partial x}\right) \\ &\quad -2\frac{k_-}{a}\mathbf{f}_0^{limit}(x,t) \\ &\Leftrightarrow \\ \kappa_+ X'(x)e^{\kappa_+ t} &= -\nu_0 X''(x)e^{\kappa_+ t} - \frac{k_-}{a}xX'(x)e^{\kappa_+ t} - 3\frac{k_-}{a}X(x)e^{\kappa_+ t} \\ &\Leftrightarrow \\ 0 &= -\nu_0 X''(x) - \left(\kappa_+ + \frac{k_-}{a}x\right)X'(x) - 3\frac{k_-}{a}X(x) \end{aligned} \tag{14}$$

This equation can be solved to obtain the following solution for $X(x)$:

$$\begin{aligned} X(x) &= C_1 \exp\left(-\frac{x}{2\nu_0}\left(\frac{k_- x}{a}+2\kappa_+\right)\right)\left(\left(\frac{k_- x}{a}+\kappa_+\right)^2-\frac{k_-\nu_0}{a}\right) \\ &\quad +C_2\left[-\left(\frac{\frac{k_- x}{a}+\kappa_+}{2\left(\frac{k_-}{a}\right)^2\nu_0}\right)\right. \\ &\quad \left.+\,\mathrm{Da}\left(\frac{\frac{k_- x}{a}+\kappa_+}{\sqrt{2\frac{k_-\nu_0}{a}}}\right)\frac{\left(\left(\frac{k_- x}{a}+\kappa_+\right)^2-\frac{k_-\nu_0}{a}\right)}{\sqrt{2\left(\frac{k_-}{a}\right)^5\nu_0^3}}\right] \end{aligned}$$

where $\mathrm{Da}(u)$ denotes the Dawson integral. The first part of the solution (proportional to C_1) equals zero for $x = n_c a$, is positive for all other x and approaches zero as $x \to \infty$. The second part of the solution (proportional to C_2) is negative for $x = n_c a$ becomes positive for (much) larger values of x and then approaches zero as $x \to \infty$. As we must have $f(n_c a, t) = 0$ and $f(x, t) \geqslant 0$ for all $x > n_c a$, we must put $C_2 = 0$. We now define the limiting distribution for the open system

$$\begin{aligned} \mathbf{f}_0^{\mathrm{limit}}(x,t) &= X_0^{\mathrm{limit}}(x)e^{\kappa_+ t} \\ &= C_1 e^{\kappa_+ t} e^{\left(-\frac{x}{2\nu_0}\left(\frac{k_- x}{a}+2\kappa_+\right)\right)}\left(\left(\frac{k_- x}{a}+\kappa_+\right)^2-\frac{k_-\nu_0}{a}\right) \end{aligned} \tag{15}$$

where

$$X_0^{\mathrm{limit}}(x) \equiv C_1 e^{\left(-\frac{x}{2\nu_0}\left(\frac{k_- x}{a}+2\kappa\right)\right)}\left(\left(\frac{k_- x}{a}+\kappa\right)^2-\frac{k_-\nu_0}{a}\right) \tag{16}$$

and (from equation (13)) the constant C_1 is given by :

$$C_1 = \frac{A_1 e^{\frac{n_c a(k_- n_c+2\kappa_+)}{2\nu_0}}}{a^2\nu_0(k_- n_c+\kappa_+)}$$

Full Solution for the Open System

The underlying idea for obtaining a full solution for $\mathbf{f}_0(x,t)$ comes from the use of Picard's iteration method, that allows self-consistent solutions with increasing precision to be derived from an initial guess for $\mathbf{f}_0(x,t)$. Here, we only perform a single iteration, and use $\mathbf{f}_0^{\mathrm{limit}}(x,t)$ as an approximation for $\mathbf{f}_0(x,t)$ in the integral term of equation (10):

$$\frac{\partial \mathbf{f}_0(x,t)}{\partial t} \approx -\nu_0 \frac{\partial \mathbf{f}_0(x,t)}{\partial x} - k_- x \mathbf{f}_0(x,t) + 2k_- \int_x^\infty \mathbf{f}_0^{limit}(z,t)\,dz \tag{17}$$

By substituting:

$$\int_x^\infty \mathbf{f}_0^{limit}(z,t)\,dz = C_1 \exp\left(\kappa_+ t - \frac{k_- x^2 + 2ax\kappa_+}{2a\nu_0}\right) \nu\left(\frac{k_- x}{a} + \kappa_+\right)$$

into equation (17) we obtain the approximate result for $\mathbf{f}_0(x,t)$:

$$\mathbf{f}_0(x,t) = \mathbf{f}_0^{limit}(x,t) + \exp\left(-\frac{k_- x^2}{2a\nu_0}\right) \Psi(x - \nu_0 t)$$

where $\Psi(u)$ is an arbitrary function, which must be determined from the initial condition. Assuming that the initial condition corresponds to seeds of a relatively well defined length, we can approximate the initial distribution as a narrow gaussian centered around the mean seed length x_0. Our initial condition is therefore given by the expression:

$$\mathbf{f}_0(x,0) = \frac{P_0(0)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)$$

and we now solve the equation:

$$\mathbf{f}_0(x,0) = \mathbf{f}_0^{limit}(x,0) + \exp\left(-\frac{k_- x^2}{2a\nu_0}\right) \Psi(x)$$

and obtain the following solution for $\Psi(x - \nu_0 t)$:

$$\Psi(u) = \exp\left(\frac{k_-(u)^2}{2a\nu_0}\right) \times \left[\frac{P_0(0)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(u-x_0)^2}{2\sigma^2}\right) - X_0^{limit}(u)\right]$$

The function $\Psi(u)$ consists of a "peak" (the gaussian term) and a "background" (the term including $X_0^{limit}(u)$) times a pre-factor which sets the amplitude:

$$\Psi(u) = \text{"prefactor"} \times [\text{"peak"} - \text{"background"}]$$

When time increases the overall shape of $\Psi(x - \nu_0 t)$ is preserved, and everything simply shifts to higher x-values with speed ν_0. We note that, since the "background"-term (and hence also $\Psi(x - \nu_0 t)$) diverges towards $-\infty$ for negative $(x - \nu t)$, we multiply this term

by a heavisidefunction $\theta(x - v_0 t)$, in order to avoid errors for times $t > 0$:

$$\Psi(x - v_0 t) = \exp\left(\frac{k_-(x - v_0 t)^2}{2av}\right) \times \left[\frac{P_0(0)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - x_0 - v_0 t)^2}{2\sigma^2}\right) - \theta(x - v_0 t) X_0^{limit}(x - v_0 t)\right] \tag{18}$$

and finally we obtain the final full analytical solution for the length distribution $f_0(x, t)$ in the open system:

$$\mathbf{f}_0(x, t) = \mathbf{f}_0^{limit}(x, t) + \exp\left(-\frac{k_- x^2}{2av_0}\right) \Psi(x - v_0 t) \tag{19}$$

where $\mathbf{f}_0^{limit}(x, t)$ is given by equation (15) and $\Psi(x - v_0 t)$ is given by equation (18).

The approximate solution (19), gives a very good fit to the numerical solution of the master equation - see Figure 2. The two terms of $\mathbf{f}_0(x, t)$ can be understood individually: the term proportional to the limiting distribution $\mathbf{f}_0^{limit}(x, t)$ is a relatively broad distribution, that increases exponentially with time (Figure 2B). Secondly, the term proportional to $\Psi(x - v_0 t)$ consists mainly of the relatively narrow peak, which moves towards larger lengths with speed v_0 (Figure 2C). The prefactor, $\exp\left(-\frac{k_- x^2}{2av_0}\right)$, ensures that the amplitude of the "Ψ"-term decreases as it shifts to higher x-values. As $t \to \infty$, $\mathbf{f}_0(x, t)$ approaches $\mathbf{f}_0^{limit}(x, t)$.

Note About Diffusion in Length Space

The expression for $\Psi(x - v_0 t)$ is derived from the master equation (10), which we initially obtained by approximating finite differences with partial derivatives which is equivalent of taylor expanding up to first order (recall equation (2)). If we had instead used the taylor expanded up to second order we would instead have the following approximation for finite differences:

$$\mathbf{f}(x \pm a, t) - \mathbf{f}(x, t) \approx \pm\frac{\partial \mathbf{f}(x, t)}{\partial x} a + \frac{1}{2}\frac{\partial^2 \mathbf{f}(x, t)}{\partial x^2} a^2 \tag{20}$$

And by defining $v(t) = 2a(k_+ m(t) - k_{off})$ (as before) and $D(t) = a^2(k_+ m(t) + k_{off})$ the continuous mean-field master equation would have read:

$$\frac{\partial \mathbf{f}(x, t)}{\partial t} = -v(t)\frac{\partial \mathbf{f}(x, t)}{\partial x} + D(t)\frac{\partial^2 \mathbf{f}(x, t)}{\partial x^2} - \frac{k_-}{a} x\mathbf{f}(x, t) + 2\frac{k_-}{a}\int_x^\infty \mathbf{f}(z, t)\,dz + k_n m(t)^{n_c}\delta(x - n_c a) \tag{21}$$

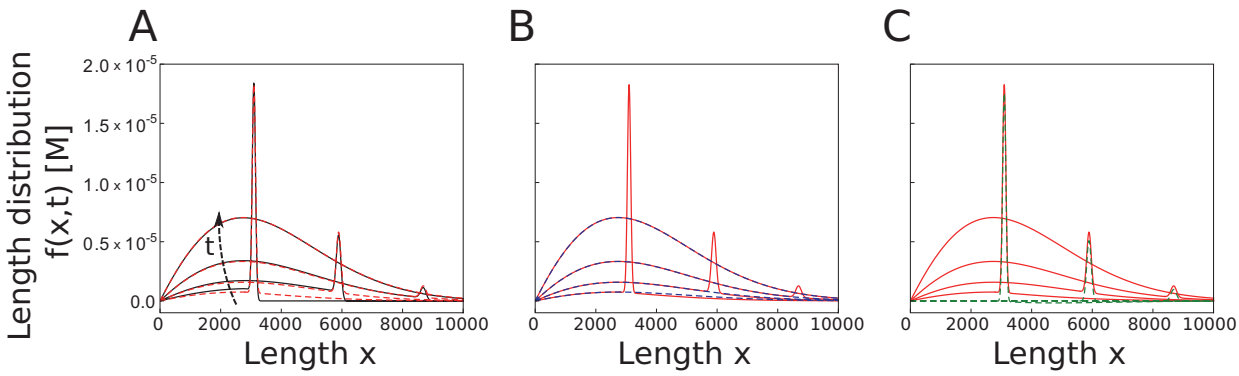


Figure 2: Time evolution of the fibril length distribution in an open system $\mathbf{f}_0(x,t)$. A: The black solid line is the numerically calculated solution of the master equation (10). The red dashed line is the predicted length distribution for the constant monomer case given in equation (19). B: The first term of equation (19), $f_0^{limit}(x,t)$ (dashed blue line) is compared with equation (19) (solid red line). B: The contribution from the second term in equation (19) describing the advection of the initial distribution (dashed green line) is compared with equation (19) (solid red line). The parameters used are: $k_+ = 3.5 \times 10^{-2}$ M^{-1}minutes^{-1}, $k_- = 5 \times 10^{-6}$ minutes^{-1}, $k_{off} = 0$, $n_c = 2$, $x_0 = 300$, $\sigma_0 = 10$, $M(0) = 1$ M, $P(0) = M(0)/x_0$, $m_0 = 999$. Curves are shown for the following times (from bottom to top): $t = 40, 80, 120, 160$ minutes.

from which it can be seen that the term proportional to $D(t)$ describes diffusion in length space. In order to account for the error that this assumption might introduce we re-introduce diffusion by substituting $\sigma^2 \rightarrow \sigma_0^2 + 2D_0 t$, which equals the behavior we expect from solving the system consisting only of drift and diffusion.

Discussion of the Analytical Solution in the Open System

The expression given by equation (19) describes in closed form the time evolution of the length distribution of an open system characterised by nucleated polymerisation and fragmentation. A comparison between the analytical expression provided by equation (19) and the numerical solution of the master equation (10) is shown in Figure 2. The analytical expression gives the correct shape of the filament distribution and matches the numerical solution for large times exactly. As it can be seen from the plot, for early times equation (19) overestimates the number of fibres with length greater than $x_0 + v_0 t$. Qualitatively, this effect arises because the filament distribution entering the fragmentation source term (in equation (17)) is from the late time regime and introduces long fibres that would not have time to grow in the period of time considered, and therefore the aggregates population is overestimated for large x. Figure 2B and 2C show the contributions from the two terms in equation (19) to the shape of the filament distribution. The first term of equation (19) is an exponentially growing biased Gaussian distribution that accounts for the increase in the population of short fibres caused by the fragmentation of longer ones. The second term of equation (19) describes the advection in size space of the initial filament distribu-

tion. As a result of fibril elongation, the Gaussian peak moves in size space with velocity v_0 and spreads out with diffusion coefficient D_0. As $t \to \infty$ the advective Gaussian peak vanishes and equation (19) recovers $f_0^{limit}(x,t)$ completely. As the fibres increase in length, their fragmentation probability increases and the filament distribution is shifted from large to small lengths. This behaviour is captured by the exponential pre-factor which ensures that the amplitude of the advective Gaussian peak decreases with increasing length x.

1.5 CLOSED SYSTEM

For the closed system we can not make the approximation $v(t) = v_0$, but must use the full expression $v(t) = 2a(k_+ m(t) - k_{off} = 2a(k_+(m_{tot} - M(t)) - k_{off})$. This makes the differential equations describing the principal moments (9) and the partial differential equation describing the length distribution (5) much more complicated.

Moments - Closed System

From equations (9) we obtain the following the formal solutions for $P(t)$ and $M(t)$:

$$\begin{aligned} P(t) &= P(0)e^{-2k_- n_c t} + \frac{k_-}{a}\int_0^t e^{-2k_- n_c (t-\tau)} M(\tau)\,d\tau \\ M(t) &= M(0)e^{-2k_+ \int_0^t P(\tau)\,d\tau} + M_\infty \left(1 - e^{-2k_+ \int_0^t P(\tau)\,d\tau}\right) \end{aligned} \tag{22}$$

where M_∞ is defined as described below. As previously described[16, 38] solutions for $P(t)$ and $M(t)$ can also be obtained using Picard's iteration method: by re-writing equations (22) in a matrix form:

$$\begin{pmatrix} P(t) \\ M(t) \end{pmatrix} = A \begin{pmatrix} M(t) \\ P(t) \end{pmatrix}$$

we can obtain approximate solution by iterating the equations in the following way: the solutions $P_0(t)$ and $M_0(t)$ (11) are substituted into the right hand side to obtain the first iteration $P_1(t)$ and $M_1(t)$:

$$\begin{pmatrix} P_1(t) \\ M_1(t) \end{pmatrix} = A \begin{pmatrix} M_0(t) \\ P_0(t) \end{pmatrix}$$

Again $P_1(t)$ and $M_1(t)$ are substituted into the right hand side to obtain the second iteration $P_2(t)$ and $M_2(t)$:

$$\begin{pmatrix} P_2(t) \\ M_2(t) \end{pmatrix} = A \begin{pmatrix} M_1(t) \\ P_1(t) \end{pmatrix}$$

The solutions will approach the true solutions of (22) as the iterations are repeated:

$$P_n(t) \to P(t) \quad \cap \quad M_n(t) \to M(t) \qquad \text{for} \quad n \to \infty$$

It turns out that $M_1(t)$ and $P_2(t)$ are already very good approximations to the true solutions for $M(t)$ and $P(t)$ [16, 38]. The solution for $M_1(t)$ is obtained as follows:

$$\begin{aligned} M_1(t) &= M(0)e^{-2k_+\int_0^t P_0(\tau)d\tau} + M_\infty\left(1 - e^{-2k_+\int_0^t P_0(\tau)d\tau}\right) \\ &= (M(0) - M_\infty)e^{-2k_+\int_0^t P_0(\tau)d\tau} + M_\infty \\ &= (M(0) - M_\infty)e^{-2k_+\left(\frac{A_1}{\kappa_+}(e^{\kappa_+ t}-1)+\frac{A_2}{\kappa_-}(e^{\kappa_- t}-1)\right)} + M_\infty \end{aligned} \tag{23}$$

The solution for $P_2(t)$ is obtained by substituting equation (23) into (22):

$$\begin{aligned} P_2(t) &= P(0)e^{-2k_- n_c t} + \frac{k_-}{a}\int_0^t e^{-2k_- n_c(t-\tau)} M_1(\tau)d\tau \\ &= e^{-2k_- n_c t}\left(P(0) + \frac{k_-}{a}\int_0^t e^{2k_- n_c \tau} M_1(\tau)d\tau\right) \\ &= e^{-2k_- n_c t}\left(P(0) + \frac{k_-}{a}\int_0^t e^{2k_- n_c \tau} M_\infty d\tau\right. \\ &\quad \left. + \frac{k_-}{a}\int_0^t e^{2k_- n_c \tau - 2k_+\left(\frac{A_1}{\kappa_+}(e^{\kappa_+\tau}-1)+\frac{A_2}{\kappa_-}(e^{\kappa_-\tau}-1)\right)}(M(0) - M_\infty)d\tau\right) \end{aligned}$$

The linear and exponentially decaying terms are neglected in front of the growing exponential and the final solution for $P_2(t)$ is obtained as follows:

$$\begin{aligned} P_2(t) &= e^{-2k_- n_c t}\left(P(0) + \frac{k_-}{a}\int_0^t e^{2k_- n_c \tau} M_\infty d\tau\right. \\ &\quad \left. + \frac{k_-}{a}\int_0^t e^{\frac{-2k_+ A_1}{\kappa_+} e^{\kappa_+\tau}}(M(0) - M_\infty)d\tau\right) \\ &= e^{-2k_- n_c t}\left(P(0) + \frac{M_\infty\left(e^{2k_- n_c t} - 1\right)}{2an_c}\right. \\ &\quad \left. + \frac{k_-\left(M_\infty - M(0)\right)}{a\kappa_+}\left[\mathrm{Ei}\left(\frac{-2A_1 k_+}{\kappa_+}\right) - \mathrm{Ei}\left(\frac{-2A_1 k_+ e^{\kappa_+ t}}{\kappa_+}\right)\right]\right) \end{aligned} \tag{24}$$

where $\mathrm{Ei}(z) = -\int_{-z}^{\infty}\frac{e^{-t}}{t}dt$ is the exponential integral function.

The expressions for $M_1(t)$ and $P_2(t)$ given by equations (23) and (24) are plotted in Figure 3, together with the analytical expressions for $M_0(t)$ and $P_0(t)$ (11), as well as the numerical solutions obtained from the master equation (5).

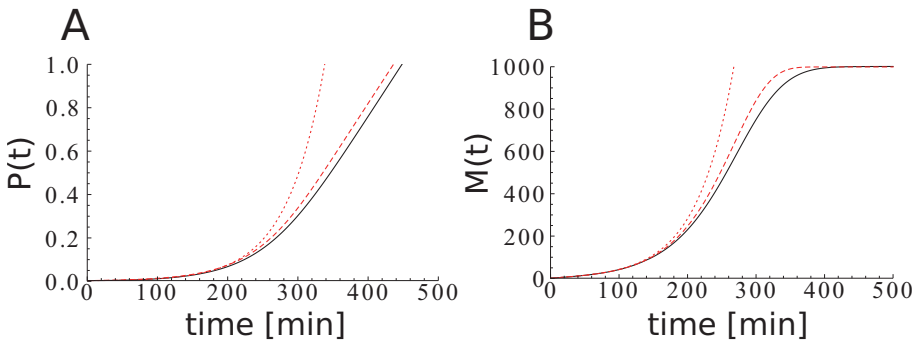


Figure 3: Time evolution of $P(t)$ and $M(t)$. A: Numerical solution for $P(t)$ obtained from the master equation (5) (solid black line), analytical solution to $P_2(t)$ (24) (dashed red line) and to the open system $P_0(t)$ (11), which is also a good approximation for early times (dotted red line). B: Numerical solution for $M(t)$ obtained from the master equation (5) (solid black line), analytical solution to $M_1(t)$ (23) (dashed red line) and to the open system $M_0(t)$ (11), which is a good approximation for early times (dotted red line). The parameters used are: $k_+ = 3.5 \times 10^{-2}$ M^{-1}minutes^{-1}, $k_- = 5 \times 10^{-6}$ minutes^{-1}, $k_{off} = 0$, $n_c = 2$, $M(0) = 1$ M, $P(0) = M(0)/300$, $m_{tot} = 1000$.

Limiting Behaviour of the Closed System

In the limit $t \to \infty$ the system will approach the steady state distribution $\mathbf{f}_\infty(x)$. We define the steady state solutions for $P(t)$ and $M(t)$:

$$\begin{aligned} P_\infty &= \frac{2k_+ m_{tot} - 2k_{off} - k_- n_c^2}{4k_+ n_c a} \\ M_\infty &= \frac{2k_+ m_{tot} - 2k_{off} - k_- n_c^2}{2k_+} \end{aligned} \tag{25}$$

and also $m_\infty = \frac{2k_{off}+k_- n_c^2}{2k_+}$ and $\nu_\infty = k_- n_c^2 a$. We note that the steady state is not a true equilibrium but a dynamic steady state in which the growth is balanced by fragmentation such that the length distribution is constant.

We solve the length distribution from the following equation:

$$0 \quad = \quad -\nu_\infty \mathbf{f}''_\infty(x) - k_- x \mathbf{f}'_\infty(x) - 3k_- \mathbf{f}_\infty(x)$$

with the result:

$$\mathbf{f}_\infty(x) \quad = \quad D_1 e^{\left(\frac{-x^2}{2(n_c a)^2}\right)} \left(x^2 - (n_c a)^2\right) \tag{26}$$

The constant D_1 is found by normalizing the total number of filaments to P_∞ $\int_{n_c a}^{\infty} \mathbf{f}_\infty(x)dx = P_\infty$:

$$D_1 = \frac{\sqrt{e}(2k_+ m_{tot} - 2k_{off} - k_- n_c^2)}{4k_+ (n_c a)^4}$$

Full Solution to the Closed System

We expect that in the early stages of the polymerisation process, the length distribution will evolve according to $\mathbf{f}_0(x,t)$, obtained in the

presence of a constant monomer pool, shifting at later times into the steady-state distribution given by $\mathbf{f}_\infty(x)$.

An interesting observation to be drawn from $\mathbf{f}_0^{limit}(x,t)$ and $\mathbf{f}_\infty(x)$, is that the distribution of fibril lengths share the same functional form at the beginning and at the end of the reaction (except for the advective peak, which is only present at early stages). In both cases the form of $\mathbf{f}(t,x)$ is given in terms of the function $X^{limit}(x)$, where the values $\nu(0)$ and $\kappa_+(0)$ are used in the early time limit and the values $\nu(\infty)$ and $\kappa_+(\infty)(=0)$ in the steady state.

Thus, an appropriate ansatz for $\mathbf{f}(t,x)$ is given by

$$\mathbf{f}(x,t) = A(t)X_0^{approx}(x) + \exp\left(-\frac{k_- x^2}{2a\nu_0}\right)\Phi\left(x-\tau(t)\right) \tag{27}$$

where $X_0^{approx}(x)$ is obtained by replacing $\nu(0)$ in equation (16) by some value ν^{approx} to be determined below:

$$X_0^{approx}(x) = X_0^{limit}(x, \nu(0) \to \nu^{approx}) \tag{28}$$

The function Φ is given by

$$\Phi(\xi) = \exp\left(\frac{k_-\xi^2}{2\nu(0)}\right) \times \left[\frac{P(0)}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\xi-x_0)^2}{2\sigma^2}\right)\right] \tag{29}$$

where $\sigma = \sigma_0 + 2D_0 t$. The amplitude $A(t)$ is calculated from the normalisation condition

$$A(t) = \frac{P(t)}{\int_{n_c a}^{\infty} X_0^{approx}(x)dx} \tag{30}$$

and $\tau(t)$ describes the position of the peak of the initial distribution at time t, as a result of advection in length space:

$$\begin{aligned}\tau(t) = \int_0^t \nu(t')dt' &= 2(k_+ m_{tot} - k_+ M(\infty) - k_{off})t \\ &- 2k_+\frac{M(0)-M(\infty)}{\kappa_+}\left[\mathrm{Ei}\left(-C_+ e^{\kappa_+ t}\right) - \mathrm{Ei}\left(-C_+\right)\right]\end{aligned} \tag{31}$$

Equation (28) corrects the late time behaviour of the length distribution by choosing ν^{approx} such that the solution reproduces the correct mean value $\mu(t) = M(t)/P(t)$ of the length distribution at all times. Using:

$$\mu(t) = \frac{\int_{n_c a}^{\infty} x X_0^{approx}(x)dx}{\int_{n_c a}^{\infty} X_0^{approx}(x)dx} = n_c a + \sqrt{\frac{k_-\nu^{approx}}{a}} \tag{32}$$

we have the following expression for ν^{approx}:

$$\nu^{approx} = \frac{k_-}{a}\left(\frac{M(t)}{P(t)} - n_c a\right)^2 \tag{33}$$

where $M(t)$ and $P(t)$ are given by equations (24) and (23).

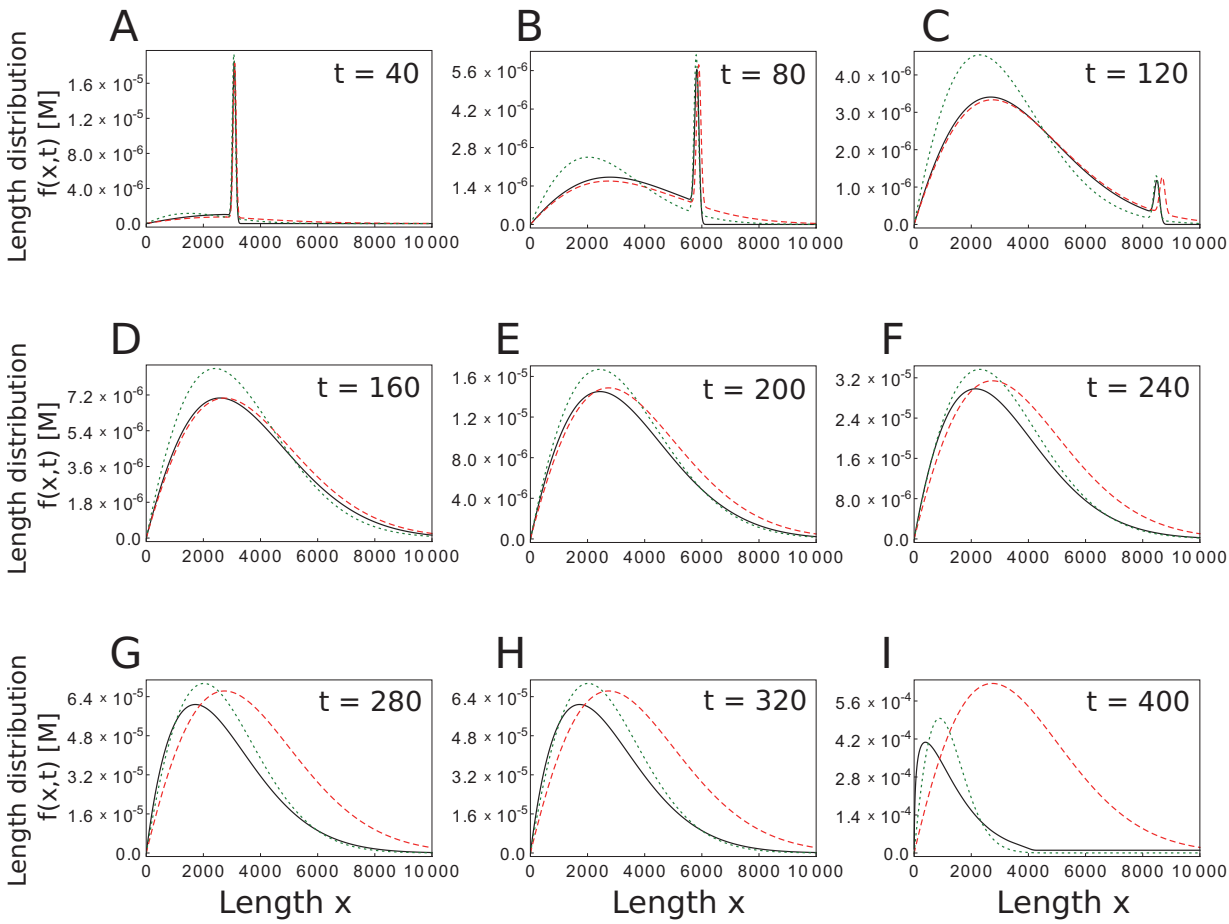


Figure 4: Time evolution of the fibril length distribution in a closed system. The black solid line is the numerically calculated solution of the master equation (1). The red dashed line is the predicted length distribution $\mathbf{f}_0(x,t)$ for the constant monomer case given by equation (19). The green dotted line is the predicted length distribution $\mathbf{f}(x,t)$ for the constant mass given in by equation (27). The parameters used are: $k_+ = 3.5 \times 10^{-2}$ M^{-1}minutes^{-1}, $k_- = 5 \times 10^{-6}$ minutes^{-1}, $k_{off} = 0$, $n_c = 2$, $x_0 = 300$, $\sigma_0 = 10$, $M(0) = 1$ M, $P(0) = M(0)/x_0$, $m_{tot} = 1000$. The times of the snapshots are: $t = 40, 80, 120, 160, 200, 240, 280, 320, 400$ minutes.

Discussion of the Analytical Solution in the Closed System

The expression given by equation (27) describes in closed form the time evolution of the length distribution of a closed system characterised by nucleated polymerisation and fragmentation. A comparison between the analytical expression provided by equation (27) and the numerical solution of the master equation (5) is shown in Figure 4, where we also plot the analytical solution to the open system (19). While the analytical expression for the open system (19) provides a better solution for early times, the analytical expression for closed systems (27) provides a better solution for late times and captures the overall behaviour of the length distribution.

At early times the aggregation process is dominated by elongation, and the length distribution evolves towards large lengths x, where the advective peak describes a large part of the population (Figure

4A-C). At later times the aggregation process is dominated by fragmentation and the population described by the advective peak disappears (Figure 4C-D). At even later times the length distribution shifts to shorter lengths x, as the filaments continue to break (Figure 4E-I). At these late times, the analytical expression for the open system (19), does no longer provides a good approximation, whereas the analytical expression for closed systems (27) provides a good approximation, although not perfect.

1.6 FIT TO EXPERIMENTAL DATA

Insulin filament length distributions were measured experimentally from seeded fibril growth experiments, monitored by changes in fluorescence and subsequently measured using AFM imaging of the filaments at different times. The experiments showed that, starting from the initial filament distribution, filaments grew to to form longer filaments after 90 minutes, and then, after 190 minutes, it appeared that the length distribution shifted towards shorter filaments. The experimental results for the length distribution of the insulin fibres are plotted in Figure 5.

In order to compare the model predictions of equation (27) with the AFM data, we first fitted the measured total mass concentration $M(t)$ of fibres to the theoretically predicted curve given by equation (23) to obtain values for the rate constants for elongation and breakage - see inset of Figure 5A. The obtained parameter values of $k_+ = 5.5 \times 10^{-2}$ M^{-1}minutes^{-1} and $k_- = 7 \times 10^{-6}$ minutes^{-1}, were subsequently used to produce both an analytical fit given by equation (27) and a numerical fit from the master equation (5) - see Figure 5.

The overall agreement between the model and the experimental data is very good, considering the parameters fit the time evolution of both the total mass of filaments and the length distribution. The experimental and theoretical length distributions both reflect a filament population, which initially grows to span a large range of lengths, before at later times, shifts to a narrower distribution at shorter lengths. Hence the filament do indeed seem to break, and the fragmentation model is therefore to some extend confirmed (recall that fragmentation was initially proposed, based solely on experimental data for the growth of the total mass of filaments).

Our results do however, also show considerable discrepancies between the experimental and theoretical length distributions. Especially the advective peak predicted by the theory is not visible in the experimental data - see Figure 5B. Instead the experimental data displays a larger build-up of shorter lengths at early times. On the other hand, the theoretical predictions for late times, seems to overestimate the extend to which the filaments break, and the theoretical length distribution shifts to shorter lengths, than what is observed experimentally. A possible explanation for these discrepan-

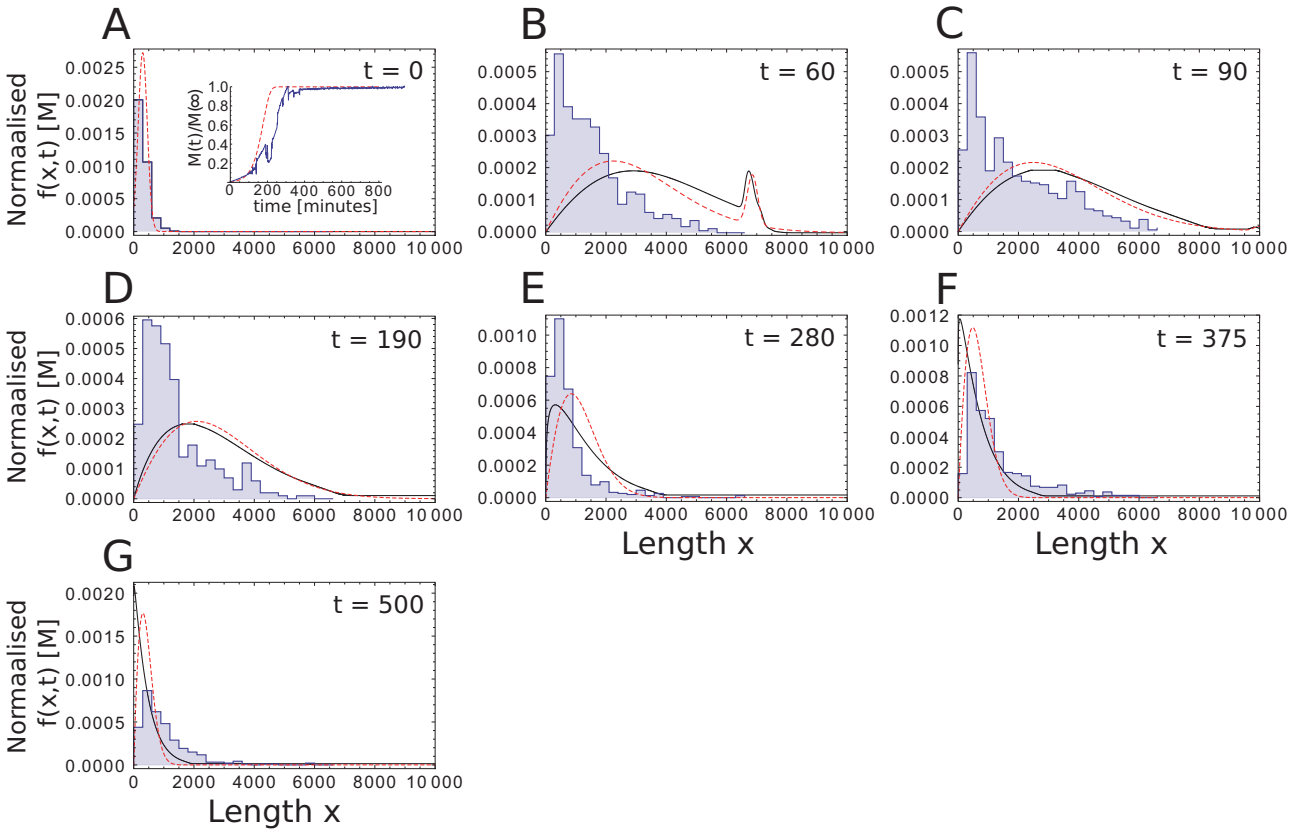


Figure 5: Predicted length distributions at 0, 60, 90, 190, 280, 375 and 500 minutes. In each plot, the columns represent the experimental data of insulin filaments with histogram bin widths of 300, the solid black curves represent the theoretical predictions from the numerical solution of the master equation (5), while the dashed red curves represents the prediction of equation (27). The parameters used were those that correspond with the fit of the polymer mass concentration in the inset of panel A: $k_+ = 5.5 \times 10^{-2}$ M^{-1}minutes^{-1}, $k_- = 7 \times 10^{-6}$ minutes^{-1}, $k_{off} = 0$, $n_c = 2$, $x_0 = 300$, $\sigma_0 = 150$, $M(0) = 1$ M, $P(0) = M(0)/x_0$, $m_{tot} = m(0) + M(0) = 1000$.

cies, could be that the fragmentation rate is not constant during the entire aggregation process. Both discrepancies could arise as a result of a fragmentation rate, which is more frequent in the early stages and less frequent in the late stages - as would be highly imaginable for growing filaments, which are more fragile in the initial stages of the growth, but matures and stabilises at later times.

2

LOGISTIC-LIKE GROWTH AND MULTIPLE AGGREGATE STRUCTURES OF p25α

In this chapter I shall present a mathematical model for the aggregation of the protein p25α influenced by the presence of different concentrations of heparin. The model applies a logistic-like growth, which is motivated in the beginning of the chapter, but the main purpose of the model is to capture the effects of different concentrations of heparin. Low, intermediate and high heparin concentrations were experimentally observed to induce formation of structurally different aggregate forms, and we solve the mathematical model numerically in order to fit the experimental data for p25α aggregation.

RELEVANT PUBLICATION:
Søren Bang Nielsen, **Pernille Yde**, Lise Giehm, Sabrina Sundbye, Gunna Christiansen, Joachim Mathiesen, Mogens Høgh Jensen, Poul Henning Jensen and Daniel E. Otzen. Multiple Roles of Heparin in the Aggregation of p25α. *Journal of Molecular Biology*, 421, 601-615 (2012)

2.1 GROWTH OF DIFFERENT AGGREGATE STRUCTURES

As mentioned in the previous chapter, the growth of amyloid fibres has been proposed to be dominated by *secondary nucleation events*. It was previously argued that lateral growth (branching) describes the secondary pathway better than fragmentation [8], i.e. the accelerated self-enhanced growth of amyloid fibres may arise due to the spatial structure of the aggregates.

The growth-rate of linear filaments is proportional to the total number of filaments, as the linear filaments grow through addition of monomers onto the ends of the filaments. One could also consider a growth scenario in which free monomers could add onto *any* monomer already embedded in the aggregate. In this case the growth-rate would be proportional to the total mass of the aggregates. For simplicity we shall for a moment consider a single growing aggregate of mass $M(t)$. The growth of a linear filament would be proportional to the number of available monomers: $m(t) = m_{tot} - M(t)$, where m_{tot} is the total number of monomers in the system:

$$\frac{dM_{linear}}{dt} \propto (m_{tot} - M_{linear}) \tag{34}$$

An aggregate which could add free monomers onto any monomer in the aggregate would grow in a logistic-like fashion:

$$\frac{dM_{logistic}}{dt} \propto M_{logistic}(m_{tot} - M_{logistic}) \tag{35}$$

but if we assume the free monomers add onto the aggregate through physical contact, such a growth scenario is unrealistic, as part of the monomers within the aggregate would be physically shielded from contact with free monomers.

The growth-rate of a spherical aggregate would be proportional to the surface area of the aggregate, which is again proportional to $M(t)^{2/3}$:

$$\frac{dM_{sphere}}{dt} \propto M_{sphere}^{2/3}(m_{tot} - M_{sphere}) \tag{36}$$

Following this line of thought one could image, that the overall growth-rate of an aggregate, is proportional to the mass to the power of some exponent α, depending on the spatial structure of the aggregate:

$$\frac{dM}{dt} \propto M^{\alpha}(m_{tot} - M) \tag{37}$$

where $\alpha = 0$ equals the linear growth scenario, $\alpha = 2/3$ equals the spherical growth scenario and $\alpha = 1$ equals the logistic growth scenario.

In Figure 6 the mass $M(t)$ is plotted as a function of time, for different exponents α. We note that the growth scenarios with $\alpha > 0$, captures the accelerated growth and the initial *lag-time*, which is typical for amyloid fibres [5, 8, 24, 25, 38, 62]. We also note that the growth curve for exponent $\alpha = 1$ is symmetrical, while the growth curves for exponents $0 < \alpha < 1$ are asymmetrical, which is also a typical characteristic for amyloid growth.

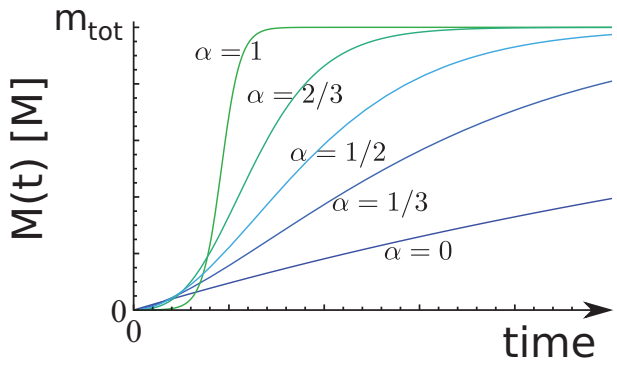


Figure 6: The mass, $M(t)$, of a single growing aggregate as a function of time, given by equation (37), for different exponents α.

Logistic-like Growth

Although we have argued above, that the logistic-like growth with $dM/dt \propto M(m_{tot} - M)$, is somewhat unrealistic, logistic-like growth may still be regarded as a good first approximation of mass dependent (and therefore self-enhanced) growth. Additionally, one can imagine many other scenarios, that result in logistic-like growth, which are not only dependent on the spatial structure of the aggregates. Here, we shall assume that the p25α fibres aggregate in a linear structure (see Figure 7, top right) and that the elongation itself is dependent on the already aggregated mass. We justify this assumption by imagining, that the free monomers are influenced to unfold or are in other ways primed for aggregation upon physical contact with the existing aggregate.

2.2 p25α and heparin

The 219-residue protein p25α (tubulin polymerization-promoting protein, also known as TPPP) is known to stimulate the aggregation of amyloid filaments and co-localise to the protein α-synuclein which under disease conditions forms the amyloid filaments known as Lewy Bodies. Aggregates of p25α have been observed in relation to disease *in vivo*, but p25α does not fibrillate by itself under native conditions *in vitro*. The aggregation of p25α can be induced by certain environmental conditions as e.g. the presence of polyanionic biomolecules such as heparin, polyglutamate, arachidonic acid micelles and RNA. Here we focus on the effect of heparin on the aggregation of p25α.

Heparin is a highly sulfated glycosaminoglycan which is known to promote fibrillation of several proteins including p25α. Experiments have shown that p25α does not fibrillate to any significant extent in the absence of heparin, but intermediate heparin concentrations lead to fibrillar aggregates, whereas large heparin concentrations lead to less structured amorphous aggregates. These findings were further supported by experimental ThT flourescence measurements for different heparin concentrations. ThT (thioflavin T) is a fluorescent chemical dye which is widely used to visually quantify the amount of protein aggregates and the ThT signal is assumed to be linearly proportional to the total mass of fibrillar aggregates, while amorphous aggregates are not expected to produce a significant ThT signal.

Experimental Findings

Experiments showed that solutions of pure p25α did not lead to ThT-positive aggregates of any significant extent whereas addition of heparin to the protein solution had dramatical impact on the ThT signal. As little as 0.4 μg/mL had an impact and led to an increase in

ThT flourescence after approximately 16 hours of incubation. With increasing heparin concentrations of up to 10 µg/mL, the lag-time is further shortened to 8-10 hours and the end-point ThT fluorescence also increases - clearly indicating that heparin induces aggregation of p25α in a concentration dependent manner. At heparin concentrations above 10 µg/mL, the lag-time further decreases to 1-3 hours, but although there is a spike in the ThT fluorescence around 10 hrs, the endpoint ThT intensity rapidly decays to a level corresponding to the absence of heparin and the observations therefore suggest an optimal heparin concentration for the induction of amyloid fibres around 10 µg/mL - see Figure 7. AFM images also revealed that the aggregates formed at heparin concentration 10 µg/mL, yielded linear fibrous aggregates, whereas at heparin concentration 100 µg/mL yielded aggregates of a more amorphous structure (Figure 7).

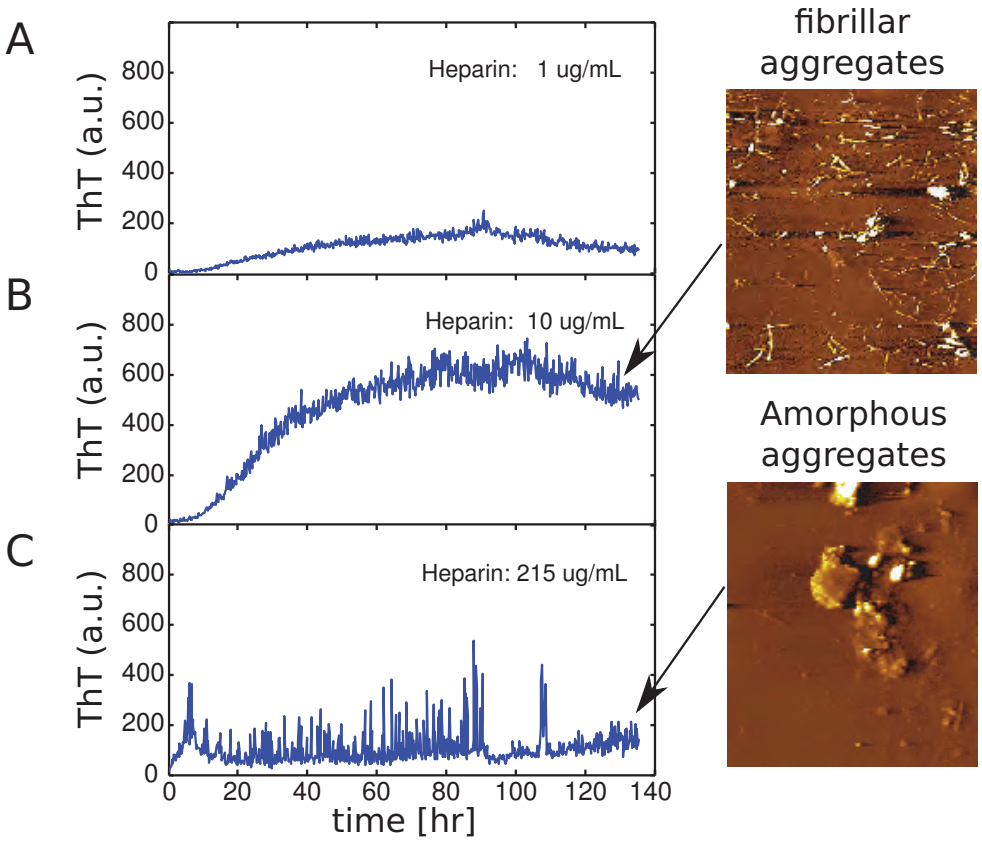


Figure 7: Representative examples of experimental measurements of ThT intensity over time, for different concentration of added heparin plus AFM pictures of the resulting aggregates. The experiments indicate an optimal heparin concentration for fibril production around 10µg/mL, and an existence of amorphous aggregates formed at higher heparin concentrations.

2.3 MATHEMATICAL MODEL

A mathematical model of the aggregation kinetics of p25α and heparin was constructed - see sketch in Figure 8. The model was con-

tructed to capture the biphasic feature of the aggregation process, namely, that p25α can form amyloid fibres at intermediate heparin concentrations, while high heparin concentrations lead to amorphous aggregates after a transient build-up of amyloid material. Accordingly, the model consists of three basic reactions: (i) nucleation of p25α onto heparin, (ii) logistic-like growth through addition of p25α-monomers onto the existing fibres, and (iii) heparin-dependent formation of an amorphous aggregate through addition of an extra heparin molecule to the existing p25α:heparin fibre. For simplicity we assume that all reactions are effectively irreversible.

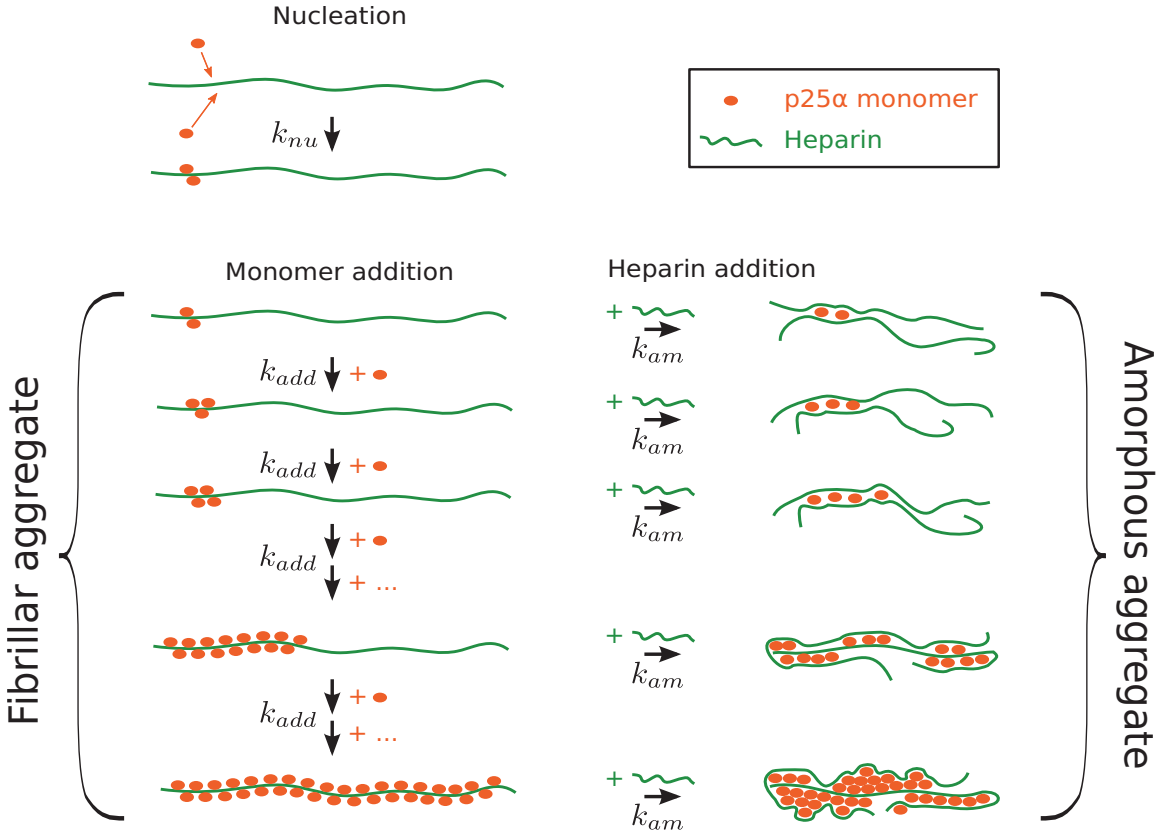


Figure 8: Sketch of the model for aggregation of fibrillar and amorphous aggregates. The model includes three basic reaction: (i) nucleation, (ii) monomer addition (resulting in fibrillar aggregates) and (iii) heparin addition (resulting in amorphous aggregates). See details in text.

Equations

The rate at which p25α nucleates onto heparin is proportional to the rate constant k_{nu}, and we assume, that this process involves two p25α monomers and one heparin molecule.

The rate at which the p25α:heparin fibres grow is proportional to the rate constant k_{add}, and as argued above we shall also assume that this rate is proportional to the length l - i.e. the number of p25α monomers, which have already aggregated onto the fibre. This

assumption is equal to assuming that the single fibres grow in a logistic-like manner.

The model also includes a maximal capacity of p25α-monomers per heparin molecule L_{max}, and hence we have the following equation for the concentration of p25α monomers:

$$\frac{d[Mono]}{dt} = -2 \cdot k_{nu} \cdot [Mono]^2 \cdot [Hep] - k_{add} \cdot [Mono] \cdot \sum_{l=2}^{L_{max}} l \cdot [F_l] \tag{38}$$

Once the fibres are formed, and as they continue to grow through addition of p25α monomers, there is a change that the fibres will add a second heparin molecule, which we assume disrupts the structure of the fibre, such that an amorphous aggregate is formed. The rate at which this happens is proportional to the rate constant k_{am}. The concentration of fibre "nuclei" (F_2) is given by the equation:

$$\frac{d[F_2]}{dt} = k_{nu} \cdot [Mono]^2 \cdot [Hep] - k_{add} \cdot [Mono] \cdot l \cdot [F_2] - k_{am} \cdot [Hep] \cdot [F_2] \tag{39}$$

the concentration of longer fibres (F_l with $2 < l < L_{max}$):

$$\frac{d[F_l]}{dt} = k_{add} \cdot [Mono] \cdot (l-1) \cdot [F_{l-1}] - k_{add} \cdot [Mono] \cdot l \cdot [F_l] - k_{am} \cdot [Hep] \cdot [F_l] \tag{40}$$

and the concentration of amorphous aggregates (of size l) is given by the equation:

$$\frac{d[A_l]}{dt} = k_{am} \cdot [Hep] \cdot [F_l] \qquad (\text{for } l \geqslant 2\,) \tag{41}$$

In order to compare the model results with the ThT data, we define the total amount of fibres and amorphous aggregates:

$$[Fibres] = \sum_{l=2}^{L_{max}} [F_l] \tag{42}$$

$$[Amorphous] = \sum_{l=2}^{L_{max}} [A_l] \tag{43}$$

and assume that the ThT signal is directly proportional to the total amount of fibres plus the total amount of amorphous aggregates, but where the amorphous aggregates only produce a ThT signal, which is 5% the intensity of the ThT signal produced by the fibres - i.e. $ThT \propto [Fibres] + 0.05[Amorphous]$. Finally, the concentration of heparin is given by the equation:

$$\frac{d[Hep]}{dt} = -k_{nu} \cdot [Mono]^2 \cdot [Hep] - k_{am} \cdot [Hep] \cdot \sum_{l=2}^{L_{max}} [F_l] \tag{44}$$

2.4 RESULTS

The model was implemented in Matlab and solved with the function "ode45", which uses fourth order Runge Kutta integration. The model was fitted to the experimental data shown in Figure 7, the result of which can be seen in Figure 9. The sigmoidal shape of the ThT curve is reproduced due to the logistic-like growth of the fibres (i.e. the growth is proportional to the length l) and - depending on the heparin concentration - fibrillar and/or amorphous aggregates are formed.

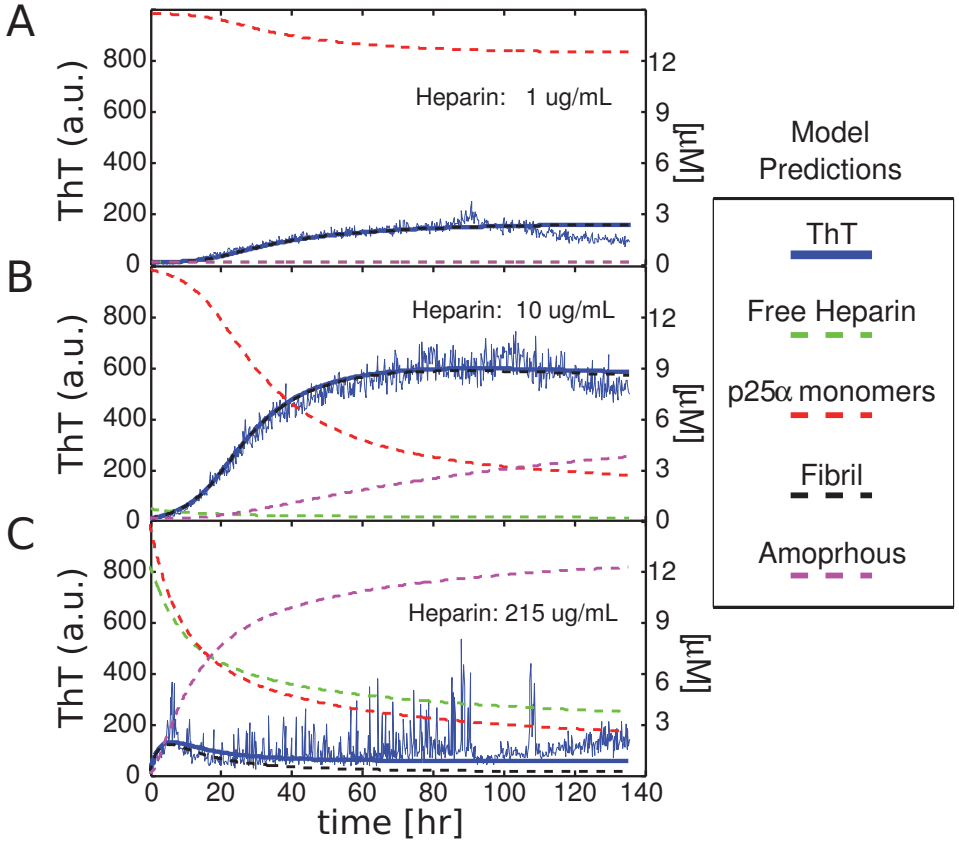


Figure 9: Result of a global fit of the model described by the equations (38-44) to experimental data for three different heparin concentrations. Additional to the resulting ThT intensity (solid blue lines) we plot here the relative concentrations of free heparin (not to scale), p25α monomers, total amount of fibres and total amount of amorphous aggregate. Parameters used for this fit: $k_{nu} = 0.6\mu M^{-2}s^{-1}$, $k_{add} = 36\mu M^{-1}s^{-1}$, $k_{am} = 120\mu M^{-1}s^{-1}$, $L_{max} = 13$.

As shown in Figure 10A, the model clearly reproduces an optimal heparin concentration and the ThT intensity at the end of the aggregation process is reproduced in very good agreement with the experimental data (Figure 10B). Also the model predictions of the concentration of p25α monomer at the end of the aggregation process is compared to experimental measurements (Figure 10C), and we observe that the model reproduces the correct overall effect, although the experimental data indicates a higher plateau level than

the model. We note that this discrepancy, may to some extend be explained by the fact that the experimental measurements plotted in Figure 10C, corresponds to the concentration of all soluble p25α, which way include both monomers and small oligomers, whereas the corresponding values from the model corresponds to monomers only.

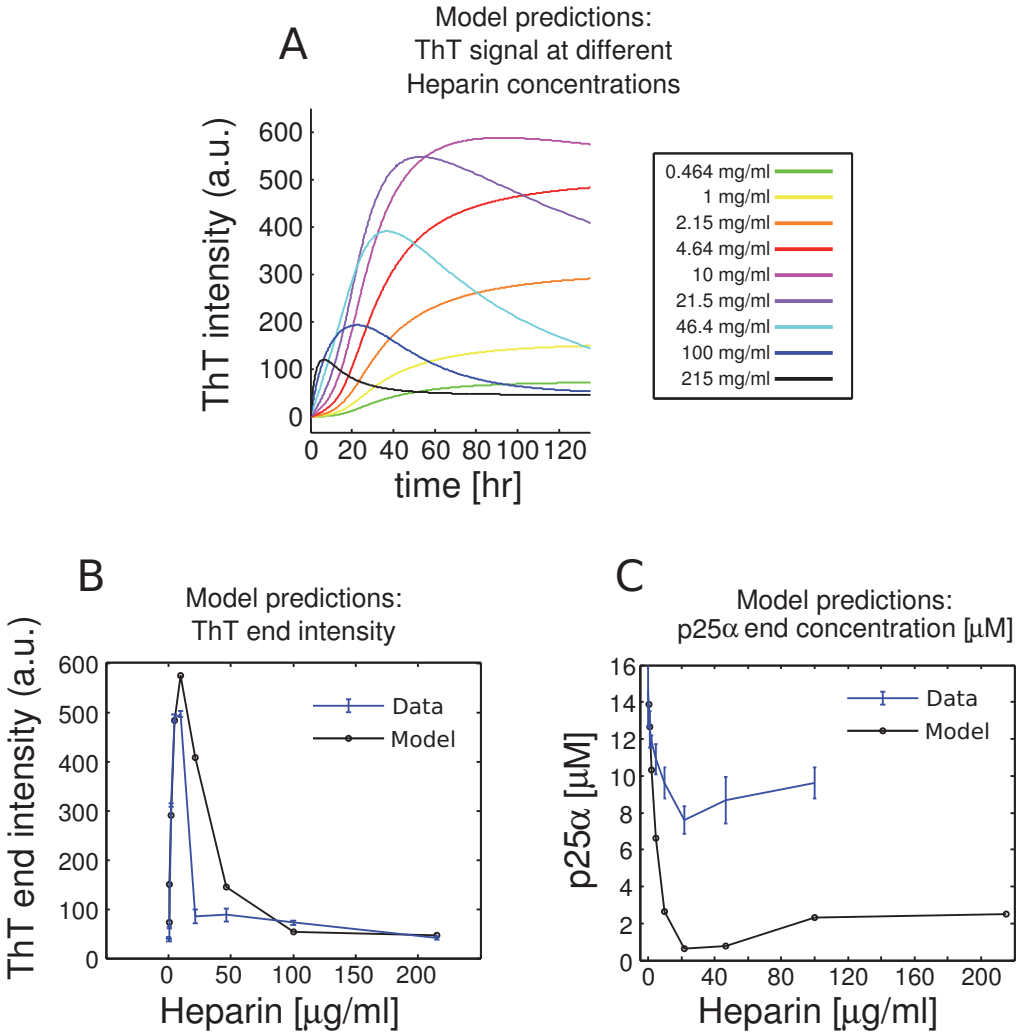


Figure 10: Model predictions for the ThT intensity at different concentrations of heparin. A: The ThT profiles as a function of time indicates clearly an optimal concentration of heparin around 10µg/mL. B: The end-point ThT level readout is compared to experimental measurements. C: The end-point concentrations of p25α monomers is compared to experimental measurements of soluble p25α complexes.

Finally, we test the agreement of the model predictions for a fixed heparin concentration, and varying initial concentrations of p25α - see Figure 11. The heparin concentration was fixed at the optimal concentration 10µg/mL, and the model produces "optimal" growth curves for each p25α concentration - again we observe an excellent agreement between the end-point ThT levels of the model and the experimental data.

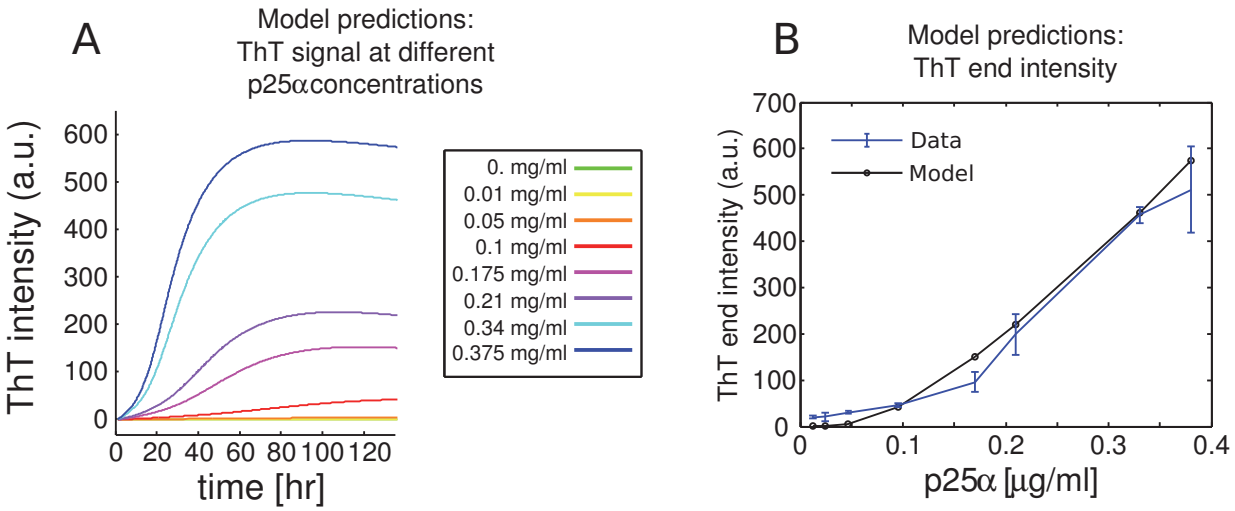


Figure 11: Model predictions of ThT profiles for optimal heparin conditions (10µg/mL) at different initial concentrations of p25α. A: Model predictions of ThT profiles as a function of time. B: Endpoitn Tht level is compared to experimental measurements.

2.5 DISCUSSION

As mentioned above, the logistic-like growth implemented in the model, may be justified by imagining several different scenarios, and is therefore to some extend characterised as a phenomenological model. We therefore note that the sketch shown in Figure 8, should not be recognised as the only possible scenario of the current model. However, the model (which is uniquely defined by its equations), very successfully reproduces the experimental measurements. The model parameters were fitted to the kinetic data shown in figure 9, and the agreement between model and experimental data shown in Figure 10 and 11 is therefore not trivial.

Other models, which were also explored, were not able to reproduce the experimental data equally well. Especially, we found that including individual simulation of the different aggregate sizes (i.e. distinguishing between different F_l), contributed significantly to the performance of the model. This finding emphasises the importance of structural aspects for the aggregation process - i.e. a mean-field model, which does not differentiate between different aggregate sizes, does not perform equally well.

3

COMPLEX AGGREGATION PATTERNS OF WHEY PROTEIN

In the previous chapters we have considered the highly ordered amyloid fibres, but other biological aggregates may be very complex and could consist of several different proteins, which adhere together in a non-ordered fashion. Such aggregation processes can be difficult to model in great detail, as there are many unknown factors - including the specific structures and compositions, or the rates of the different binding and unbinding reactions. In such cases a mathematical model including too many details becomes less meaningful, and it can be favourable to develop models which describe only the overall processes and exclude specific details.

In this chapter I shall present a mathematical model for the aggregation of the whey proteins β-lactoglobulin (bLG) and α-lactalbumin (aLA) influenced by different pH and calcium concentrations. Aggregates of whey proteins comprises a good example of a complex aggregation process, which includes several proteins and yields different aggregate structures depending on the external conditions. These aggregates are much less ordered than the amyloid aggregates described in Chapter 1 and 2. Here we investigate a system consisting of two different whey proteins, subject to variations in two external conditions (pH and calcium concentration). The aggregates form two seemingly different types of aggregates and in order to understand and distinguish the different steps of the aggregation process, we construct a simple mathematical model which includes the overall reaction kinetics, but excludes as many unknown details as possible.

3.1 WHEY PROTEIN

Whey constitutes a valuable and versatile source of proteins for nutritional purposes and as functional ingredients. β-lactoglobulin (bLG) and α-lactalbumin (aLA) constitute the principal components of whey protein and represents approximately 54% and 21% of the protein mass [10, 37]. While the native form of aLA is monomeric, the native form of bLG is predominantly dimeric. Heating causes the dimeric forms to dissociate and monomers to unfold into non-native states [11, 60]. bLG may expose a thiol and hydrophobic groups and disulphide-linked oligomers of mixed bLG and aLA may form. Under proper conditions (pH, salt, heat-load etc.) soluble aggregates and larger protein networks, gels or precipitates may form [1, 43, 56, 57, 64]. The thermal denaturation and following aggregation of whey proteins is a highly relevant event in the process-

ing of whey proteins in the dairy industry, serving either to increase shelf-life through microbial destruction or to achieve new functionalities such as gelation, foaming ability, increased thermal stability etc. [4, 10, 30, 47, 50, 59, 63, 67, 68, 72, 77].

Much work has been done in defining and explaining the mechanism of how solutions and heating conditions lead to the formation of soluble aggregates, and most model studies have focused on the most abundant whey protein, bLG, as a model for whey protein aggregation and a good first approximation to the denaturation and aggregation patterns observed for more complex protein mixtures.

The present model includes both bLG and aLA, and was constructed in order to model experimental data of pH and calcium dependent aggregation of whey protein. The experimental data was obtained for 2:1 weight-ratio aLA:bLG protein mixtures, subjected to controlled heating conditions at 90°C for 600 seconds. The samples were exposed to different pH ($6.5 - 8.1$) and calcium concentrations ($0 - 4$mM) and it was observed that the proteins form two seemingly distinguishable aggregate forms: one aggregate species consists of relatively small aggregates (SAs) of approximately 10-15 monomers, and the other aggregate species consists of relatively large aggregates (LAs) of many hundreds of monomers. Different pH and calcium conditions influenced the resulting mass-ratio of small and large aggregates, and the data suggests, that low pH and high calcium concentrations favour formation of large aggregates, whereas high pH and low calcium concentration favour formation of small aggregates. The data also suggests that bLG monomers are consumed faster when pH is high (see Figure 14).

3.2 MATHEMATICAL MODEL

In order to understand and distinguish different steps of the aggregation process, a mathematical model of the reaction kinetics was constructed. The model is based on knowledge from previous studies of whey protein, and on the experimental findings described above. The mathematical model can be used as a means of testing hypotheses and assumptions about the different processes, that occur in the system. The model is sketched in Figure 12 and is based on the following basic assumptions:

- Initially the system consists of bLG and aLA monomers in a solution with a given pH and calcium concentration.
- The monomers can combine to form non-native dimers. Only bLG:bLG and aLA:bLG dimers are allowed in the model as aLA:aLA dimers are very unlikely, due to the lack of a free thiol in aLA.

- Small aggregates (SAs) are formed when the dimers grow to become trimers, quatremers etc. Growth happens by adding additional monomers onto the existing oligomer.
- Larger aggregates (LAs) may form when oligomers lump together, in a way which is fundamentally different from regular oligomer growth. Whereas the regular oligomer growth can happen only by addition of monomers, formation of "Large Aggregates" may involve direct aggregation of oligomers of all sizes.
- Small aggregates and large aggregates are treated as two distinct types of aggregates in the model. The model assumes no specific spatial arrangement of the different aggregates although the small aggregates are thought to have a somewhat linear structure.
- Any calcium is present in the solution may adhere to any monomer or oligomer in the solution.
- The different processes in the system are can be dependent on calcium or pH, and we model these processes consistently with the findings of the experimental results.

The model is constructed keeping in mind the principle of Occam's razor: We want the model to be as simple as possible, while still reproducing the experimental results. By minimizing the number of assumptions, we also minimize the number of unknown rate-constants, which will have to be fitted to the data. It is very possible that several models will be able to produce acceptable fits to the data, and hence we cannot rule out other possible processes, but we can hope to capture the dominating processes of the system, and rule out models that cannot produce acceptable fits.

Reactivity of Monomers

The reactivity of bLG is dependent on the current tertiary structure. When heated above 65 degrees bLG denaturates partially and a conformational change results in exposure of the free thiol group as well as parts of the hydrophobic core of the protein. The reactive species of bLG is the thiolate form, and since the conformational transition is reversible, only a fraction of bLG is truly available for aggregation. The reactivity of bLG increases with increasing pH (as also indicated by the experimental data), and therefore we include in our model a "reactivity" R_{bLG}, which is proportional to the fraction of bLG in the basic form. We calculate this fraction using the pK_a value of bLG, which is ~ 8.7.

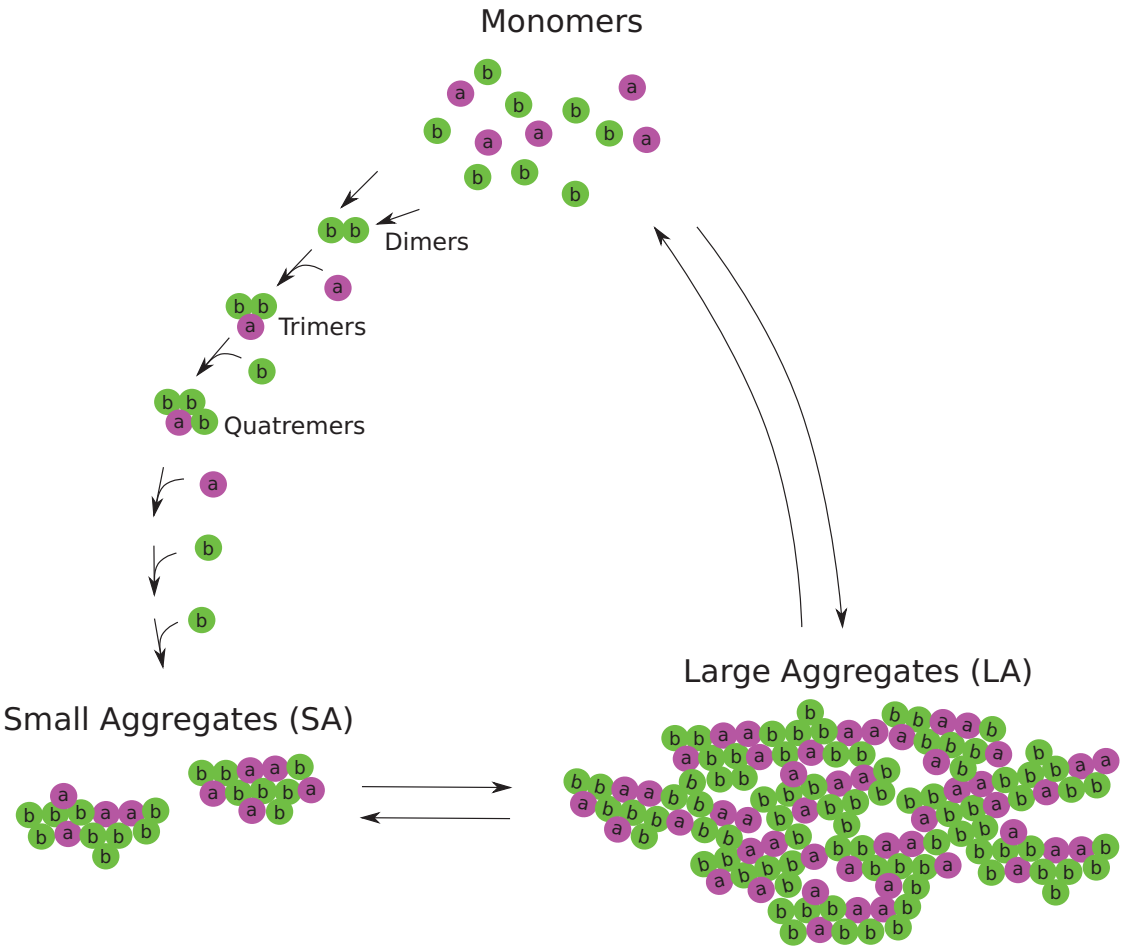


Figure 12: Sketch of the different aggregation paths of the model.

In practice this means, that all reactions which are proportional to the amount of bLG monomer is also proportional the basic fraction f_{base} given by:

$$R_{bLG} = f_{base} = \frac{B}{B+A} = \frac{B/A}{B/A+1} \tag{45}$$

where

$$(B/A) = 10^{(pH-pK_a)} \tag{46}$$

aLA, on the other hand, van be considered to be largely denaturated under all the relevant conditions and therefore has a reactivity of 100%:

$$R_{aLA} = 1 \tag{47}$$

Binding of Calcium

The amount of added calcium has a large impact on the aggregation processes although the specific reaction processes are unknown. By measuring the amount of free calcium with a calcium electrode (data

not shown) it has been observed, that free calcium concentration decays relatively fast, and that the calcium therefore must adhere to the protein in the solution. In order to simplify our model we assume that all the calcium adheres to the protein effectively immediately after denaturation, and we also assume that each monomer can bind up to two calcium ions (see Figure 13).

In the model all calcium is distributed at random amongst the monomers in the initial solution: if there are C calcium ions and M monomers in the solution there must be 2M binding sites for calcium of which C sites are occupied. The probability of each binding site to be occupied is given by $p = C/(2M)$. The amount of monomers with two bound calcium ions is therefore Mp^2 amount of monomers with no bound calcium is $M(1-p)^2$ and the amount of monomers with one bound calcium ion is $2Mp(1-p)$ (where the factor 2 is a combinatorial factor). Once the calcium is distributed amongst the monomers in the initial solution, we do not allow for unbinding of calcium. The calcium ions therefore follow the monomers as they aggregate, but we do not keep track of the specific positions of a calcium ion within an aggregate - we only keep track of the total amount of calcium within a certain aggregate.

Reaction Rates

The model includes several chemical reactions for which the rate constants are unknown. These rates must all be fitted to the experimental data and we denote the rate constants as follows: the rates at which dimers are formed are proportional the rate constants k_{bb} for bLG:bLG dimers and k_{ab} for aLA:bLG dimers. The rate at which small aggregates add additional monomers is assumed to be independent of the size and composition of the existing small aggregate, but depend only on the kind of monomer, which is added. Hence rate at which the small aggregates grow are proportional to the rate constants k_b for addition of a bLG monomer and k_a for addition of an aLA monomer. The rate at which large aggregates are formed is proportional to the rate constant k_{LA}. This process is also pH and calcium dependent and the details are described in a separate section below. We also include an unbinding rate by which a small aggregate can unbind from a large aggregate - this rate is proportional to the rate constant k_u.

Large Aggregates

The large aggregates cannot simply be very large oligomers as this would suggest a continuous range of sizes - and this is not observed experimentally. There are a number of different ways one can imagine that the large aggregates form: maybe the monomers can form two different kinds of aggregate - one being the small aggregates (oligomers) that grow from dimers to trimmers etc. and the other

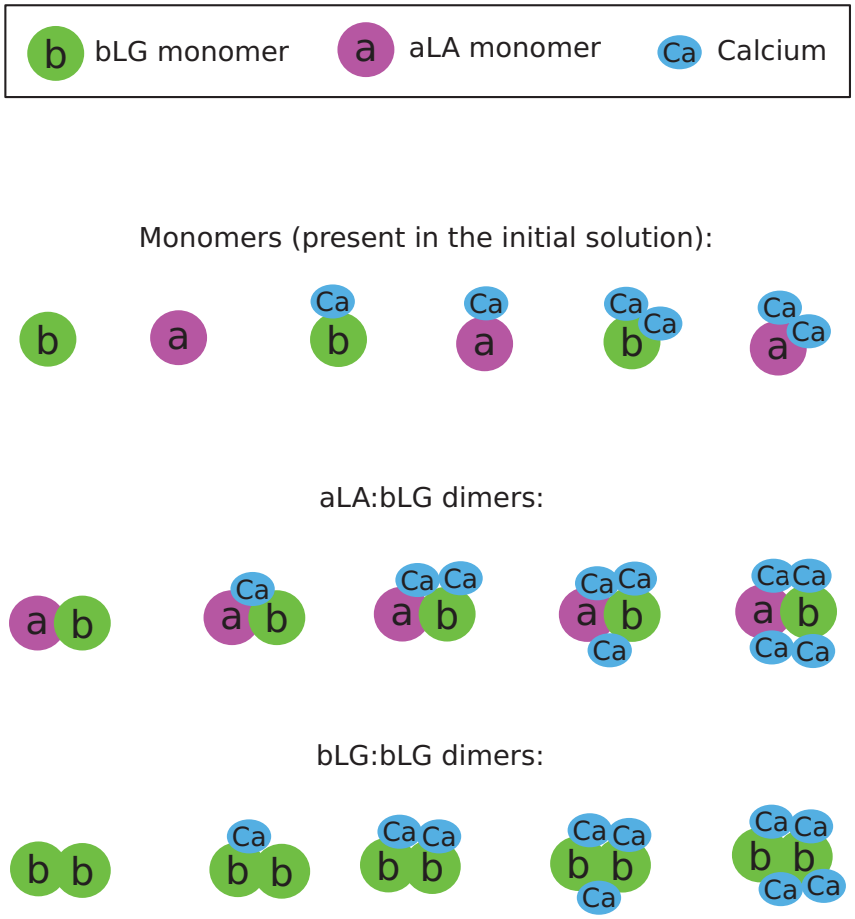


Figure 13: Our model assumes that all monomers can bind up to two calcium ions. This is implemented by distributing all calcium on the initial monomer mixture, and as a result there are 6 different monomers present in the initial solution - three of each monomer-type: bLG and aLA (in this Figure: b and a). The monomers form dimers of the type aLA:bLG and bLG:bLG and there are 5 different dimers of each type.

being large aggregates which could possibly grow faster or in a less ordered fashion. In this scenario there would therefore be two competing pathways, that deplete the monomer concentration - one pathway leading to formation of small aggregates and another pathway leading to formation of large aggregates. Another possibility would be that only oligomers of a certain size can form large aggregates by "lumping together" i.e. adhering to each other once they have grown to a certain size. This scenario would therefore describe a linear pathway where the monomer concentration would be depleted due to formation of small aggregates, and the concentration of small aggregates would be depleted due to formation of large aggregates. In order to minimize the number of assumptions, we allow both possibilities: oligomers of all sizes (small aggregates *and* monomers) can then form large aggregates by "lumping together", in a less ordered fashion. We call this model the "Mixed Pathways Model".

While we keep track of the particular sizes of small aggregates, we do not specify the sizes of large aggregates, but simply model a pool of large aggregates, which can bind and release (unbind) small aggregates (with the rates k_{LA} and k_u). In order to incorporate the observed influence of calcium and pH on aggregate formation we introduce an aggregation process which depends on calcium and pH in the following way:

$$k_{LA} = k_1(Ca/Length) + k_2(pH_{threshold} - pH) \tag{48}$$

In this way the large aggregates are favoured by high calcium concentrations and low pH in a linear and mutually independent fashion. The calcium dependence is chosen to be such that it is the calcium density in the oligomer and not the total amount of calcium which is important. Here 'Length' is the length of the given oligomer measured in number of monomers and Ca/Length thus represents the calcium density. The pH dependence is proportional to $(pH_{threshold} - pH)$, where $pH_{threshold}$ is an unknown parameter.

Model Implementation

If each protein can bind up to two calcium ions this means there are three version of each monomer:

(bLG)	(aLA)
(bLG)-(Ca)	(aLA)-(Ca)
(bLG)-$(Ca)_2$	(aLA)-$(Ca)_2$

With 6 different monomers it is in principle possible to form 18 different dimers. But we do not allow for pure aLA dimers and we do not keep track of the explicit composition but only to total number of calcium ions in the dimer. This leaves 10 different possibilities:

(bLG)-(bLG)	(bLG)-(aLA)
(bLG)-(bLG)-(Ca)	(bLG)-(aLA)-(Ca)
(bLG)-(bLG)-$(Ca)_2$	(bLG)-(aLA)-$(Ca)_2$
(bLG)-(bLG)-$(Ca)_3$	(bLG)-(aLA)-$(Ca)_3$
(bLG)-(bLG)-$(Ca)_4$	(bLG)-(aLA)-$(Ca)_4$

In order to implement the model we denote the small aggregates $SA_{(i,j,c)}$ where i equals the number of aLA monomers, j equals the number of bLG monomers and c equals the number of calcium ions bound to the oligomer:

$$SA_{(i,j,c)} = (aLA)_i(bLG)_j(Ca)_c$$

Although the monomers are not thought of as small aggregates we use for implementation purposes the notation:

$SA_{(0,1,c)}$ = bLG monomers
$SA_{(1,0,c)}$ = aLA monomers

and for dimers:

$SA_{(0,2,c)}$ = bLG-bLG dimers
$SA_{(1,1,c)}$ = bLG-aLA dimers

We also define the total number of bLG monomers:

$$bLG_{sum} = \sum_{c=0}^{2} SA_{(0,1,c)}$$

the total number of aLA monomers:

$$aLA_{sum} = \sum_{c=0}^{2} SA_{(1,0,c)}$$

and the total number of small aggregates which are larger than monomers:

$$SA_{sum} = \left[\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sum_{c=0}^{2} SA_{(i,j,c)} \right] - bLG_{sum} - aLA_{sum}$$

When small aggregates bind to a large aggregate we store the information about the small aggregate (i.e. the amount of aLA, bLG and calcium in the small aggregate) but denote it $LA_{(i,j,c)}$ instead of $SA_{(i,j,c)}$.

$$SA_{(i,j,c)} \rightarrow LA_{(i,j,c)} \qquad \text{SA is converted to (binds to) LA}$$

We do not model explicitly the size distribution of the large aggregates but simply keep track of the total amount of protein (and calcium) which is in the large aggregate "state". The total amount of large aggregate is given by:

$$LA_{sum} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{c=0}^{2} LA_{(i,j,c)}$$

Equations

The equations governing the dynamics of monomers are given by:

$$\begin{aligned}
\frac{dSA_{(0,1,c)}}{dt} &= -k_{bb}SA_{(0,1,c)}\left(bLG_{sum} + SA_{(0,1,c)}\right) \\
&\quad - k_{ab}SA_{(0,1,c)}aLA_{sum} \\
&\quad - k_{b}SA_{(0,1,c)}SA_{sum} \\
&\quad - k_{LA}SA_{(0,1,c)} + k_{u}LA_{(0,1,c)} \\
\frac{dSA_{(1,0,c)}}{dt} &= -k_{aa}SA_{(1,0,c)}\left(aLA_{sum} + SA_{(1,0,c)}\right) \\
&\quad - k_{ab}SA_{(1,0,c)}bLG_{sum} \\
&\quad - k_{a}SA_{(1,0,c)}SA_{sum} \\
&\quad - k_{LA}SA_{(1,0,c)} + k_{u}LA_{(1,0,c)}
\end{aligned} \tag{49}$$

where the second term in the parenthesis comes from the fact that two monomers are used when a homogenic dimer is formed. The equations governing the dynamics of dimers are given by:

$$\begin{aligned}
\frac{dSA_{(0,2,c)}}{dt} &= \sum_{p+q=c} k_{bb}SA_{(0,1,p)}SA_{(0,1,q)} \\
&\quad - k_{b}bLG_{sum}SA_{(0,2,c)} \\
&\quad - k_{a}aLA_{sum}SA_{(0,2,c)} \\
&\quad - k_{LA}SA_{(0,2,c)} + k_{u}LA_{0,2,c} \\
\frac{dSA_{(1,1,c)}}{dt} &= \sum_{p+q=c} k_{ab}SA_{(1,0,p)}SA_{(0,1,q)} \\
&\quad - k_{b}bLG_{sum}SA_{(1,1,c)} \\
&\quad - k_{a}aLA_{sum}SA_{(1,1,c)} \\
&\quad - k_{LA}SA_{(1,1,c)} + k_{u}LA_{(1,1,c)}
\end{aligned} \tag{50}$$

The equations governing the dynamics of small aggregates that are larger than dimers are given by:

$$\begin{aligned}
\frac{dSA_{(i,j,c)}}{dt} = &\quad \sum_{k=0}^{2} k_{a}SA_{(1,0,k)}SA_{(i-1,j,c-k)} \\
&+ \sum_{k=0}^{2} k_{b}SA_{(0,1,k)}SA_{(i,j-1,c-k)} \\
&- k_{a}aLA_{sum}SA_{(i,j,c)} \\
&- k_{b}bLG_{sum}SA_{(i,j,c)} \\
&- k_{LA}SA_{(i,j,c)} + k_{u}LA_{(i,j,c)}
\end{aligned} \tag{51}$$

And finally the equations governing the dynamics of the large aggregates are given by:

$$\frac{dLA_{(i,j,c)}}{dt} = k_{LA} SA_{(i,j,c)} - k_u LA_{i,j,c} \tag{52}$$

3.3 RESULTS

The mathematical model was implemented in Matlab and solved with the function "ode45", which uses fourth order Runge Kutta integration. The results of fitting the model to kinetic data for different combinations of pH and calcium can be seen in Figure 14. The rate parameters of the model have been globally fitted to match the experimental data, and as can be seen from Figure 14, the model successfully captures both the dynamics of a particular aggregation experiment (for a given pH and calcium concentration) and the overall effects of changing pH and calcium. High pH and low calcium concentration results in aggregates dominated by small aggregates (SAs). Also high pH leads to fast consumption of monomers, due to the high reactivity of bLG. Both low pH and high calcium concentration results in aggregates dominated by large aggregates (LAs), where high calcium concentration and high pH is dominated by fast kinetics and a clear overshoot effect in the concentration of small aggregates, the situation with low pH and low calcium concentration is dominated by a slower kinetics, although with a similar end result after 600 seconds of heating. As we keep track of the sizes of small aggregates in the model, we can use the model to predict differences in the size distribution of small aggregates for different values of pH and calcium concentration. The results agree with the overall expected sizes (average 10-12 monomers). Further verification of the model might be obtained if a higher resolution could be obtained for size measurements. This was, however, unfortunately not possible experimentally due to overlap of peaks from small- and large aggregates, respectively. Other versions of the model have also been tested (data not shown), but we have not been able to produce equally successful fit using models of "Competing Pathways" or "Linear Pathways" (as described above).

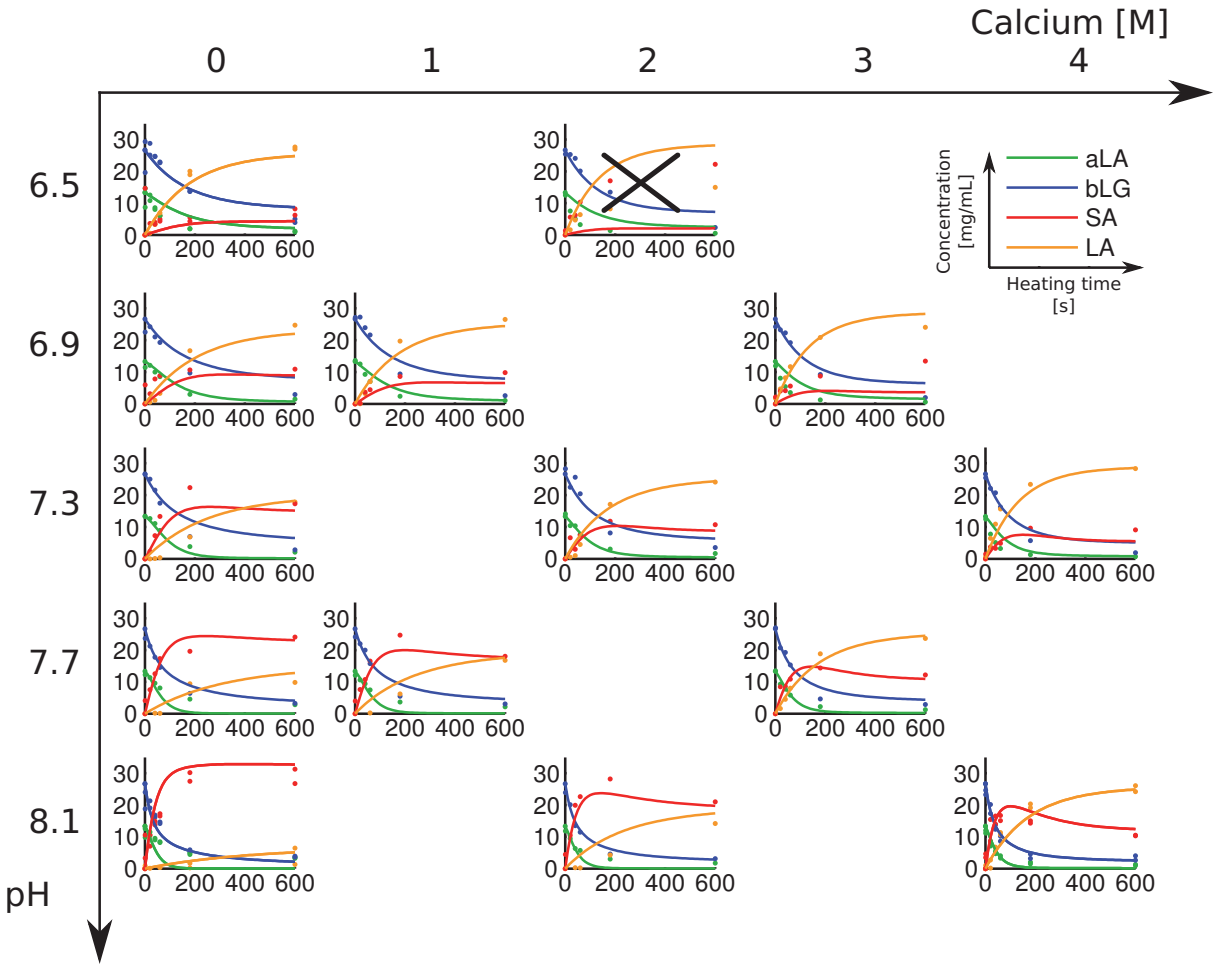


Figure 14: Model fit (lines) to experimental data (dots), for the "Mixed Pathways Model". The experimental data at $pH = 6.5$ and $Ca = 2$, has been disregarded as it deviates significantly from the expected output, and has possibly been mis-recorded. Parameters used for this fit: $k_a = 0.05\ s^{-1}$, $k_b = 0.0008\ s^{-1}$, $k_{ab} = k_{bb} = 0.03\ (Ms)^{-1}$, $k_1 = k_2 = k_u = 0.02\ s^{-1}$, $pK_a = 8.7$, $pH_{thres} = 8.3$.

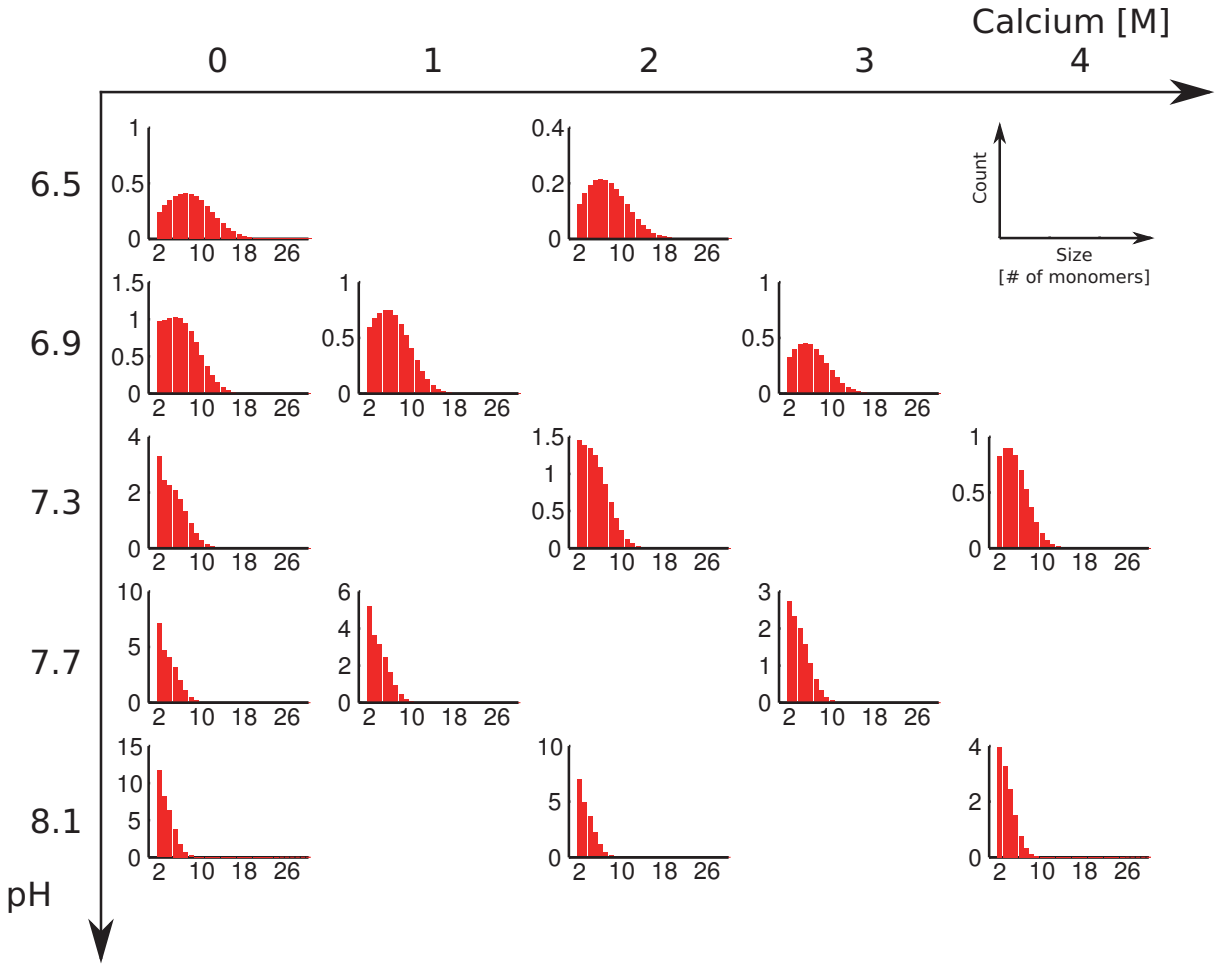


Figure 15: Size distribution of the small aggregates (oligomers), as theoretically predicted by the "Mixed Pathways Model", at the end of the aggregation process (at time t = 600 seconds). Parameters are the same as in Figure 14.

Part II

SPATIAL REGULATION IN BIOLOGICAL SYSTEMS

INTRODUCTION

Many mathematical models of biological systems have focused on the dynamics of a single cell or another small and confined system for which space can readily be neglected. In these cases a completely well mixed model can be applied with great success. It is however, important to consider the possible functions and restrictions that may be introduced by different spatial arrangements - especially when modelling large systems such as multicellular organisms or other systems for which the relevant time and length scales are large compared to the effective diffusion constant.

In this part of the thesis I shall consider a biological system for which different spatial arrangements are very important. Specifically, we shall consider different aspects of inflammatory response, both in general and in relation to type II diabetes.

Spatial Restrictions and Advantages

There are many interesting biological systems, which have evolved to overcome certain spatial restrictions. In multicellular organisms a typical challenge is to send information over long distances, in a fast and reliable manner. Well known examples which overcome this challenge are the nervous system, the blood stream and the venous system. But there are also many biological systems, which exploit spatial arrangements to their advantage. Well known examples are the flagella of bacteria or our own body parts (legs, fingers, teeth, etc.) - different species have evolved to perform different tasks. On the single cell scale, spatial arrangements are also very important. Inside an eukaryotic cell, biological membranes serve to organise different chemical reactions into specialised compartments and in many cases spatial restrictions are actively constructed in order to perform different functions and maintain homoeostasis.

Inflammation

As an example of a system, for which spatial organisation plays a key role, we shall in Chapter 4 consider different aspects of inflammatory response. During inflammatory response many different components need to interact in order to localise and eliminate intruding pathogens, and we shall focus on the signal transmission on tissue level, which is needed in order to attract and lead white blood cells to the site of infection. In order to study the cell communication on tissue level we construct a simple mathematical model of the tissue cells. The model is based on the regulatory network of the transcription factor NF-κB, which is known to play a key role in

inflammatory response, and which regulates the local concentration of cytokines, that function as chemoattractants for migrating white blood cells.

Type II Diabetes

Inflammatory response is also highly involved in the pathology of type II diabetes, where the insulin producing β-cells, often display chronic low-grade inflammation. The β-cells strongly regulate the local concentration of the cytokine IL-1β, and previous studies of the regulatory system inside β-cells have pointed out the possibility of a so-called *vicious cycle*, in which β-cells maintain an inflamed state in a self-sustaining manner. The mathematical model developed in Chapter 4, is therefore easily transferable to the β-cells and in Chapter 5 we extend the model in order to study inflammation in relation to type II diabetes.

The β-cells are localised in the pancreas, where they are organised into small clusters called pancreatic islets (or islets of Langerhans). In Chapter 5 we study how such spatial arrangements may add to the overall regulation of IL-1β, and consequently lead to different islet fates.

4

CYTOKINE WAVES IN THE INFLAMMATORY RESPONSE

The inflammatory response is a good example of a complex process, in which many different components need to interact in order to achieve a common goal: eliminate an intruding pathogen. The interactions involve several aspects that need spatial coordination: localization of the intruding pathogens, signalling to the bloodstream and recruitment of white blood cells. In this chapter I shall present a mathematical reaction-diffusion model, which has been especially developed to study signal transmission in the tissue. The model is based on the regulatory network of the transcription factor NF-κB, where the internal details of the single cell have been simplified in order to construct a multicellular model, which investigates the system on tissue level. Before I describe the model in detail, some general aspects of inflammation and NF-κB are introduced.

RELEVANT PUBLICATIONS:

Pernille Yde, Benedicte Mengel, Mogens H Jensen, Sandeep Krishna and Ala Trusina. Modeling the NF-κB mediated inflammatory response predicts cytokine waves in tissue. *BMC Systems Biology* 5:115 (2011)

Pernille Yde, Mogens Høgh Jensen, and Ala Trusina Analyzing inflammatory response as excitable media *Physical Review E* 84:051913 (2011)

4.1 INFLAMMATION

Inflammation is the body's first reaction to local damage caused either by infectious agents or by other injury such as burns, frostbites or chemical irritants. Inflammation is non-specific and is part of the innate immune system. Its goal is to eliminate the injurious agents and remove damaged tissue components before healing and repair can begin.

The inflammatory response is initiated by damaged cells in the connective tissue which release a range of signaling molecules. These signalling molecules mediate (i) a vascular response which increases the blood flow around the site of injury and (ii) a cellular response which recruits white blood cells to the cite of injury - mainly neutrophils are recruited. Neutrophils are phagocytotic cells that can engulf and digest damaged cell parts and foreign components (e.g. bacteria and viruses) in a process called phagocytosis. As a result all injurious agents are removed and the injured tissue is prepared for healing[55].

The Chemoattractant Signal

When the neutrophils exit the blood stream they must actively move in a directed manner in order to reach the site of injury. This directed movement is called chemotaxis and the neutrophils follow a chemical signal of chemoattractants such as chemokines and cytokines. The chemoattractant molecules are released from the tissue cells at the site of injury, however it is unknown how the chemoattractant signal is transmitted through the tissue. In order to signal a direction for neutrophils to follow, a local gradient of chemoattractant is needed - and this gradient must be sharp enough for neutrophils to detect.

In principle there are multiple ways a chemoattractant signal can be transmitted through the tissue, however one can argue, that not all of them are equally efficient and reliable. In the simplest scenario the chemoattractant molecules diffuse passively from the site of infection. In this case the resulting concentration gradient decays exponentially with the distance from the source [74], and the range of the signal will be limited by the typical short half-life of the chemoattractant molecules. The gradient signal is therefore relatively short-ranged, and can only recruite neutrophils that are already relatively close to the site of infection - this scenario is visualised in Figure 16A.

Interestingly, some neutrophil chemoattractants, e.g. the cytokines TNF and IL-1, have the unique ability to self-amplify by means of tissue cells or tissue resident macrophages [49]. Such an active participation of tissue cells, could potentially amplify the cytokine concentration across the tissue, and create a sustained high concentration of cytokines. The blood vessels would serve as sinks where cytokines are carried away by the bloodstream, and as a result a sharp gradient would develop near the blood vessels. However, away from the blood stream, the concentration of cytokines will reach a local steady state, in which production is balanced by degradation, and as a result there is no chemoattractant gradient for the neutrophils to follow - this scenario is visualised in Figure 16B.

It is important to note that both scenarios ("No Amplification", Figure 18A and "Continuous Amplification", Figure 18B) create static gradients, i.e. once in steady state, the gradients are not changing in time. Such static gradients are intrinsically toxic for the tissue as, for example, continuous exposure of tissue cells to high TNF levels triggers apoptosis [7]. Interestingly, some experimental evidence suggest that static gradients are also suboptimal for neutrophil chemotaxis; neutrophils seem to orient themselves better when exposed to temporally varying gradients [2, 21, 28, 69, 70].

If the cytokine concentration is not amplified continuously, but *transiently* (i.e. with a peak-like profile), the tissue cells will avoid sustained exposure to toxic cytokines, while the signal - the chemoattractant gradient - can still penetrate far in the tissue. Such transient

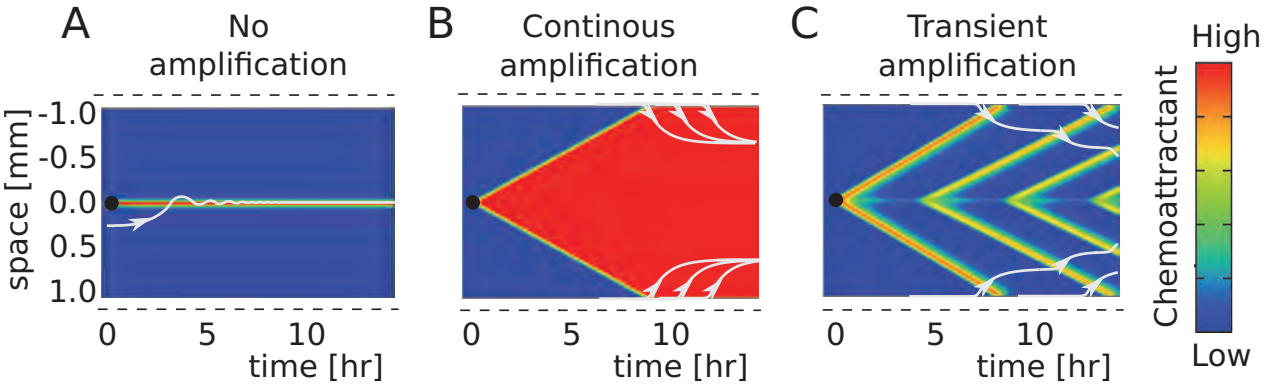


Figure 16: Three scenarios for signal propagation from the site of infection and neutrophil recruitment. The site of infection is marked by the black circle. Cells are aligned vertically, and changes of chemoattractant concentrations in time, are developing from left to right. White lines represent trajectories of neutrophils chemotaxing from blood vessels (dashed lines at the boundaries). The trajectories were calculated using a model described in [76]. In panel A the chemoattractant is simply diffusing from the site of infection. The signal is short-ranged, and neutrophils can only be recruited if they are not too far away - here no neutrophils are recruited from the distant blood vessels. In panel B the diffusing chemoattractant is continuously amplified by the tissue cells. The gradient is sharp close to the blood vessels, but disappears deeper in the tissue - thus leaving neutrophils devoid of direction. In panel C diffusing chemoattractant is amplified *transiently*, resulting in propagating waves, that recruit neutrophils from the blood stream and also serve as a directing signal for the neutrophils that are deeper in the tissue.

amplification can result in single or re-emerging propagating waves, that transiently recruit neutrophils - such a scenario is visualised in Figure 18C.

Cytokines induce activation of the transcription factor NF-κB, which is a key regulator of the inflammatory response. In turn active NF-κB upregulates cytokine production [3, 41, 61], thus constituting a positive feedback (amplification). The NF-κB response to stimuli is indeed transient, and the synthesis and secretion of inflammatory cytokines from tissue cells were shown to be parallel to the transient NF-κB activation [31].

4.2 NF-κB

The transcription factor NF-κB ("Nuclear factor kappa B") induces transcription of many genes, and besides playing a key role in immune response, NF-κB is also involved in other processes such as cellular growth, survival and apoptosis.

NF-κB is a family of transcription factors and consists of 5 different proteins, named p65 (or RelA), RelB, c-Rel, p50 and p52. The 5 different proteins bind to each other and form dimers. The most common dimer is p65:p50, but both hetero- and homodimers exist. The different dimers all bind to the same binding domain on the

DNA, and in the following we shall refer to all versions of the dimer simply as NF-κB.

The regulatory system of NF-κB is sketced in Figure 17A. NF-κB is inhibited by several inhibitors (e.g. $I\kappa B_\alpha$, $I\kappa B_\delta$ and $I\kappa B_\epsilon$ - all referred to as IκB), which bind to NF-κB and form an NF-κB-IκB complex. In a resting cell almost all NF-κB is present in this inhibitor-bound form, which is located outside of the nucleus. Here NF-κB is *in-active*, meaning that it does not bind to DNA and hence does not induce transcription. The amount of NF-κB which is localized outside the nucleus is some times referred to as cytoplasmic NF-κB.

Upon extracellular stimulation the IκB kinase, IKK, is converted into its active form, which can phosphorylate the inhibitor and thereby label it for active degradation. When the inhibitor is degraded, NF-κB is released from the NF-κB-IκB complex and NF-κB is therefore free to enter the nucleus. Inside the nucleus NF-κB is *active*, meaning that it can bind to DNA and induce transcription. The amount of NF-κB inside the nucleus (active NF-κB) is some times referred to as nuclear NF-κB.

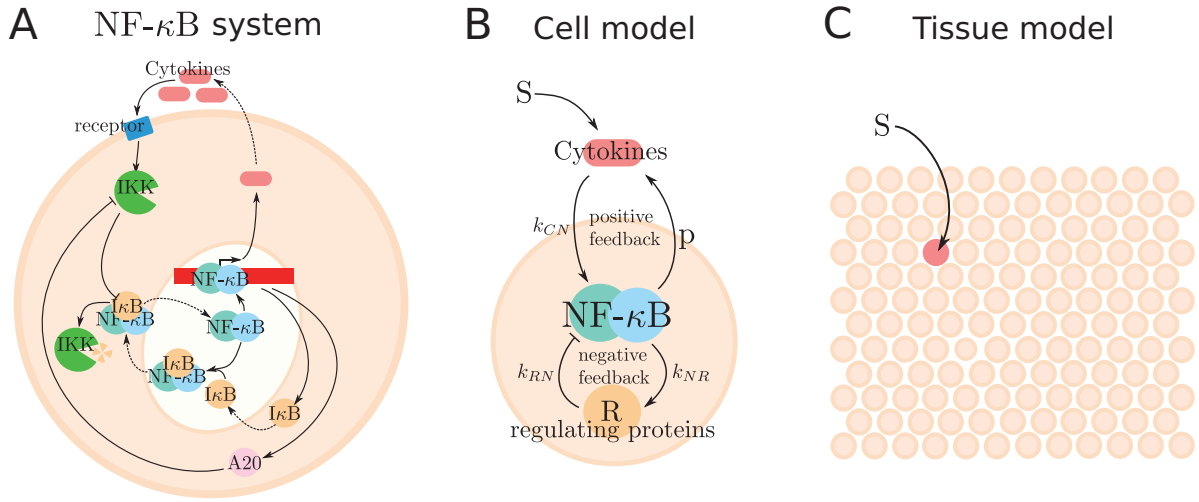


Figure 17: A: The regulatory system of NF-κB, including the inhibitors (IκB) and upstream regulators (here the upstream regulator A20 is shown). B: A sketch of the simple model of the NF-κB system, described by the equations (53)-(55). C: A 2D model of the tissue consisting of several cells in a hexagonal lattice. As indicated the cytokine source described by the parameter S (see equation (56)), is only added to a single cell in the lattice - this position resembles the site of infection.

Self-Inhibition

Amongst other genes, NF-κB induces transcription of the genes coding for its own inhibitor proteins. When this happens, mRNA is produced and transported out into the cytoplasm, where it is translated and new IκB is synthesized. The IκB proteins enter the nucleus where they bind NF-κB, and the NF-κB-IκB complex is again transported out of the nucleus. All in all the NF-κB and IκB system

constitutes a negative feedback loop, which can shuttle NF-κB in and out of the nucleus, changing from inactive to active and back again. Previous studies have shown that the amount of nuclear NF-κB tends to exhibit damped oscillations upon extracellular stimuli and these oscillations can be modeled by nested negative feedback loops[32, 33, 39, 45, 48].

Amplification of the Cytokine Signal

Extracellular cytokines and chemokines trigger activation of the NF-κB system by binding to receptors on the cell surface which facilitate IKK activation. Active NF-κB induces transcription of hundreds of genes and very interestingly NF-κB also induces transcription of many cytokines (e.g. TNF, IL-1 and IL-6). Newly synthesized cytokines are liberated into the extracellular space where they are free to bind to receptors once again. Hence there is a positive feedback between cytokines and NF-κB with the result that cytokine concentration is amplified by NF-κB activation. This positive feedback has to our knowledge not received much attention in the previous models of NF-κB and we have therefore constructed a model which includes this feedback.

4.3 MODEL

The NF-κB system is very complex and an attempt to model every biological component and every chemical reaction explicitly would be almost impossible - also it would include many unknown parameters and a fit to experimental data would probably not be unique. A previous model developed by Hoffmann et. al. includes 24 variables and this model has since been the main inspiration to the development of even more simplified models including 7 or even as little as 3 variables[33, 39, 45].

For the present project the main goal is to develop a model, which can be used to model cytokine regulation on tissue level and therefore we construct a very simplified model of the NF-κB system but add the positive feedback on cytokine production as this could be important for intercellular cytokine regulation. The NF-κB system is simplified as sketched in Figure 17B.

Equations

The model consists of the three variables N, R and C. All three variables are in principle functions of both time and space, but for simplicity we abbreviate $N(r,t)$, $R(r,t)$ and $C(r,t)$ by the shorthand notation N, R and C.

N represents active (i.e. nuclear) NF-κB, R represents regulating proteins such as inhibitors (e.g. $I\kappa B_\alpha$, $I\kappa B_\delta$ and $I\kappa B_\epsilon$) and upstream regulators (e.g. A20 and cesanne) and C represents cytokines (e.g.

TNF, IL-1, IL-6, etc.). The model consists of a negative feedback between N and R and a positive feedback between N and C - see Figure 17B. The feedbacks are modelled by the equations:

$$\frac{dN}{dt} = k_{CN} \cdot \frac{C^2}{C^2 + K_A^2} \cdot (N_T - N) - k_{RN} \cdot R \tag{53}$$

$$\frac{dR}{dt} = k_{NR} \cdot N - R/\tau_R \tag{54}$$

$$\frac{dC}{dt} = p \cdot N - C/\tau + D\nabla^2 C + S^* \tag{55}$$

The first term in the equation for N represents cytokine mediated activation of NF-κB. The term is nonlinear in C and the Hill-function (with Hill coefficient 2) comprises a threshold effect in the NF-κB activation. The parameter K_A represents roughly the threshold cytokine concentration C^* that must be exceeded in order to effectively activate NF-κB. N_T is the total amount of NF-κB (nuclear and cytoplasmic: $N_T = N + N_c$) which is (to a good approximation) constant within the cell [39]. The term representing activation of NF-κB is proportional to $(N_T - N)$, because this denotes the amount of cytoplasmic NF-κB, which is available for activation. The second term in the equation for N is a simple linear term which is proportional to R, due to the negative feedback. Note that this negative term is (for simplicity) independent of N. As a consequence of this, there is a chance that N may become negative - which would be un-physical. In order to avoid this we impose the special condition $\partial N/\partial t \geqslant 0$ when $N = 0$. Numerically, this is implemented by resetting $N = 0$ every time N becomes negative.

The first term in the equation for R represents production of inhibitors and regulators through transcription and translation and is to a first approximation linearly proportional to N. The second term in the equation for R represents simple decay with the half-life τ_R

The first term in the equation for C represents production of cytokines through transcription and translation and is linear proportional to N. The second term in the equation for C represent simple decay with the cytokine half-life τ and the third term represents diffusion of cytokines between cells. Note that only cytokines are allowed to diffuse between cells whereas NF-κB and inhibitors are of course contained within the cell membrane. Finally, the fourth term S^* represents an external production of cytokines. This external cytokine source serves as the initial stimuli in our simulations and is meant to represent cytokines produced by macrophages or other first responders at the site of infection - or in a laboratory setup the external cytokine source could represent cytokines which are added to the system via pipetting. In general S^* is a transient cytokine source in our simulations and for simplicity we choose a step-like function:

$$S^* = \begin{cases} S > 0 & \text{for } 0 < t < t_S \\ 0 & \text{otherwise} \end{cases} \tag{56}$$

In order to model the tissue we set up a 2D reaction-diffusion model of an islet consisting of a number of cells arranged on a hexagonal grid - see Figure 17C. In the tissue model the external cytokine source S^* only becomes non-zero at a specific position in the tissue. This position in meant to simulate the site of infection where cytokines are initially released from macrophages, and the aim of our model is then to study how the cytokine signal is transmitted through the rest of the tissue.

Parameters

The rate-constants k_{CN}, k_{RN}, k_{NR} and τ_R have been fitted to match the typical timescale of the NF-κB initial peak and are therefore fixed parameters for all our simulations ($k_{CN} = k_{RN} = k_{NR} = 5.0\ \mathrm{hr}^{-1}$ and $\tau_R = 2$ hr) [74]. The halflife and diffusion constant of cytokines are fixed based on experimental measurements: we use the diffusion constant $D = 2 \cdot 10^{-7} \frac{\mathrm{cm}^2}{\mathrm{sec}}$ and halflife, $\tau = 10$ minutes [12, 29, 40]. The cytokine concentration has been rescaled such that $K_A = 1$ and the cytokine concentration is measured in units of K_A (i.e. $K_A \to 1$ and $C \to C/K_A$). Also the diameter of a tissue cell is set to $d_{cell} = 15\mu$m. Similarly the NF-κB concentration has been rescaled in units of N_T (i.e. $N_T \to 1$ and $N \to N/N_T$) [39, 75, 76], thus the only free parameters of the model are p and S (and t_S). The parameter p sets the strength of the positive feedback, and therefore reflects how strongly the cytokine concentration is amplified by NF-κB regulation. The parameter S reflects the strength of the external cytokine source which is present at times $0 < t < t_S$.

Parameters		
d_{cell}	15	μm
k_{CN}	5	hr^{-1}
k_{RN}	5	hr^{-1}
k_{NR}	5	hr^{-1}
τ_R	2	hr
K_A	1	M/K_A
N_T	1	M/N_T
D	$2 \cdot 10^{-7}$	$\frac{\mathrm{cm}^2}{\mathrm{sec}}$
τ	10	min
p	variable	hr^{-1}
S	variable	$M/K_A\ \mathrm{hr}^{-1}$
t_S	variable	hr

Alternative Versions of the Model

The equations (53)-(55) resemble the "minimal version" of a model which has been developed over time. Earlier versions of the model

have slightly different equations but give the same overall results [74, 75, 76]. In earlier versions terms representing NF-κB-induced production were modeled not linearly proportional to N, but proportional to a term of the form $N^2/(N^2 + K_H)$. This alternative form takes into account cooperative effects caused by the dimeric nature of NF-κB, but the linear form is preferred because it is simpler yet gives similar results. The term regarding inhibition of NF-κB in equation (53) is only proportional to R, but should of course also depend on N. In earlier versions the term was therefore multiplied by a factor of the form $N/(N - K_N)$, however, we found that a very small K_N gave the best results and therefore we choose to ignore the term $N/(N - K_N) \approx 1$ in order to keep our equations as simple as possible. Effectively, this approximation assumes that the inhibition process is always saturated in N. Lastly, the hill coefficient in equation (53) has in earlier versions of the model had the value 3 instead of 2. A higher hill coefficient narrows the threshold effect slightly, but a hill coefficient $H = 2$ yields similar results.

4.4 RESULTS

This section will review the main results obtained from numerically integrating equations (53)-(55). This section will to some extend follow the results published in [75] and [76] and we shall firstly consider a single cell before moving on to a tissue model consisting of several cells.

SINGLE CELL MODEL

A single cell is simulated by numerically integrating equations (53)-(55), disregarding the diffusion term in equation (55). The equations have been implemented using a fourth-order Runge Kutta method with fixed stepsize. The system is initialized in the steady state $N = R = C = 0$, which is also interpreted as the resting state. Initially there is no stimuli, meaning no external source of cytokines: $S^* = 0$.

Response to Different Stimuli

At time $t = 0$ the system is stimulated by "turning on" the external cytokine source ($S^* = S > 0$) and the response to different stimuli is investigated - see Figure 18. The external cytokine source S causes C to increase to a new steady state level where the positive term S is balanced by the degradation with halflife τ. If the stimuli, S, is strong enough and the cytokine concentration exceeds the threshold level required to activate NF-κB (roughly speaking if C exceeds the level $C^* \approx K_A = 1$) the system will respond with an increase in N which in turn amplifies C further. Depending on the value of S, three qualitatively different scenarios can be achieved: (i) if S is small the increase in C will not activate N (Figure 18A) and C only

increases slightly. (ii) If on the other hand S is large (Figure 18C) the increase in C will cause an increase in N which in turn causes R to increase. As a result a new steady state will be obtained in which R is high and both N and C are balanced at levels significantly higher than pre-stimulation values. (iii) Intermediate values of S (Figure 18B) will very interestingly lead to oscillations. As we shall see below this behavior arises due to the bistable nature of the system and the system oscillates between low and high steady states. In all three cases the system settles back to the resting state after the initial stimuli vanishes - i.e. if the external source of cytokines is "turned off" after some time ($S^* = 0$).

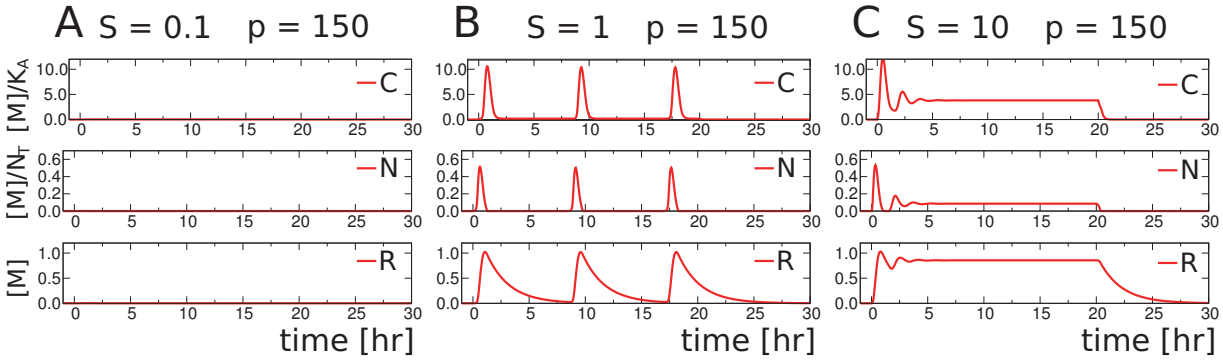


Figure 18: Numerical simulations of a single cell described by equations (53)- (55). Different values of the parameter S, result in three qualitatively different responses. A: A small S does not lead to any significant response. B: Intermediate S leads to an oscillating NF-κB (N) response and therefore also an oscillating amplification of the cytokine concentration (C). C: A large S leads to a high steady state, in which both N and C are balanced at levels significantly higher than pre-stimulation values. Once the initial source is removed ($S = 0$ at time $t = t_S = 20$ hrs), the system return to the resting state.

Locked States and Chronic Inflammation

The output of the model also depends strongly on the parameter p. Also three qualitatively different scenarios can be obtained keeping S constant and varying p - see Figure 19. The three scenarios have a lot in common with the scenarios seen in Figure 18, although there are some appreciable differences. For the scenarios shown in Figure 19 the external source S is kept at the same value as the one used in Figure 18B - i.e. the external cytokine source *is* strong enough to activate NF-κB. But as seen in Figure 19A if p is very small the positive feedback does not reach its full capacity and as a result almost nothing happens. Intermediate p leads to oscillations (Figure 19B) and high p leads to a new steady state in which both N, R and C are balanced at levels significantly higher than pre-stimulation values (Figure 19C). There is a very interesting difference between the situation in Figure 18C and Figure 19C: when the external cytokine source is removed ($S^* = 0$ at $t = t_S$), the system with high p does not return to the resting state. If p is high enough (as in Figure 19C) the system will maintain the high-level fixed point even after removal of

stimuli and we shall refer to such situations as "locked states". Physiologically, these situations are very interesting as they correspond to situations of constant NF-κB activation which does not resolve by it self. A locked state may therefore resemble situation of chronic inflammation.

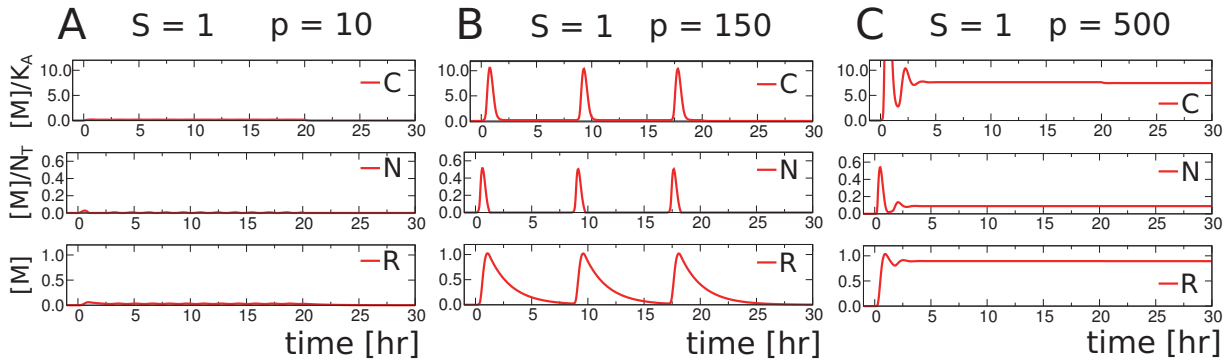


Figure 19: Numerical simulations of a single cell described by equations (53)- (55). Different values of the parameter p, result in three qualitatively different responses. A: A small p does not lead to any significant response. B: Intermediate p leads to an oscillating NF-κB (N) response. This panel is exactly equal to panel B in figure 18 C: A large p leads to a high steady state, in which both N and C are balanced at levels significantly higher than pre-stimulation values. In contrast to the situation shown in figure 18C, the system of high p, does not return to the resting state once the initial source is removed ($S = 0$ at time $t = t_S = 20$ hrs). This situation is referred to as a "locked state".

Phase Space Analysis

The system described by the equations (53)-(55) has many things in common with other excitable media as e.g. the Belousov Zhabotinsky reaction, and in order to analyse the system we shall follow an approach similar to the one described by Meron [46]. We shall see that plotting the nullclines and the phase space of the system leads to an easy apprehensible understanding of the system, and confirms for us, the notion that the different versions of the model (described in the end of Section 4.3) will in fact lead to similar qualitative results.

We notice that N and C are fast variables whereas R is a relatively slow variable due to the long halflife τ_R. Thus the model contains two effective timescales and as a first approximation we assume that N and C will effectively reach steady state and adiabatically follow changes in R. In order to understand the system we therefore investigate the phase space and nullclines of N and C for fixed values of R in the N-C-plane. The nullclines are defined by the equations:

$$\text{N-nullcline:} \quad 0 = k_{CN} \cdot \frac{C^2}{C^2 + K_A^2} \cdot (N_T - N) - k_{RN} \cdot R \tag{57}$$

$$\text{C-nullcline:} \quad 0 = p \cdot N - C/\tau + S^* \tag{58}$$

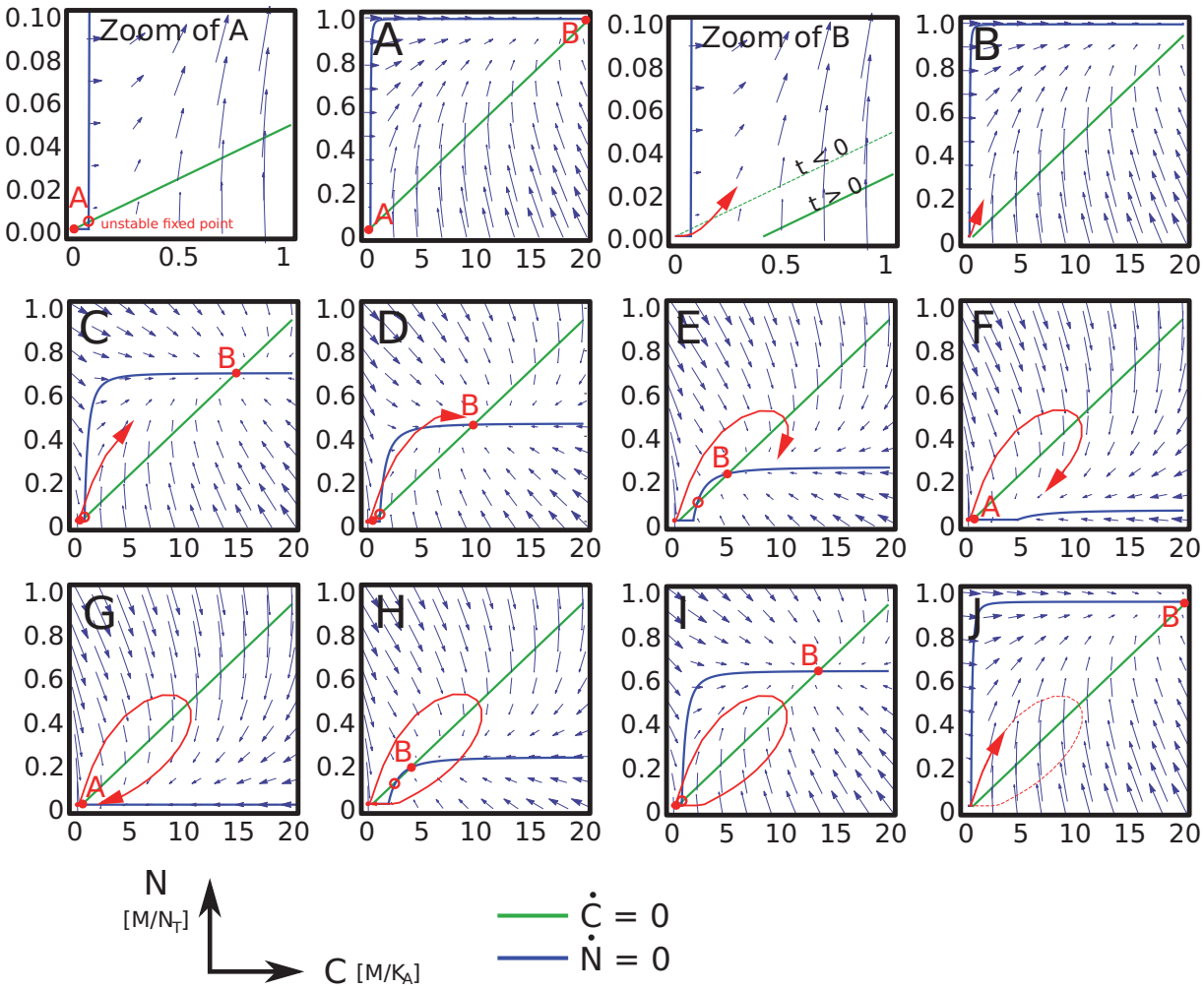


Figure 20: The situation shown in Figure 18B is here shown in the phase plane of N and C. Nullclines are plotted in blue ($dN/dt = 0$) and green ($dT/dt = 0$) lines. Stable fixed points are indicated with solid red dots. Unstable fixed points are indicated with red circles. See text for a description of the different panels. The times corresponding to the panels are: A) $t = 0.0$ to $t = 1.0$, B) $t = 1.1$, C) $t = 1.2$, D) $t = 1.4$, E) $t = 1.6$, F) $t = 2.0$, G) $t = 2.7$, H) $t = 3.7$, I) $t = 9.1$ hours.

or expressing N as a function of C:

$$\text{N-nullcline:} \quad N = 1 - \frac{k_{RN} R(C^2 + K_A^2)}{k_{CN} C^2} \tag{59}$$

$$\text{C-nullcline:} \quad N = \frac{C}{\tau p} - \frac{S^*}{p} \tag{60}$$

In Figure 20 we plot different time instances of the system corresponding to the scenario seen in Figure 18B. Before stimulation ($t < 0$ and $S^* = 0$) the nullclines intersect in 3 distinct points in the N-C-plane. We notice that these intersections do not necessarily represent *global* fixed points of the system, but nevertheless, we shall refer to the intersections of the nullclines in the N-C-plane as fixed points.

Before stimulation ($t < 0$ and $S^* = 0$) there are two stable fixed points separated by an unstable fixed point in-between, and the sys-

tem is therefore bistable. We shall refer to the two stable fixed points as "Fixed Point A" and "Fixed Point B" - see Figure 20A. For $t < 0$ (and $S^* = 0$) Fixed Point A and the unstable fixed point lie very close to each other in the N-C-plane and both have relatively low levels of N and C - see the zoom of panel A in Figure 20. Fixed Point A corresponds to the resting state and is in fact a global fixed point of the system.

At time $t = 0$ the external cytokine source is "turned on" ($S^* = S$) causing the C-nullcline to shift to the right by an amount $\delta = S \cdot \tau$, as shown in the zoom of panel B in Figure 20. Hence, if S is large enough, Fixed Point A and the unstable fixed point will disappear in a saddle-node bifurcation, and now the system will begin to evolve towards Fixed Point B, causing both N and C to increase (Figure 20B). An increase in N causes R to increase correspondingly. As this happens the N-nullcline is gradually altered - dynamically changing the phase-space as shown in Figure 20B-D. The system will continuously evolve towards Fixed Point B which is gradually moving in the N-C-plane (Figure 20B-D) - eventually making N and C decrease (Figure 20E). While the N-nullcline moves, Fixed Point A and the unstable fixed point have re-established in a new saddle-node bifurcation (since Figure 20C). Eventually R will increase to such high values that Fixed Point B coalesces with the unstable fixed point and disappears in a third saddle-node bifurcation (Figure 20E-F). Now, the system will evolve towards Fixed Point A, causing N and C to decrease back to almost pre-stimulation values (Figure 20F-H). As N is no longer high, R will begin to decrease because of spontaneous degradation - a relatively slow process set by the timescale of τ_R. This will cause the N-nullcline to move "back" (as shown in Figure 20G-I). As the N-nullcline moves, Fixed Point B and the unstable fixed point are re-established (Figure 20H) - but now the system is caught in the basin of attraction of Fixed Point A (Figure 20H). As R slowly decreases, the system rests in Fixed Point A (Figure 20I). Eventually, the N-nullcline has moved such that Fixed Point A and the unstable fixed point once again disappear in a saddle-node bifurcation - and the system will once again make a round in the phase space (Figure 20J).

Influence on the Phase Space by Different Values of S

The qualitatively different scenarios of Figure 18A and 18C can be well understood from an investigation of the phase space. In order to exhibit oscillations the system must be able to undergo all the saddle-node bifurcations described above: firstly, Fixed Point A and the unstable fixed point coalesce and, secondly, Fixed Point B and the unstable fixed point coalesce. The value of S sets the size of the C-nullcline shift, $\delta = S \cdot \tau$ (recall Figure 20, zoom of panel B). In Figure 18A, S is so small that the first bifurcation between Fixed

Point A and the unstable fixed point does not happen. Hence Fixed Point A is only moved very slightly towards higher values of N and C, but the system rests in this fixed point (which is still a global fixed point of the system). In Figure 18c, S is very large and the shift of the C-nullcline is correspondingly larger. In this scenario the system comes to rest in Fixed Point B *before* it disappears in a saddle note bifurcation - hence we conclude, that in this case, Fixed Point B must be a *global* fixed point of the system. If the external cytokine source is later removed ($S^* = 0$ at $t = t_S$) the C-nullcline shifts back to its original position and this may cause Fixed Point B and the unstable fixed point to coalesce such that the system can settle back to the resting state once again - as is the case in Figure 18c.

It is important to note that Fixed Point B must be a global fixed point in order to obtain a situation as the one seen in Figure 18c. Global fixed points of the system can be calculated by setting all three equations (53)-(55) equal to zero. However, just because there *exists* a global fixed point, it is not certain that the system will ever *reach* this fixed point. Therefore, a calculation of the global fixed points of the system only gives insight to whether the possibility is there. But the fate of the system also depends strongly on the initial conditions and the actions that perturb the system (e.g. the addition of an external cytokine source).

Influence on the Phase Space by Different Values of p

The slope of the C-nullcline is given by $(p\tau)^{-1}$, and in Figure 21 we plot the C-nullcline for different values of p together with the N-nullcline for two different values of R (low and high).

If p is very large the slope will be very flat - shifting Fixed Point B to larger C-values. A large p increases the possibility that the system will come to rest in Fixed Point B, and if p is very large, removal of the external cytokine source ($S^* = 0$ at $t = t^*$), which causes the C-nullcline to shift "back", may not be enough to cause Fixed Point B and the unstable fixed point to coalesce. Hence the system is locked in Fixed Point B.

If, on the other hand, p is very small the slope of the C-nullcline is very steep and even a small increase in R will cause Fixed Point B to lie very close to the resting state. Hence the system will come to rest in Fixed Point B, but seemingly the system will not react very much, because the fixed points are now so close.

Influence on the Phase Space by Different Values of τ

In our simulations we keep τ fixed, because its size is (roughly) known from the literature. However, changes in the parameter τ has effects similar to both S and p: the size of the shift in the C-nullcline is given by $\delta = S \cdot \tau$, and the slope of the C-nullcline is given by $(p\tau)^{-1}$. The effect of τ makes intuitive sense, as all parameters (S, p

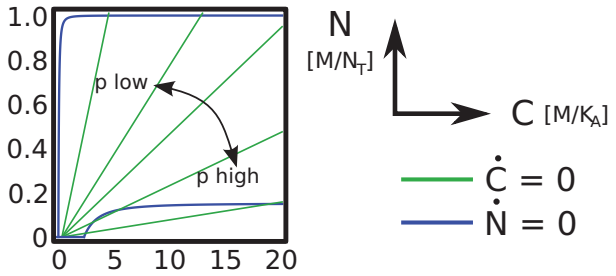


Figure 21: Visualisation of how changes in the parameter p influences the nullclines. Nullclines are plotted in blue ($dN/dt = 0$) and green ($dT/dt = 0$) lines. The N-nullcline is plotted for two different values of R (high and low - compare with Figure 20). The C-nullcline is plotted for a range of different p-values.

and τ) influence the total amount of cytokines in the system. Large S and p both contribute to a large production of cytokines, whereas a large τ slows down the decay of cytokines.

Alternative Versions of the Model

By plotting the nullclines in phase space, it is possible to directly understand, how small changes in the equations (53)-(55) would qualitatively affect the results. We shall now investigate how the nullclines would look, if we had used an alternative version of the model - as the ones described by the end of Section 4.3. Originally, this analysis was performed in the opposite order: starting with a more sophisticated version of the model, investigating the nullclines inspired us to simplify the model to the final version given by equations (53)-(55).

The N-nullcline is given by equation (59), expressing N as a function of C. For $C >> K_A (= 1)$ the N-nullcline approaches a constant: $N = 1 - k_{RN} R / k_{CN}$. For smaller values of C the N-nullcline curves sharply and approaches $N = -\infty$. However - as mentioned in Section 4.3 - we also impose the special condition $\partial N/\partial t \geqslant 0$ for $N = 0$ (in order to avoid negative values of N). This part of the nullcline may seem a bit awkward, as it results in a non-differentiable kink - this is best visible in the N-nullcline in Figure 20E. However, the kink would have appeared as a smooth curve, if the last term of equation (53) had been given by $k_{RN} R \cdot N/(N + K_1)$ instead of the simpler version $k_{RN} R$. The corresponding N-nullcline is plotted in Figure 22 (compare with Figure 20E), where we see the kink is now replaced by a smooth curve. The curve increases its sharpness as the parameter K_1 decreases, and the simple version given by equations (53)-(55), corresponds to $K_1 = 0$. Physically, this resembles a situation in which inhibition of N is always saturated in N - therefore only depending on the concentration of R (the inhibitors).

The C-nullcline in Figure 22 also represents an alternative model. Here, the first term in equation (55) is given by $pN^2/(N^2 + K_2^2)$ in-

stead of pN. We see that the alternative version, which is designed to take into account the dimeric nature of NF-κB (as described by the end of Section 4.3), results in a slightly curved C-nullcline, instead of a straight line. However, we know from the phase space analysis, that such a sophisticated form of the C-nullcline is not crucial for the qualitative results and therefore we choose the simple linear form in order to minimize the number of unknown parameters.

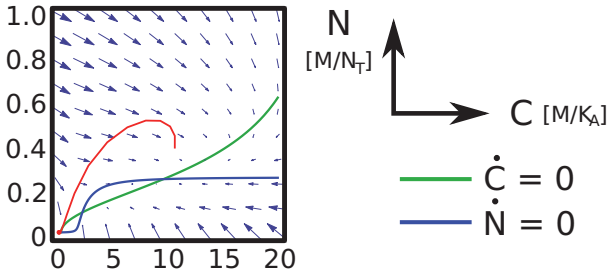


Figure 22: Visualisation of how the nullclines would look for the alternative model described in the text (end of Section 4.3). Nullclines are plotted in blue ($dN/dt = 0$) and green ($dT/dt = 0$) lines. The N-nullclines plotted here can be directly compared to Figure 20E. The alternative version of the model produces nullclines with softer curves, but as can be inferred from the plot, the alternative model leads to similar qualitative results.

TWO DIMENSIONAL TISSUE MODEL

In order to develop a spatial model of the tissue, we construct a two-dimensional lattice of cells - see Figure 17C. Every cell in the lattice is able to regulate cytokine production as described by equations (53)-(55). Cytokines are allowed to diffuse between cells and - as we shall see below - this gives rise to interesting spatio-temporal patterns of high cytokine concentration.

Numerically the spatial model is implemented by updating the temporal and spatial parts of the differential equations in turn. In practice this means the equations (53)-(55) are solved in two steps: all terms except the diffusion term in equation (55) are updated using a fourth-order Runge Kutta method with fixed stepsize. The diffusion term is updated using a simple Euler integration scheme, where cytokines are allowed to diffuse to the six nearest neighbors in the hexagonal grid. We use open boundary conditions.

Cytokine Waves

During inflammatory response only cells at the site of infection would be subject to an external stimulus, and - as sketched in Figure 17C - we simulate this by adding the external stimulus S *only to one cell* in the two-dimensional lattice. The specific values of S need to be somewhat larger than the ones used in the single cell simulations, because diffusion makes the effective removal of cytokines much larger.

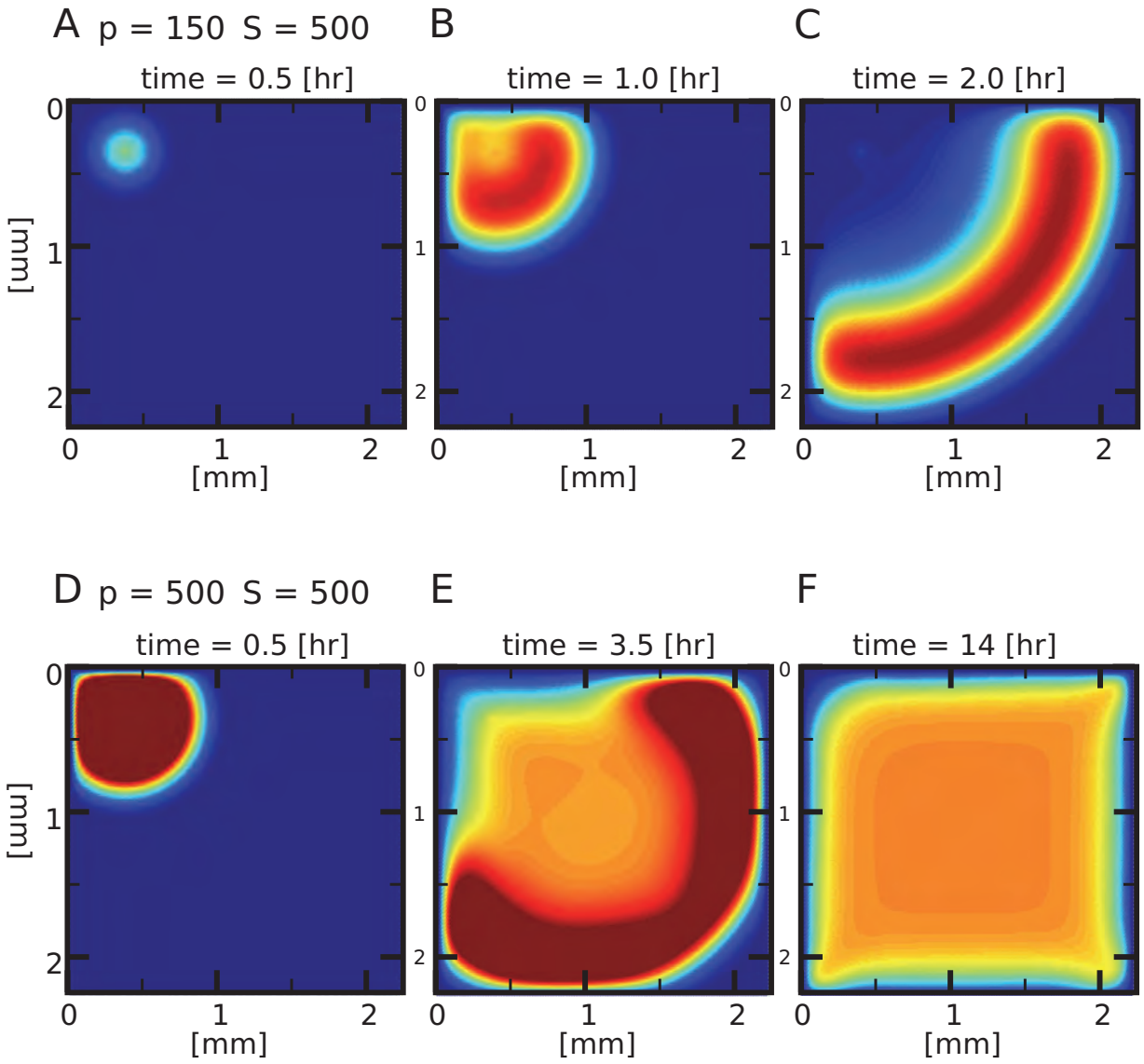


Figure 23: Numerical simulations of the 2D tissue model sketched in Figure 17C. A: Intermediate values of the parameter p leads to cytokine waves propagating from the site of infection, where the initial source S is added. When the initial source is removed at time $t = t_S = 10$ hr, the system settles back to the resting state (not shown). B: High values of the parameter p leads to locked states, where the cytokine concentration (C) is sustained at an elevated level. Initially the system propagates a wave of very high cytokine concentration and the situation shown in the second panel ($t = 3.5$ hr) corresponds to the second wave. Hereafter the system sustains an elevated level of cytokines, which only decays close to the absorbing boundary. The initial source is removed at time $t = t_S = 10$ hr, and as can be seen from the third panel, the locked state is sustained after the initial source is removed.

In Figure 23A we show a scenario, in which the stimulated cell oscillates. At time $t = 0$ an external cytokine source is added to the stimulated cell (i.e. $S^* = S > 0$ *only* at the stimulated cell), which starts to amplify the cytokine concentration. The cytokines diffuse to neighboring cells which consequently also get stimulated - and thus a wave of high cytokine concentration is created. The oscillatory behavior of the central cell will initialize new waves until the

external stimulus is removed (by resetting $S^* = 0$ at the stimulated cell). The cells which have $S^* = 0$ will only get stimulated when they feel a spill-over of cytokines from their neighbors and hence the situation is indeed cooperative in the sense that the cytokine wave is truly propagated from one cell to the next.

As mentioned in section 4.1 and sketched in Figure 16C the propagating cytokine waves are very interesting, because they may represent an optimal spatio-temporal chemoattractant signal.

Locked States on the Tissue Level

Locked states similar to the ones described above for the single cell (Figure 19C) can also exist in the spatial model. If the parameter p is high enough all the cells in the system will "lock" - see Figure 23B. As a result the cytokine concentration is constantly elevated and a gradient is only present close to the absorbing boundaries. again the locked state is very interesting because it may resemble chronic inflammation in the tissue.

4.5 DISCUSSION OF THE MODEL VALIDITY AND PHYSIOLOGICAL RELEVANCE

The simple model presented in this chapter neglects many details of the regulatory system of a true cell, and can therefore not be expected to reproduce the entire truth. However, the simplified model does capture many interesting features that are also observed experimentally.

Experimental studies of inflammatory response following myocardial infarction (heart attacks), have shown that while all infarctions lead to robust amplification of cytokine concentrations within the early response, the cytokine concentrations may return to baseline levels if the infarction is small. If, on the other hand the infarction is large (in our model: large S), or if the inflammatory response of the host is exuberant (in our model: large p), there can be either "*sustained cytokine upregulation or a second wave of cytokine upregulations*" [49]. These observation agree very well with the locked states and the oscillating waves produced by our simple model. Additionally, the studies of myocardial infarction have demonstrated that, a region with high cytokine concentration is able to activate cytokine upregulation in remote regions, which were not affected by the initial infarction [35, 49, 52]. Again this observation agrees very well with our simple model, which demonstrates that NF-κB mediated cytokine amplification is able to spread through tissue as an excitable media.

5 ISLETS OF LANGERHANS

Islets of Langerhans are tiny clusters of pancreatic cells, that sense and regulate blood sugar levels by secreting glucagon and insulin (as well as some other hormones). Glucagon is produced and secreted by α-cells and insulin is produced and secreted by β-cells. The shape and size of the islets of Langerhans varies from between species, but they are often somewhat spherical in shape and many species (including humans and rodents) have the β-cells arranged in the core of the islet, and α-cells arranged around the periphery of the islet. Rodents have very spherical islets, while human islets have a more complex and folded structure.

In this section we shall investigate inflammation and cytokine regulation in the islets of Langerhans. The finite size of the islets imposes certain restrictions on the spreading of the inflammatory signal and as we shall see, the geometry of the islets, may also play a functional role. To model the β-cells, we use the simple mathematical model described in Chapter 4, and the main focus of this chapter will be to investigate different spatial arrangements of the islets.

5.1 TYPE II DIABETES

Dysfunctional, chronic or systemic inflammation is related to a vast variety of human diseases, many of which are characterised as lifestyle diseases, that develop over a long time period, due to a set of very complex conditions. These diseases have become increasingly dominant during the 20th century, as the average lifetime has increased, and advances in technology has allowed us to treat many infectious diseases, which were previously the most dominant lethal diseases.

Dysfunctional regulation of the inflammatory response also plays a key role in the pathology of type II diabetes[17, 19]. Type II diabetes is characterized by high blood glucose (hyperglycemia) as a consequence of insulin resistance or insulin deficiency. Insulin production is governed by β-cells, which respond to insulin deficiency by enhancing insulin production. However, if the β-cells fail to compensate sufficiently the islets of Langerhans may enter a state of chronic inflammation. Over time chronic inflammation in the islets of Langerhans may lead to β-cell death, which worsens the situation further.

5.2 β-CELLS AND IL-1β

The key mediator of inflammatory activity in β-cells is the cytokine IL-1β which is known to play a dual role: while transient exposure to elevated IL-1β concentrations improves insulin secretion and promotes β-cell survival, prolonged exposure to high levels of IL-1β leads to chronic inflammation, impaired insulin secretion and β-cell death[19, 20, 65].

IL-1β activates NF-κB and the main signaling pathways of IL-1β in a β-cell are shown in Figure 24A. Active NF-κB induces production of pro-IL-1β, which needs further processing before mature IL-1β is finally liberated into the extracellular space. The NLRP3 inflammasome (which is activated by high glucose levels as well as other danger signals) cleaves pro-caspase-1, increasing amounts of active caspase-1, which in turn cleaves pro-IL-1β and forms mature IL-1β.

Hence the system constitutes an amplifying feedback loop in which IL-1β induces its own expression in a self-sustaining manner. This feedback has previously been referred to as a *vicious cycle*[17, 20, 42].

5.3 MODEL

The signaling network of IL-1β in a β-cell (Figure 24A) is very similar to the one described in chapter 4 (Figure 17A). As a first approximation we therefore model the β-cells using a similar model (Figure 24B) - but this time the cytokines (C) represent the specific cytokine IL-1β (I). A pancreatic islet is modeled as a small cluster of β-cells in a two-dimensional model - see Figure 17C.

Equations

The β-cells regulate the concentration of IL-1β according to the equations:

$$\frac{dN}{dt} = k_{IN} \cdot \frac{I^2}{I^2 + K_A^2} \cdot (N_T - N) - k_{RN} \cdot R \tag{61}$$

$$\frac{dR}{dt} = k_{NR} \cdot N - R/\tau_R \tag{62}$$

$$\frac{dI}{dt} = p \cdot N - I/\tau + D\nabla^2 I + S^* \tag{63}$$

which are explained in more detail in Chapter 4. The β-cell cluster is surrounded by tissue cells, which do not respond noticeably to IL-1β. Therefore, this part of the model consists of cells which do not regulate IL-1β, but simply allows IL-1β to diffuse and decay. These cells (indicated by the grey cells in Figure 17C) are therefore described by the equation:

$$\frac{dI}{dt} = -I/\tau + D\nabla^2 I \tag{64}$$

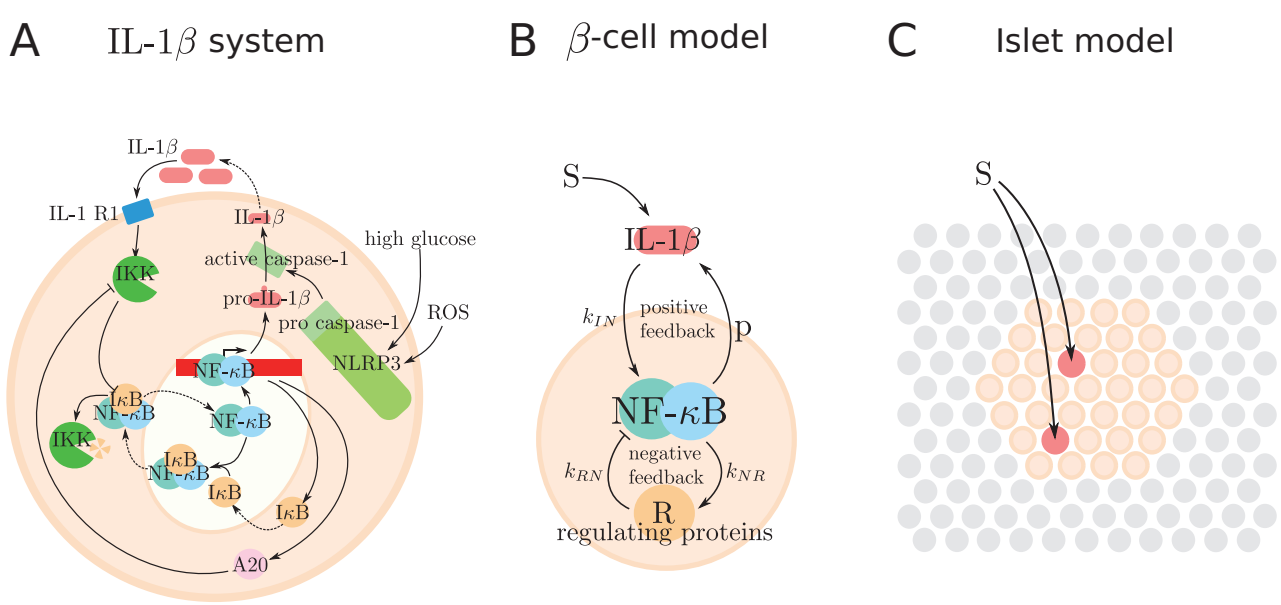


Figure 24: A: The regulatory system of NF-κB and IL-1β in β-cells. The main NF-κB system is describes in Chapter 4, and can also be seen in Figure 17. Here we also show the extra steps of regulating IL-1β, through activation of the NLRP3 inflammasome and caspase-1. The NLRP3 inflammasome is activated, by high glucose levels, reactive oxygen species (ROS) as well as other stress and danger signals. B: A sketch of the simple model of the NF-κB system and IL-1β regulation, described by the equations (61)-(63) (the model is similar to the model described in Chapter 4, but sketched again here for convenience). C: 2D model of an islet. The β-cells are arranged in a round cluster and surrounded by cells, which do not amplify IL-1β (indicated by the grey color). The surrounding cells allow for diffusion and degradation of IL-1β (as described by equation (64)). Also indicated (by red cells) are the IL-1β sources, described by the parameter S (see equation (65)). Here 2 sources are sketched, but the number and configuration of sources may vary.

Initial IL-1β Sources

In order to model the early stages of the disease we add a number of IL-1β sources in the islet. It was previously assumed that the initial increase in islet IL-1β was produced by islet-invading immune cells (e.g. macrophages), but it has recently been hypothesized that the initial increase in islet IL-1β is mainly produced by stressed β-cells, and that immune cells invade the islet later (increasing IL-1β levels further)[20]. In our model we do not specify the originators of the initial increase of IL-1β but simply add to the model a number of spatially distributed IL-1β-sources - i.e. a discrete number of cells within the islet have positive $S^* = S > 0$ for a finite time ($0 < t < t^*$):

$$S^* = \begin{cases} S > 0 & \text{for } 0 < t < t_S \\ 0 & \text{otherwise} \end{cases} \tag{65}$$

NLRP3 *Inflammasome Activity*

The severity of hyperglycemia is reflected through the NLRP3 inflammasome and (in turn) caspase-1 activity - see Figure 24A. The caspase-1 activity is implicitly reflected through the model parameter p, which also reflects the strength of the positive feedback (recall Chapter 4).

Parameters

The number of initial IL-1β sources is described by the new parameter NS (Number of Sources). In order to easily investigate the effect of changing the number of sources, we fix the parameters S and t_S. We also introduce the islet radius R (in terms of an integer number β-cell diameters) and the total number of β-cells within the islet N. All remaining parameters are unchanged and equal to the ones used in Chapter 4.

Parameters		
$d_{\beta cell}$	15	μm
k_{CN}	5	hr^{-1}
k_{RN}	5	hr^{-1}
k_{NR}	5	hr^{-1}
τ_R	2	hr
K_A	1	M/K_A
N_T	1	M/N_T
D	$2 \cdot 10^{-7}$	$\frac{cm^2}{sec}$
τ	10	min
p	variable	hr^{-1}
NS	variable	integer
S	500*	$M/K_A\ hr^{-1}$
t_S	12*	hr
R	variable	β-cell diameters
N	variable	integer

* = fixed value.

5.4 RESULTS

As expected from the results described in Chapter 4, we shall see below, that the islets can enter states of oscillating or of chronically elevated IL-1β concentrations. As the islets are not very large, propagating waves of IL-1β will not be visible - instead the entire islet will oscillate almost synchronously. We relate the oscillating state of the model to a healthy inflammatory response and the locked state to a pathological state of chronic inflammation - the so-called *vicious*

cycle. Over time chronic inflammation leads to β-cell death, and the locked state is therefore lethal for the islets. The two qualitatively different states can therefore be related to two entirely different islet fates.

The aim of this section is to investigate which situations - and especially which spatial conditions - that lead to different cell fates. We have already seen in Chapter 4 that high p lead to locked states, and we shall confirm this behaviour for the islets. In order to study various spatial effects, we shall simulate islets of varying "internal" and "external" spatial arrangements. Internal spatial arrangements refer to the number of initial sources and the exact configuration of these within the islet. External spatial arrangement refer to varying sizes and shapes of the islet as a whole.

Pulsing and Locked Islets

Simulations of islets with increasing values of the parameters p, can be related to islets subject to increasing severity of hyperglycemia, as the positive feedback is to some extend controlled by the NLRP3 inflammasome (recall Figure 24A and 24B).

The results of simulating identical islets, but with two different values of p, can be seen in Figure 25. Here we have chosen an islet, with 5 IL-1β sources, and each source has a fixed strength $S^* = S$ for the finite time $0 < t < t_S$. If p is small there is effectively no response (not shown). For intermediate p the islet responds with a pulsating NF-κB activity and a similar pulsating amplification of IL-1β. Once the sources are removed ($S^* = 0$ at $t = t_S$) the islet stops pulsing and returns to the resting state, and this scenario is interpreted as healthy and well-functioning. On the other hand a relatively large p results in a locked state, from which the islet does not recover, when the initial sources are removed. This scenario is interpreted as dangerous chronic inflammation, which is possibly lethal for the islet. We do not see examples of half- or partially locked islets. The islets either survive or get locked completely - a result which is also consistent with experimental observations [22].

Analytical Solution of the IL-1β Concentration Profile

We can derive an analytical expression for the concentration of IL-1β as a function of space, by considering the steady state values of equations (61)- (62):

$$0 = k_{IN} \cdot \frac{I_{ss}^2}{I_{ss}^2 + K_A^2} \cdot (N_T - N_{ss}) - k_{RN} \cdot R_{ss}$$

$$0 = k_{NR} \cdot N_{ss} - R_{ss}/\tau_R$$

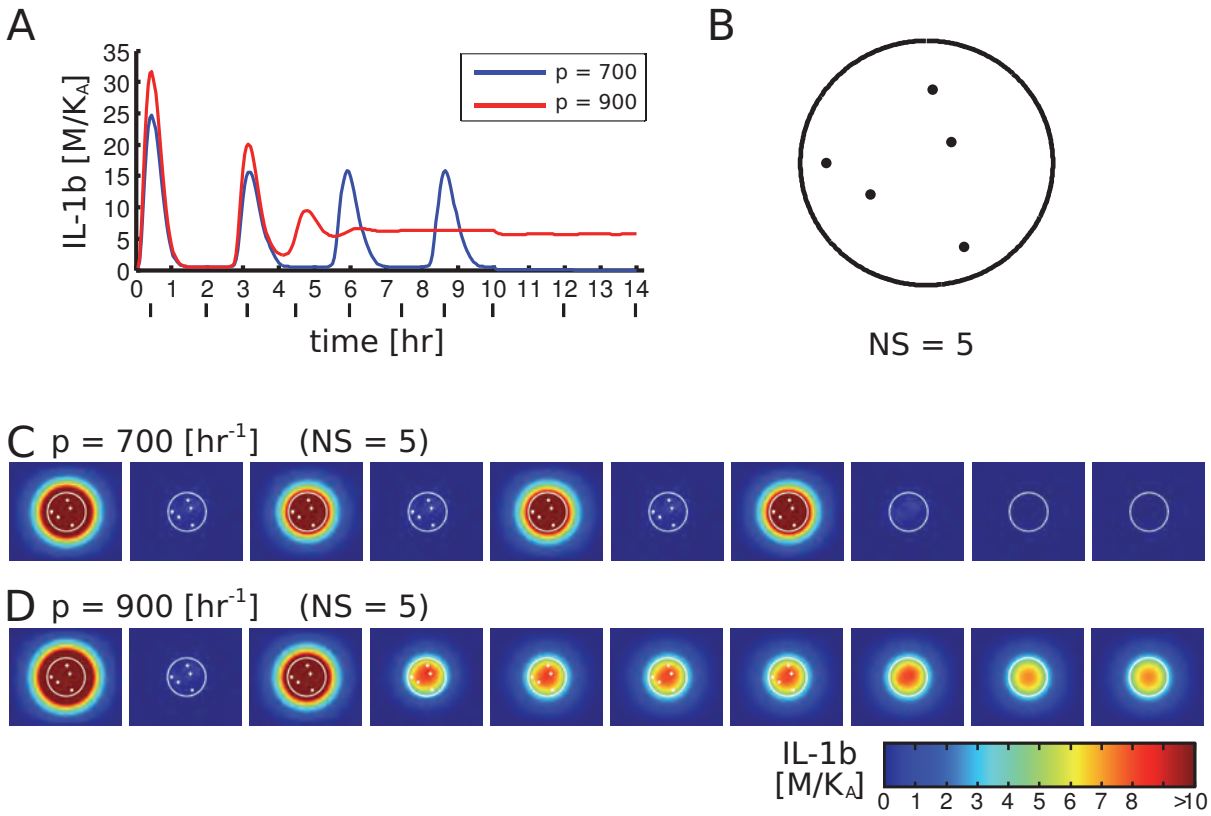


Figure 25: Simulations of similar islets with different values of the parameter p (see equation (63)). A: The average concentration of I (IL-1β) as a function of time. Both islets are subject to 5 IL-1β sources and sources are removed after 10 hours ($t_S = 10$ hr). B: Configuration of the 5 IL-1β sources. C-D: Snapshots of the concentration of I for the two islets, with different p-values (C: $p = 700$ and D: $p = 900$). The times of the different snapshot are indicated by the vertical bars in A. The islet periphery and positions of the IL-1β sources are indicated with white. The radius of all islets is $R = 10$ cell diameters.

we obtain the steady state solution for NF-κB:

$$N_{ss} = \frac{I^2}{I^2(1+\gamma)+\gamma} \tag{66}$$

where we have used $N_T = 1$, $K_A = 1$ and $\gamma = \frac{k_{NR}k_{RN}\tau_R}{k_{IN}}$. Substituting equation (66) into the steady state equation for IL-1β (recall equation (63)), we obtain the following expression:

$$0 = p \cdot \frac{I_{ss}^2}{I_{ss}^2(1+\gamma)+\gamma} - I_{ss}/\tau + D\nabla^2 I_{ss} \tag{67}$$

where we have set $S^* = 0$, in order to investigate cases of self-sustained locked states.

In locked islets, we expect the concentration of IL-1β to be relatively large. If $I_{ss} >> 1$, we can approximate $\frac{I_{ss}^2}{I_{ss}^2(1+\gamma)+\gamma} \approx \frac{1}{(1+\gamma)}$, leading to the following equation:

$$0 = p \cdot \frac{1}{(1+\gamma)} - I_{ss}/\tau + D\nabla^2 I_{ss} \tag{68}$$

In the 2D case I_{ss} is a function of the spatial coordinates x and y, or in cylindrical coordinates a function of r. Equation (68) has the following solution:

$$I_{ss}(r) = \left(I_{ss}(0) - \frac{p\tau_I}{1+\gamma}\right) \text{BesselI}_0\left(\frac{r}{\sqrt{\tau_I D}}\right) + \frac{p\tau_I}{1+\gamma} \tag{69}$$

Equation (69) describes the concentration of IL-1β, in a locked islet, and is true for all $r < R$, where R equals the islet radius.

Outside the islet (for $r > R$), the following equation describes the IL-1β concentration:

$$0 = -I_{ss}/\tau + D\nabla^2 I_{ss} \tag{70}$$

This equation has the following solution:

$$I_{ss}(r) = A \cdot \text{BesselK}_0\left(\frac{r}{\sqrt{\tau_I D}}\right) \tag{71}$$

which describes the concentration of IL-1β, outside the islet.

We can now determine the two unknown constants $I_{ss}(0)$ and A, by demanding that the profile of $I_{ss}(r)$ is both continuous and differentiable at the islet boundary $r = R$. From these two conditions we obtain the following expressions:

$$A = \frac{p\tau_I R}{\sqrt{\tau_I D}(1+\gamma)} \text{BesselI}_1\left(\frac{R}{\sqrt{\tau_I D}}\right) \tag{72}$$

$$I_{ss}(0) = \frac{p\tau_I}{1+\gamma}\left(1 - \frac{R}{\sqrt{\tau_I D}} \text{besselK}_1\left(\frac{R}{\sqrt{\tau_I D}}\right)\right) \tag{73}$$

The IL-1β concentration given by equation (69) and (71)-(73) is plotted in Figure 26.

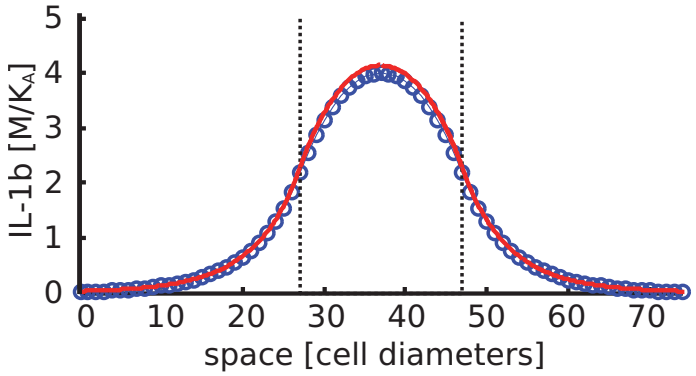


Figure 26: Cross section of the IL-1β concentration profile in a locked islet, as given by equations (69) and (71)-(73) (full red lines) and obtained from numerical simulations (blue circles). Dashed vertical lines indicate the islet boundary. Note: this plot is calculated using a hill-coefficient $H = 3$ (instead of $H = 2$ in equation (61)). Figure courtesy of T. H. Hansen [34].

INTERNAL SPATIAL ARRANGEMENTS

In this section we shall investigate the effects of changing the number of sources, and the configuration of these, within the islet.

Response to Different Number of Sources (parameter NS*)*

The results of simulating islets with varying number of initial IL-1β sources can be seen in Figure 27. If the islet is stimulated by a relative small number of sources the islet responds with a pulsating NF-κB activity and a similar pulsating amplification of IL-1β (Figure 27C) - a scenario which is interpreted as healthy and well-functioning. If a similar islet is stimulated by a relatively high number of sources, the islet may transition into a locked state, from which it does not recover once the sources are removed (Figure 27D) - a scenario which is interpreted as dangerous chronic inflammation.

It is also possible to obtain scenarios in which the islet transition into a state that looks like a locked state, however, once the sources are removed ($S^* = 0$ at $t = t^*$) the islet "un-locks" and returns to the resting state (Figure 27E). This happens if the number of sources is high, but the parameter p is not quite high enough, to maintain the locked state. This scenario does, however, exhibit a transient a IL-1β increase and is therefore *not* interpreted as dangerous chronic inflammation.

Response to Different Source Configurations

Interestingly, there are some situations, which can lead to either pulsing or locked islets, depending *only* on the exact configuration of the sources within the islet. Examples of islets with equal parameters ($p = 700$ hr^{-1} and $NS = 7$), but different source configurations can be seen in Figure 28A and 28B, and we see that while some of the islets display transient IL-1β amplification the others get locked. This result is very interesting, as it implies, that a pancreas with many islets, subject to equal conditions (or from a modelling perspective equal parameters), may exhibit coexistence of islets that display (healthy) pulsating amplification of IL-1β and locked islets that display chronic inflammation. In Figure 28C we show how the probability of "getting locked" increases as the number of sources increases (averaged over 10 random source configurations). Only some parameters give rise to coexistence of pulsing and locked islets and we note that the transition (in parameter-space) from pulsing to locked is quite sharp (i.e. for a large part of the parameter space there is no coexistence - all configurations will lead to the same islet fate).

As can be seen from configuration example '3' and '4' (Figure 28B), changing only the position of a single source (out of 7 sources in total), can be enough to change the cell fate completely (see Figure

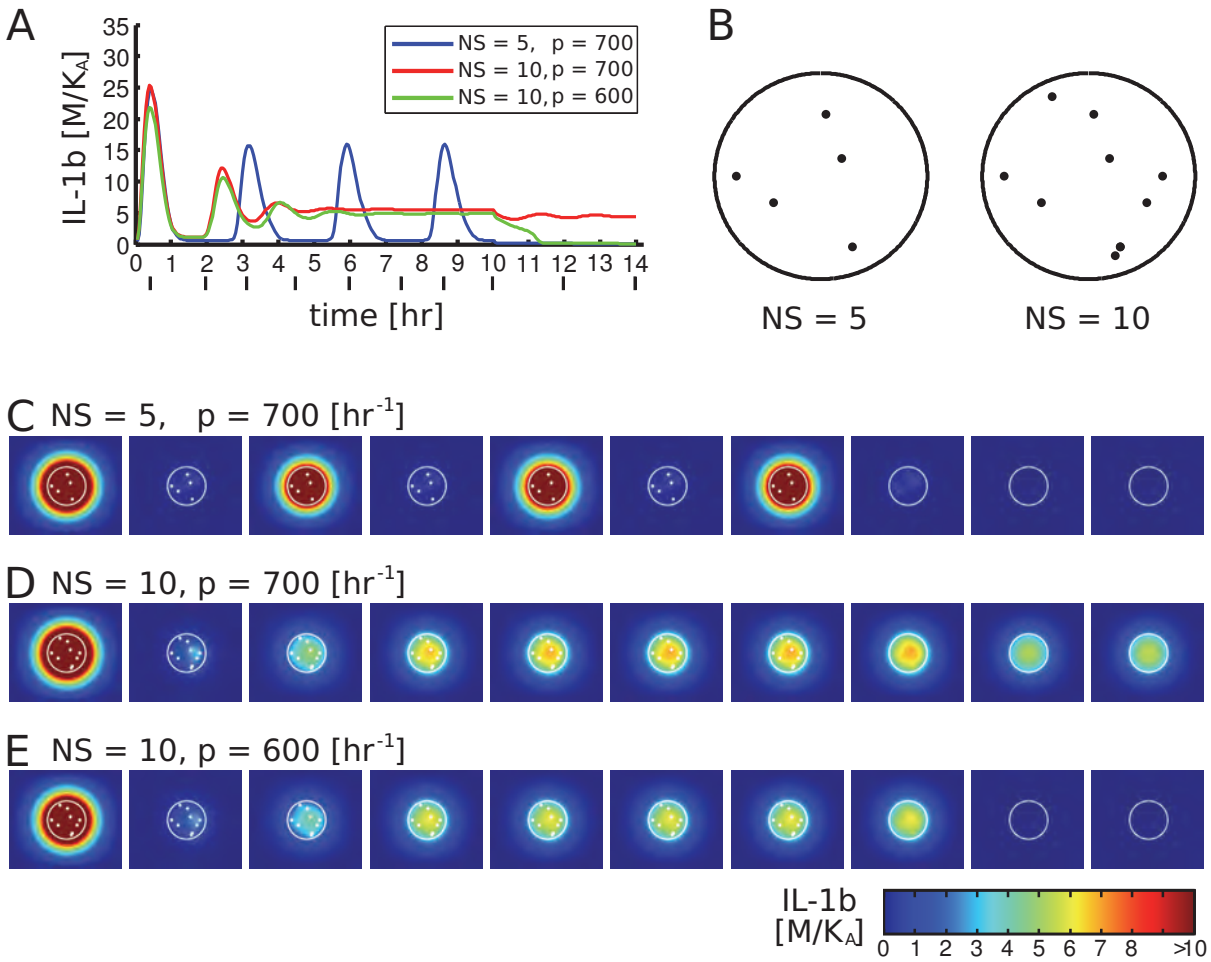


Figure 27: Simulations of similar islets with different Number of sources NS (and different p-values). A: The average concentration of I (IL-1β) as a function of time. One islet is subject to 5 IL-1β sources and two islets are subject to 10 IL-1β sources. The sources are removed after 10 hours ($t_S = 10$ hr). B: Configurations of the IL-1β sources. C-D: Snapshots of the concentration of I for the two islets, with similar p-value, but different number of sources (NS) (C: NS = 5 and D: NS = 10). E: Snapshots of the concentration of I for an islet, similar to the one shown in panel D, but with slightly lowered p-value. C-EThe times of the different snapshot are indicated by the vertical bars in A. The islet periphery and positions of the IL-1β sources are indicated with white. The radius of all islets is R = 10 cell diameters.

28A). From the shown examples of source configurations, it seems that a clustered configuration of sources leads to a locked state, more often than a situation with more dispersed configurations. We tested this assumption by measuring the root-mean-square displacement of the sources for 100 random configurations of otherwise equal islets ($p = 700$ hr^{-1} and $NS = 7$), however we conclude that the the root-mean-square itself does not clearly distinguish, which islets will get locked and which will not - see Figure 28D.

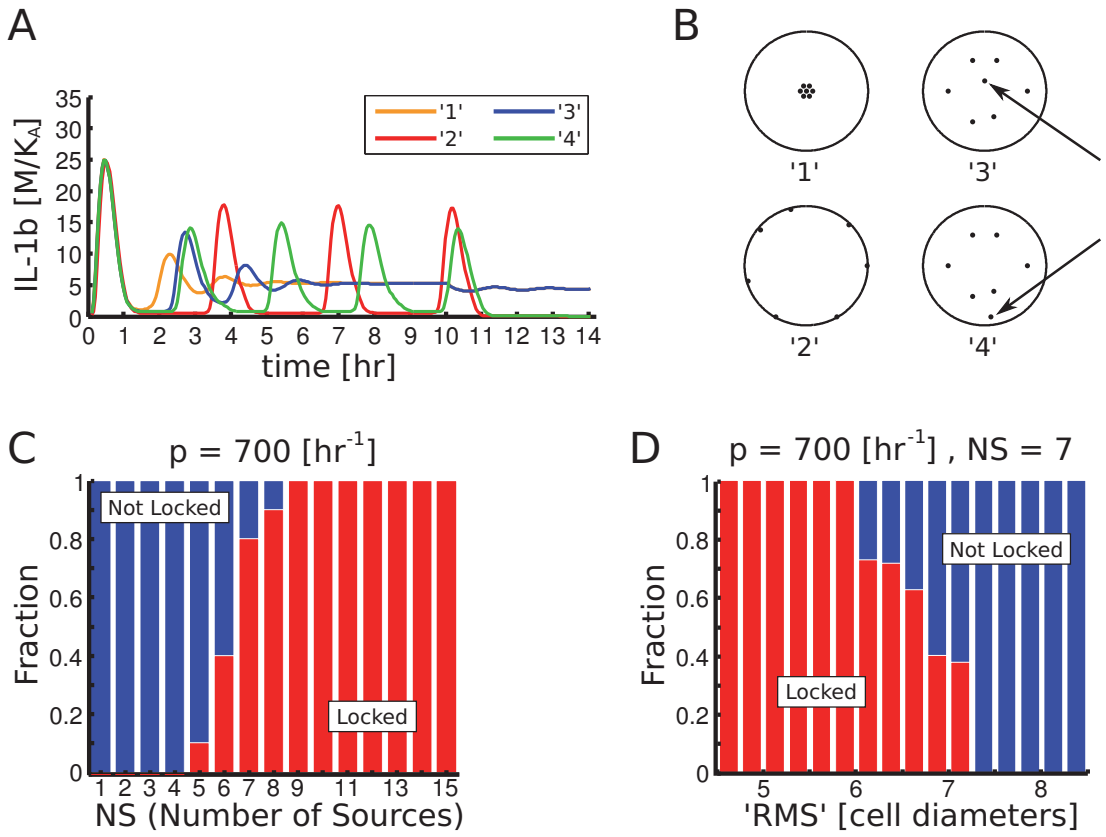


Figure 28: Simulations of similar islets with different source configurations. A: The average concentration of I (IL-1β) as a function of time. All islets are subject to 7 IL-1β sources (NS = 7) and p = 700 hr^{-1}. The sources are removed after 10 hours (t_S = 10 hr), and the number '1' to '4', shown in the legend corresponds to the source configurations shown in panel B. B: Configurations of the IL-1β sources, for the four different examples shown in panel A. C: The probability of getting "Locked" vs. "Not Locked", as a function of the number of sources, averaged over 10 random source configurations. D: The probability of getting "Locked" vs. "Not Locked", as a function of the root-mean-square displacement of the sources, average over 100 random configurations of 7 sources. The p-value and radius of all islets are p = 700 hr^{-1} and R = 10 cell diameters.

EXTERNAL SPATIAL ARRANGEMENTS

As mentioned above different species display differences in the exact spatial structure of the islets of Langerhans. Rodent islets are very spherical with the β-cells arranged in the middle and the α-cells arranged as an outer core. Human islets have a more folded structure, but share the tendency of having the β-cells in the middle and α-cells arranged peripherally [9, 23]. Some species have the opposite structure with α-cells in the middle and β-cells peripherally.

Despite the differences between species there are also many similarities and the structural arrangement of the islets may play a functional role. In this section we shall investigate the effects of changing size and shape of the islet.

Islet Size

In order to test whether the size of the islets may play a functional role we have performed simulations in which we changed the radius of the islets. We have scanned different combinations of the parameters p and NS and have evaluated 10 random source configurations for each parameter-set. In Figure 29, we plot both the probability of getting locked and the average IL-1β level in the islet, some time (here 4 hours) after the initial sources have been removed (both are averaged over the 10 random source configurations). The model predicts that small islets are less prone to transition into the locked state. This is an extremely interesting observation as it implicates, that the the reason why β-cells are grouped into small islets - as opposed to being arranged in a single large cluster - could be that the separation into smaller islets gives a protective advantage. From a modelling perspective, the reason that small islets are less prone to get locked, is that they have a higher chance of extruding IL-1β though diffusion, and therefore also a higher change of resetting to the resting state ("un-locking"). The arrangement of β-cells into smaller islets may therefore resemble a survival strategy which protects the islets from the very potent paracrine signalling of β-cells, which can cause a whole islet to "lock" if just a few cells start to over-produce IL-1β.

Critical p*-value as a Function of Islet Size*

From the analytical expression given by equation (69), we can obtain approximate values for the diffusion term ($D\nabla^2 I_{ss}$) in equation (67). Using this approximation, equation (67), becomes a polynomial in I_{ss} and the number of solutions for I_{ss} resembles the number of possible steady states for IL-1β. The number of solutions depends on the value of the parameter p: high p leads to steady state solutions of high I_{ss} - i.e. high p leads to the existence of locked states. If p is too small the locked states (high steady states) disappear and the corresponding solution to the polynomial given by equation (67) becomes imaginary. We define the critical p (p_c) as the smallest value of the parameters p, for which locked states exist. In Figure 30, we plot p_c as a function of islet size (R), as obtained from both numerical simulations and from numerical solutions to the polynomial given by equation (67), where the diffusion term ($D\nabla^2 I_{ss}$), has been evaluated both in the centre of the islet ($r = 0$) and at the boundary of the islet ($r = R$). Consistent with our findings in Figure 29 the critical p is highly dependent of islet size (radius R), and decreases sharply as islet size increases.

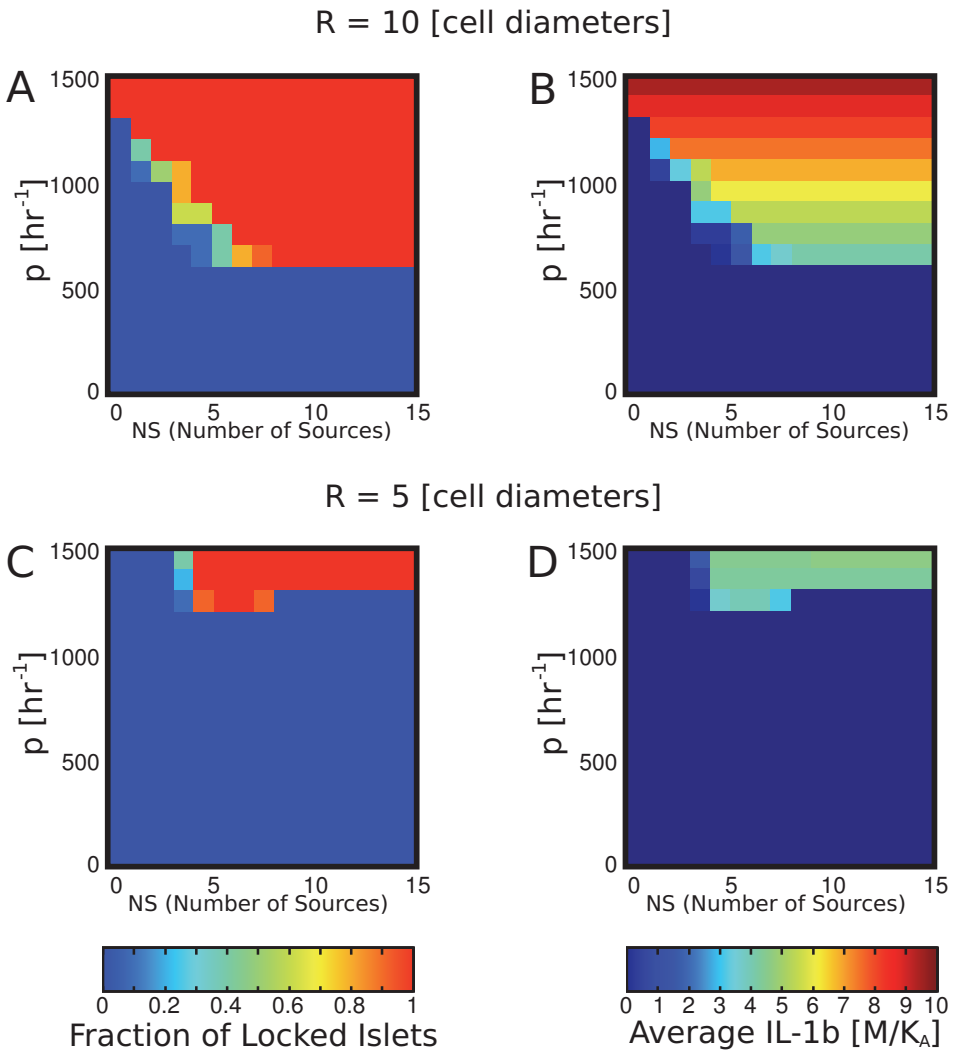


Figure 29: Parameters scans of p and NS for islets of different sizes. A-B: Islets of radius R = 10 cell diameters (corresponding to a total of N = 367 β-cells), transistion into locked states for a relatively large subset of parameters. C-D: Islets of radius R = 5 cell diameters (corresponding to a total of N = 91 β-cells), transistion into locked states for a relatively small subset of parameters. In panel A and C, we plot the fraction of islets which transition into the locked state, as averaged over 10 random source configurations for each parameter combination. In panel B and D, we plot the average concentration of I (IL-1β) inside the islet some time after the sources are removed. Specifically the sources were removed at $t_S = 10$ hrs and we plot the average concentration of I at time $t = 14$ hrs. Also, the values are averaged over 10 random source configurations for each parameter combination.

Islet Shape

The shape of the islet plays a similar role as the size: if the islets are not spherical but have folded shapes as observed in humans, the β-cells effectively become more dispersed, and therefore have better chances of extruding IL-1β and resetting to the resting state. In Figure 31 we compare different shapes of roughly the same total number of β-cells and conclude that a more complex shape may introduce a higher probability of avoiding the locked state. We should

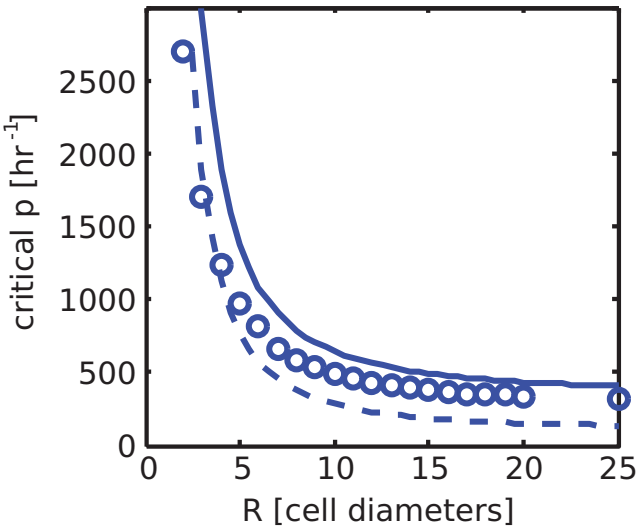


Figure 30: The critical p (p_c) as a function of islet radius R. The blue circles indicate solution obtained from numerical simulations. The blue lines represent the numerical solutions to the polynomial given by equation (67), where the diffusion term ($D\nabla^2 I_{ss}$), has been evaluated using the analytical expression (69) at $r = 0$ (full blue line) and $r = R$ (dashed blue line). Note: these plots are calculated using a hill-coefficient $H = 3$ (instead of $H = 2$ in equation (61)). Figure courtesy of T. H. Hansen [34].

however note, that the folded shape may also play other roles. In human islets almost all β-cells are in direct contact with at least one α-cell - hence it is possible that the glucagon-insulin regulation is more effective or sensitive.

5.5 DISCUSSION OF THE MODEL VALIDITY AND PHYSIOLOGICAL RELEVANCE

The simple model described by equations (61)-(63), does not capture all the details of the regulatory system within a β-cell. Especially, the notion that, the parameter p implicitly reflects the severity of hyperglycemia is not completely correct (work is currently being conducted in order to refine the model). However, the model poses a simple system which is capable of producing characteristic features of the IL-1β response in β-cells. As such, the 2D islet model poses an excellent tool for investigating the spatial aspects, described in this chapter.

Very interestingly the simple model captures some spatial features, that are also observed experimentally. Records of islet sizes and compositions, in cadaveric pancreatic sections from diabetic (type II) and non-diabetic humans, have shown a preferential loss of large islets in T2D patients [36]. This observation agrees well with our result showing that large islets get locked easier than the small islets, and therefore the β-cells in the large islets should also have a higher risk of dying. Additionally, it was recorded that large islets have a smaller fraction of β-cells, and that the composition of α and

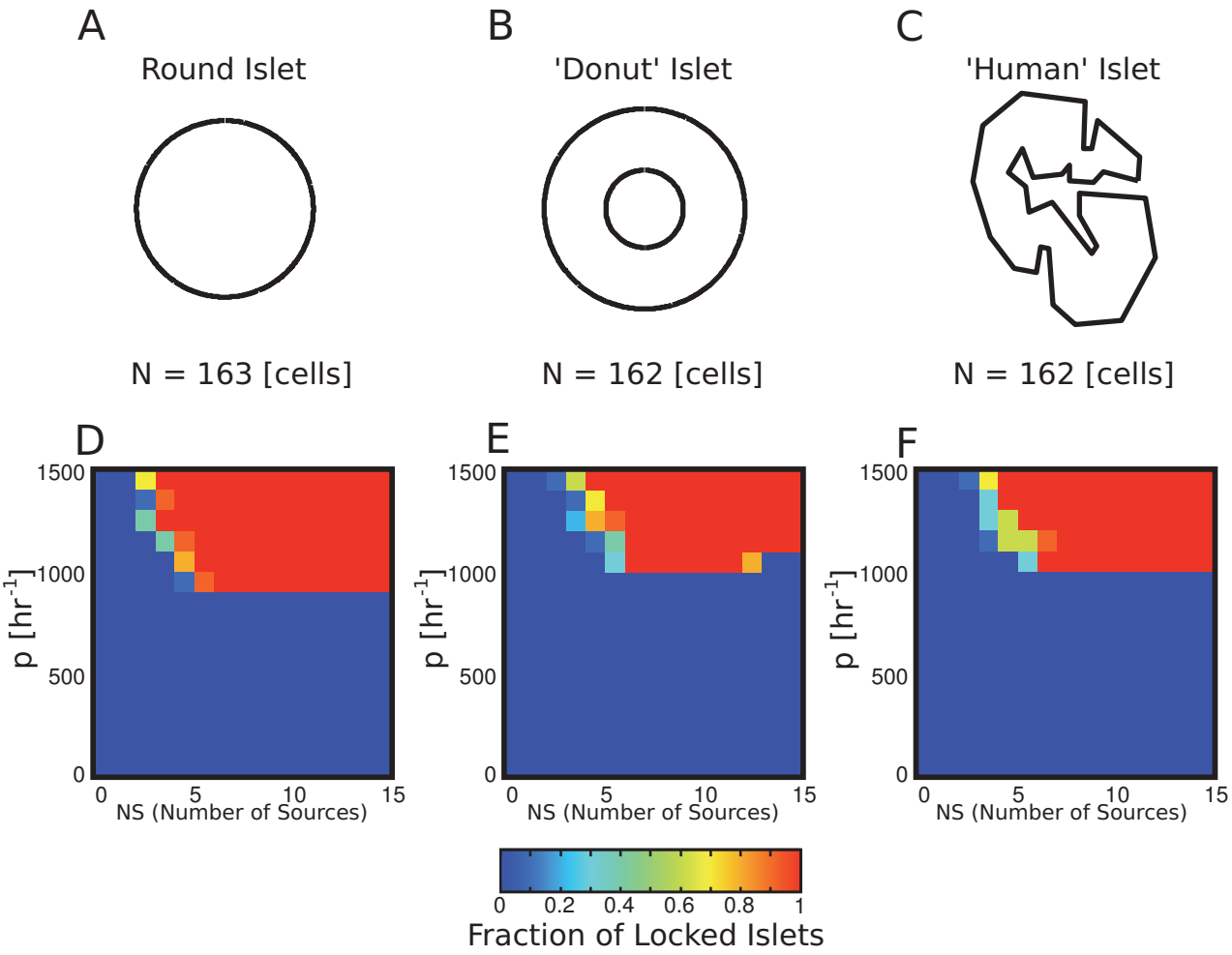


Figure 31: Parameters scans of p and NS for islets of different shapes. A-C: Peripheries of the three different shapes. A: A round islet of radius R = 6.6 cell diameters, corresponding to a total of N = 163 β-cells. B: A 'donut' islet of outer radius $R_2 = 7.5$ and inner radius $R_1 = 3$ cell diameters, corresponding to a total of N = 162 β-cells. C: A 'human' islet [9], with a total of N = 162 β-cells. In panel D-F, we plot the fraction of islets which transition into the locked state, as averaged over 10 random source configurations for each parameter combination.

β-cells in the large islets are more intermingled, than in the small islets [36]. This observation also agrees well with our results, as we predict the more dispersed arrangement of β-cells introduces a protective advantage with respect to getting locked. However - and as mentioned above - the complex intermingled compositions may also have other advantages, for example it has been hypothesised that a high α-β-contact may be an advantage for a sensitive glucagon-insulin regulation.

CONCLUDING REMARKS

CONCLUDING REMARKS

Mathematical modelling presents a great tool for performing all kinds of thought experiments. It can be used both as a tool for exploring ideas and for testing existing hypothesis. During my PhD I have often encountered results and model behaviours, which I did not expect when constructing the model. I therefore conclude that even very simple models can often lead to unforeseen insight. In my opinion, experimental observations should therefore always be accompanied by a mathematical model, that contributes to the understanding of the system, or - also very importantly - can be used to accentuate the features of the observations, which are not readily understood.

In the future I hope mathematical modelling will be increasingly integrated in the research of biological systems, as I believe it will contribute significantly to the understanding of experimental findings as well as concepts and complex systems, that we do not yet have the tools to investigate experimentally.

In this thesis I have presented a number of mathematical models of biological systems - all somehow related to spatial structures and regulation. In my experience the inclusion of spatial conditions in the mathematical models, can often lead to interesting results - even in cases where I initially thought that dynamics and timing were the only relevant factors to consider. I therefore also believe that spatial aspects are often very important for biological system, and when applicable they should therefore be included in the mathematical models.

PAPERS

The length distribution of frangible filaments: an analytical study in the continuum approximation.

Pernille Yde*

Niels Bohr Institute, University of Copenhagen,

Blegdamsvej 17, 2100 Copenhagen, DK

Thomas C. T. Michaels,† Julian Willis,‡ Alexander K. Buell,§ Mogens Høgh Jensen,¶ and Tuomas P. J. Knowles**

Department of Chemistry, University of Cambridge,

Lensfield Road, Cambridge CB2 1EW, U.K.

(Dated: January 11, 2014)

Abstract

Nucleated polymerisation phenomena are general linear growth processes that are alternatively part of the normal functioning of nature or are encountered as the mechanisms in the development of certain neurodegenerative disorders. The growth of these linear structures has been shown to be dominated by secondary rather than primary mechanisms, including filament fragmentation. In some of our previous work we derived self-consistent solutions for the time evolution of the polymer mass concentration of filaments that undergo internal breakage in addition to elongation and primary nucleation. Due to the non-linear nature of the master equation of filamentous growth, however, the time evolution of the full length distribution has been challenging to access and to date analytical solutions for the filament distribution are known only in certain special cases. We describe here an analytical approach based on Picard's iteration method, which provides self-consistent solutions for the length distribution of breakable filamentous structures. We use the presented theoretical framework to analyse AFM data of the length distribution of insulin filaments.

I. INTRODUCTION

The linear self-assembly of filamentous structures from soluble proteins is a process of fundamental importance to the normal functioning of nature,[9–13] as well it is observed in relation to a number of neurodegenerative disorders, including Parkinson's and Alzheimer's diseases.[15–23] A key parameter that determines the severity and progression of such diseases is represented by the length of the fibrils produced through the growth process. Thus, understanding the mechanisms that yield to a particular form of the filament length distribution represents an important tool in the quest to design of new therapeutic strategies.

As a result, numerous experimental and theoretical studies[9,10,24–27,31,33–42,65] focussed on the kinetics of the self-assembly of filamentous protein structures. In our previous work,[39,40] we have demonstrated that the growth kinetics of these important structures is often dominated by secondary mechanisms, including filament fragmentation, whereby fibers break into shorter ones which then act as new seed aggregates that accelerate the growth reaction. By introducing the principal moments of the filament distribution and by using mathematical techniques based on fixed-point mappings, we obtained self-consistent solutions for the polymer number and mass concentrations.[39–42] The availability of these solutions provided preliminary information about the shape of the length distribution, including the average length of fibrils.[40–42] However, due to the non-linear nature of the elementary mechanisms involved in the growth process, closed-form solutions of the full master equation have been challenging to obtain and analytical expressions for the length distribution of breakable filaments have been derived only in certain special cases,[30,42,52–54] including systems with constant monomer supply, or by assuming the steady-state. Consequently, most theoretical studies of length distributions of fragmenting filaments relied on numerical integration for solving the master equation.[27,28,43–46]

Building on earlier work, we present here an analytical study of the time evolution of the length distribution for the growth of fragmenting filamentous structures in the continuum limit approximation and obtain self-consistent solutions for the distribution of filament sizes. We test our results against numerical solutions of the master equation and use the resulting analytical expressions to fit experimental length distributions obtained from measurements of growing insulin filaments.

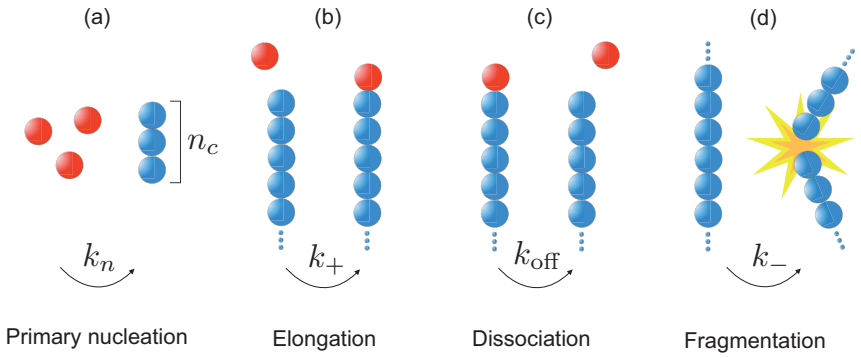


FIG. 1. Schematic representation of the elementary mechanisms of the nucleated polymerisation of breakable protein filaments.

II. MODEL

We consider an ensemble of monomeric polypeptide molecules and aggregates that undergo the growth processes sketched in Fig. 1. Fibrils can increase in size through elongation processes, i.e. the addition of single monomers onto the ends of the filament, or shrink through dissociation from either ends and through filament fragmentation. Here, we consider only the formation of linear filaments i.e. no branching or clumping processes are allowed and we define the aggregation number j as the size of the filaments meassured in number of monomers. We also include a minimal aggregation number n_c which defines the smallest stable filament size. Filaments of aggregation number smaller than n_c will disintegrate into monomers and thereby re-enter the monomer pool. We assume that the formation of these aggregates takes place through a primary nucleation processes in which n_c monomers spontaneously bind together to form a nucleus.

The behaviour of the ensemble of monomers and aggregates can be formulated within the formalism of chemical kinetics. On accounting for the elementary mechanisms of Fig. 1, the temporal evolution of the concentration $f(t, j)$ of filaments of aggregation number j is

described by the following master equation equation[9,10,39,40,55]

$$\begin{aligned}\frac{\partial f(j,t)}{\partial t} &= 2k_+ m(t)\left[f(j-1,t) - f(j,t)\right] \\ &+ 2k_{\rm off}\left[f(j+1,t) - f(j,t)\right] \\ &- k_-(j-1)f(j,t) + 2k_- \sum_{i=j+1}^{\infty} f(i,t) \\ &+ k_n m(t)^{n_c}\delta_{j,n_c},\end{aligned} \tag{1}$$

where the evolution of the concentration of monomers, $m(t)$, is obtained by exploiting the conservation of the total mass, $m_{\rm tot}$,

$$\frac{dm(t)}{dt} = -\frac{d}{dt}\sum_{j=n_c}^{\infty} j f(t,j). \tag{2}$$

The terms in Eq. (1) that are proportional to k_+ describe the growth of fibrils through elongation, where the factor two accounts for the fact that monomers can attach at either end of the filament. Similarly terms proportional to $k_{\rm off}$ account for the dissociation of monomers from the ends of the aggregates. The third line of Eq. (1) pertains to filament fragmentation. The first term is a loss term and describes the breakage of a filament of aggregation number j at one of the $(j-1)$ internal bonds. The second term, $2k_-\sum_{i=j+1}^{\infty} f(i,t)$, includes the contributions from the creation of filaments of aggregation number j due to the fragmentation of longer filaments. Finally, the last term of Eq. (1) describes the primary nucleation of new filaments of aggregation number $j = n_c$ proceeding at a rate that is proportional to the n_c-th power of the monomer concentration.

III. MASTER EQUATION IN THE CONTINUUM LIMIT

In the limit of large aggregation numbers, the filament length distribution varies sufficiently slowly and the discrete system of equations (1) can be approximated by its continuum version. In the transition from a discrete to a continuum description of protein aggregation, we treat the discrete index j as a continuum variable x and expand the finite differences in Eq. (1) in terms of partial derivatives at leading order

$$f(j \pm 1, t) \approx f(x,t) \pm \frac{\partial f(x,t)}{\partial t} + \mathcal{O}(f''), \tag{3}$$

$$\sum_{i=j+1}^{\infty} f(j,t) \approx \int_x^{\infty} f(x,t)dx. \tag{4}$$

By introducing the polymerisation drift coefficient $v(t) = 2(k_+ m(t) - k_{\text{off}})$, the master equation, Eq. (1), can be formulated in the continuum limit as follows

$$\frac{\partial f(x,t)}{\partial t} = -v(t)\frac{\partial f(x,t)}{\partial x} - k_- x f(x,t) + 2k_- \int_x^\infty f(z,t)dz + k_n m(t)^{n_c}\delta(x - n_c), \tag{5}$$

where the boundary conditions $f(x,t) = 0$ for $x \leq n_c$ and $f(x,t) \to 0$ for $x \to \infty$ are imposed.[54] For the initial condition of Eq. (5) we consider here the situation where a certain number of seed aggregates with well-defined lengths is present initially. Thus, at $t = 0$ the length distribution corresponds to a narrow Gaussian distribution centred around the average seed length x_0 and with standard-deviation σ_0

$$f(x, t = 0) = \frac{P(0)}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(x - x_0)^2}{2\sigma_0^2}\right), \tag{6}$$

where $P(0)$ denotes the number concentration of seed aggregates initially present.

We note that in the transition from a discrete to a continuum formulation of Eq. (1), the multiplicative pre-factor $(x - 1)$ in the loss term related to fragmentation has been replaced by x. This fact makes intuitive sense not only because rewriting $(x - 1) = x$ is justified for large aggregation numbers, but also because it corresponds to a scenario in which the filaments can break anywhere along the continuous chain, even infinitely close to the ends.

The continuum approximation provided by Eq. (5) was derived replacing finite differences with first-order derivatives. In general, one will get better approximations by considering higher order terms in the Taylor expansion. For example, if derivatives up to second order are considered in the expansion, an additional term in the continuum master equation is obtained, which is of the form

$$D(t)\frac{\partial^2 f(x,t)}{\partial x^2}, \tag{7}$$

where $D(t) = k_+ m(t) + k_{\text{off}}$ describes the diffusion of the filament distribution in length space. As we will see below, the effect of this term can to a good approximation be accounted for by including diffusion in the same way as it enters the solution to the drift-diffusion equation with constant coefficients and no fragmentation.

Our goal is to solve Eq. (5) subject to the above initial and boundary conditions. As a strategy for the solution, we shall consider two different scenarios: open and closed systems. We consider the open system scenario first, in which the monomer pool is infinitely large

and the monomer concentration is kept constant to the initial value, $m(t) = m(0)$. This situation is unrealistic for many experimental systems, but it provides useful information about the initial behaviour of the system. We then build on the knowledge of the solution for the open system for constructing expressions for the length distribution that are valid in the more realistic closed system scenario, in which the monomer concentration varies with time and the total mass is conserved.

IV. PRINCIPAL MOMENTS AND STEADY-STATE DISTRIBUTION

A. Principal moments

Preliminary insight into the form of the solution of Eq. (5) can be obtained by considering the principal moments of the length distribution, defined as

$$I_n(t) = \int_{n_c}^{\infty} x^n f(x,t) dx. \tag{8}$$

The zeroth moment, $I_0(t) \equiv P(t)$ gives information about the number concentration of fibrils, whereas the first moment of the distribution, $I_1(t) \equiv M(t)$, is the polymer mass concentration. The equations describing the temporal evolution of $P(t)$ and $M(t)$ are obtained by integrating the master equation, Eq. (5), on both sides, yielding

$$\begin{aligned} \frac{dP(t)}{dt} &= k_- \left[M(t) - 2n_c P(t)\right] \\ \frac{dM(t)}{dt} &= \left[v(t) - k_- n_c^2\right] P(t), \end{aligned} \tag{9}$$

where we have neglected $\mathcal{O}(k_n)$ terms in front of the contributions from the elongation and fragmentation of aggregates. We note that in the limit of large nucleus sizes, n_c, Eq. (9) recovers the moment equation in the discrete limit.[39,40,42]

B. Steady-state distribution

In the limit $t \to \infty$ the system approaches a dynamic steady state in which the growth of filaments is balanced by fragmentation and the length distribution stays constant in time. We note, however, that this dynamic steady state does not correspond to an equilibrium state, because in the absence of fibril association the detailed balance condition is satisfied

only with respect to the polymerisation-depolymerisation of fibrils, but not with respect to breakage.

The steady-state values for the principal moments are obtained by setting time derivatives equal to zero in Eq. (9), yielding

$$P(\infty) = \frac{2k_+ m_{\text{tot}} - 2k_{\text{off}} - k_- n_c^2}{4k_+ n_c}, \quad M(\infty) = 2n_c P(\infty). \tag{10}$$

Note that while both $P(\infty)$ and $M(\infty)$ depend on the details of the system, the average length of fibrils, $L(\infty) = M(\infty)/P(\infty) = 2n_c$, solely depends on the nucleus size, n_c.

The steady-state length distribution, $f(x, \infty)$, is given by the solution to the following differential equation

$$v(\infty) f''(x, \infty) + k_- x f'(x, \infty) + 3k_- f(x, \infty) = 0, \tag{11}$$

which is obtained by differentiating the continuous master equation, Eq. (5), with respect to x and subsequently setting time derivatives equal to zero. Here, $v(\infty) = k_- n_c^2$ is the polymerisation drift coefficient at steady state calculated from Eq. (10). The solution of Eq. (11) is given by a narrow, biased Gaussian distribution

$$f(x, \infty) = C\left(x^2 - n_c^2\right) \exp\left(\frac{-(x^2 - n_c^2)}{2n_c^2}\right), \tag{12}$$

where the constant C is fixed by the normalisation condition $\int_{n_c a}^{\infty} f(x, \infty) dx = P(\infty)$ and reads

$$C = \frac{2k_+ m_{\text{tot}} - 2k_{\text{off}} - k_- n_c^2}{4k_+ n_c^4}. \tag{13}$$

C. Self-consistent solutions for principal moments

We exploited analytical techniques based on Picard's iteration methods to extend the validity of linearised early-time solutions of the moment equations, Eq. (9), by performing one step of self-consistent fixed point iteration,[39,40] yielding the following closed-form expressions for the polymer number and mass concentrations in the continuum limit

$$\begin{aligned} P(t) = e^{-2k_- n_c t} \Bigg[& P(0) + \frac{M(\infty)}{2n_c}(e^{2k_- n_c t} - 1) \\ & + \frac{k_-(M(\infty) - M(0))}{k_+} \left(\text{Ei}\left(-C_+\right) + \text{Ei}\left(-C_+ e^{\kappa t}\right)\right) \Bigg] \\ M(t) = (M(0) - M(\infty)) & \exp\left(-C_+ e^{\kappa t} + C_- e^{-\kappa t}\right) + M(\infty), \end{aligned} \tag{14}$$

where $\kappa = \sqrt{k_- v(0)}$ and

$$C_\pm = \frac{k_+}{\kappa}\left(\sqrt{\frac{k_-}{v(0)}}[M(0) - n_c P(0)] \pm P(0)\right). \tag{15}$$

V. SELF-CONSISTENT SOLUTION FOR THE LENGTH DISTRIBUTION IN OPEN SYSTEMS

In this section we focus on open systems, where the concentration of monomers is constant. This scenario emerges either at the early stages of the reaction or when particular regulating mechanisms are present, such as in some *in vivo* systems where protein synthesis and degradation are active.[31,32] The underlying idea for solving Eq. (5) comes from the use of Picard's iteration method, that allows self-consistent solutions with increasing precision to be derived starting from an initial guess for the filament distribution. According to the contraction mapping principle, Eq. (13) can be solved iteratively using the following algorithm

$$\frac{\partial f_0^{(n+1)}(x,t)}{\partial t} = -v(0)\frac{\partial f_0^{(n+1)}(x,t)}{\partial x} - k_- x f_0^{(n+1)}(x,t) + 2k_- \int_x^\infty f_0^{(n)}(z,t)dz \tag{16}$$

for a starting value $f_0^{(0)}(x,t)$ chosen to be sufficiently close to the exact solution $f_0(x,t)$. As the starting point of our self-consistent iteration scheme, we choose the long-time limit for length distribution for open systems, which has been studied before[42] and reads

$$f_0^{(0)}(x,t) = X_0^{\text{limit}}(x) e^{\kappa(0)t}, \tag{17}$$

where $\kappa(0) = \sqrt{k_- v(0)} - k_- n_c$ and, for convenience, we have introduced the function

$$X_0^{\text{limit}}(x) = C_1\left[(k_- x + \kappa(0))^2 - k_- v(0)\right] \exp\left(-\frac{k_-(x^2 - n_c^2) + 2\kappa(0)(x - n_c)}{2v(0)}\right) \tag{18}$$

with

$$C_1 = \frac{P(0) + \sqrt{k_-/v(0)}\,[M(0) - n_c P(0)]}{2v(0)(k_- n_c + \kappa(0))}. \tag{19}$$

The late time solution, Eq. (17), describes a short-fibril biased Gaussian distribution, whose amplitude grows exponentially with time, while the shape is time independent, as a result of the concurrent actions of growth and fragmentation.

The first order self-consistent length distribution is obtained by substituting Eq. (17) into Eq. (16). On integration, Eq. (17) yields

$$\int_x^\infty f_0^{(0)}(z,t)dz = C_1 \exp\left(\kappa(0)t - \frac{k_- x^2 + 2x\kappa(0)}{2v(0)}\right) v\left(k_- x + \kappa(0)\right). \tag{20}$$

Consequently, the following result is obtained after one iteration step in Eq. (16)

$$f_0^{(1)}(x,t) = f_0^{(0)}(x,t) + \exp\left(-\frac{k_- x^2}{2v(0)}\right) g\left(x - v(0)t\right). \tag{21}$$

The function g is determined by implementing the initial condition, Eq. (6)

$$g\left(\xi\right) = \exp\left(\frac{k_- \xi^2}{2v(0)}\right) \times \left[\frac{P(0)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\xi - x_0\right)^2}{2\sigma^2}\right) - \theta(\xi) X_0^{\text{limit}}(\xi)\right], \tag{22}$$

where the Heaviside function $\theta(\xi)$ has been introduced in order to avoid non-physical divergence for $t \to \infty$ and the effects from the second order diffusion term, Eq.(7), have been incorporated by replacing σ_0 by $\sigma = \sigma_0 + 2D(0)t$.

The expression Eq. (21) describes in closed form the time evolution of the length distribution of an open system characterised by nucleated polymerisation and fragmentation. A comparison between the analytical expression provided by Eq. (21) and the numerical solution of the master equation is shown in Fig. 2a. The analytical expression gives the correct shape of the filament distribution and matches the numerical solution for large times exactly. As it can be seen from the plot, for early times Eq. (21) overestimates the number of fibrils with length greater than $x_0 + v(0)t$. Qualitatively, this effect arises because the filament distribution entering the fragmentation source term in Eq. (16) in the first iteration is from the late time regime and introduces long fibrils that would not have time to grow in the period of time considered, and therefore the aggregates population is overestimated for large aggregation numbers. This error can be systematically reduced by performing successive iterations of Eq. (16) beyond the first order solution discussed here. In Figs. 2b and 2c we show the contributions from the two terms in Eq. (21) to the shape of the filament distribution. The first term of Eq. (21) is an exponentially growing biased Gaussian distribution that accounts for the increase in the population of short fibrils caused by the fragmentation of longer ones. The second term of Eq. (21) describes the advection in size space of the initial filament distribution. As a result of fibril elongation, the Gaussian peak moves in size space

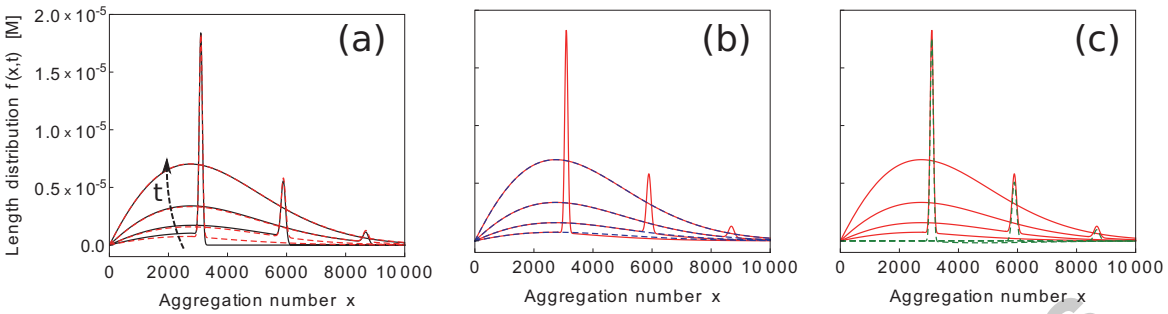


FIG. 2. Time evolution of the fibril length distribution in an open system. (a) The black solid line is the numerically calculated solution of the master equation, Eq. 5. The red dashed line is the predicted length distribution for the constant monomer case given in Eq. (21). (b) The first term of Eq. (21), $f_0^{(0)}(x,t)$, (dashed blue line) is compared with Eq. (23) (solid red line). (c) The contribution from the second term in Eq. (23) describing the advection of the initial distribution (dotted green line) is compared with Eq. (23) (solid red line). The parameters used are: $k_+ = 3.5 \times 10^{-2}$ M^{-1}min^{-1}, $k_- = 5 \times 10^{-6}$ min^{-1}, $k_{off} = 0$, $n_c = 2$, $x_0 = 300$, $\sigma_0 = 10$, $M(0) = 1$ M, $P(0) = M(0)/x_0$, $m(t) = m_{tot} - M(0) = 999$. Curves are shown for the following times (from bottom to top): $t = 40, 80, 120, 160$ minutes.

with velocity $v(0)$ and spreads out with diffusion coefficient $D(0)$. As $t \to \infty$ the advective Gaussian peak vanishes and Eq. (21) recovers $f_0^{(0)}(x,t)$ completely. As the fibrils increase in length, the their fragmentation probability increases and the filament distribution is shifted from large to small lengths. This behaviour is captured by the exponential pre-factor which ensures that the amplitude of the advective Gaussian peak decreases with increasing length x.

VI. LENGTH DISTRIBUTIONS IN CLOSED SYSTEMS

In this section we focus on the growth of filaments within a closed system scenario, where the concentration of monomers varies with time but the total mass concentration is constant. Under these circumstances, a qualitative picture of the evolution of the length distribution can be drawn from the knowledge of the filament distribution of aggregates growing in an open system, by noticing that at earlier times the concentration of available monomers is approximately constant in time. Thus, we expect that in the early stages of

the polymerisation process, the length distribution will first evolve according to the biased Gaussian, Eq. (21), obtained in the presence of a constant monomer pool, shifting at later times into the steady-state distribution given by Eq. (12). An interesting observation to be drawn from Eqs. (21) and (12), is that the distribution of fibril lengths share the same functional form at the beginning and at the end of the reaction. In both cases the form of $f(t,x)$ is given in terms of the function $X^{\rm limit}(x)$, where the values $v(0)$ and $\kappa(0)$ are used in the early time limit and the values $v(\infty)$ and $\kappa(\infty)=0$ in the steady state. Thus, an appropriate ansatz for $f(t,x)$ is given by

$$f(x,t) = A(t) X_0^{\rm approx}(x) + \exp\left(-\frac{k_- x^2}{2v(0)}\right) h\left(x-\tau(t)\right), \tag{23}$$

where $X_0^{\rm approx}(x)$ is obtained by replacing $v(0)$ in Eq. (18) by some value $v^{\rm approx}$ to be determined in the sequel. The function h is given by

$$h(\xi) = \exp\left(\frac{k_-\xi^2}{2v(0)}\right) \times \left[\frac{P(0)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\xi-x_0)^2}{2\sigma^2}\right)\right], \tag{24}$$

where $\sigma = \sigma_0 + 2D(0)t$. The amplitude $A(t)$ is calculated from the normalisation condition

$$A(t) = \frac{P(t)}{\int_{n_c}^{\infty} X_0^{\rm approx}(x)dx} \tag{25}$$

and

$$\begin{aligned}\tau(t) = \int_0^t v(t')dt' &= 2(k_+ m_{\rm tot} - k_+ M(\infty) - k_{\rm off})t \\ &- 2k_+ \frac{M(0)-M(\infty)}{\kappa}\left[{\rm Ei}\left(-C_+ e^{\kappa t}\right) - {\rm Ei}\left(-C_+\right)\right]\end{aligned} \tag{26}$$

describes the position of the peak of the initial distribution at time t as a result of advection in length space associated with monomer depletion through fibril elongation. Equation (23) preserves the early time behaviour of Eq. (21), while the late time behaviour is corrected by choosing $v^{\rm approx}$ such that the solution reproduces the correct mean value $\mu(t) = M(t)/P(t)$ of the length distribution at all times, yielding

$$v^{\rm approx} = k_- \left(\frac{M(t)}{P(t)} - n_c\right)^2, \tag{27}$$

where $M(t)$ and $P(t)$ are given by Eqs. (14). A comparison between the numerical solution of the master equation and the analytical expression of Eq. (23) is shown in Fig. 3 for different times. The numerical solution of the master equation verifies the conclusion, that

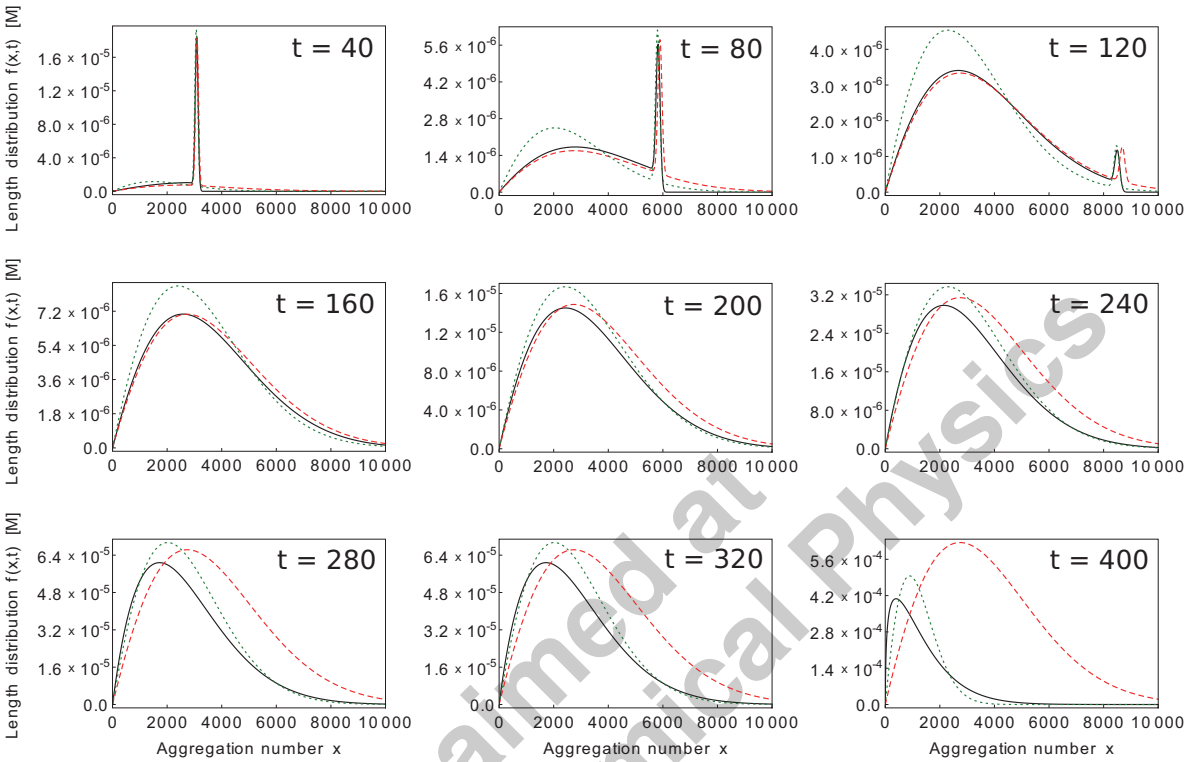


FIG. 3. Time evolution of the fibril length distribution in a closed system. The black solid line is the numerically calculated solution of the master equation, Eq. 5. The red dashed line is the predicted length distribution for the constant monomer case given in Eq. (21). The green dotted line is the predicted length distribution for the constant mass given in Eq. (23). The parameters used are the same as in Fig. 2. The times of the snapshots are: $t = 40, 80, 120, 160, 200, 240, 280, 320, 400$ minutes.

the evolution of the length distribution follows that of constant monomer systems at early times, shifting to the prediction of Eq. (23) at late times.

We now show that the theoretical framework of filamentous growth provided by Eq. (23) can be used to fit experimental measurements of the length distribution of growing insulin filaments. Fibril length distributions were measured from seeded fibril growth kinetic runs monitored by changes in fluorescence and subsequent AFM imaging at different times. The experiment showed that, starting from the initial filament distribution, filaments steadily grew longer through recruitment of free polypeptide molecules at the ends of aggregates until, after 190 minutes, it appeared that the length distribution shifted towards shorter fibrils

as a consequence of fragmentation overcoming the effect of elongation. The experimental results for the length distribution of the insulin fibrils are plotted in Fig. 4. In order to compare the model predictions of Eq. (23) with the AFM data, we first fitted the measured total mass concentration $M(t)$ of fibrils to the theoretically predicted curve Eq. (14) to obtain values for the rate constants for elongation and breakage. For the obtained values of $k_+ = 5.5 \times 10^{-2}$ M^{-1}min^{-1} and $k_- = 7 \times 10^{-6}$ min^{-1} the comparison of Eq. (23) with the insulin length measurements in Fig. 4 shows overall good agreement between theory and experiment. The solid black lines in Fig. 4 represents the theoretical prediction for the length distribution as obtained from the numerical solution of the master equation, Eq. (1), while the dashed red curves represent the theoretical prediction from Eq. (23).

VII. MATERIALS AND METHODS

Fibril length distributions were measured from seeded fibril growth kinetic runs monitored by changes in fluorescence. Bovine insulin monomer (Gemini Bio-Products, USA) was dissolved at a concentration of 0.5 mg ml^{-1} in a solution of 10 mM HCl and 30 mM NaCl in Milli-Q water with 60 μM Thioavin T (ThT) fluorescent dye and 0.5 μg ml^{-1} seed fibrils at a total volume of 1.5 ml. Precise protein monomer concentration was determined by absorbance spectroscopy using a Cary 400 Scan UV- Visible Spectrophotometer or ThermoScientific Nanodrop 2000 Spectrophotomer with molar extinction coefficient 1.0 cm^{-1} for 1.0 mg ml^{-1} at 276 nm 33 . Immediately before starting the kinetic run, seed fibres were added to the cuvettes containing the monomer solution and mixed by inversion several times. Thermostat-controlled temperature was maintained at 45^{o}C and the progress of fibril growth was followed by measuring ThT fluorescence at 1 minute intervals with excitation wavelength 440 nm and emission wavelength 480 nm. After each 30 minute period the reaction mixture was mixed by inverting the cuvette several times before a 10 μl reaction aliquot was removed and diluted in 10 mM HCl to appropriate concentrations for AFM imaging. Both the reaction aliquots and reaction aliquot dilutions were stored at 4^{o}C until preparation of AFM slides to slow fibril growth and breakage kinetics. A fluorescence measurement and reaction aliquot were also taken after 24 hours to provide information about the long-time limit at equilibrium. Taken as a value corresponding to 100% completion of aggregation, this final fluorescence value was used to normalise the fluorescence measure-

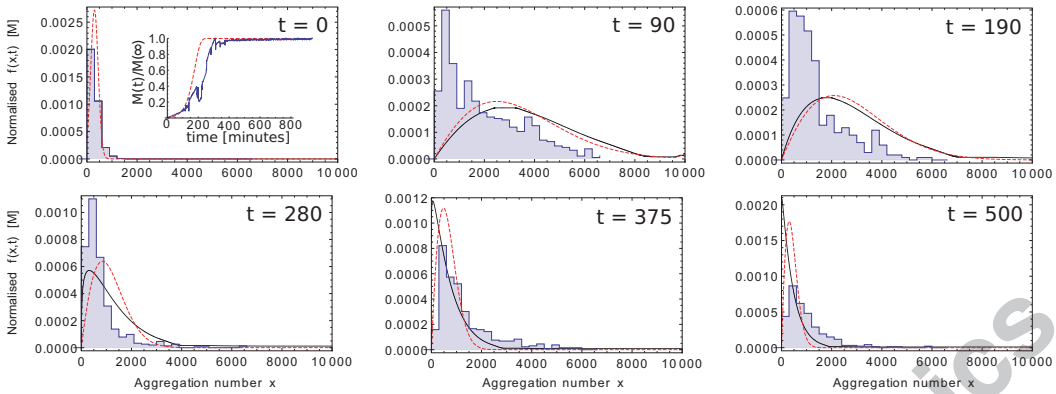


FIG. 4. Predicted length distributions at 0, 90, 190, 280, 375 and 500 minutes. In each plot, the columns represent the experimental data of insulin filaments with histogram bin widths of 300, the solid black curves represent the theoretical predictions from the numerical solution of the master equation, while the dashed red curves represents the prediction of Eq. (23). The parameters used were those that correspond with the fit of the polymer mass concentration in the inset: $k_+ = 5.5 \times 10^{-2}$ M^{-1}min^{-1}, $k_- = 7 \times 10^{-6}$ min^{-1}, $k_{off} = 0$, $n_c = 2$, $x_0 = 300$, $\sigma_0 = 150$, $M(0) = 1$ M, $P(0) = M(0)/x_0$, $m_{tot} = m(0) + M(0) = 1000$.

ments of each sample to give the fraction of protein incorporated into amyloid fibrils as a function of time. This allows sample comparison and the extent of the aggregation process at each time point when a reaction aliquot was removed to be determined. A kinetic run with strong shear forces was conducted with stirring using a magnetic stirring bar to induce high fibril breakage rates, while an unstirred kinetic run without a magnetic stirring bar was conducted to investigate the lower intrinsic fibril breakage rates. As suggested by DePace *et al.*,[69] to avoid shear forces causing fibril fragmentation during pipetting, wide-bore pipette tips were used by cutting all pipette tips 1 cm from the end to create a width $1 - 2$ mm in diameter. Furthermore, mixing of solutions during dilution was achieved only by inversion to avoid shearing during all manipulations.

VIII. CONCLUSIONS

In this paper, we have provided self-consistent solutions of the master equation of linear self-assembly in the continuum limit to study the length distribution of fragmenting filamentous structures. We have checked the performance of the analytical solutions by comparison

with numerical solutions of the master equation. Furthermore, we have demonstrated that the model yields good agreement with experimental data of the length distribution of growing insulin fibrils measured using AFM.

ACKNOWLEDGMENTS

This work has been supported by St John's College, Cambridge (TCTM, TPJK), Magdalene College, Cambridge (AKB), the Leverhulme Trust (AKB), the Newman foundation and the Danish National Research Foundation through the Center for Models of Life (PY, MHJ).

* pernille.yde@gmail.com

† tctm3@cam.ac.uk

‡ jcww2@cam.ac.uk

§ ab761@cam.ac.uk

¶ mhjensen@nbi.ku

** tpjk2@cam.ac.uk

[1] P. Flory, *Statistical Mechanics of Chain Molecules*, (Interscience, New York, 1969).

[2] E. Gazit, Chem. Soc. Rev. **36**, 1263 (2007).

[3] T. P. J. Knowles and M. J. Buehler, Nature Nanotechnol. **6**, 469 (2011).

[4] A. Klug, Angew. Chem., Int. Ed. Engl. **22**, 565582 (1983).

[5] D. L. Caspar, Biophys. J. **32**, 103138 (1980).

[6] K. van Workum, J. F. Douglas, Macromol. Symp. **227**, 116 (2005).

[7] A. Y. Morozov, R. F. Bruinsma and J. Rudnick, J. Chem. Phys. **131** 2, 155101 (2009).

[8] D. Endres and A. Zlotnick, Biophys. J. **83**, 1217 (2002).

[9] F. Oosawa and S. Asakura, *Thermodynamics of the polymerisation of proteins*, (Academic Press: Waltham, MA, USA, 1975).

[10] F. Oosawa and M. Kasai, J. Mol. Biol. **4**, 10-21 (1962).

[11] L.S. Tobacman and E.D. Korn, J. Biol. Chem. **258**, 3207 (1983).

[12] C. Frieden and D.W. Goddette, Biochemistry **22**, 5836 (1983).

[13] A. Wegner and J. Engel, Biophys. Chem. **3**, 215 (1975).

[14] C.M. Dobson, Nature **426**, 884-890 (2003).

[15] C.M. Dobson, Nature **426**, 884-890 (2003).

[16] F. Chiti and C.M. Dobson, Annu. Rev. Biochem. **75**, 333-336 (2006).

[17] S.B. Prusiner, Science **252**, 1515-1522 (1991).

[18] J.H. Come, P.E. Fraser, and P.T. Lansbury, Proc. Natl. Acad. Sci. U S A **90**, 5959-5963 (1993).

[19] A. Aguzzi, and T. O'Connor **9**, 237-248 (2010).

[20] A. Aguzzi, and C. Haass, Science **302**, 814-818 (2003).

[21] J. Hardy, and D.J. Selkoe, Science **297**, 353-356 (2002).

[22] D.J. Selkoe, Nature **426**, 900-904 (2003).

[23] P.T. lansbury, and H.A. Lashuel, Nature **443**, 774-779 (2006).

[24] K. Adachi and T. Asakura, J. Biol. Chem. 254, 7765 (1979).

[25] S. Prigent, A. Ballesta, F. Charles, N. Lenuzza, P. Gabriel, L. M. Tine, H. Rezaei, and M. Doumic, PLoS One **79**, e43273 (2012).

[26] D. Hall and H. Edskes, J. Mol. Biol. **336**, 775 (2004).

[27] F. A. Ferrone, J. Hofrichter, and W. A. Eaton, J. Mol. Biol. **183**, 611 (1985).

[28] P. Arosio, M. Beeg, L. Nicoud, and M. Morbidelli, Chem. Eng. Sci. **78**, 2132 (2012).

[29] F. Ferrone, Methods Enzymol. **309**, 256 (1999).

[30] G. B. Ermentrout, and L. Edelstein-Keshet, Bull. Math. Biol. **60**, 477 (1998).

[31] J. Masel, V. A. A. Jansen, and M. A. Nowak, Biophys. Chem. **77**, 139 (1999).

[32] T. Pöschel, N. V. Brilliantov, and C. Froemmel, Biophys. J. **85**, 3460 (2003).

[33] A. Vitalis and R. V. Pappu, Biophys. Chem. **159**, 14 (2011).

[34] K. C. Kunes, D. L. Cox, and R. R. P. Singh, Phys. Rev. E **72**, 051915 (2005).

[35] J. M. Andrews and C. J. Roberts, J. Phys. Chem. B **111**, 7897 (2007).

[36] A. M. Morris, M. A. Watzky, J. N. Agar, and R. G. Finke, Biochemistry **47**, 2413 (2008).

[37] S. Schreck and J.-M. Yuan, J. Phys. Chem. B **117**, 6574 (2013).

[38] A. Lomakin, D. B. Teplow, D. A. Kirschner, and G. B. Benedek, Proc. Natl. Acad. Sci. U.S.A. **94**, 7942 (1997).

[39] T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, Science **326**, 1533 (2009).

[40] S. I. A. Cohen, M. Vendruscolo, M. E. Welland, C. M. Dobson, E. M. Terentjev, and T. P. J. Knowles, J. Chem. Phys. **135**, 065105 (2011).

[41] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, J. Chem. Phys. **135**, 065106 (2011).

[42] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, J. Chem. Phys. **135**, 065107 (2011).

[43] A. Wegner, J. Mol. Biol. **161**, 217 (1982).

[44] T. Medkour, F. Ferrone, F. Galacttros, and P. Hannaert, Acta Biotheor. **56**, 103 (2008).

[45] W.-F. Xue, S. W. Homans, and S. E. Radford, Proc. Natl. Acad. Sci. U.S.A. **105**, 8926 (2008).

[46] C. F. Wright, S. A. Teichmann, J. Clarke, and C. M. Dobson, Nature (London) **438**, 878 (2005).

[47] M. Ataka, Prog. Cryst. Growth Charact. **30**, 109 (1995).

[48] H. Flyvbjerg, E. Jobs, and S. Leibler, Proc. Natl. Acad. Sci. U.S.A. **93**, 5975 (1996).

[49] E.T. Powers and D. L. Powers, Biophys. J. **91**, 122 (2006).

[50] J.A. Cooper, E. L. Buhle, S. B. Walker, T. Y. Tsong, and T. D. Pollard, Biochemistry **22**, 2193 (1983).

[51] T.P.J. Knowles, T.W. Oppenheim, A.K. Buell, D.Y. Chirgadze, and M.E. Welland, Nat. Nanotechnol. **5**, 204 (2010).

[52] M.P. Firestone, R. De Levie, S.K. Rangarajan, J. Theor. Biol. **104**, 535552 (1983).

[53] S.K. Rangarajan, R. De Levie, J. Theor. Biol. **104**, 553570 (1983).

[54] H. Flyvbjerg, T. E. Holy, S. Leibler, Phys. Rev. E **54**, 5538 (1996).

[55] F. Oosawa, J.Theor. Biol. **27**, 6986 (1970).

[56] S.I.A. Cohen, M. Vendruscolo, C.M. Dobson, and T.P.J. Knowles, Int. J. Mol. Sci. **12**(9), 5844-5852 (2011).

[57] E. Rhoades, J. Agarwal and A. Gafni, Biochim. Biophys. Acta **1476**, 2308 (2000).

[58] Y. Li and C.J. Roberts, J. Phys. Chem. B **113**, 702032 (2009).

[59] T.R. Serio, A.G. Cashikar, A.S. Kowal, G.J. Sawicki, J.J. Moslehi, L. Serpell, M.F. Arnsdorf and S.L. Lindquist, Science **289**, 131721 (2000).

[60] J.W. Kelly, Nature Struct. Biol. **7**, 8246 (2000).

[61] A.B. Padrick and A.D. Miranker, Biochemistry **41**, 4694703 (2002).

[62] T.P.J. Knowles, D.A. White, A.R. Abate, J.J. Agresti, S.I.A. Cohen, R.A. Sperling, E.J.D. Genst, C.M. Dobson and D.A. Weitz, Proc. Natl. Acad. Sci. USA **106**, 1474651 (2011).

[63] J.A.D. Wattis and J.R. King, J. Phys. A: Math. Gen. **31**, 71697189 (1998).

[64] M. F. Bishop and F. A. Ferrone, Biophys. J. **46**, 631 (1984).

[65] F. Ferrone, Methods Enzymol. **309**, 256 (1999).

[66] L. Y. Chen, N. Goldenfeld, and Y. Oono, Phys. Rev. E **54**, 376 (1996).

[67] E.J. Hinch, *Perturbation Methods*, (Cambridge University Press, 1991).

[68] M. Holmes, *Introduction to Perturbation Methods*, (Springer-Verlag, New York, 1995).

[69] A. H. DePace, and J.S. Weissman, Nature Structural Biology **9**, 389-396 (2002).

doi:10.1016/j.jmb.2012.01.050 J. Mol. Biol. (2012) 421, 601–615

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Multiple Roles of Heparin in the Aggregation of p25α

Søren Bang Nielsen[1], Pernille Yde[2], Lise Giehm[1], Sabrina Sundbye[3], Gunna Christiansen[4], Joachim Mathiesen[2], Mogens Høgh Jensen[2], Poul Henning Jensen[3] and Daniel E. Otzen[1]*

[1]*Interdisciplinary Nanoscience Research Center (iNANO), Center for Insoluble Structures (inSPIN), Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10C, DK-8000 Aarhus C, Denmark*
[2]*Niels Bohr Institute, Copenhagen University, Blegdamsvej 17, DK-2100 Copenhagen, Denmark*
[3]*Department of Biomedicine - Medical Biochemistry, Aarhus University, Ole Worms Allé 1170, DK-8000 Aarhus C, Denmark*
[4]*Department of Biomedicine - Medical Microbiology and Immunology, Wilhelm Meyers Allé 4, 8000 Aarhus C, Denmark*

Received 30 September 2011; received in revised form 29 December 2011; accepted 30 January 2012
Available online 3 February 2012

Edited by S. Radford

Keywords:
amyloid;
Parkinson's disease;
modeling;
anionic surface;
thioflavin T

The 219-residue protein p25α stimulates the fibrillation of α-synuclein (αSN) *in vitro* and colocalizes with it in several α-synucleinopathies. Although p25α does not fibrillate by itself under native conditions *in vitro*, αSN-free p25α aggregates have also been observed *in vivo* in, for example, multiple system atrophy. To investigate which environmental conditions might trigger this aggregation, we investigated the effect of polyanionic biomolecules on p25α aggregation. Heparin, polyglutamate, arachidonic acid micelles, and RNA all induce p25α aggregation. More detailed studies using heparin as template for aggregation reveal that a minimum of 10–14 heparin monosaccharide units per heparin polymer are required. Bona fide fibrils are only formed at intermediate heparin concentrations, possibly because an excess of heparin binding sites blocks the inter-p25α contacts required for amyloid formation. Other polyanions also show an optimum for amyloid formation. Aggregation involves only modest structural changes according to both spectroscopic and proteolytic experiments. The aggregates do not seed aggregation of heparin-free p25α, suggesting that heparin is required in stoichiometric amounts to form organized structures. We are able to reproduce these observations in a model involving two levels of binding of p25α to heparin. We conclude that the modest structural changes that p25α undergoes can promote weak intermolecular contacts and that polyanions such as heparin play a central role in stabilizing these aggregates but in multiple ways, leading to different types of aggregates. This highlights the role of non-protein components in promoting protein aggregation *in vivo*.

*Corresponding author. E-mail address: dao@inano.au.dk.

Present address: L. Giehm, ZEALAND Pharma A/S, Smedeland 36, DK-2600 Glostrup, Denmark.

Abbreviations used: αSN, α-synuclein; PD, Parkinson's disease; PDB, Protein Data Bank; GAG, glycosaminoglycan; ThT, thioflavin T; HW, high molecular weight; AFM, atomic force microscopy; PBS, phosphate-buffered saline; CAC, critical aggregation concentration; DP, degree of polymerization; EM, electron microscopy; FTIR, Fourier transform infrared; ITC, isothermal titration calorimetry; ATR, attenuated total reflectance; DOPG, dioleoyl phosphatidylglycerol.

Introduction

α-Synuclein (αSN) is directly involved in the development of familial forms of Parkinson's disease (PD) and Lewy body dementia as well as other so-called α-synucleinopathies. αSN aggregates accumulate as cytoplasmic inclusions known as Lewy bodies. Several proteins are known to stimulate αSN aggregation, including the brain-specific protein p25α, which is effective at substoichiometric concentrations *in vitro*.[1] Although p25α can interact with tubulin and induce aberrant tubulin assembly,[2]

0022-2836/$ - see front matter

its function remains unknown. p25α is a heat-stable 219-residue protein that we have shown to be relatively flexible but natively folded.[3] The N-terminal region (residues 3–43) is missing in the two human paralogues p25β and p25γ and is probably unstructured, as its removal has no effect on folding or tubulin binding.[3] Furthermore, the C-terminal part is most likely also unstructured, as the corresponding region of p25γ in the *Caenorhabditis elegans* homologue [residues 102–176, Protein Data Bank (PDB) ID: 1PUL] is unfolded.[4] The core region (residues 44–156) consists of a five-helix bundle, which, in mice (PBD ID: 1WLM) and humans (PDB ID: 2JRF), has a small β-sheet exposed to one side. p25α is present in all parts of the brain but is predominantly expressed in oligodendrocytes,[5] unlike αSN, which is expressed in neurons. p25α only colocalizes with αSN under disease conditions such as PD, Lewy body dementia, and multiple system atrophy.[1] In Alzheimer's disease, Pick's disease, and multiple system atrophy, p25α inclusions without αSN were observed in neuronal cytoplasm and nuclei, suggesting an aggregated state of p25α.[6,7]

To uncover possible environmental cues to the development of this state, we here explore physiologically relevant conditions that may stimulate p25α aggregation. Polyanionic assemblies such as the sulfated glycosaminoglycan (GAG) heparin, anionic phospholipid vesicles, and RNA are likely candidates as they trigger aggregation of tau protein, in heparin's case via 1:1 tau monomer:heparin complexes.[8] These lead to neuronal aggregates in Alzheimer's disease and a range of other diseases collectively named tauopathies.[9] Tau–GAG interactions have even been suggested to be a central event in the development of Alzheimer's disease.[10] Tau occurs in the cytosol, clearly demonstrating that heparin can affect aggregation in this part of the cell *in vivo*. The *in vitro* fibrillation of recombinant tau by polyanionic compounds is hypothesized to be mediated by anionic condensing agents that stabilize a fibrillation-competent intermediate state.[11] Heparin stimulates amyloid formation of many different proteins, including αSN,[12] Aβ,[13] apomyoglobin,[14] and gelsolin.[15] GAGs also stimulate fibrillation of β_2-microglobulin,[16] possibly by scaffolding the fibrillation process.[17] Similarly, heparin accelerates the association of transthyretin oligomers to fibrils[18] but has no effect on upstream processes such as transthyretin tetramer dissociation. Although heparin and other GAGs are found in the extracellular matrix, they also occur intracellularly where they may play a role in storage granules, the nucleus, and other intracellular organelles.[19] For example, heparin and related polymers have recently been shown to promote aggregation of many intracellularly stored peptide hormones.[20,21] This makes it reasonable to investigate heparin's interactions with p25α. The ability to induce structural changes is often dependent on the length of the heparin molecule,[22,23] consistent with the fact that expression of heparanase reduces amyloid deposition.[24,25] Nucleic acids can also stimulate aggregation of, for example, prion proteins.[26] αSN, β_2-microglobulin, and Aβ can also fibrillate in the presence of SDS (see references in Ref. 27), though it should be noted that this requires the presence of monomeric surfactant to stimulate protein aggregation through shared micelles.[27]

Here, we provide evidence that p25α does not fibrillate to any significant extent by itself but readily forms aggregates in the presence of polyanionic biopolymers such as heparin, RNA, and anionic lipids. However, the aggregates contain a mixture of amyloid and non-amyloid structures, and the distribution of the two structural classes is dictated by the amount of polyanionic compound present. This indicates that p25α can form a spectrum of different aggregates in response to specific environmental conditions. We are able to capture these phenomena in a model where p25α binds to heparin in two stages. Our observations may have implications for an interpretation of the biological distribution of p25α aggregates.

Results

p25α aggregation is stimulated by heparin of minimum length 6 monosaccharide units

We have previously shown that p25α accelerates the ability of αSN to form thioflavin T (ThT)-binding amyloid fibrils when co-incubated for several days under physiological buffer conditions.[1] In those experiments, p25α did not form any insoluble aggregates on its own. When we repeat this experiment in a plate reader under conditions known to promote aggregation, namely, shaking and the presence of glass beads,[28] we also observe that 15 μM (0.38 mg/ml) p25α by itself does not form ThT-positive aggregates to any significant extent (Fig. 1a).

However, the addition of high molecular weight (HW) heparin (average molecular mass, 17 kDa, corresponding to ~60 monosaccharide units, as described in Supplementary Information, p. 2) dramatically changes this. As little as 0.4 μg/ml heparin has an impact, leading to an increase in ThT fluorescence after approximately 16 h of incubation (Fig. 1a). With increasing heparin concentrations of up to 10 μg/ml, the lag phase is further shortened to 8–10 h (Fig. 1a) and the endpoint ThT fluorescence increases (Fig. 1b). This clearly indicates that heparin induces aggregation of p25α in a concentration-dependent manner. At concentrations above 10 μg/ml heparin, the lag phase further decreases to

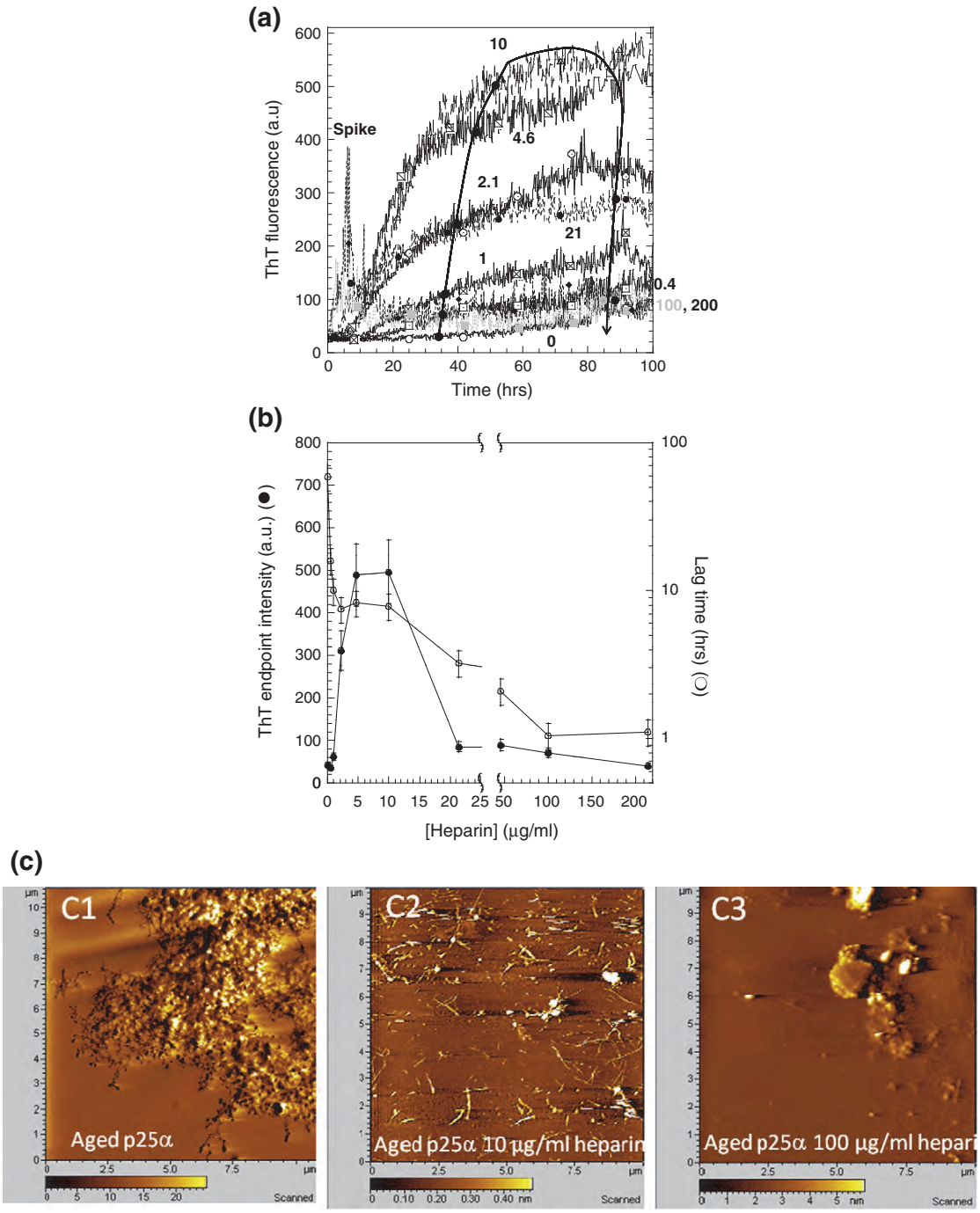


Fig. 1. (a) Aggregation of p25α in the presence of 0–200 μg/ml heparin over time followed by ThT fluorescence intensity. The data depict representatives of triplicate experiments. Numbers indicate heparin concentration in micrograms per milliliter. (b) Final plateau levels (●) and lag periods (□) from the data set, of which a sample is shown in (a). (c) AFM images of samples from (a).

1–3 h, but although there is a spike in the ThT fluorescence around 10 h, the endpoint ThT intensity rapidly decays to a level corresponding to the absence of heparin (Fig. 1b). Atomic force microscopy (AFM) images (Fig. 1c) show that full-length p25α in both 0 (Fig. 1c1) and 100 μg/ml heparin (Fig. 1c3) forms amorphous aggregates, whereas short fibrils are formed at intermediate

(10 μg/ml) heparin levels (Fig. 1c2) and thus correspond nicely with the ThT levels reached in Fig. 1a. Note that much smaller amounts of aggregates are formed at 0 than at 100 μg/ml heparin (Fig. 2, see below). These observations clearly suggest an optimal heparin concentration for the induction of amyloid fibrils around 10 μg/ml.

Aggregates formed at different concentrations after incubation in round-bottomed flasks were spun down to evaluate how heparin affected the amount of aggregated material, and the amount of supernatant and pellet was quantified by SDS-PAGE and densitometric scanning (Fig. 2a). To remove bias in the aggregation approach, the concentration of soluble p25α was also determined by UV absorbance at 280 nm after incubation in a plate reader (Fig. 2b). As the heparin concentration increases to 10 μg/ml, UV absorbance data showed a steep decrease in the remaining soluble p25α fraction to a plateau around 0.2 mg/ml, which remains relatively constant as the heparin concentration increases to 100 μg/ml. This is in excellent agreement with densitometric scanning data (Fig. 2a). The insoluble aggregates were insoluble not only in phosphate-buffered saline (PBS) buffer but also in mild surfactant (Triton X-100) but were completely solubilized by SDS (data not shown).

These data suggest the existence of a critical aggregation concentration (CAC) of ~0.2 mg/ml p25α. Below the CAC, there is not enough monomer to form fibrils. Above the CAC, the monomer concentration remains essentially constant irrespective of the total protein concentration. To confirm the existence of such a threshold concentration, we followed the ThT signal during fibrillation of p25α at various concentrations using 10 μg/ml heparin (Fig. 2c). Indeed, the endpoint ThT levels only rises gradually at low concentrations and then shows a steep increase above ~0.17 mg/ml (6.8 μM) p25α.

The other two p25α constructs with missing C- or N-termini (p25α Δ3–43 in Fig. S1a and p25α Δ156–219 in Fig. S1b) show a small degree of fibrillation by themselves (2- to 10-fold increase in ThT fluorescence), and fibrillation levels peak already at 1 μg/ml heparin, though the lag times are shortest at 10 and 100 μg/ml heparin. No

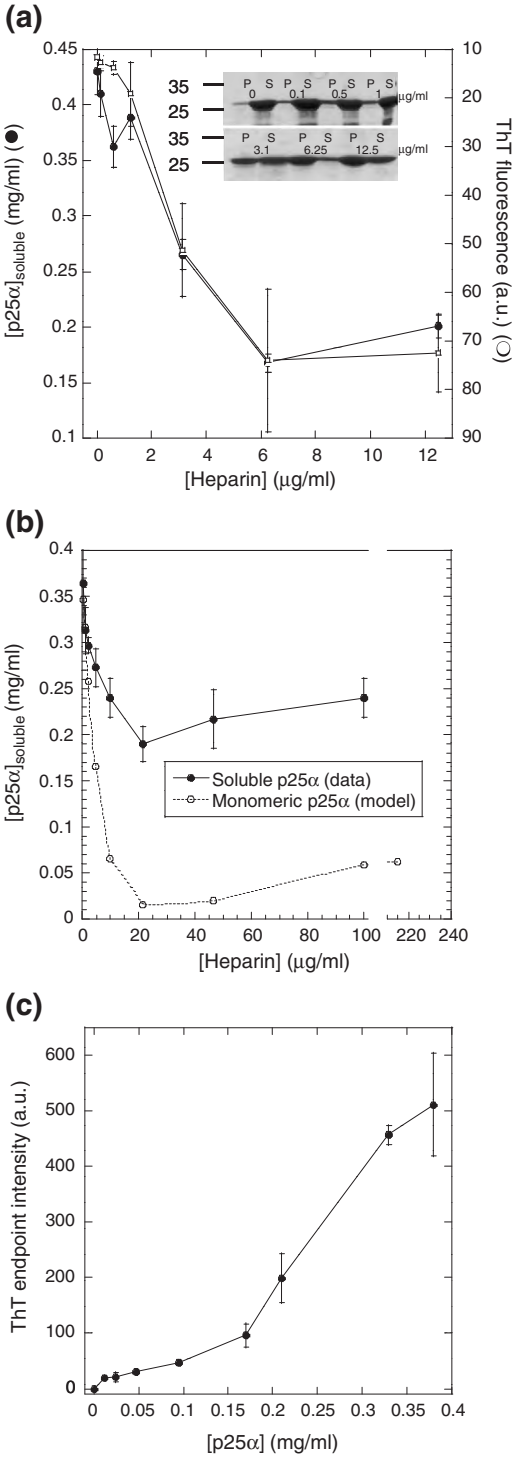


Fig. 2. (a) Amount of soluble p25α left after 3 days of incubation of 20 μM p25α with different amounts of heparin at 37 °C in round-bottomed flasks, estimated by centrifugation followed by SDS-PAGE densitometric analysis of supernatant (S) and pellet (P) using ImageJ (inset). Data are compared with endpoint ThT fluorescence (note that the second *y*-axis is inverted to facilitate comparison). (b) Filled circles: amount of soluble p25α left after 6 days of incubation of 15 μM p25α at 37 °C in a plate reader, estimated by OD_{280} of the supernatant left after centrifugation. Open circles: Prediction of monomeric p25α concentration after 6 days of incubation using the theoretical model developed in this work. (c) Endpoint ThT fluorescence of p25α incubated at various concentrations with 10 μg/ml heparin at 37 °C in a plate reader for 5 days.

overshoot is seen at 100 μg/ml heparin, unlike full-length p25α. Thus, the two unstructured terminal regions of p25α influence, but are not critical for, aggregation.

Next, we examined the effect of heparin polymer length on the stimulation of p25α aggregation. We monitored the aggregation propensity of p25α in the presence of smaller heparin species with a degree of polymerization (DP) of 2, 6, 10, and 14 heparin monosaccharide units, respectively (Fig. 3). DP2 did not increase ThT fluorescence beyond that in the absence of heparin at any concentration. Similarly, the addition of 1 μg/ml DP6 did not result in further increases in the ThT levels. Hardly any material could be spun down, corroborating the lack of aggregation (data not shown). In contrast, the addition of 10 and 100 μg/ml DP6 to p25α led to an increase in ThT intensity of ~19-fold and ~37-fold compared to the background, respectively. However, unlike p25α incubated with 100 μg/ml HW heparin, no overshoot in ThT signal was observed using the 1–100 μg/ml DP6 heparin evaluated here.

When increasing the DP to 10 and 14 units, respectively, we observed a time profile with characteristic lag phases and order of final ThT signals resembling those observed in the presence of high MW heparin. With 1 μg/ml DP10 heparin, a lag phase of ~32 h was observed, whereas the longer DP14 decreased the lag phase to ~23 h, which is close to that of HW heparin. The addition of 10 μg/ml DP10 and DP14 shortens the lag phase to ~20 and ~14 h, respectively, and led to the highest ThT fluorescence signals. Further shortening of the lag phases to ~10 and ~4 h was observed when the concentration of DP10 and DP14 was increased 10-fold to 100 μg/ml, respectively. A gradual decline in ThT fluorescence over time to levels below those of 10 μg/ml heparin was also observed at high DP10 and DP14 concentrations, although without the spike-like appearance seen for full-length heparin (Fig. 1a). Fibrils were seen at all three heparin concentrations according to EM (shown as insets in Figs. 3a–c for DP14), typically interspersed with more amorphous material. The DP10–DP14 samples had aggregated sufficiently to allow material to be spun down for Fourier transform infrared (FTIR) analysis, yielding a shoulder around 1627 cm^{-1} characteristic of amyloid structure (data not shown).

Summing up, DP2 does not stimulate p25α aggregation, DP6 results in an aggregation pattern similar to high MW heparin in the presence of p25α Δ156–219 and p25α Δ3–43 mutants, whereas DP10 and DP14 result in aggregation patterns similar to high MW heparin although the overshoot at 100 μg/ml heparin concentration is less pronounced. Furthermore, an increasing DP decreases the lag phase of aggregation. Thus, a polymerization degree of 10–14 heparin units is sufficient to recapitulate the aggregation-stimulating properties of full-length heparin.

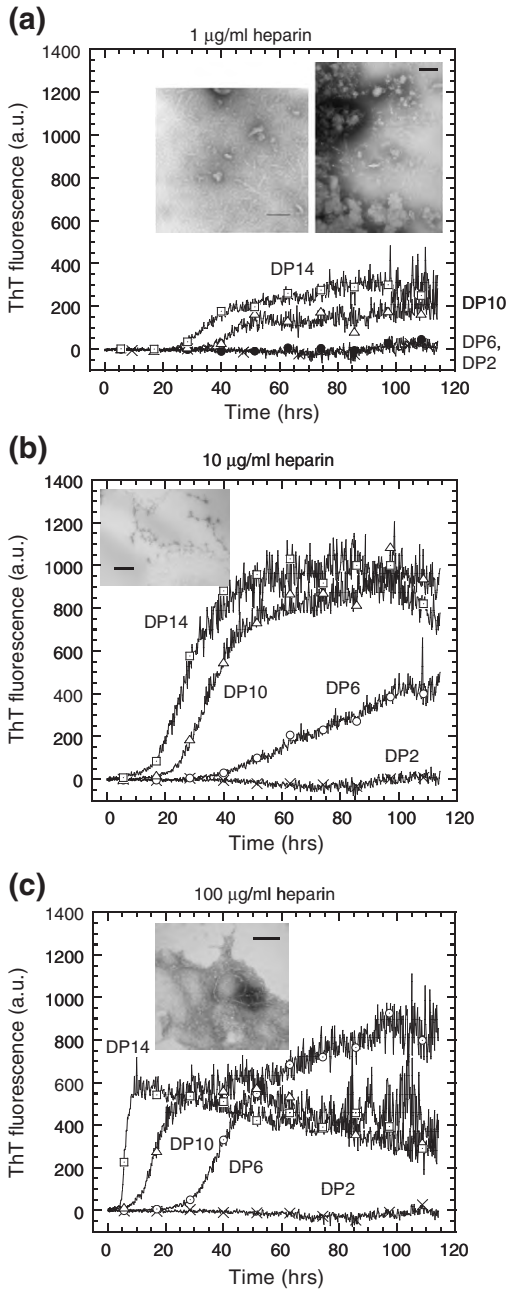


Fig. 3. Dependence on heparin DP on the aggregation propensity of p25α as shown by the ThT fluorescence intensity as a function of time using heparin of DP2 (×), DP6 (○), DP10 (△), and DP14 (□) units at (a) 1 μg/ml, (b) 10 μg/ml, and (c) 100 μg/ml concentrations. Insets show representative EM images of aggregates visualized after incubation. The scale bar represents 100 nm.

Heparin binds stably to both fibrillar and monomeric p25α

To elucidate whether heparin only interacted transiently with p25α or was involved in a more stable complex, we incubated 10 μg/ml fluorescein-labeled HW heparin with 15 μM p25α under the same aggregation conditions as in the previous section. After 160 h of incubation at 37 °C, centrifugation or filtration removed 40±10% of fluorescein fluorescence from solution, similar to the 55±10% of p25α removed from solution in Fig. 2a. This indicates that heparin forms a stable complex with p25α in the aggregated state.

We next sought to elucidate to what extent HW heparin interacts with soluble p25α. Accordingly, we applied isothermal titration calorimetry (ITC) to obtain a thermodynamic description of the heparin:p25α interaction. ITC measurements of the binding of HW heparin to p25α were done by stepwise titration of 200 μM HW heparin into 15 μM p25α at 37 °C while recording the evolved heat (Fig. 4, inset). Clearly, the interaction of heparin with p25α is exothermic. The peaks were integrated over time to obtain the heat evolved per mole of heparin and corrected for the contribution of heparin into buffer and buffer into protein (Fig. 4). Thermodynamic binding parameters were obtained by fitting data to a model with n binding sites for p25α per heparin molecule, yielding the following parameters: $n=2.44\pm0.06$, dissociation constant $K_d=2.14\pm0.64$ μM, $\Delta H=-14.0\pm1.7$ kcal/mol, $-T\Delta S=5.99$ kcal/mol, and thus a Gibbs free energy of interaction of −8.1 kcal/mol. Clearly, the reaction is driven by enthalpy (binding interactions between p25α and heparin) and opposed by entropy (immobilization of the two components). Similar thermodynamic values ($K_d=3.73\pm0.93$ μM, $\Delta H=-4.5\pm1.7$ kcal/mol, $-T\Delta S=3.2$ kcal/mol, and thus a Gibbs free energy of interaction of −7.7 kcal/mol) were obtained using the more homogeneous DP14 heparin preparation (Fig. 4). This indicates that our HW heparin binding data were not skewed by, for example, preferential binding to higher molecular weight species.

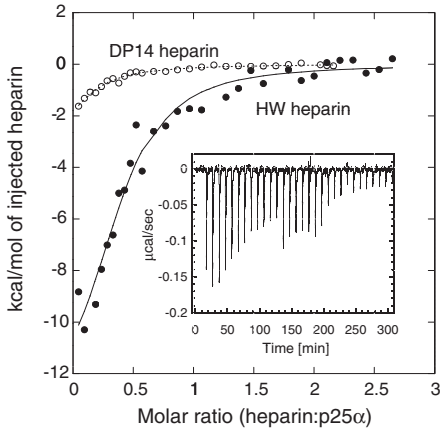


Fig. 4. ITC analysis of p25α:heparin interactions by titration of 200 μM HW heparin (filled circles) or DP4 heparin (open circles) into 15 μM p25α. The inset shows the heats of interaction upon injection of HW heparin, which was integrated over time to obtain the heat evolved per mole of heparin added and corrected for the contribution of heparin into buffer and buffer into protein. The two sets of data were fitted to a model with n binding sites per p25α (continuous line for HW heparin, broken line for DP14 heparin) to obtain thermodynamic binding parameters.

The high affinity compares well to that of αSN wt and A30P mutant (K_d values of 0.19 μM and 0.6 μM, respectively)[12] and other heparin binding proteins.[29] The exothermic interaction likely arises from multiple interactions of basic amino acid residues with the negatively charged sulfate and carboxyl groups of heparin.[29,30] This is further emphasized by the high pI of p25α (calculated pI, ~9.48). We have previously reported that monomeric p25α does not undergo significant structural changes over shorter time scales in the presence of heparin.[3] The present report shows that aggregation only occurs after prolonged incubation with heparin. Electrostatic interactions may therefore coordinate the protein in an aggregation-prone yet largely native conformation.

The fitted parameters are consistent with a simple model in which ~2 p25α molecules bind per heparin. However, in view of the repetitive polymeric nature of heparin, we do not rule out a shift in the heparin:p25α stoichiometry as we titrate heparin into p25α solution; such a "sliding model" cannot be excluded from the data. Remarkably, amyloid formation is observed only when binding enthalpies close to the maximum of −14.0 kcal/mol are observed. This indicates that essentially all added heparin bind to p25α, and each heparin molecule may thus potentially interact with multiple p25α monomers. AF4 separation of 15 μM p25α further showed that the protein peak (A205 nm) shifted toward higher retention times in the presence of 10 μg/ml heparin (data not shown), indicating that a significant proportion of p25α in solution interacts with heparin. The molar ratio at 15 μM p25α and 10 μg/ml heparin is 1 heparin:25 p25α, which leads to ~1.2 disaccharide units per p25α (assuming an average disaccharide molecular mass of 580 Da based on the disaccharide repeat 2-*O*-sulfated iduronic acid and 6-*O*-sulfated, *N*-sulfated glucosamine).

p25α aggregates contain amyloid structure

We decided to investigate the nature of these ThT-positive aggregates in more detail. The CD spectra

of p25α incubated in the presence of DP2–DP14 and HW heparin all show a minimum centered around 205 nm with a shoulder around ~220 nm (Fig. S3). For heparin-free p25α, the minimum is at ~220 nm and the shoulder is around 205 nm. At the lowest heparin concentrations (1 μg/ml), the amplitude of the CD signal is similar for samples incubated with DP2, DP6, and DP10, whereas the amplitude is slightly reduced in the presence of DP14 (Fig. S3a). On increasing the concentration to 10 μg/ml, the CD signal of DP2 remains unchanged whereas the CD signal of DP6, DP10, and DP14 is reduced (Fig. S3b). This effect is even more prominent in p25α samples incubated in the presence of 100 μg/ml heparin (Fig. S3c), leading to CD signals with amplitudes decreasing in the order DP2>DP6>DP10>DP14, the latter of which approaches the spectrum obtained in the presence of HW heparin.

The fact that the CD signal intensity decreases both with increasing heparin concentration and with decreasing polymerization can be rationalized in several ways. Either there could be a partial conversion of largely unstructured p25α into β-sheet structures (which has a lower signal intensity than random coil), or there could be a loss in overall signal intensity because increased aggregation leads to more light scattering and thus less signal. The latter is consistent with ThT measurements in which an increased aggregation propensity is observed at both increasing DP and increasing concentration. We therefore turned to attenuated total reflectance (ATR)-FTIR, which is not sensitive to light-scattering artifacts. The spectra of the p25α aggregates generally display a broad peak in the amide I region around 1656 cm^{-1} and a shoulder at approximately 1627 cm^{-1}, whereas freshly prepared p25α present a single peak at ~1656 cm^{-1} (Fig. 5a). The center and breadth of the ~1656 cm^{-1} peak are likely to arise from α-helical structures (1653±4 cm^{-1})[31] with a certain degree of conformational freedom,[32] in agreement with previous studies of p25α in which the protein was identified as flexible yet natively folded.[33] The shoulders identified at ~1627 cm^{-1} in aggregated p25α samples arise from amyloid structures (which manifest a peak at 1611–1630 cm^{-1}).[34]

Consistent with ThT measurements, we observed an increase in the amyloid shoulder from 0 to 10 μg/ml heparin after which the shoulder appears to decrease in intensity. This is clearly manifested in the amount of amyloid structure determined by deconvolution of the FTIR spectra by Lorentzian line fitting in Fig. 5b (example of fit shown in inset). Clearly, the largest amount of amyloid structure of ~25–30% is observed at a heparin concentration of 0.4–10 μg/ml, whereas a loss in amyloid structure is observed at higher heparin concentrations (~10–19%) but remain higher than heparin-free p25α (~4%). We conclude that heparin favors

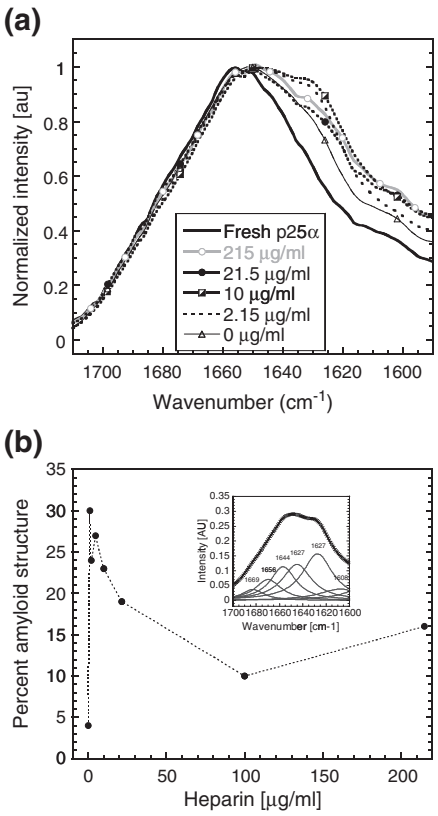


Fig. 5. (a) FTIR spectra of p25α aggregates formed in the presence of high MW heparin. (b) Estimation of amyloid structure content by deconvolution of FTIR spectra at different heparin concentrations by fitting of Lorentzian line shapes as shown in the inset.

amyloid formation by p25α in a dose-dependent manner at concentrations of up to around 10 μg/ml heparin, after which the binding of heparin may immobilize p25α in a state from which it may not take part in the fibrillation process.

Together with our CD data, this emphasizes that the presence of heparin favors amyloid formation by p25α. Further, the coexistence of amyloid and flexible structures in the FTIR analysis and the decrease in CD amplitude combine to suggest that the secondary structure of aggregates formed by p25α in the presence of heparin consists of an amyloid core surrounded by one or more flexible regions.

To determine in a non-spectroscopic manner if aggregation leads to a significant rearrangement of p25α, we subjected free p25α as well as heparin incubated in the presence of 100 μg/ml heparin to trypsin digestion. Both types of p25α are degraded by trypsin, yielding similar bands (Fig. S4). However, the extent of proteolysis was

reduced for aggregated p25α as shown already at the lowest trypsin concentration of 2 nM employed since clear bands persist in the aggregated samples, whereas intact free p25α had almost entirely disappeared and turned into lower molecular weight fragments. This may be due to simple steric inaccessibility in aggregated p25α and does not indicate a very substantial rearrangement of the protein structure. Incubation with lower concentrations of heparin led to smaller protection levels but no changes in the band distribution, suggesting that the heparin reduces proteolytic cleavage by simple steric protection.

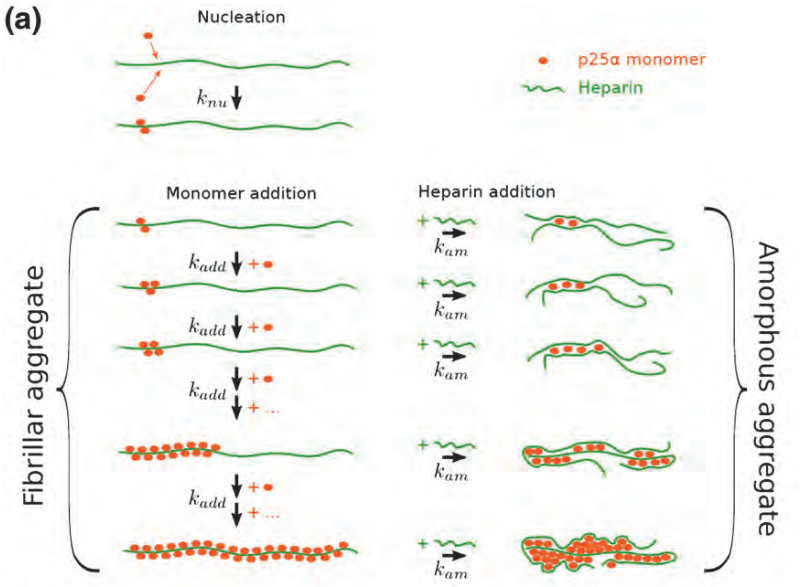
We have previously shown that monomeric p25α incubated in the presence of αSN stimulates the latter's ability to fibrillate.[1] αSN fibrils do not stimulate aggregation of monomeric p25α (data not shown). To investigate whether the p25α aggregates formed in the presence of heparin could stimulate αSN aggregation, we sonicated these p25α–heparin aggregates (to fragment the aggregates and thus increase the number of available growing ends for potential fibril elongation) and incubated them together with αSN. However, we observed no significant change in either the kinetics or overall yield of αSN fibrils (data not shown). Furthermore, these sonicated aggregates did not lead to any significant fibrillation of p25α in the absence of anionic components. The implication of this observation is that heparin is needed in stoichiometric amounts to stabilize the p25α fibril-like aggregates, and p25α is not able to incorporate into existing aggregates by itself.

Modeling p25α aggregation in the presence of heparin

We have attempted to recapitulate the kinetic data for p25α–heparin interactions in a mechanistic model outlined in Fig. 6a. The model has to capture the biphasic feature of these interactions, namely, that at 15 μM, p25α can form amyloid fibrils at an optimal heparin concentration of 10 μg/ml, while higher heparin concentrations lead to amorphous aggregates after a transient buildup of amyloid material. Accordingly, the model consists of three basic reactions: (1) nucleation of p25α onto heparin, (2) addition of p25α-monomers onto the nucleus to form amyloid structures, and (3) (at higher heparin concentrations) heparin-dependent formation of amorphous aggregates through addition of an extra heparin molecule to the existing p25α:heparin complex. Our model only requires one additional heparin molecule to bind at stage 3, though binding of a larger number of bound molecules is also possible. Details are provided in Supplementary Information.

The resulting differential equations describing the system are shown in Fig. 6b. The equations have been solved numerically in MATLAB. As shown in Fig. 7a, the model nicely fits the experimental data from Fig. 1a (fit parameters indicated in the legend to Fig. 7a) and recaptures the existence of an optimal heparin concentration for amyloid production (Fig. 7b). The simple model captures a great deal of the experimentally observed phenomena. At low heparin concentrations, all heparin polymers are quickly bound to p25α, and additional monomers are incorporated until they reach the maximal capacity, $N=43$. At higher heparin concentrations, excessive free heparin can bind to the p25α:heparin units and thereby form amorphous aggregates, leading to a lower ThT signal. Importantly, the concentration of free (uncomplexed) p25α declines slowly to what appears to be a plateau level. In our model, prolonged incubation will eventually lead to complete complexation due to the postulated irreversibility of the reaction, but a small backward reaction cannot be ruled out, leading to true equilibrium. The general shape of this curve is similar to the experimental data for free p25α (Fig. 2b) down to the minimum around 20 μg/ml heparin, although the overall plateau levels are lower in the model. This discrepancy in plateau levels can, to some extent, be explained by the fact that the model data only show free p25α, whereas some p25α:heparin complexes with small numbers of bound p25α may remain free in solution. The minimum around 20 μg/ml in Fig. 2b only shows borderline statistical significance but may be rationalized as follows in the model: The minimum indicates an optimum concentration for sequestering p25α as insoluble amyloid structures. p25α sequestering occurs through two processes, namely, nucleation onto free heparin and elongation onto growing p25α:heparin complexes. At 20 μg/ml heparin, there is enough heparin to allow p25α to bind and nucleate to a significant extent before the second heparin binding step occurs; at higher heparin concentrations, binding of additional heparin to the growing p25α:heparin complex occurs so rapidly as to block further incorporation of p25α.

We have also tested how the model responds to different concentrations of p25α at a fixed heparin concentration [Hep] = 10 μg/ml. The model nicely reproduces the experimental results (Fig. 2c). The simulated data make it more obvious that the curve is close to horizontal at low concentrations before starting to rise at higher concentrations, leading to an apparent threshold concentration of ~0.05 mg/ml p25α required for amyloid formation. Within the framework of our model, this threshold is not a consequence of the second-order p25α concentration dependence in Fig. 6b (where a minimum of two p25α molecules bind to each heparin molecule), since first- or third-order p25α concentration dependences give a similar shape though they do not fit the experimental data nearly as well (data not



(b)

p25α monomers

$$\frac{d[Mono]}{dt} = -2 \cdot k_{nu} \cdot [Mono]^2 \cdot [Hep] - k_{add} \cdot [Mono] \cdot \sum_{l=2}^{N_{max}} l \cdot [U_l] \quad (1)$$

Fibrillar units

$$\frac{d[U_l]}{dt} = \begin{cases} k_{nu} \cdot [Mono]^2 \cdot [Hep] & -k_{add} \cdot [Mono] \cdot l \cdot [U_l] \quad -k_{am} \cdot [Hep] \cdot [U_l] \quad \text{for} \quad l = 2 & (2) \\ k_{add} \cdot [Mono] \cdot (l-1) \cdot [U_{l-1}] & -k_{add} \cdot [Mono] \cdot l \cdot [U_l] \quad -k_{am} \cdot [Hep] \cdot [U_l] \quad \text{for} \quad 2 < l < N_{max} & (3) \\ k_{add} \cdot [Mono] \cdot (l-1) \cdot [U_{l-1}] & -k_{am} \cdot [Hep] \cdot [U_l] \quad \text{for} \quad l = N_{max} & (4) \end{cases}$$

Heparin

$$\frac{d[Hep]}{dt} = -k_{nu} \cdot [Mono]^2 \cdot [Hep] - k_{am} \cdot [Hep] \cdot \sum_{l=2}^{N_{max}} [U_l] \quad (5)$$

Amorphous units

$$\frac{d[A_l]}{dt} = k_{am} \cdot [Hep] \cdot [U_l] \quad \text{for} \quad 2 \le l \le N_{max} \quad (6)$$

$$[\text{Fibril}] = \sum_{l=2}^{N_{max}} [U_l]$$

$$[\text{Amorphous}] = \sum_{l=2}^{N_{max}} [A_l]$$

Fig. 6. (a) Model for p25α aggregation on heparin. The three processes included in the model are as follows: (1) Nucleation, in which two free p25α monomers bind to heparin. (2) Addition of free p25α monomers to the p25α:heparin units that are already formed. The p25α:heparin units can keep adding monomers until they have reached the maximum capacity of N monomers. (3) Addition of another heparin molecule, which will cause the p25α:heparin unit to restructure into an amorphous aggregate. (b) Differential equations describing time evolution of the mass concentration of p25α monomer, heparin, fibrillar p25α:heparin units (U_l), and amorphous p25α:heparin units (A_l). The total amount of amyloid fibrils is at all times given by the sum of fibrillar p25α:heparin units (Total fibril mass concentration = $\sum[U_l]$), and similarly, the total amount of amorphous aggregates is given by the sum of amorphous p25α:heparin units (Total amorphous mass concentration = $\sum[A_l]$).

shown). The threshold is retained (although the fit is reduced in quality) if the maximal number of p25α molecules binding per heparin molecule (N) is reduced. Removal of the logistical aspect (growth proportional to the number of incorporated p25α molecules) has the same consequences. All these modifications (except the third-order [p25α] dependence) reduce the steepness of the concentration dependence beyond the threshold value and the overall dynamics of fibril growth. Thus, the threshold is likely a cumulative consequence of the many steps involved in the nucleation and elongation processes.

Other anionic assemblies can also stimulate aggregation to different extents

Having observed the strong effect of heparin on p25α, we reasoned that other anionic assemblies might also induce aggregates in p25α. Indeed, the anionic surfactant SDS and the anionic lipid dioleoyl phosphatidylglycerol (DOPG) could induce ThT-positive aggregates to some extent, and this effect is attenuated when DOPG is diluted out with the zwitterionic lipid DOPC, which by itself does not induce any aggregates (Fig. S5a). However, the aggregates were in all cases amorphous according to

AFM images (Fig. S5b) and, unlike heparin, did not show any significant changes in CD spectra apart from a general decrease because of aggregate formation. More fibril-like aggregates were obtained using polyglutamate, where we observed an optimum ratio of polyE needed to stimulate aggregation, similar to heparin (Fig. S6a and b). Other polyanionic species such as RNA (Fig. S6c) and

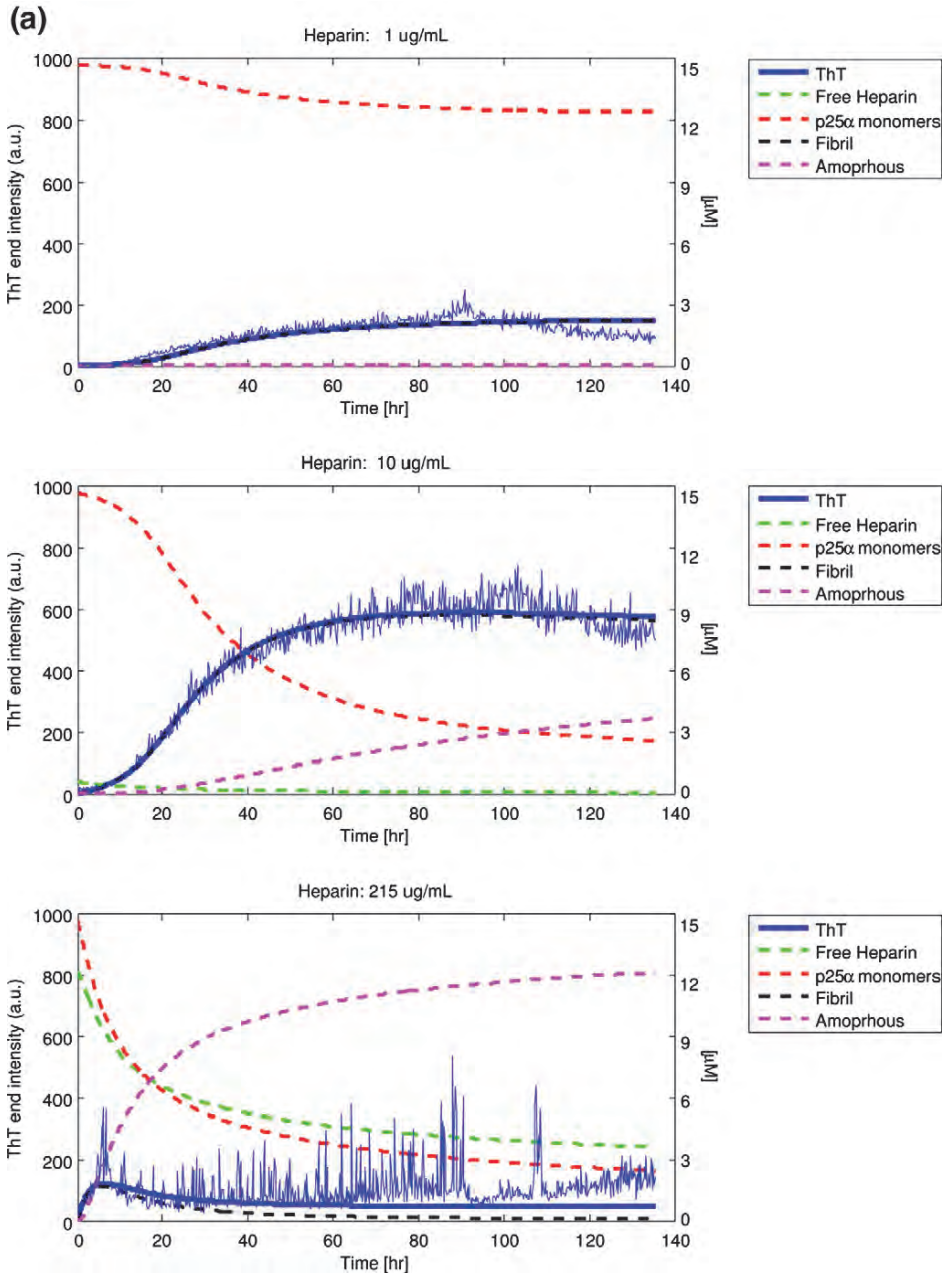


Fig. 7. (a) Model fit to three different heparin concentrations (1, 10, and 215 μg/ml). The parameters that gave the best fit were the following: $k_{nu}=0.6\ \mu M^{-2}\ s^{-1}$, $k_{add}=36\ \mu M^{-1}\ s^{-1}$, and $k_{am}=120\ \mu M^{-1}\ s^{-1}$. (b) Model predictions of ThT end intensities as a function of heparin concentration at 15 μM p25α. Experimental data from Fig. 1b were included for comparison. Inset shows modeled time course of aggregation (ThT intensities); the stippled arrow line shows the direction of increasing heparin concentration. (c) Predicted ThT signal for different concentrations of p25α, with heparin concentration fixed to [Hep] = 10 μg/ml. Experimental data from Fig. 2c were included for comparison. Inset: ThT time profiles for the different concentrations of p25α indicated in the figure (in milligrams per milliliter).

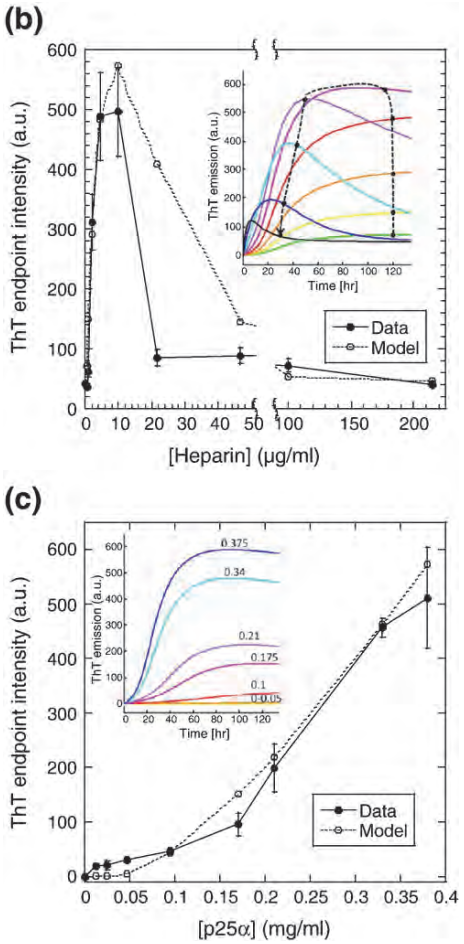


Fig. 7 (*legend on previous page*)

arachidonic acid (Fig. S6d) were also able to stimulate the aggregation process.

Discussion

A model for heparin-induced aggregation of p25α

Here, we demonstrate that low concentrations of the polyanionic polymers heparin and polyglutamate induce the formation of fibril-like aggregates of p25α. Anionic self-assembling amphiphiles such as SDS and DOPG also lead to aggregates but they have a less fibrillar character. In general, however, amyloid formation does not seem to involve rearrangement of a large part of the protein. Although ThT fluorescence signals of the resulting aggregates are relatively low in intensity compared to bona fide fibrils of αSN, deconvoluted FTIR spectra of heparin-induced aggregates indicate up to 26% amyloid structure depending on heparin concentration, and the proteolysis experiments suggest that the same overall fold is retained in the aggregate as in the monomeric protein. Furthermore, the fact that heparin molecules of polymerization 6–10 (corresponding to 26–43 Å in length for fully extended heparin, cf. PDB structure 1HPN and,[35] which is equivalent to 8–14 amino acid residues in β-sheet conformation[36]) are sufficient to stimulate aggregation suggests that only a small part of the protein needs to be aligned.

Based on the existing data and the close agreement with our model outlined in Fig. 6a, we propose the following scenario for aggregation:

A few residues of the protein (possibly those corresponding to the small β-sheet structure in the human and murine homologues) align with the heparin or polyglutamate structure. The interaction is probably driven by electrostatic (exothermic) interactions, just as has been suggested for many other heparin–protein interactions including apomyoglobin.[14] Based on ITC data, we estimate that each heparin molecule binds 2.4 p25α molecules, corresponding to 20–24 monosaccharide units and consistent with our observations that 10–14 monosaccharide units are sufficient to recapitulate the effect of full-length heparin. The initial binding of two p25α molecules serves as template for the subsequent alignment of additional p25α molecules, and this core structure is sufficient to propagate the amyloid structure, presumably surrounded by more or less native-like structure (explaining the rather modest spectroscopic changes observed by CD and FTIR). This is similar to the ability of a grafted Gln10 sequence to induce fibrillation of RNase where native domains align along a central amyloid zipper.[37] If too much heparin is added, it will bind to existing p25α:heparin complexes and block further buildup of p25α aggregates. We can rule out the alternative scenario that free heparin immobilizes p25α molecules and reduces the amount of free p25α that is needed to build onto templated p25α and form proper fibrils, since this will not produce the observed highly reproducible overshoot at the beginning of the fibrillation process (Fig. 1a). The amorphous aggregates represent association of heparin-complexed p25α molecules where the large concentration of heparin sterically prevents large-scale fibrillar alignment. In this model, the overshoot observed at higher heparin concentrations represents transient amyloid formation, after which more heparin binds to reorganize the aggregates to a more amorphous structure (leading to a more amorphous appearance).

The apparent ease of reorganization is consistent with our minimal amyloid model, in which only a small part of p25α participates in intermolecular cross-β contacts. The small number of intermolecular contacts will likely make binding weak and

dynamic, allowing the fibrils to form quickly but then rearrange to form amorphous aggregates in the presence of more heparin. The fact that the overshoot does not happen in the truncated versions of p25α, where increasing amounts of heparin uniformly leads to a lowering of the ThT fluorescence, suggests that both termini are required to facilitate the transient formation of this amyloid-like aggregate, possibly by protecting against too strong binding to heparin and allowing monomeric p25α to congregate on templated p25α. Although both terminal regions contain a small preponderance of positive charge, they are both highly charged (a third of all residues contain acidic or basic regions), and this high charge density may reduce binding to heparin.

Interestingly, our model confirms the existence of a residual monomer concentration corresponding to CAC of p25α after a plateau ThT level has been reached. This is an inherent property of the model and is probably a consequence of the many steps involved in nucleation and elongation, which leads to a gradually decreasing rate of incorporation of p25α into fibrils or amorphous aggregates. CACs also play a role in aggregation processes that require formation of oligomeric assemblies as precursors or nucleators of fibrillation.[38] However, we have no evidence for specific oligomeric structures of p25α; rather, heparin appears to serve as the template to start the aggregation reaction by stabilizing a continuum of different heparin:p25α complexes.

The fact that p25α aggregates do not significantly seed αSN aggregation indicates that stimulation of αSN aggregation is a separate phenomenon. Monomeric p25α undergoes a conformational change when complexed with monomeric αSN prior to aggregation,[3] and this possibility is obviously reduced when p25α is aggregated. We have previously observed that many other polyanionic structures do not have an effect on the structure of monomeric p25α,[3] implying that structural changes for p25α are coupled to aggregation. SDS can also induce aggregation of αSN to form structures midway between fully rigid classical fibrils and more amorphous aggregates.[39] However, the mechanism is likely to be different from heparin, since SDS draws proteins together by forcing them to cluster around a shared micelle. This leaves a small part of the protein free to extend from the micelle and possibly bridge to other micellar clusters via overlapping amyloid sequences.[39]

Role of polyanionic biomolecules *in vivo*

It is as yet unclear if the p25α-positive inclusions detected in neurons affected in Alzheimer's disease, Pick's disease, and multiple systems atrophy contain aggregated p25α, but they definitely present as inclusions with a high content of p25α without the presence of αSN or tau.[6,7] However, we have demonstrated that biological polyanionics can trigger aggregation. Some are more obvious candidates than others for stimulation of p25α aggregation *in vivo*. Heparin is produced by mast cells and basophilic cells and released only into the vasculature at sites of tissue injury. It is commonly isolated from the mucosal lining. However, the closely related heparan sulfate is reported to colocalize with tau aggregates in nerve cells of the Alzheimer's disease brain,[40] suggesting that these GAGs also can play a role in stimulating p25α aggregation in the brain. Other GAGs are also found intracellularly.[19] Zwitterionic phospholipids, which have no effect on p25α aggregation, are ubiquitous, but their anionic counterparts, which can stimulate p25α aggregation in this study and as reported elsewhere for many other proteins,[41,42] are typically only exposed on cell surfaces upon apoptosis. It is of particular interest that arachidonic acid also triggers the aggregation, as polyunsaturated fatty acids can trigger oligomerization of αSN,[43] and perturbations in polyunsaturated fatty acids have been demonstrated in models of PD.[44,45]

Materials and Methods

Materials

Heparin sodium salt from bovine intestinal mucosa (33 mg/ml stock solution; size range, ~6–80 kDa with a center at approximately 17 kDa, see Supplementary Information and Fig. S2), polyglutamate (molecular mass, 15–50 kDa), and RNA (R6750) were from Sigma-Aldrich (St. Louis, MO). Fluorescein–heparin was from Life Technologies (Grand Island, NY). Heparin oligosaccharides (dp2, dp6, dp10, and dp14 produced by enzymatic cleavage) were from Dextra (Reading, UK). All lipids were from Avanti Polar Lipids (Alabaster, AL).

Production of different p25α constructs

p25α (219 residues)[1] and p25 Δ3–43 (lacking residues 3–43)[3] were produced as previously described. The p25α constructs lacking the C-terminus (residues 156–219) were constructed from the pET-11d expression vector harboring the p25α gene as previously described.[46] Correct insertion was verified by DNA sequencing (MWG-Biotech). Prior to use, all p25α solutions were filtered through a 0.22-μm filter. AF4-MALS analyses (see experimental details in Supplementary Information, p. 2) indicated no species larger than monomeric p25α.

Aggregation of recombinant p25α constructs

p25α (20 μM; full-length or truncated) was incubated with PBS buffer (150 mM NaCl and 20 mM sodium phosphate, pH 7.4) in round-bottomed tubes and

incubated for 3 days at 37 °C with orbital shaking at 300 rpm. Aggregation was induced by addition of heparin (0–12.5 μg/ml), arachidonic acid (0–1 mM), polyglutamate (0–0.175 μg/ml), or RNA (0–1 mg/ml). Aggregation efficiency was evaluated by ThT fluorescence and/or centrifugation. For ThT analysis, p25α samples of 100 μl were pipetted in triplicate into a 96-well plate and 100 μl of 40 μM ThT diluted in 90 mM glycine–NaOH, pH 8.5, was added. The fluorescence intensity was measured with a Wallac Victor 1420 multilabel counter (Perkin Elmer, Waltham, MA), using excitation at 450 nm and fluorescence emission at 486 nm. To separate the supernatant and pellet, 100 μl of the sample was spun down in a Beckman airfuge at 14,000 rpm for 20 min, and the supernatant and pellet were separately run on SDS-PAGE.

Plate reader assays

These were used to follow p25α aggregation in real time. Conditions included 50 min/h orbital shaking at 300 rpm with one 3-mm-diameter glass microsphere (Glaswarenfabrik Karl Hecht GmbH & Co. KG, Sondheim, Germany) to increase reproducibility[28] in a sealed clear-bottom 96-well plate (Nunc, Thermo Fisher Scientific, Roskilde, Denmark). p25α (15 μM) in PBS buffer was supplemented with 40 μM ThT and appropriate amounts of heparin, SDS or DOPC, DOPG, and DOPC:DOPG vesicles. Fluorescence measurements were carried out every 12 min using excitation at 448 nm and emission at 485 nm in a Tecan GENios Pro plate reader (Männedorf, Switzerland). To determine the residual p25α concentration in solution, we collected samples from the plate after 140 h and centrifuged them at 13,000 rpm for 10 min. The concentration of soluble p25α species was determined by UV absorbance at 280 nm using a Nanodrop ND-1000 (Thermo Fisher Scientific, Wilmington, DE) using a calculated extinction coefficient of 0.425 mg cm^{-1} ml^{-1}. Errors are estimated to ~15% for end plateau ThT fluorescence levels and 20% for lag times based on visual inspection of triplicates.

ATR-FTIR spectroscopy

ATR-FTIR analysis of p25α aggregates was performed on a Bruker Tensor 27 FTIR (Bruker, Karlsruhe, Germany) equipped with a Golden Gate single reflection diamond ATR unit (Specac Ltd., Orpington, UK). Aggregates were concentrated by centrifugation at 13,000 rpm for 5 min by removal of the supernatant and resuspended in a minimum volume of MilliQ H_2O by gentle pipetting. Aliqouts of sample (typically 4–10 μl) were transferred to the ATR cell and dried in a stream of nitrogen gas until the water signal had stabilized. The adsorption spectrum was recorded in the interval 1000–4000 cm^{-1} using a nominal resolution of 2 cm^{-1} and noise was reduced through 128 accumulations. Peak fitting was done by the use of Lorentzian lines between 1575 and 1725 cm^{-1} wave numbers and the percentage of individual secondary-structure elements was calculated from the integral of peaks centered between 1600 and 1700 cm^{-1}. Data treatment was done using OPUS version 5.5 (Bruker).

Electron microscopy

Two hundred microliters of a 15-μM p25α aggregation sample was produced as described above using 1, 10, or 100 μg/ml DP2–DP10 or HW fraction heparin. The aggregated p25α was isolated by centrifugation as described above, and the pelleted aggregates were resuspended in 21 μl of ddH_2O. Grids were washed in two drops of double-distilled water and stained with 1% phosphotungstic acid (pH 6.8) and blotted dry on filter paper. Samples were viewed with a JEOL 1010 transmission electron microscope.

Atomic force microscopy

The morphology of p25α aggregates was visualized on a PicoSPMI apparatus (Molecular Imaging Corporation, Tempe, AZ). Aggregated sample (5 μl) was pipetted onto freshly cleaved mica, air dried for ~10 min, washed with water, and blown dry. The images were recorded in contact mode using a cantilever (Sini AFM probe, NanoAndMore GMBH), and the images were visualized by the software Picoscan 5.5.3 (Molecular Imaging Corporation).

Solubilization of p25α aggregates

p25α aggregates were isolated by centrifugation. The pellet was washed in PBS, pH 7.2, and then resuspended in 1% Triton X-100 followed by incubation at 37 °C for 15 min. The suspension was centrifuged again and the pellet was resuspended in 1% SDS and incubated at 37 °C for 15 min. After another round of centrifugation, the supernatants and pellets were analyzed by SDS-PAGE.

Isothermal titration calorimetry

The binding of heparin to p25α was monitored using a VP-ITC form MicroCal (Northampton, MA) at 37 °C. The solutions were degassed under vacuum with a ThermoVac accessory from MicroCal. Heparin (200 μM; HW heparin or DP14 heparin) was titrated into 1.4 ml of 15 μM p25α in PBS, pH 7.2, using 12 injections of 5 μl and 22 injections of 10 μl, and the peak areas were integrated using the ITC data analysis module in MicroCal Origin 7.0. Reference titrations in which PBS buffer, pH 7.2, was titrated into p25α and heparin titrations into PBS, pH 7.2, were subtracted but did not give rise to significant signals.

Acknowledgements

This work was supported by the Innovation Consortium CureND via a grant from the Danish Ministry of Science, Technology and Innovation (D.E.O., L.G., S.S., and P.H.J.) and from the Lundbeck Foundation (D.E.O., S.B.N., and M.H.J.).

D.E.O. is supported by the Danish Research Foundation (inSPIN).

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2012.01.050

References

1. Lindersson, E., Lundvig, D., Petersen, C., Madsen, P., Højrup, P., Moos, T. *et al.* (2005). P25a is co-expressed with α-synuclein in α-synucleinopathies and stimulates its aggregation. *J. Biol. Chem.* **280**, 5703–5715.
2. Hlavanda, E., Kovacs, J., Olah, J., Orosz, F., Medzihradszky, K. F. & Ovadi, J. (2002). Brain-specific p25 protein binds to tubulin and microtubules and induces aberrant microtubule assemblies at substoichiometric concentrations. *Biochemistry*, **41**, 8657–8664.
3. Otzen, D. E., Lundvig, D., Wimmer, R., Hatting, L., Pedersen, J. R. & Jensen, P. H. (2005). p25alpha is flexible but natively folded and binds tubulin in an oligomeric complex. *Protein Sci.* **14**, 1396–1409.
4. Monleón, D., Chiang, Y., Aramini, J. M., Swapna, G. V. T., Macapagal, D., Gunsalus, K. C. *et al.* (2004). Backbone ^{1}H, ^{15}N and ^{13}C assignments for the 21 kDa *Caenorhabditis elegans* homologue of "brain-specific" protein. *J. Biomol. NMR*, **28**, 91–92.
5. Takahashi, M., Tomizawa, K., Ishiguro, K., Sato, K., Omori, A., Sato, S. *et al.* (1991). A novel brain-specific 25 kDa protein (p25) is phosphorylated by a Ser–Thr–Pro kinase (Tpk-Ii) from Tau protein-kinase fractions. *FEBS Lett.* **289**, 37–43.
6. Baker, K. G., Huang, Y., McCann, H., Gai, W. P., Jensen, P. H. & Halliday, G. M. (2006). P25α immunoreactive but α-synuclein immunonegative neuronal inclusions in multiple system atroph. *Acta Neuropathol.* **111**, 193–195.
7. Kovacs, G. G., Laszlo, L., Kovacs, J., Jensen, P. H., Lindersson, E., Boton, G. *et al.* (2004). Natively unfolded tubulin polymerization promoting protein TPPP/p25 is a common marker of alpha-synucleinopathies. *Neurobiol. Dis.* **17**, 155–162.
8. Zhu, H. L., Fernández, C., Fan, J. B., Shewmaker, F., Chen, J., Minton, A. P. & Liang, Y. (2010). Quantitative characterization of heparin binding to tau protein. *J. Biol. Chem.* **285**, 3592.
9. Gendron, T. F. & Petrucelli, L. (2009). The role of tau in neurodegeneration. *Mol. Neurodegener.* **11**, 13.
10. Goedert, M., Jakes, R., Spillantini, M. G., Hasegawa, K., Smith, M. J. & Crowther, R. A. (1996). Assembly of microtubule-associated protein tau into Alzheimer-like filaments induced by sulphated glycosaminoglycans. *Nature*, **383**, 550–553.
11. Kuret, J., Chirita, C. N., Congdon, E. E., Kannanayakal, T., Li, G., Necula, M. *et al.* (2005). Pathways of tau fibrillization. *Biochim. Biophys. Acta*, **1739**, 167–178.
12. Cohlberg, J. A., Li, J., Uversky, V. N. & Fink, A. L. (2002). Heparin and other glycosaminoglycans stimulate the formation of amyloid fibrils from [alpha]-synuclein in vitro. *Biochemistry*, **41**, 1502–1511.
13. McLaurin, J., Franklin, T., Zhang, X., Deng, J. & Fraser, P. E. (1999). Interactions of Alzheimer amyloid-beta peptides with glycosaminoglycans: effects on fibril nucleation and growth. *Eur. J. Biochem.* **266**, 1101–1110.
14. Vilasi, S., Sarcina, R., Maritato, R., De Simone, A., Irace, G. & Sirangelo, I. (2011). Heparin induces harmless fibril formation in amyloidogenic W7FW14F apomyoglobin and amyloid aggregation in wild-type protein in vitro. *PLoS One*, **6**, e22076.
15. Suk, J. Y., Zhang, F., Balch, W. E., Linhardt, R. J. & Kelly, J. W. (2006). Heparin accelerates gelsolin amyloidogenesis. *Biochemistry*, **45**, 2234–2242.
16. Yamamoto, S., Yamaguchi, I., Hasegawa, K., Tsutsumi, S., Goto, Y., Gejyo, F. & Naiki, H. (2004). Glycosaminoglycans enhance the trifluoroethanol-induced extension of β_2-microglobulin-related amyloid fibrils at a neutral pH. *J. Am. Soc. Nephrol.* **15**, 126–133.
17. Zhang, X. & Li, J. P. (2010). Heparan sulfate proteoglycans in amyloidosis. *Prog. Mol. Biol. Transl. Sci.* **93**, 309–334.
18. Bourgault, S., Solomon, J. P., Reixach, N. & Kelly, J. W. (2011). Sulfated glycosaminoglycans accelerate transthyretin amyloidogenesis by quaternary structural conversion. *Biochemistry*, **50**, 1001–1015.
19. Kolset, S. O., Prydz, K. & Pejler, G. (2004). Intracellular proteoglycans. *Biochem. J.* **379**, 217–227.
20. Maji, S. K., Schubert, D., Rivier, C., Lee, S., Rivier, J. E. & Riek, R. (2008). Amyloid as a depot for the formulation of long-acting drugs. *PLoS Biol.* **6**, e17.
21. Nielsen, S. B., Franzmann, M., Basaiawmoit, R. V., Wimmer, R., Mikkelsen, J. D. & Otzen, D. E. (2010). β-Sheet aggregation of kisspeptin is stimulated by heparin but inhibited by amphiphiles. *Biopolymers*, **93**, 678–689.
22. Valnickova, Z., Petersen, S. V., Nielsen, S. B., Otzen, D. E. & Enghild, J. J. (2007). Heparin binding induces a conformational change of pigment epithelium-derived factor. *J. Biol. Chem.* **282**, 6661–6667.
23. Sandwall, E., O'Callaghan, P., Zhang, X., Lindahl, U., Lannfelt, L. & Li, J. P. (2010). Heparan sulfate mediates Aβ internalization and cytotoxicity. *Glycobiology*, **20**, 533–541.
24. Zacharia, E., Metzger, S., Chajek-Shaul, T., Aingorn, H., Elkin, M., Firedmann, Y. *et al.* (2004). Transgenic expression of mammalian heparanase uncovers physiological functions of heparan sulfate in tissue morphogenesis, vascularization and feeding behavior. *FASEB J.* **18**, 252–263.
25. Li, J. P., Galvis, M. L. E., Gong, F., Zhang, X., Zacharia, E., Metzger, S. *et al.* (2005). *In vivo* fragmentation of heparan sulfate by heparanase overexpression renders mice resistant to amyloid protein A amyloidosis. *Proc. Natl Acad. Sci. USA*, **102**, 6473–6477.
26. Nandi, P. K. & Nicole, J. C. (2004). Nucleic acid and prion protein interaction produces spherical amyloids which can function in vivo as coats of spongiform encephalopathy agent. *J. Mol. Biol.* **344**, 827–837.
27. Otzen, D. E. (2010). Amyloid formation in surfactants and alcohols: membrane mimetics or structural switchers? *Curr. Protein Pept. Sci.* **11**, 355–371.
28. Giehm, L. & Otzen, D. E. (2010). Strategies to increase the reproducibility of α-synuclein fibrillation in plate reader assays. *Anal. Biochem.* **400**, 270–281.
29. Capila, I. & Linhardt, R. J. (2002). Heparin–protein interactions. *Angew. Chem., Int. Ed.* **41**, 390.

30. Hileman, R. E., Fromm, J. R., Weiler, J. M. & Linhardt, R. J. (1998). Glycosaminoglycan–protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *BioEssays*, **20**, 156–167.
31. Kong, J. & Yu, S. (2007). Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochim. Biophys. Sin.* **39**, 549.
32. Chirgadze, Y., Brazhnikov, E. & Nevskaya, N. (1976). Intramolecular distortion of the [alpha]-helical structure of polypeptides. *J. Mol. Biol.* **102**, 781–792.
33. Otzen, D., Lundvig, D., Wimmer, R., Nielsen, L., Pedersen, J. & Jensen, P. (2005). p25α is flexible but natively folded and binds tubulin with oligomeric stoichiometry. *Protein Sci.* **14**, 1396–1409.
34. Zandomeneghi, G., Krebs, M. R. H., McCammon, M. G. & Fandrich, M. (2004). FTIR reveals structural differences between native β-sheet proteins and amyloid fibrils. *Protein Sci.* **13**, 3314–3321.
35. Gandhi, N. S. & Mancera, R. L. (2008). The structure of glycosaminoglycans and their interactions with proteins. *Chem. Biol. Drug Des.* **72**, 455–482.
36. Creighton, T. E. (1993). *Proteins. Structures and Molecular Properties*, 2nd edit. W.H. Freeman & Co., New York, NY.
37. Sambashivan, S., Liu, Y., Sawaya, M. R., Gingery, M. & Eisenberg, D. (2005). Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature*, **437**, 266–269.
38. Powers, E. T. & Powers, D. L. (2006). The kinetics of nucleated polymerizations at high concentrations: amyloid fibril formation near and above the "supercritical concentration". *Biophys. J.* **91**, 122–132.
39. Giehm, L., Oliveira, C. L. P., Christiansen, G., Pedersen, J. S. & Otzen, D. E. (2010). SDS-induced fibrillation of α-synuclein: an alternative fibrillation pathway. *J. Mol. Biol.* **401**, 115–133.
40. Snow, A. D., Willmer, J. P. & Kisilevsky, R. (1987). Sulfated glycosaminoglycans in Alzheimer's disease. *Hum. Pathol.* **18**, 506–510.
41. Zhao, H., Tuominen, E. K. J. & Kinnunen, P. K. J. (2004). Formation of amyloid fibers triggered by phosphatidylserine-containing membranes. *Biochemistry*, **43**, 10302–10307.
42. Knight, J. D. & Miranker, A. D. (2004). Phospholipid catalysis of diabetic amyloid assembly. *J. Mol. Biol.* **341**, 1175–1187.
43. Perrin, R. J., Woods, W. S., Clayton, D. F. & George, J. M. (2001). Exposure to long chain polyunsaturated fatty acids triggers rapid multimerization of synucleins. *J. Biol. Chem.* **276**, 41958–41962.
44. Sharon, R., Bar-Joseph, I., Mirick, G. E., Serhan, C. N. & Selkoe, D. J. (2003). Altered fatty acid composition of dopaminergic neurons expressing alpha-synuclein and human brains with alpha-synucleinopathies. *J. Biol. Chem.* **278**, 49874–49881.
45. Assayag, K., Yakunin, E., Loeb, V., Selkoe, D. J. & Sharon, R. (2007). Polyunsaturated fatty acids induce α-synuclein-related pathogenic changes in neuronal cell. *Am. J. Pathol.* **171**, 2000–2011.
46. Kleinnijenhuis, A. J., Hedegaard, C. J., Lundvig, D., Sundbye, S., Issinger, O. G., Jensen, O. N. & Jensen, P. H. (2008). Identification of multiple post-translational modifications in the porcine brain specific p25 α. *J. Neurochem.* **106**, 925–933.

Yde et al. BMC Systems Biology 2011, 5:115
http://www.biomedcentral.com/1752-0509/5/115

BMC Systems Biology

RESEARCH ARTICLE Open Access

Modeling the NF-κB mediated inflammatory response predicts cytokine waves in tissue

Pernille Yde, Benedicte Mengel, Mogens H Jensen, Sandeep Krishna and Ala Trusina*

Abstract

Background: Waves propagating in "excitable media" is a reliable way to transmit signals in space. A fascinating example where living cells comprise such a medium is Dictyostelium D. which propagates waves of chemoattractant to attract distant cells. While neutrophils chemotax in a similar fashion as Dictyostelium D., it is unclear if chemoattractant waves exist in mammalian tissues and what mechanisms could propagate them.

Results: We propose that chemoattractant cytokine waves may naturally develop as a result of NF-κB response. Using a heuristic mathematical model of NF-κB-like circuits coupled in space we show that the known characteristics of NF-κB response favor cytokine waves.

Conclusions: While the propagating wave of cytokines is generally beneficial for inflammation resolution, our model predicts that there exist special conditions that can cause chronic inflammation and re-occurrence of acute inflammatory response.

Background

Inflammatory response (IR) in higher organisms requires efficient chemotaxis of neutrophils to sites of infection [1]. At the same time excessive neutrophil accumulation has been shown to play a role in diseases such as asthma, atherosclerosis, multiple sclerosis, inflammatory bowel disorder and arthritis [2]. It however remains an open question how the chemoattractant signal is transmitted through the tissue. "Propagating waves" present an optimal way of transmitting a signal across large distances and occur in many biological systems [3], [4]. In particular, propagating waves of chemoattractant are utilized by the social amoeba Dictyostelium D. - a model system for neutrophil chemotaxis [5]. While neutrophils *can* efficiently chemotax through chemoattractant waves [5] it is unclear if they ever encounter such situations. Unlike Dictyostelium, neutrophils do not generate the waves themselves and it remains an open question if there exists a mechanism that could initiate and propagate waves of chemoattractant during IR.

We here suggest that NF-κB is the missing link relating IR in tissue cells to the propagation of a chemoattractant signal. NF-κB upregulates transcription of many cytokines which serve as chemoattractants for neutrophils e.g. TNF, IL-1, IL-6, IL-8 and IL-11 [6-8]. At the same time these cytokines activate NF-κB response. We show that a simple model of spatially coupled tissue cells contains all the necessary components to initiate and propagate waves of chemoattractant cytokines. This model behaves as an "excitable medium" [9] and relies on the following well-known characteristics of IR: 1) fast transient response of NF-κB, 2) a positive feedback from NF-κB to cytokines and 3) short half-life of cytokines. Using mathematical modeling, we find that all these properties favor formation and propagation of cytokine waves.

Propagating waves - an optimal strategy for signaling in the tissue?

In principle there are multiple ways a chemoattractant signal can be transmitted through the tissue, however not all of them are equally efficient and reliable. In the simplest scenario the chemoattractant molecule passively diffuses from the site of infection. This will result in a short-ranged signal where the concentration decays exponentially with the distance from the source (see Figure 1A). The range of the signal will be further limited by the typical short half-life of the chemoattractant molecules.

* Correspondence: trusina@nbi.dk
Niels Bohr Institute, Blegdamsvej 17, DK 2100, Copenhagen, Denmark

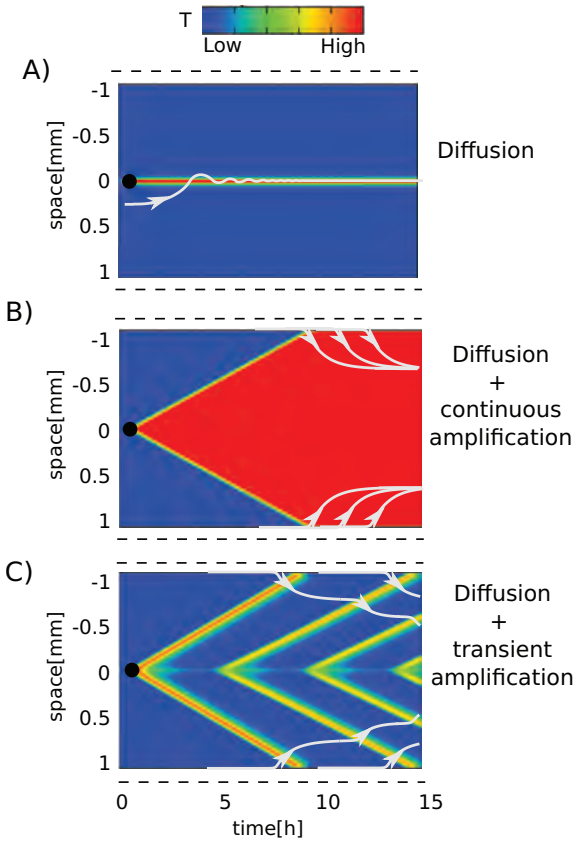


Figure 1 Three scenarios for signal propagation from the site of infection. Three scenarios for signal propagation from the site of infection marked by the black circle. Cells are aligned vertically and changes of chemoattractant concentrations in time are developing from left to right. The blue (red) encode low (high) concentrations of chemoattractant (T). White lines represent trajectories of neutrophils chemotaxing from blood vessels (dashed lines). The trajectories were calculated using a "Local Excitation - Global Inhibition" model, described in the additional file 1. In A) the chemoattractant is diffusing from the site of infection. The signal is short-ranged, and no neutrophils are recruited from the distant blood vessels. In both B) and C) the signal is long-ranged. In B) the diffusing chemoattractant is continuously amplified by tissue cells. The gradient is sharp at the blood vessel but it disappears deeper in the tissue thus leaving neutrophils devoid of direction. In C) diffusing chemoattractant is amplified *transiently*. The waves recruit neutrophils from the blood stream and also serve as a directing signal for the neutrophils that are already in the tissue.

Some neutrophil chemoattractants, e.g. the cytokines TNF and IL-1, have the unique ability to self-amplify by means of tissue cells or tissue resident macrophages. For example, the upregulation of TNF in a localized area of myocardium can easily induce TNF upregulation in neighboring normal myocardium [10]. This active participation of tissue cells *amplifies* the cytokine concentration across the tissue as illustrated in Figure 1B. Blood vessels serve as a sink for cytokines where they are carried away by the bloodstream. As a result a sharp gradient develops near the blood vessel. It is important to note that both the "diffusion" and the "amplification" (Figure 1A and 1B) scenarios create static gradients, i.e. once in steady state the gradients are not changing in time. Such static gradients are intrinsically toxic for the tissue as, for example, continuous exposure of tissue cells to high TNF levels triggers apoptosis [11]. Interestingly, some experimental evidences suggest that static gradients are also suboptimal for neutrophil chemotaxis; neutrophils seem to orient themselves better when exposed to temporally varying gradients [5,12-15].

If the cytokine concentration is not amplified continuously but *transiently* (i.e. with a peak-like profile), the tissue cells will avoid sustained exposure to toxic cytokines while the signal - the chemoattractant gradient - can still penetrate far in the tissue. Such transient amplification can result in single or re-emerging "propagating waves" as shown in Figure 1C.

It turns out that cytokines are indeed amplified only transiently [16]. Cytokines induce activation of NF-κB - a key regulator of IR in tissue cells. In turn active NF-κB upregulates cytokine production [6-8], thus constituting an amplifying *positive feedback*. If NF-κB-response to inflammatory stimuli is transient, so will be the amplification. And indeed, the synthesis and secretion of inflammatory cytokines from tissue cells were shown to be parallel to the the NF-κB transient activation [16].

Results and Discussion

Model

The model consists of spatially distributed cells each containing NF-κB-like circuits. The circuit in each cell thus contains 3 variables: An NF-κB-like variable N, a *regulator* variable R that combines the effects of all inhibitors in one variable and a cytokine-like variable T see Figure 2A. The key features of our model which we shall explain in more detail below are: 1) slow inhibition due to negative feedback from inhibitors (R), 2) fast amplification due to positive feedback from cytokines (T) and 3) spatial coupling of the NF-κB-like circuits due to diffusion of extracellular cytokines.

Transient response through negative feedback

Upon extracellular stimulation the level of active NF-κB increases to reach a peak value after approximately 30 minutes and has decreased again after approximately one hour [17-19]. It has been shown that this *transient* NF-κB response is assured by multiple negative feedback loops where active NF-κB induces expression of its own inhibitors. Some of these inhibitors act directly on NF-κB, e.g. IκBα, β and ε. Others, as for example the members of A20 family proteins A20 [20] and Cesanne

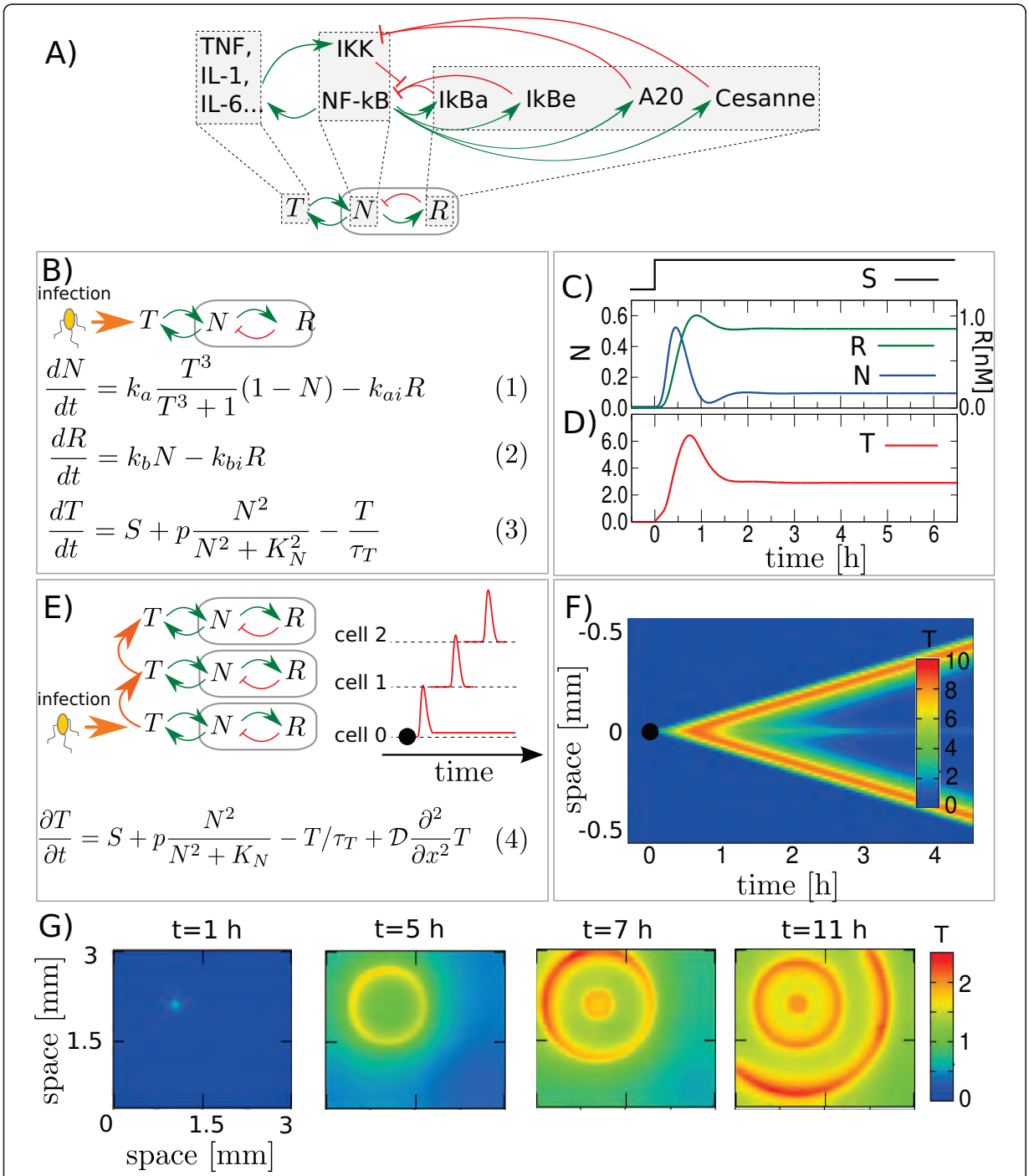


Figure 2 Model. Model construction: A) Multiple feedbacks in NF-κB regulatory network are reduced to the core circuit consisting of positive cytokine feedback, $T \rightarrow N \rightarrow T$, that allows signal amplification and negative feedback, $N \rightarrow R \dashv N$, that captures NF-κB transient dynamics. B) Mathematical description of the NF-κB-like circuit. N denotes the fraction of active NF-κB; *T* is the fold induction in cytokine concentration and R is a negative regulator of *N*. C) The transient response of N to the stimuli S, and D) corresponding *T* profile in single isolated cell. E) The model of spatially coupled NF-κB-like circuits. The coupling is through the diffusion of cytokines, *T*, and results in "propagating waves " of *T* shown in space-time plot in F) and G). In G) we illustrate the 2D variant of the model simulated with the same parameters.

[21] inhibit upstream signaling components at the level of IKK or above.

The main intent with our model is not to capture all the intricacies of the NF-κB system, but rather to focus on the few mechanisms we believe are crucial for linking the transient response of NF-κB to spatio-temporal profiles of cytokines. With that in mind we choose to reduce the multiple nested negative feedbacks comprising the NF-κB regulatory network to a single negative feedback, see Figure 2A and 2B, equations 1 and 2. The main purpose of this negative feedback, $N \rightarrow R \dashv N$, is to reproduce the transient dynamics of NF-κB in response to TNF-stimulation. It is important to note that this part of the model is a phenomenological generalization, i.e. it aims at a simple mathematical reproduction of observed results while not relating to the exact mechanisms. We have tested and confirmed that the main results hold if the nested negative feedbacks are modeled explicitly, see Figure S2 in additional file 1.

Amplification through positive cytokine feedback

The positive feedback where nuclear NF-κ*B* up-regulates production of cytokines and they in turn induce NF-κB nuclear localization is described by equations 1 and 3 in Figure 2A. The strength of the N→T upregulation is governed by the parameter p, and thus captures the strength of the entire positive feedback (the strength of the T→N upregulation, k_a is fixed, see Methods section). For simplicity we replace the double negative NF-κB activation pathway - where cytokines (e.g., TNF) activates IKK, which in turn inhibits the NF-κB inhibitor - by T directly activating N. Mathematically this $T \rightarrow N$ activation term is modeled by a Hill function. It encodes an activation *threshold* in NF-κB system, as recently reported by Turner et al. and Tay et al. [19,22].

Spatial coupling of single cells

The spatial coupling of cells is schematically shown in Figure 2E. Newly synthesized cytokines are secreted into the extracellular space where they diffuse and induce NF-κB response in neighboring cells. The diffusion of T is described by the diffusion term in equation (4). Note that T is the only variable that diffuses. The variables N and R are bounded inside the individual cells. The blood vessels are placed at the two ends of the line of cells and are modeled as absorbing (open) boundaries. Absorbing boundaries take into account that the cytokines are carried away by the blood flow, thus producing a sink for the variable T. We further assume that the small blood vessels have negligible effect on cytokine diffusion (the results also hold if the small blood vessels are taken into account, see Figure S1 in additional file 1).

In our analysis we assume the parameters estimated for TNF to be characteristic of multiple cytokines constituting the positive feedback loop. Our model consists of 3 variables for each cell and 8 parameters. Among the 8 parameters there is only one that is essentially unconstrained: The strength of positive feedback, p. The other parameters are estimated as described in Methods section.

Propagating Waves Arise from Spatially Coupled NF-κB-like circuits

The dynamics of a single isolated NF-κB-like circuit is shown in Figure 2C and 2D. The results are obtained by numerical integration of the ordinary differential equations (1)-(3) (Figure 2B). To trigger the response we add an extracellular stimulus S (see equation (3)). Here and in all following simulations the stimulus, S, is added at time 0 and is present at all times unless otherwise mentioned. The stimulus S should be thought of as bacterial endotoxin or initiating cytokines secreted by macrophages.

Shortly after stimulus is added the cell responds by increasing the level of active NF-κB (N). Because of the negative feedback from inhibitors (R), NF-κB (N) decreases back to lower values after approximately an hour. The positive feedback from NF-κB (N) amplifies the concentration of cytokines (T) to become many orders of magnitude larger than it would have reached by the small stimulus S alone (Figure 2D). Note that the cytokine concentration (T) also peaks on a timescale similar to that of NF-κB.

Interestingly, our model, although simplified, can capture characteristic biphasic response in both the NF-κB-like variable, N (Figure 2C, blue line), and the cytokine-like variable, T (Figure 2D, red line). Here an acute phase - with a well pronounced peak - is followed by a late phase - where the concentrations are lower than in the acute phase, but are above the pre-stimulus concentrations. In our model the late-phase of the response is entirely due to cytokine positive feedback. This observation agrees well with in-vivo experiments by Han et al., where authors demonstrated that both TNF and IL-1 are required for the late-phase response [16].

When multiple cells are aligned next to each other (in a one-dimensional lattice) a peak in T propagates from cell-to-cell, see Figure 2E and 2F. These results are obtained using equations (1), (2) and (4).

The main focus of this study is the response to a well localized source of inflammation, i.e. the source of bacterial endotoxin or the cytokine-secreting macrophages accumulated at the location of the damaged cells. We model this by adding a small external stimuli, S, at time $t \geq 0$ only to the cell in the middle (Figure 2E, F black circle). The cytokines produced by the stimulated cell at the site of infection will diffuse away and thereby trigger the transient response of the NF-κB system in neighboring cells. The result of this cell-to-cell coupling is a "propagating wave" of NF-κB induction followed by

cytokine production and hence a wave of high T-concentration propagating through the tissue, see Figure 2F and 2G. Note that the result above relies only on three requirements: 1) Transient response in NF-κB-like variable N (slow negative feedback), 2) amplification of the cytokine-like variable T (positive feedback) and 3) diffusion of T in extracellular space (spatial coupling). Variations in the parameter p, show that strong positive feedback generates a more pronounced wave. Both the amplitude and the speed of the wave increase with p (see Figure 3).

Inflammatory response exhibits the characteristics of an "excitable media"

Propagating waves are employed by many other biological systems, which share the need of sending information over relatively large distances, where passive diffusion is insufficient. Similar phenomena are observed in movement of calcium in differentiating Xenopus oocytes [3] and the rapidly propagating action potentials of neurons [4]. These systems, as well as the spatially coupled NF-κB-like circuits, share the properties of an "excitable media". An excitable medium is comprised of locally excitable regions - in our case it is a tissue cell - which all have the ability to get *induced* (excited) and *inhibited*. Such systems are characterized by the "excitation threshold", so that sub-threshold stimuli are rapidly damped, and the system persists in a *resting state* (low T, N and R). Super-threshold stimuli induce sharp local response and the system transits into the *excited state* (high T and N). Shortly after the response occurs, the region becomes insensitive to further perturbation and is said to be in a *refractory period* (high R), after which it can relax back to the resting state where it is again sensitive to perturbations [9].

We have performed a detailed mathematical analysis of the mechanism behind the excitable-media properties of NF-κB-like circuits (see Figure 1S in additional file 1). The analyses confirm that the NF-kB responses coupled in space present a novel example of excitable media.

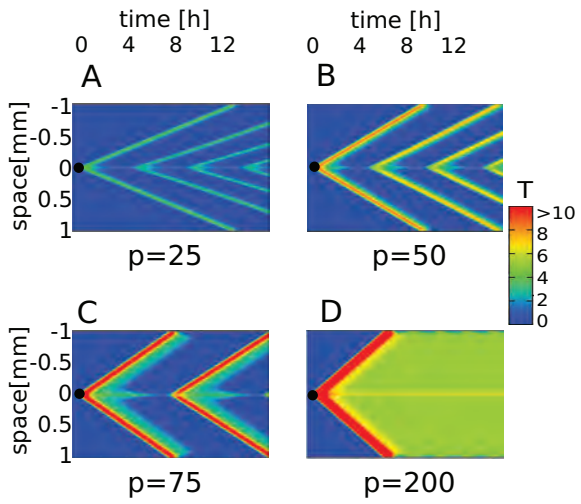


Figure 3 The role of positive feedback. Propagating waves re-emerge A)-C) at low to intermediate strengths of positive feedback, p = 25, 50, 75. If the amplification is too strong the system is locked in "high T" state, D) p = 200.

Predictions and physiological relevance

It has recently been shown that the circuits combining positive and negative feedbacks allow for robust oscillations [23,24]. We find that our model, where such circuits are coupled in space, can indeed produce *re-emerging* waves.

The model predicts that the conditions for re-establishment depend strongly on two parameters: The strength of cytokine positive feedback, p and the cytokine half-life, τ_T. These parameters control the amount of cytokines, T, and have to be inversely related, i.e. $p \propto 1/\tau_T$ to minimize the exposure of tissue cells to cytokines (e.g., a strong positive feedback, that we found to favor wave formation, can be compensated by short cytokine half-life). Remarkably, the reported cytokine half-life is indeed short and ranges between 3-25 minutes [25-27].

The analyses of the parameter p (Figure 3A-D) show that: 1) The frequency of waves can be modulated by the strength of positive feedback and 2) the dependence is non-monotonic, i.e. the frequency is maximal (corresponding to period of 4.5 hours) at intermediate values of p (Figure 3A-D) and 3) there exists a "locked" state with high T concentration and infinite refractory period - see Figure 4D.

The frequency of the cytokine waves can have a direct implication on the amounts of neutrophils recruited to the site of infection (i.e., more frequent waves recruit more neutrophils). As the NF-κB response is modulated at multiple levels (e.g., cytokine receptor desensitization, cooperative transcriptional regulation, etc.) one can imagine a scenario where the strength of positive feedback can be modulated to encode the severity of infection by e.g. increasing the transcription, translation or secretion rates of cytokines.

Chronic and recurrent acute inflammation

Surprisingly, the "locked" state with continuously high T (Figure 3D), appears to be self-sustained. Unlike the repetitive waves in Figure 3A-C - which disappear once stimuli is removed - the sustained high production of T can be triggered by just a short pulse of stimuli (Figure 3D and 4A). In the context of IR this situation resembles chronic inflammation - meaning that the response

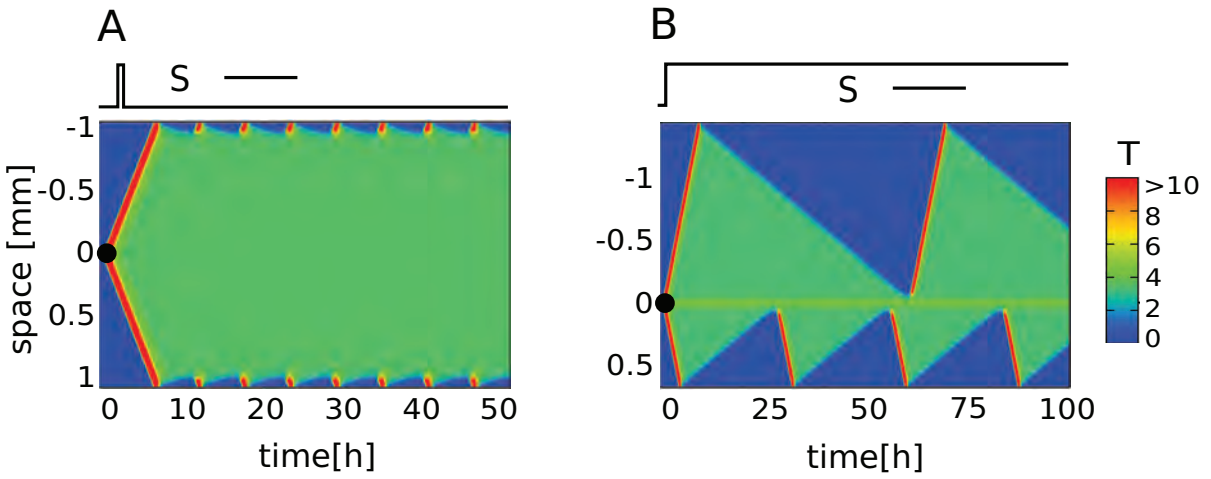


Figure 4 The effects of absorbing boundaries at the blood vessels. The effects of absorbing boundaries at the blood vessels. A) Waves persistently re-emerge close to the boundary in response to a short, 30 minutes, pulse of stimuli when $p >= 105$. The group of cells neighboring the source is locked in "high T"- state as shown by green color. B) The frequency of re-initiation is higher if the source is located closer to the absorbing boundary, here $p = 83$.

does not resolve even after the damage has been repaired [2].

In Figure 4A it is apparent that the "locked" state is not uniform for all tissue cells. There are roughly two groups of cells: The cells neighboring the source are in a locked state, with T above the excitation threshold. Close to the blood vessels the cytokine concentration will always be low and the nearby cells are able to return to their resting state. At these cells waves can re-emerge, hence creating the situation as in Figure 4A. A similar partitioning of Dictyostelium D. cells into *sustained pacemakers* (the cites where waves originate) at the center of aggregation and *signal transducing cells* elsewhere has been first theoretically predicted by Geberth et al. [28] and later experimentally shown by Gregor et al. [29]. In the case of Dictyostelium D. the state with self-sustained and self-organized pacemakers (which corresponds to "chronic inflammation" state in our model) is desired and so the population works towards reaching high concentration of inducer, i.e. towards a locked state. In the case of NF-κB, the self-sustained pacemakers are undesirable. We expect this system to function at intermediate concentrations of inducer, i.e., a concentration that allows the cells near the cite of infection to be "inducible pacemakers". While these would oscillate in response to the external stimuli, the rest of the cells will propagate the signal (acute response). It is interesting to note that the apoptosis induced by sustained high concentrations of inducer might serve a mechanism that further limits the establishment of self-sustained pacemakers in inflammatory response.

If we consider cytokine waves as a signature for acute inflammatory response and sustained cytokine levels as characteristic of chronic inflammation, then Figure 4A predicts that the two coexist. That is, the chronic inflammation will be accompanied by the recurrent acute IR. Although somewhat contra-intuitive it is frequently observed that, acute and chronic inflammation coexist over long periods, implying continual reinitiation. Examples are found in rheumatoid arthritis, asthma, chronic obstructive pulmonary disease, multiple sclerosis, Crohn's disease, ulcerative colitis, and cancers [2].

Another observation, which is interesting from a physiological perspective is that the frequency of propagating waves can depend on the distance to the boundary for some values of p (Figure 4B). The frequency is higher at the nearby boundary, which suggests that recruitment of neutrophils will be more frequent from the closest blood vessels - a mechanism that can potentially contribute to a faster and a more localized IR.

The physiological significance of each of the above profiles might be determined by e.g. vascularization properties of the tissue as well as the severity of the infection. Thus if the source of infection is small a single wave might be enough to attract a sufficient amount of neutrophils, whereas a larger damage would benefit from repetitive and more frequent waves. In fact this tendency *is* experimentally observed during IR to myocardial injury [10]: In rodent models of myocardial infarction, the cytokines, IL-1, TNF and IL-6 are upregulated up to 50 fold within first hours. They can return to baseline levels if the infarction is small or, if the

infarction is large, there is either sustained cytokine upregulation or a second wave of cytokine upregulation [30,31].

Conclusions

The ability of sending information from one point in space to another is crucial for multicellular organisms, and biological systems have developed many different strategies that address this challenge.

In case of inflammation, the information about the insult - i.e. inflammatory cytokines - can be carried either *passively*, i.e. through diffusion or *actively* amplified by tissue cells. An active transmission of the inflammatory signal is supported by numerous experimental observations [10,32,33] (e.g., local RNA transcription and translation is required for efficient neutrophil emigration [32]).

We here show that the characteristics of the NF-κB regulatory network - fast spatially coupled positive feedback combined with slow negative feedback - are necessary for active propagation of the cytokine wave. Additionally, the characteristic short half-life of cytokines and the recently discovered threshold in NF-κB activation in single cells, are both conditions favoring emergence of the cytokine waves.

While there exists extensive literature on mathematical models addressing the role of the multiple negative feedbacks in NF-κB dynamics [17-19,27,34-36], the positive cytokine feedback has only been considered by Werner et al. [37]. To our knowledge, this is the first time that mathematical modeling addresses the role of cytokine positive feedback in the context of spatially distributed cells.

The cessation or "resolution" of the inflammatory response is as important as its initiation. While moderate and appropriately timed inflammatory response is beneficial - excessive, delayed or prolonged inflammation was shown to be a primary cause in many inflammatory diseases [2,38]. Furthermore, in cases such as tuberculosis, it is the host inflammatory response and not bacterial toxins that are responsible for the damage to the host. In this regard, both the transient nature of cytokine waves and the resulting transient neutrophil recruitment (also observed experimentally [16,33,39-41]) are the mechanisms that naturally minimize inflammatory tissue damage. Additionally, the presence of an obligate "refractory period" following the wave and lasting 5-12 hours will impose further constraints on the IR. Furthermore, our results in Figure 4 showing the persistent IR to transient damage predict that the chronic inflammation ("locked" region) will be accompanied by the re-current acute inflammation ("oscillatory region" close to the blood vessels). Remarkably, acute and chronic inflammation do coexist over long periods in such diseases as rheumatoid arthritis, asthma, multiple sclerosis etc. [2].

The accumulated experimental evidence together with our modeling results suggest that 1) NF-κB is a strong candidate for a mechanism generating "propagating waves" of chemoattractant cytokines. 2) The mechanism behind the propagating waves can have both beneficial and deleterious effects. While it assures reliable signal propagation and avoids long-lasting exposure to toxic cytokines, there are special conditions when the system over-reacts and generates situations which can be interpreted as inflammatory dysfunction as e.g. chronic inflammation.

Methods

The tissue cells are modeled as discrete units all containing NF-κB-like circuits consisting of the variables N, R and T, which influence each other as sketched in Figure 2E. The interactions are modeled by the following differential equations:

$$\frac{dN}{dt} = k_a \frac{T^3}{T^3 + 1}(1 - N) - k_{ai} R \tag{1}$$

$$\frac{dR}{dt} = k_b N - k_{bi} R \tag{2}$$

$$\frac{dT}{dt} = S + p\frac{N^2}{N^2 + K_N^2} - \frac{T}{\tau_T} \tag{3}$$

N is activated by T and inhibited by R. The $T \to N$ activation term is proportional to the amount of inactive NF-κB, (1 - N), and the Hill function, $\frac{T^3}{T^3+1}$, which ensures that activation only occurs when T exceeds a certain activation threshold. Here T is normalized relative to its activation threshold which is thus given by $T^* = 1$. The $R \dashv N$ inhibition term is proportional to R and is assumed to be saturated in N (equation 1). See additional file 1 for details on the choice of parameters and normalization of the variables. In equation (2) R is activated by N ($N \to R$ activation term is proportional to N) and decays with the half-life $1/k_{bi}$. Finally, in equation (3), T is activated by N and decays with the half-life τ_T . The $N \to T$ activation term is modeled with the Hill function, $\frac{N^2}{N^2+K_N^2}$, since NF-κB was reported to form dimers, but the results hold if it is replaced by a term linear in N.

The system is induced by the small stimulus, S, which represents a small influx of T. In-vivo this influx could correspond to cytokines secreted by macrophages or bacterial endotoxin. We have used $S = 5 \cdot 10^{-4} h^{-1}$ and this influx of T is only added at the site of infection (black dot in Figure 2F) and only at times $t \geq 0$. To

account for T diffusion between cells, we add a diffusion term to equation (3), which then becomes:

$$\frac{dT}{dt} = S + p\frac{N^2}{N^2 + K_N^2} - \frac{T}{\tau_T} + D\frac{\partial^2}{\partial x^2}T \quad (4)$$

The distance between the tissue cells is set to $\delta x \approx 15\mu m$ corresponding approximately to the cell size. We set $D = 2 \cdot 10^{-7}\frac{cm^2}{sec}$, a numerical value estimated for the diffusible factors of similar size [42] (we have also tested that our main results hold against several fold variation in D). The cytokine half-life in bloodstream has been experimentally measured to range between 3-15 minutes [25,26] and was estimated to be 25 minutes in [27]. As short cytokine half-life promotes wave propagation, we use a conservatively long τ_T = 25 min. The results are qualitatively unchanged in the entire range of τ_T = 3 - 25 min.

The parameters for the negative feedback were chosen to be such that the output qualitatively reproduces the transient dynamics of NF-κB, i.e. peak around 30 minutes and the response is decreased by 1 hour; $k_a = k_{ai} = k_b = 5h^{-1}$ and $k_{bi} = 0.5h^{-1}$. The parameter K_N = 0.3 is chosen relatively to the peak-hight of N which can maximally obtain the value N = 1 (See normalization in additional file 1). The Hill coefficient H = 2 used in the term $\frac{N^2}{N^2+K_N^2}$ is chosen because NF-κB is known to form dimers. The model however also works if we use a simple linear regulation. Thus the only free parameter remaining is the strength of positive feedback, p.

Generating the alternative scenarios, considered in Figure 1A and 1B

In the scenario seen in Figure 1A there is no regulation of T. The system is modeled by a single differential equation:

$$\frac{dT}{dt} = S - \frac{T}{\tau} + D\frac{\partial^2 T}{\partial x^2} \quad (5)$$

In the scenario seen in Figure 1B, T is amplified by N, but there is no transient response of N-corresponding to no R. The system is modeled by the two differential equations:

$$\frac{dN}{dt} = k_a\frac{T^3}{T^3 + 1}(1 - N) - k_{ai}N \quad (6)$$

$$\frac{dT}{dt} = S + p\frac{N^2}{N^2 + K_N^2} - \frac{T}{\tau} + D\frac{\partial^2 T}{\partial x^2} \quad (7)$$

All results are found using 4th-order Runge-Kutta integration. In Figures 1C, 2F, 3 and 4 we modeled a row of 400 cells and always add the stimulus S, to the middle cell.

Additional material

Additional file 1: Supplementary results. In this file we provide the details of the model derivation and its robustness to modifications. Please use Acrobat Reader to open this file.

Aknowledgements

We would like to thank Ian Dodd, Sebastian Bernardsson, Kim Sneppen, Namiko Mitarai and Philip Merksamer for the discussions and critical reading of the manuscript.

Authors' contributions

AT conceived the project; PY, BM, AT, MHJ and SK designed the research and wrote the manuscript; PY, BM and AT performed the research. All authors have read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 24 February 2011 Accepted: 19 July 2011
Published: 19 July 2011

References

1. Medzhitov R: **Inflammation 2010: new adventures of an old flame.** *Cell* 2010, **140(6)**:771-776[http://dx.doi.org/10.1016/j.cell.2010.03.006].
2. Nathan C, Ding A: **Nonresolving inflammation.** *Cell* 2010, **140(6)**:871-882 [http://dx.doi.org/10.1016/j.cell.2010.02.029].
3. Lechleiter J, Girard S, Peralta E, Clapham D: **Spiral calcium wave propagation and annihilation in Xenopus laevis oocytes.** *Science* 1991, **252(5002)**:123-126.
4. HODGKIN AL, HUXLEY AF: **Propagation of electrical signals along giant nerve fibers.** *Proc R Soc Lond B Biol Sci* 1952, **140(899)**:177-183.
5. Geiger J, Wessels D, Soll DR: **Human polymorphonuclear leukocytes respond to waves of chemoattractant, like Dictyostelium.** *Cell Motil Cytoskeleton* 2003, **56**:27-44[http://dx.doi.org/10.1002/cm.10133].
6. Loike JD, el Khoury J, Cao L, Richards CP, Rascoff H, Mandeville JT, Maxfield FR, Silverstein SC: **Fibrin regulates neutrophil migration in response to interleukin 8, leukotriene B4, tumor necrosis factor, and formyl-methionyl-leucyl-phenylalanine.** *J Exp Med* 1995, **181(5)**:1763-1772.
7. Rekdal O, Konopski Z, Svendsen JS, Winberg JO, Espevik T, Osterud B: **The TNF Receptors p55 and p75 Mediate Chemotaxis of PMN Induced by TNFalpha and a TNFalpha 36-62 Peptide.** *Mediators Inflamm* 1994, **3(5)**:347-352[http://dx.doi.org/10.1155/S0962935194000487].
8. Aleksic T, Baumann B, Wagner M, Adler G, Wirth T, Weber CK: **Cellular immune reaction in the pancreas is induced by constitutively active IkappaB kinase-2.** *Gut* 2007, **56(2)**:227-236[http://dx.doi.org/10.1136/gut.2005.084665].
9. Meron E: **Pattern formation in excitable media.** *Physics Reports* 1992, **218**:1-66[http://www.sciencedirect.com/science/article/B6TVP-46SXPNW-7W/2/abaed74d9adfb66d4e4e6177ba4354d2].
10. Nian M, Lee P, Khaper N, Liu P: **Inflammatory cytokines and postmyocardial infarction remodeling.** *Circ Res* 2004, **94(12)**:1543-1553 [http://dx.doi.org/10.1161/01.RES.0000130526.20854.fa].
11. Baud V, Karin M: **Signal transduction by tumor necrosis factor and its relatives.** *Trends in Cell Biology* 2001, **11(9)**:372-377[http://www.sciencedirect.com/science/article/B6TCX-43RK8XD-K/2/233bb0952eb0afb21c6a7db67cef4314].
12. Albrecht E, Petty HR: **Cellular memory: neutrophil orientation reverses during temporally decreasing chemoattractant concentrations.** *Proc Natl Acad Sci USA* 1998, **95(9)**:5039-5044.
13. Ebrahimzadeh PR, Högfors C, Braide M: **Neutrophil chemotaxis in moving gradients of fMLP.** *J Leukoc Biol* 2000, **67(5)**:651-661.
14. Vicker MG: **Ideal and non-ideal concentration gradient propagation in chemotaxis studies.** *Exp Cell Res* 1981, **136**:91-100.
15. Vicker MG, Lackie JM, Schill W: **Neutrophil leucocyte chemotaxis is not induced by a spatial gradient of chemoattractant.** *J Cell Sci* 1986, **84**:263-280.

16. Han SJ, Ko HM, Choi JH, Seo KH, Lee HS, Choi EK, Choi IW, Lee HK, Im SY: **Molecular mechanisms for lipopolysaccharide-induced biphasic activation of nuclear factor-kappa B (NF-kappa B).** *J Biol Chem* 2002, **277(47)**:44715-44721[http://dx.doi.org/10.1074/jbc.M202524200].
17. Hoffmann A, Levchenko A, Scott ML, Baltimore D: **The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation.** *Science* 2002, **298(5596)**:1241-1245[http://dx.doi.org/10.1126/science.1071914].
18. Ashall L, Horton CA, Nelson DE, Paszek P, Harper CV, Sillitoe K, Ryan S, Spiller DG, Unitt JF, Broomhead DS, Kell DB, Rand DA, Sée V, White MRH: **Pulsatile stimulation determines timing and specificity of NF-kappaB-dependent transcription.** *Science* 2009, **324(5924)**:242-246[http://dx.doi.org/10.1126/science.1164860].
19. Tay S, Hughey JJ, Lee TK, Lipniacki T, Quake SR, Covert MW: **Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing.** *Nature* 2010, **466(7303)**:267-271[http://dx.doi.org/10.1038/nature09145].
20. Hutti JE, Turk BE, Asara JM, Ma A, Cantley LC, Abbott DW: **IkappaB kinase beta phosphorylates the K63 deubiquitinase A20 to cause feedback inhibition of the NF-kappaB pathway.** *Mol Cell Biol* 2007, **27(21)**:7451-7461[http://dx.doi.org/10.1128/MCB.01101-07].
21. Enesa K, Zakkar M, Chaudhury H, Luong LA, Rawlinson L, Mason JC, Haskard DO, Dean JLE, Evans PC: **NF-kappaB suppression by the deubiquitinating enzyme Cezanne: a novel negative feedback loop in pro-inflammatory signaling.** *J Biol Chem* 2008, **283(11)**:7036-7045[http://dx.doi.org/10.1074/jbc.M708690200].
22. Turner DA, Paszek P, Woodcock DJ, Nelson DE, Horton CA, Wang Y, Spiller DG, Rand DA, White MRH, Harper CV: **Physiological levels of TNFalpha stimulation induce stochastic dynamics of NF-kappaB responses in single living cells.** *J Cell Sci* 2010, **123(Pt 16)**:2834-2843[http://dx.doi.org/10.1242/jcs.069641].
23. Krishna S, Semsey S, Jensen MH: **Frustrated bistability as a means to engineer oscillations in biological systems.** *Phys Biol* 2009, **6(3)**:036009[http://dx.doi.org/10.1088/1478-3975/6/3/036009].
24. Tsai TYC, Choi YS, Ma W, Pomerening JR, Tang C, Ferrell JE: **Robust, tunable biological oscillations from interlinked positive and negative feedback loops.** *Science* 2008, **321(5885)**:126-129[http://dx.doi.org/10.1126/science.1156951].
25. Castell JV, Geiger T, Gross V, Andus T, Walter E, Hirano T, Kishimoto T, Heinrich PC: **Plasma clearance, organ distribution and target cells of interleukin-6/hepatocyte-stimulating factor in the rat.** *Eur J Biochem* 1988, **177(2)**:357-361.
26. Beutler BA, Milsark IW, Cerami A: **Cachectin/tumor necrosis factor: production, distribution, and metabolic fate in vivo.** *J Immunol* 1985, **135(6)**:3972-3977.
27. Cheong R, Bergmann A, Werner SL, Regal J, Hoffmann A, Levchenko A: **Transient IkappaB kinase activity mediates temporal NF-kappaB dynamics in response to a wide range of tumor necrosis factor-alpha doses.** *J Biol Chem* 2006, **281(5)**:2945-2950[http://dx.doi.org/10.1074/jbc.M510085200].
28. Geberth D, Hütt MT: **Predicting the distribution of spiral waves from cell properties in a developmental-path model of Dictyostelium pattern formation.** *PLoS Comput Biol* 2009, **5(7)**:e1000422[http://dx.doi.org/10.1371/journal.pcbi.1000422].
29. Gregor T, Fujimoto K, Masaki N, Sawai S: **The onset of collective behavior in social amoebae.** *Science* 2010, **328(5981)**:1021-1025[http://dx.doi.org/10.1126/science.1183415].
30. Irwin MW, Mak S, Mann DL, Qu R, Penninger JM, Yan A, Dawood F, Wen WH, Shou Z, Liu P: **Tissue Expression and Immunolocalization of Tumor Necrosis Factor-alpha in Postinfarction Dysfunctional Myocardium.** *Circulation* 1999, **99(11)**:1492-1498[http://circ.ahajournals.org/cgi/content/abstract/99/11/1492].
31. Ono K, Matsumori A, Shioi T, Furukawa Y, Sasayama S: **Cytokine Gene Expression After Myocardial Infarction in Rat Hearts: Possible Implication in Left Ventricular Remodeling.** *Circulation* 1998, **98(2)**:149-156[http://circ.ahajournals.org/cgi/content/abstract/98/2/149].
32. Cybulsky MI, McComb DJ, Movat HZ: **Protein synthesis dependent and independent mechanisms of neutrophil emigration. Different mechanisms of inflammation in rabbits induced by interleukin-1, tumor necrosis factor alpha or endotoxin versus leukocyte chemoattractants.** *Am J Pathol* 1989, **135**:227-237.
33. Movat HZ, Cybulsky MI, Colditz IG: **Emigration and accumulation of PMN-leukocytes induced by endotoxin, interleukin 1 and other chemotactic substances.** *Folia Histochem Cytobiol* 1986, **24(2)**:75-88.
34. Krishna S, Jensen MH, Sneppen K: **Minimal model of spiky oscillations in NF-kappaB signaling.** *Proc Natl Acad Sci USA* 2006, **103(29)**:10840-10845[http://dx.doi.org/10.1073/pnas.0604085103].
35. Mengel B, Krishna S, Jensen MH, Trusina A: **Theoretical analyses predict A20 regulates period of NF-kB oscillation.** *arxive[physics.bio-ph]* 2009, **0911.0529.**
36. Paszek P, Ryan S, Ashall L, Sillitoe K, Harper CV, Spiller DG, Rand DA, White MRH: **Population robustness arising from cellular heterogeneity.** *Proc Natl Acad Sci USA* 2010, **107(25)**:11644-11649[http://dx.doi.org/10.1073/pnas.0913798107].
37. Werner SL, Barken D, Hoffmann A: **Stimulus specificity of gene expression programs determined by temporal control of IKK activity.** *Science* 2005, **309(5742)**:1857-1861[http://dx.doi.org/10.1126/science.1113319].
38. Iwasaki A, Medzhitov R: **Regulation of adaptive immunity by the innate immune system.** *Science* 2010, **327(5963)**:291-295[http://dx.doi.org/10.1126/science.1183021].
39. Cybulsky MI, Cybulsky IJ, Movat HZ: **Neutropenic responses to intradermal injections of Escherichia coli. Effects on the kinetics of polymorphonuclear leukocyte emigration.** *Am J Pathol* 1986, **124**:1-9.
40. Colditz IG, Movat HZ: **Kinetics of neutrophil accumulation in acute inflammatory lesions induced by chemotaxins and chemotaxinigens.** *J Immunol* 1984, **133(4)**:2169-2173.
41. Colditz IG: **Two patterns of early neutrophil accumulation in acute inflammatory lesions.** *Inflammation* 1988, **12(3)**:251-263.
42. Francis K, Palsson BO: **Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion.** *Proc Natl Acad Sci USA* 1997, **94(23)**:12258-12262.

doi:10.1186/1752-0509-5-115
Cite this article as: Yde *et al.*: Modeling the NF-κB mediated inflammatory response predicts cytokine waves in tissue. *BMC Systems Biology* 2011 **5**:115.

PHYSICAL REVIEW E **84**, 051913 (2011)

Analyzing inflammatory response as excitable media

Pernille Yde, Mogens Høgh Jensen, and Ala Trusina
Niels Bohr Institute, University of Copenhagen, Denmark
(Received 8 August 2011; published 18 November 2011)

The regulatory system of the transcription factor NF-κB plays a great role in many cell functions, including inflammatory response. Interestingly, the NF-κB system is known to up-regulate production of its own triggering signal—namely, inflammatory cytokines such as TNF, IL-1, and IL-6. In this paper we investigate a previously presented model of the NF-κB, which includes both spatial effects and the positive feedback from cytokines. The model exhibits the properties of an excitable medium and has the ability to propagate waves of high cytokine concentration. These waves represent an optimal way of sending an inflammatory signal through the tissue as they create a chemotactic signal able to recruit neutrophils to the site of infection. The simple model displays three qualitatively different states; low stimuli leads to no or very little response. Intermediate stimuli leads to reoccurring waves of high cytokine concentration. Finally, high stimuli leads to a sustained high cytokine concentration, a scenario which is toxic for the tissue cells and corresponds to chronic inflammation. Due to the few variables of the simple model, we are able to perform a phase-space analysis leading to a detailed understanding of the functional form of the model and its limitations. The spatial effects of the model contribute to the robustness of the cytokine wave formation and propagation.

DOI: 10.1103/PhysRevE.84.051913 PACS number(s): 87.18.Gh, 82.40.Ck, 05.45.−a

I. INTRODUCTION

Excitable media are naturally encountered in many biological systems. A typical excitable medium behaves in a manner much resembling spectators making a wave of raised hands at a sports game. The *excitable units* (or sections) get stimulated by their neighbors and amplify the exciting stimuli. At this stage the units are said to be in an excited state. Subsequent to excitation there is a recovery period in which new excitation is not possible, referred to as the refractory period. As a result of this behavior, spatially coupled excitable units are able to propagate undamped waves of high stimuli concentration through the system.

Some biological species have evolved to utilize the undamped waves that excitable media produce as a means of sending information through the system. Two well-known examples of biological excitable media are the neuron [1,2], which is able to propagate action potentials down the axon, and colonies of the social amoeba *Dictyostelium discoideum* [3,4], which propagate spiral waves of cyclic adenosine monophosphate (cAMP) and accordingly perform self-organized directed migration toward a common center. Both systems share the need for sending information through relatively large distances, where simple processes, such as, for example, diffusion, would not be adequate.

As recently shown by the authors, the regulatory system of nuclear factor κB (NF-κB) also contains the necessary components in order to exhibit "excitability," i.e., behave as an excitable medium [5].

NF-κB is present in all mammalian cells and is known to play an important role during inflammatory response [6–8]. The NF-κB system is triggered by inflammatory cytokines and in turn amplifies the cytokine signal, thus creating an excited state in which cytokine production is high. But because NF-κB also triggers production of its own inhibitors, the excited state will not last: eventually inhibitor concentration will become abundant and bind all NF-κB, making it inactive and hence cytokine production ceases. As long as inhibitors are plentiful, new activation of NF-κB cannot result in an excitation comparable to the initial one, although NF-κB has been shown to exhibit secondary small-amplitude peaks [6,7]. Thus the state with high inhibitor concentration constitutes a refractory period.

As a result of this behavior tissue cells containing NF-κB regulatory systems should theoretically be able to propagate traveling waves of high cytokine concentration through the tissue. Since cytokines also function as a neutrophil chemoattractant, this scenario is in good agreement with the current belief that neutrophils chemotax in a similar fashion as *Dictyostelium d.*, namely, through waves of chemoattractant.

As recently shown, a simple model of spatially coupled NF-κB units (cells) naturally leads to the propagation of cytokine waves in the tissue [5]. The model is a simplification of the real NF-κB system and provides a useful tool for investigating and understanding the underlying mechanisms of the complex regulatory system. In this paper we present and analyze the model in greater detail and obtain a better understanding of the many mechanisms that the simple model captures. The findings of this paper can hence contribute to the general understanding of inflammatory response—in particular, how different components of the immune system may send and transmit information through the organism. In addition, these findings contribute to the understanding of neutrophil recruitment during inflammatory response.

II. MODEL

In order to create an excitable medium it is important that the excitable unit responds with a transient amplification of the stimuli (opposed to persistent amplification). This means that the excitable unit must be an adapter in the sense that the system must adapt to the new surroundings after a transient phase. It is experimentally observed that the NF-κB system responds with a pronounced initial peak in nuclear NF-κB and thereafter, secondary oscillation of much smaller amplitude [6,7]. The damped oscillatory behavior arises due to several

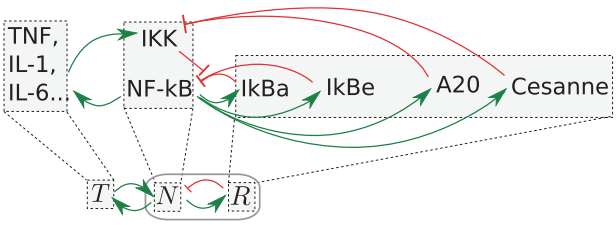


FIG. 1. (Color online) The NF-κB regulatory system is simplified as sketched here. Cytokines such as TNF, IL-1, and IL-6 activate the NF-κB system through the IKK pathway. The cytokines are simplified as a single variable denoted T. Active NF-κB is highly correlated to active IKK, and these two variables are thus also simulated as a single variable denoted N. Since cytokines activate IKK (and hence NF-κB) and NF-κB in return up-regulates production of cytokines, there is a positive feedback between the variables T and N. Inhibitors ($\mathrm{IkB}_{\alpha,\beta,\epsilon}$) and upstream regulators (A20, cesanne) all function to perform a negative feedback on either IKK or NF-κB and are hence simulated as the single regulating variable R, which performs a negative feedback on N. Activating interactions are sketched with $\rightarrow$ and inhibiting interactions are sketched with $\dashv$.

inhibitors performing negative feedback, but for our purpose it is sufficient to note that the secondary behavior is of much smaller amplitude than the initial peak, and hence the NF-κB system *is* an adapter.

In order to analyze the system we have constructed a simple model which captures the overall behavior of the NF-κB system. (We have verified our results by also simulating the system in greater detail, including several inhibitors and upstream regulators, and confirm that the qualitative behavior is also exhibited for this more sophisticated model.) The NF-κB system is simplified as sketched in Fig. 1. Cytokines such as tumor necrosis factor (TNF), interleukin-1 (IL-1), and interleukin-6 (IL-6) stimulate the NF-κB system though the IκB kinase (IKK). The cytokines are simulated by a single variable which we denote T. When IKK is activated inhibitors are degraded and NF-κB is released, translocating into the nucleus where it is active. Thus the concentrations of IKK and NF-κB follow each other and can be simulated by one variable, which we denote N. The inhibitors ($\mathrm{IkB}_{\alpha,\beta,\epsilon}$) and upstream regulators (A20, cesanne) all cause either IKK or NF-κB concentration to go down. These inhibitors and regulators are simulated by a "regulator" variable which we denote R.

The effect of inhibitors and other regulators is to perform a negative feedback on NF-κB, and is modeled by a simple negative feedback loop (see interactions between N and R in Fig. 1). These interactions can be described by the equations

$$\frac{dN}{dt} = k_{\mathrm{activate}} f(T) \times (N_T - N) - k_{\mathrm{inhibit}} R \tag{1}$$

$$\frac{dR}{dt} = r_{\mathrm{on}} N - r_{\mathrm{off}} R \tag{2}$$

The activation of N corresponds to translocating NF-κB into the nucleus. This term is proportional to some function of the cytokine concentration $f(T)$ and the amount of cytoplasmic NF-κB (we assume the total amount of NF-κB (N_T) is constant. Thus the amount of NF-κB which is available for activation is given by the amount of cytoplasmic NF-κB; $N_C = N_T - N$) [9].

Inhibition of NF-κB is proportional to the amount of inhibitors R and is considered to be saturated in N. (Simulations have shown that this approximation does not introduce an error of noticeable size.)

The activation of R is proportional to N and inactivation of R is modeled as a spontaneous degradation, only proportional to R. In order for this simple model to function as an adapter, it is important that the rate constant r_{off} is slow compared with the other rate constants of the system [5,10].

When the NF-κB network is stimulated by cytokines it responds by up-regulating hundreds of genes, including those coding for cytokine production. The newly synthesized cytokines are secreted into the extracellular matrix, where they can again stimulate the IKK pathway. Thus the interaction between NF-κB and cytokines constitutes a positive feedback (see interactions between N and T in Fig. 1).

The local concentration of cytokines (T) is modeled by the equation

$$\frac{dT}{dt} = p\frac{N^2}{N^2 + K^2} - \frac{T}{\tau} + S. \tag{3}$$

NF-κB–induced production of cytokines is proportional to the rate constant p (for positive feedback) and to the Hill function $N^2/(N^2 + K^2)$, because NF-κB is a dimeric transcription factor. As we shall see below, this term could also be modeled as a simple linear response (pN) and still give similar results. The cytokine degradation is modeled by a simple linear decay with a typical lifetime τ. The term S represents an additional cytokine production functioning as an external stimuli: during inflammatory response cytokines are secreted from nearby macrophages, which would correspond to a small flux of cytokines. This flux is "turned on" at time $t = 0$ and is modeled by a step function

$$S = \begin{cases} 0 & \text{for} \quad t < 0 \\ S_{\mathrm{on}} & \text{for} \quad t > 0, \end{cases} \tag{4}$$

where $S = 0$ corresponds to no stimuli. In the case of spatially coupled cells only the cytokines are secreted into extracellular space, and hence only the variable T is allowed to diffuse in between cells. In this case the equation describing cytokine concentration (at the ith cell) is given by

$$\frac{\partial T_i}{\partial t} = p\frac{N_i^2}{N_i^2 + K^2} - \frac{T_i}{\tau} + S_i + D\frac{\partial^2 T_i}{\partial x^2}, \tag{5}$$

with the only difference being the addition of the diffusion term.

For the system to react as an excitable medium the activation of the excitable units must be strongly thresholded. This threshold is in accordance with recent experimental findings [11,12]. We implement this by modeling the activation of NF-κB with a sigmoidal response to T (and Hill coefficient = 3):

$$f(T) = \frac{T^3}{T^3 + K_A^3}. \tag{6}$$

The variables have been renormalized in the following way: $N \rightarrow N/N_T$ and $T \rightarrow T/K_A$, which is equal to putting the parameters N_T and K_A [Eqs. (1) and (6)] equal to unity (and redefining the remaining parameters [5]). This also means that

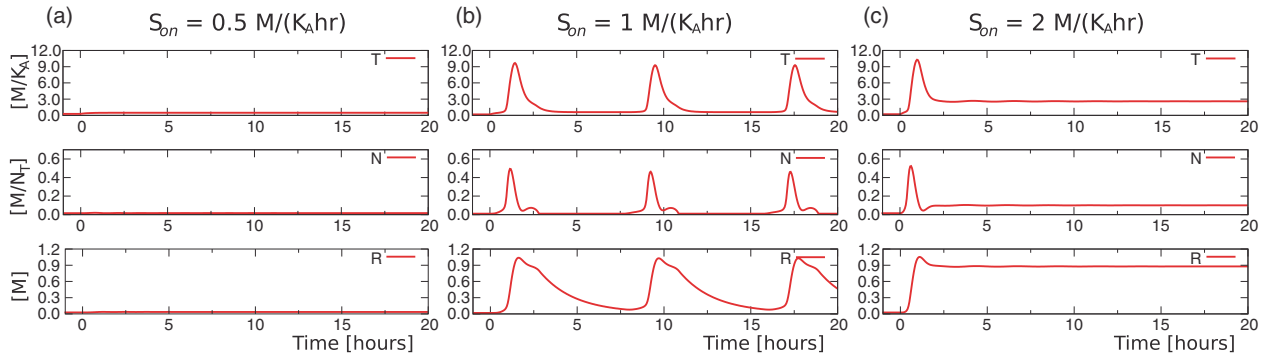


FIG. 2. (Color online) Simulation of reaction to stimulus using Eqs. (1), (2), and (3). The stimulus is turned on at time $t = 0$ [see Eqs. (3) and (4)]. Top panels: cytokine concentration (T). Middle panels: Active NF-κB concentration (N). Bottom panels: regulator concentration (R) (representing the combined effect of all inhibitors). The unit $[M]$ stands for molar concentration. K_A and N_T are normalization constants of T and N, respectively. (a) The weak stimulus ($S = 0.5\ M/(K_A\ \mathrm{hr})$) causes T to increase a little but not enough to activate N. The system comes to rest in a new steady state with low concentrations of all three variables. (b) The intermediate stimulus ($S = 1\ M/(K_A\ \mathrm{hr})$) causes the system to oscillate. The increase in T exceeds the triggering threshold for activating N and consequently, all variables rise to high levels. The high level of R inhibits N, which decreases back to almost prestimulation levels after approximately 2 hours. At low N level T and R will begin to decrease; R decreases slowly because of the slow degradation rate r_{off} [see Eq. (2)]. After approximately 9 hours R has decreased back to prestimulation levels and the system spikes again. (c) At high stimulus ($S = 2\ M/(K_A\ \mathrm{hr})$) the system will not settle back to prestimulation levels, because the inhibition from R is not enough to drive N back down, once the positive feedback is present. As a result, the system comes to rest in a new steady state in which both N and T levels are much higher than triggering levels. R is sustained at a high level, creating an infinite refractory period.

the cytokine triggering threshold for activating N is reached when T exceeds $T^* \approx 1$.

The parameter K [Eqs. (3) and (5)] describes the NF-κB positive-feedback threshold for internal transcription of cytokines. To achieve maximal sensitivity to N this parameter was chosen to match approximately half-maximum of the initial N peak, which gave $K = 0.3$. (N reaches a maximum of $\approx$0.6 in our simulation.)

The rate constants k_{activate}, k_{inhibit}, r_{on}, and r_{off} have been fitted to match the typical time scale of the NF-κB initial peak ($k_{\mathrm{activate}} = k_{\mathrm{inhibit}} = r_{\mathrm{on}} = 5.0\ \mathrm{hr}^{-1}$ and $r_{\mathrm{off}} = 0.5\ \mathrm{hr}^{-1}$). The lifetime and diffusion constant of TNF have previously been estimated [13] and are used here as the cytokine lifetime, $\tau = 25$ minutes, and diffusion constant, $D = 2 \times 10^{-7}\,\frac{\mathrm{cm}^2}{\mathrm{min}}$. Thus the only free parameter of our model is the parameter p. This parameter sets the strength of the positive feedback, and as we shall see in the Results section, this parameter can be varied to be both too small, not obtaining an adequate feedback, or too large, making the system incapable of returning to resting state.

III. RESULTS: TEMPORAL BEHAVIOR OF A SINGLE CELL

The system described by Eqs. (1), (2), and (3) is simulated starting from an initial steady state where all concentrations are low and there is no stimuli ($S = 0$). At time $t = 0$ the system is stimulated by "turning on" the small cytokine flux ($S = S_{\mathrm{on}}$). Had there been no interaction with NF-κB, T would increase to a steady-state level given by a balance between S_{on} and τ [see Eq. (3)]. But if the stimulation S_{on} is strong enough (roughly speaking, if T exceeds the threshold $T^* \approx 1$), the system will respond with an up-regulation of N, which in turn amplifies T to values manyfold larger than the initial stimulation. Depending on the value of S_{on}, three qualitatively different scenarios can be achieved: if S_{on} is too small the increase in T will not activate N [Fig. 2(a)]. If S_{on}, on the other hand, is large enough to make T exceed the triggering threshold, $T^* \approx 1$, N will increase and cause T to increase further [Figs. 2(b) and 2(c)]. As a result N will also increase to a high level and consequently activate production of its own inhibitors: R begins to increase. As R peaks the negative feedback causes N to decrease and settle back to lower values. If S_{on} is large, [Fig. 2(c)] a new steady state will be obtained in which R is high and both N and T are balanced at levels significantly higher than prestimulation values [Fig. 2(c)]. Interestingly, intermediate values of S_{on} [Fig. 2(b)] will lead to situations where N and T settle back to prestimulation values when R is high. Because N decreases to such low values, the inhibitor R will also start to decrease, although this is a slow process because of the slow degradation rate r_{off} [see Eq. (2)]. When R decreases sufficiently N is no longer inhibited and after some time N and T can peak again [Fig. 2(b)].

A. Phase-plane analysis of the system

The intermediate S_{on}, leading to oscillatory behavior, is of course a very interesting situation. The system has many things in common with classical excitable media, such as, e.g., the Belousov-Zhabotinsky reaction, and we follow an approach similar to the one described in the review by Meron [14].

Notice that N and T are fast variables whereas R is a rather slow variable. Thus the model contains two effective time scales and we can assume that N and T will effectively reach steady state and follow changes in R adiabatically. In order to understand the system in greater detail, we plot the nullclines of N and T for fixed values of R. The nullclines are plotted in N-T space (see Fig. 3). Before stimulation ($t < 0$) the

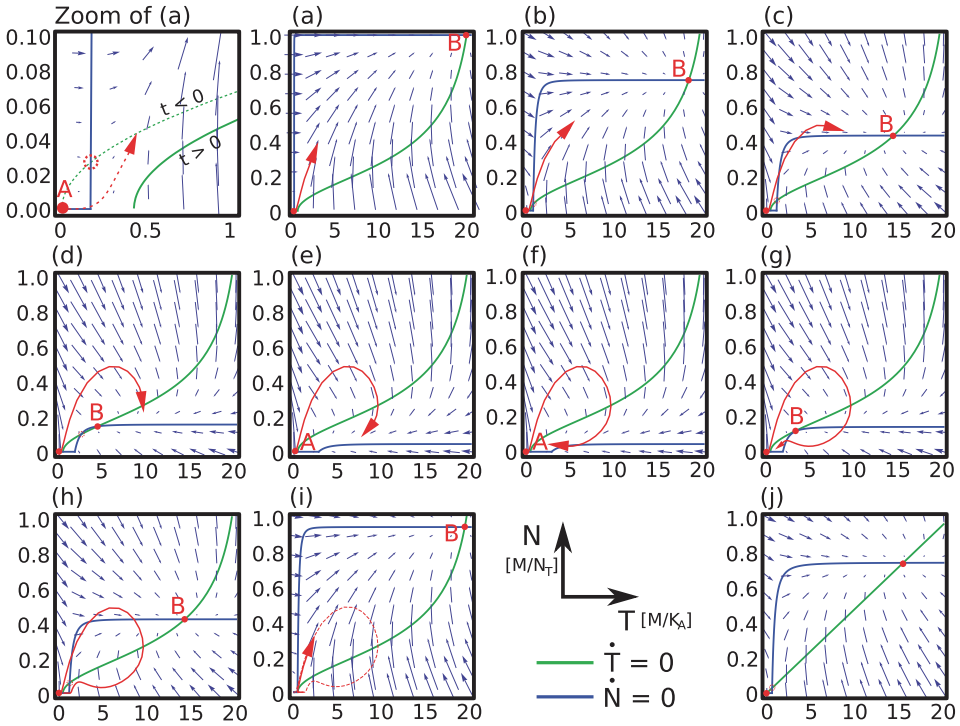


FIG. 3. (Color) The situation shown in Fig. 2(b) is here shown in the phase plane of N and T. Nullclines are plotted in blue ($dN/dt = 0$) and green ($dT/dt = 0$) lines. Stable fixed points are indicated with solid red dots. Unstable fixed points are indicated with dashed red circles. Initially ($t < 0$) the system has three fixed points located at the intersections of the nullclines. Two of these fixed points have low N and T values and are shown in the zoom of panel (a) (first panel). The system is at rest in the low stable fixed point, which we refer to as fixed point A. At time $t = 0$ the stimulus S is turned on [see Eqs. (3) and (4)], causing the T nullcline to shift to the right as shown in the zoom of panel (a) (first panel). Consequently, fixed point A and the unstable fixed point disappear in a saddle-node bifurcation and the system starts to evolve toward the fixed point with high N and T, which we refer to as fixed point B [see panel (a)]. As N increases, R will also increase [see Eq. (2)], causing the N nullcline to move as shown in panels (b)-(e). The system will dynamically change and always evolve toward the stable fixed point, eventually causing N and T to decrease [panels (d)-(f)]. At some point R becomes so large (the N nullcline has moved so far) that fixed point B disappears in a saddle-node bifurcation [panels (d) and (e)], and the system will now evolve toward fixed point A which has been re-established [since panel (b)]. At this point N has decreased back to a relatively low level and R will consequently begin to decrease, causing the N nullcline to move back [panels (f)–(h)]. Meanwhile, the system is caught in the basin of attraction of fixed point A [see panel (g)] and will move toward this fixed point [panel (h)]. Eventually, R has decreased sufficiently and the N nullcline has moved such that fixed point A disappears again and the system begins a new round in phase space [panel (i)]. The times corresponding to the panels are: (a) $t = 0.0$ to $t = 1.0$, (b) $t = 1.1$, (c) $t = 1.2$, (d) $t = 1.4$, (e) $t = 1.6$, (f) $t = 2.0$, (g) $t = 2.7$, (h) $t = 3.7$, and (i) $t = 9.1$ hours. Panel (j) shows the nullclines as they would look if cytokine production (up-regulation of T) had been modeled with a simple linear term pN instead of the sigmoidal term $(N^2/(N^2 + K^2))$ used in Eq. (3).

nullclines intersect in three distinct fixed points—two stable fixed points separated by an unstable fixed point in between. We refer to the two stable fixed points as fixed point A and fixed point B. For $t < 0$ fixed point A and the unstable fixed point lie very close to each other in N-T space, and both have relatively low levels of N and T. [See intersections of dotted green line and blue line in the first panel of Fig. 3 (zoom of panel (a))].

When S is shifted from $S = 0$ to $S = S_{on}$, the T nullcline is shifted to the right by an amount $\delta = \Delta S\tau$. Hence, if S_{on} is large enough, fixed point A and the unstable fixed point will disappear in a saddle-node bifurcation, and the only fixed point of the system is now fixed point B [Fig. 3(a)]. As the system begins to evolve toward fixed point B, N increases and causes R to increase correspondingly. As this happens the N nullcline will begin to move, dynamically changing the phase space as shown in Figs. 3(a)–3(c). The system will continuously evolve toward fixed point B as it moves "down" [Figs. 3(a)–3(c)], eventually making N and T decrease [Fig. 3(d)]. While the N nullcline moves, fixed point A and the unstable fixed point have re-established in a new saddle-node bifurcation [since Fig. 3(b)]. Eventually R will increase to such high values that fixed point B coalesces with the unstable fixed point and disappears in a second saddle-node bifurcation [Figs. 3(d) and 3(e)]. Now the system will evolve toward fixed point A, causing N and T to decrease back to almost prestimulation values [Figs. 3(e) and 3(f)]. As N is no longer high, R will no longer be up-regulated and will begin to decrease because of spontaneous degradation. This will cause the N nullcline to move "back" [as shown in Figs. 3(f)-3(h)], although as mentioned above this is a slow process (because of slow r_{off}). As the N nullcline moves, fixed point B and

the unstable fixed point are re-established [Fig. 3(g)], but now the system is caught in the basin of attraction of fixed point A [Fig. 3(g)]. As R slowly decreases, the system rests in fixed point A [Fig. 3(h)]. Eventually, the N nullcline has moved such that fixed point A and the unstable fixed point once again disappear in a saddle-node bifurcation, and the system will once again make a round in the phase space [Fig. 3(i)].

The three qualitatively different scenarios of Fig. 2 can be well understood from an investigation of the phase space. In order to exhibit oscillations the system must be able to undergo the two saddle-node bifurcations described above: first, fixed point A and the unstable fixed point coalesce, and second, fixed point B and the unstable fixed point coalesce. The value of S_{on} sets the size of the T-nullcline shift, $\delta = \Delta S\tau$ (recall Fig. 3, first panel). A too-small S_{on} will not cause the first bifurcation because the T nullcline is not shifted far enough. A too-high S_{on} will inhibit the system from undergoing the second bifurcation because the shift is too large and the system will come to rest in fixed point B.

As mentioned in the Model section, we could also choose to model the positive feedback from NF-κB on cytokine production as a simple linear response, pN, instead of the sigmoidal response, $N^2/(N^2 + K^2)$, which is only valid if NF-κB is truly a dimeric transcription factor [see Eq. (1)]. In this case the T nullcline would be a straight line and the N nullcline would remain unchanged. We plot this situation in Fig. 3(j), from which it can be inferred that such a simplification of the model would lead to similar results. From this plot we conclude that at least one of the nullclines must have a sigmoidal form in order to obtain a bistable system. This means that a smaller Hill coefficient, $H = 2$, would suffice in Eq. (6). Hence a minimal model could be obtained by modeling N and T dynamics by the equations

$$\frac{dN}{dt} = k_{\text{activate}} \frac{T^2}{T^2 + K_A^2} \times (N_T - N) - k_{\text{inhibit}} R$$
$$\frac{dT}{dt} = pN - \frac{T}{\tau} + S.$$

Compare with Eqs. (1), (3), and (6).

B. The excitability of the system depends on the strength of positive feedback p

The effect of the positive feedback can be understood by investigation of the nullclines upon variation of p [see Eq. (3)]. The slope of the T nullcline is roughly set by p [see dashed green lines in Fig. 4(a)]. Qualitatively there are three distinct behaviors with weak, intermediate, and strong feedback being similar to the three states with weak, intermediate, and strong stimuli in Fig. 2.

If p is small ($p \approx 10$) the slope of the T nullcline is very steep and hence fixed point B will have a small T value. The system cannot get excited as even a small increase in R will move fixed point B down to low N and T values and the system will have only a very small round in the phase space before reaching this fixed point. The system comes to rest in fixed point B, because R will never become large enough to cause the second saddle-node bifurcation. The resulting situation is very similar to the one in Fig. 2(a).

On the other hand, a strong positive feedback ($p > \approx 100$) allows for a single excitation followed by an infinite refractory period. Large p makes the slope of the T nullcline flatter [Fig. 4(a)]. Right after the stimulus is induced the system follows a long trajectory in the phase space, resulting in a spike in N and T. However, the system comes to rest in fixed point B because the maximal R value is not high enough to move the N nullcline sufficiently far down for fixed point B to disappear in a saddle-node bifurcation. In this case fixed point B has significantly higher N and T levels, meaning that the cytokine concentration is sustained high above the triggering level. The relatively high N level causes R to be sustained at a high level, hence creating an infinite refractory period. This situation will be very similar to the one shown in Fig. 2(c). We refer to this situation as a locked state because the nullclines are locked in fixed point B, even when the stimulus is removed. In the picture of inflammatory response the locked situation would correspond to *chronic inflammation*.

The nullclines of the system can of course also be altered by other parameters of the model, and in order to explore changes in cytokine production we have varied the parameter τ which determines the typical lifetime of the cytokines before they are degraded (the inverse degradation rate). Hence a high

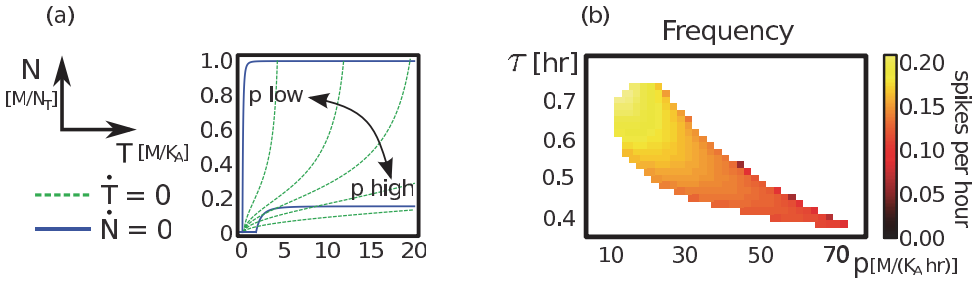


FIG. 4. (Color online) (a) The slope of the T nullcline (dashed green line) becomes steeper as p decreases and flatter when p increases. The N nullcline (solid blue line) is shown for two different values of R and will move from the high plateau to the low plateau as R increases (recall Fig. 3). In the case of small p the N nullcline will not need to move very far before fixed point B (recall Fig. 3) has moved to relatively low levels of N and T, hence creating a situation as shown in Fig. 2(a), where the system comes to rest in fixed point B. In the case of high p the system will also come to rest in fixed point B, which in this case is created at high T levels. R will never become large enough to make fixed point B disappear in a bifurcation and the system is locked in fixed point B. (b) Combinations of the parameters p and τ which lead to oscillatory behavior. The color of the graph indicates the frequency of the oscillations.

p (a high production rate of cytokines) should be counteracted by a low τ in order to keep the cytokine concentration balanced such that it can repeatedly transcend the triggering threshold at $T^* \approx 1$, corresponding to repetitive rounds in phase space as shown in Fig. 3. In other words, the nullclines must lie such that they are able to undergo saddle-node bifurcations both at fixed point A and at fixed point B. Whereas p sets the slope of the T nullcline, τ sets the size of the shift to the right when the stimulus S is introduced. The frequency at which the system can spike depends on how fast the system will undergo the two bifurcations. In Fig. 4(b) we show a plot of the combinations of p and τ which lead to self-oscillatory situations together with their spiking frequencies.

IV. RESULTS: SPATIOTEMPORAL MODEL OF THE TISSUE LEADS TO PROPAGATING WAVES

When the cells are coupled in space and cytokines are allowed to diffuse between them, waves of high cytokine concentration arise [see Fig. 5(a)]. We have constructed a spatial model consisting of a one-dimensional lattice of cells. Every cell is able to regulate cytokine production as described in Eqs. (1), (2), and (5), and only the variable T is allowed to diffuse between cells. We use open boundaries representing the bloodstream in which the cytokines (T) will be absorbed. During inflammatory response only cells at the site of infection would be subject to the external stimulus S [see Eq. (5)], and we simulate this by adding the external stimulus S only to the central cell of the one-dimensional lattice; $S_i = S \times \delta(i,0)$. Adding the diffusion term [see Eq. (5)] causes the effective removal of cytokines to become larger, and in order to counteract this we have increased S tenfold compared to the above ($S_{\text{on}} = 10\ \text{hr}^{-1}$).

At time $t = 0$ the central cell is stimulated and starts to amplify the cytokine concentration. The cytokines will diffuse to neighboring cells which consequently also get stimulated, and thus a wave is created. We stress that the second (and later) waves arise because of the oscillatory behavior of the central cell which will initialize new waves that can propagate through the system. The cells which do not feel the external stimulus S will only get stimulated when they feel a spillover of cytokines from their neighbors. Hence the situation is indeed cooperative in the sense that the cytokine wave is truly propagated from one cell to the next; the cells are not oscillating individually. If the external stimulus S is removed from the central cell, no new waves will be initialized and the system will settle back to rest as soon as the last wave has reached the absorbing boundary.

A. Space contributes to the robustness of the model

An interesting observation is that the spatial model seems more robust toward creating repetitive waves. In Fig. 5(d) we plot the combinations of p and τ which lead to propagating

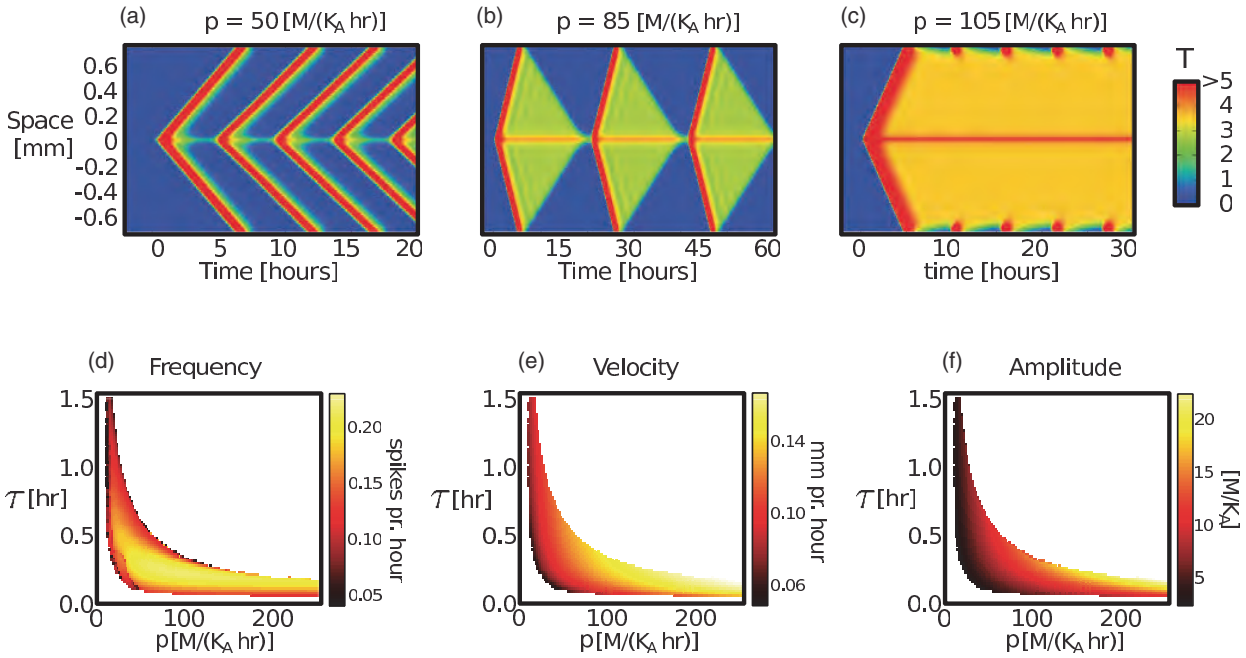


FIG. 5. (Color) (a)–(c) Space-time plots of the cytokine (T) concentration for three different values of the parameter p, which describes the strength of the positive feedback between N and T [see Eqs. (1), (2), and (5)]. (a) The central cell is stimulated at time $t = 0$ and initializes waves of high cytokine concentration which are propagated through the spatial system. (b) At higher values of p the positive feedback is so strong that the system becomes flooded with cytokines (the cells are in the locked state described in the text). But diffusion effects from the absorbing boundaries enable the system to resettle to prestimulation values and new waves can propagate. (c) If p is very large only the cells close to the boundary will be able to escape the locked state. Here oscillations will arise even though p is very high. Notice the different time scales. (d)–(f) Combinations of p and τ which lead to repetitive waves in the spatial model. We plot frequency (d), velocity (e), and amplitude (f).

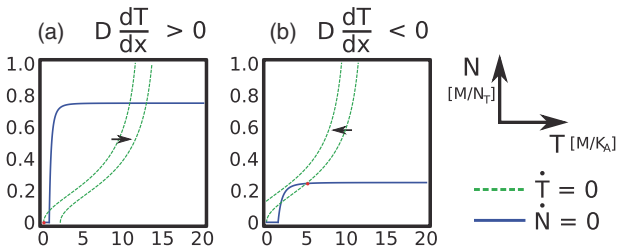


FIG. 6. (Color online) Diffusion effects will shift the T nullcline (dashed green line) horizontally and can be both positive (corresponding to a shift to the right) and negative (corresponding to a shift to the left). (a) The N nullcline (blue line) is plotted for a relatively low R level. If the diffusion term is positive, it can cause the system to bifurcate such that fixed point A disappears. This stimulates the system to move around in phase space as shown in Fig. 3. (b) The N nullcline (blue line) is plotted for a relatively high R level. If the diffusion term is negative, it can cause the system to bifurcate such that fixed point B disappears and the system is unlocked from the locked state. This effect contributes to the ability to bifurcate at both fixed points and makes the spatial model more prone to exhibit repetitive waves than the single isolated cell.

waves. As can be seen from the plot, there are far more p-τ-combinations that lead to repetitive waves than in the case of a single isolated cell [recall Fig. 4(b)]. Figures 5(d)–5(f) also display how typical wave characteristics such as frequency, velocity, and amplitude change with p and τ. Velocity and amplitude of the waves grow with increasing p (and decreasing τ), which leads to strong and fast cytokine production. On the other hand, the frequency is highest where p and τ are correctly balanced, in order to be able to undergo the saddle-node bifurcations, described above, as fast as possible.

The reason why the spatial model is more robust can be found in the effects of diffusion. In the nullcline picture, the diffusion term [see Eq. (5)] acts to shift the T nullcline horizontally (see Fig. 6). As opposed to the external stimulus S, which also shifts the T nullcline horizontally, the diffusion term can become both positive and negative. A positive diffusion term corresponds to cytokines diffusing in from the neighbors, leading to an increased positive flux of cytokines and hence a shift of the T nullcline to the right [see Fig. 6(a)]. In this situation the diffusion terms acts as a stimulus just like S, but a stimulus which travels through space and stimulates the cells one by one, creating a wave. On the other hand a negative diffusion term, meaning that cytokines diffuse away, leads to a shift of the T nullcline to the left. The spacial organization increases the chance that somewhere between the source and the absorbing boundary there will be a cell where the positive and negative diffusion terms balance such that cells can undergo saddle-node bifurcations at both fixed points. If, for example, p is high, a group of cells near the center become locked in fixed point B (locked state). For cells further away from the source a large negative diffusion term will shift the T nullcline to the left [see Fig. 6(b)]. In this situation the diffusion term unlocks the system so that it will again be able to undergo the bifurcation; hence the diffusion term expands the parameter space that can undergo both bifurcations and hence create waves.

Of course, some combinations of the parameters p and τ will lead to situations where most cells in the system cannot undergo bifurcations because diffusion is not strong enough. This can lead to situations where almost all cells become locked in fixed point B [see Fig. 5(c)]. In this situation we still observe oscillations but only close to the boundaries. These oscillations arise because the diffusion term will be very large and negative close to the boundary. Hence the cells which are close enough to the boundary will always be able to undergo bifurcations and oscillate. In Fig. 5(b) we show an intermediate situation where diffusion into the absorbing boundaries also plays a large role — it enables the system to oscillate, although with a smaller frequency.

V. CONCLUSION

The simple model presented in this paper captures many of the most important features of the NF-κB system, although it is highly simplified and consists of only three variables. The model essentially consists of a coupled positive and negative feedback, which makes it able to transiently amplify a signal of high cytokine concentration. This simple system provides a good tool for investigating and understanding the interactions between NF-κB and cytokines, especially because it makes it possible to explore the phase space, thereby achieving a greater understanding of the parameters.

The model captures how a single unit (cell) can become an oscillator if it is stimulated appropriately (close to the site of infection), but also how it can simply pass on the signal if it is stimulated transiently (in tissue farther away).

From phase-space analysis we conclude that the system is bistable and able to oscillate because it can undergo bifurcations, shifting the system between low and high fixed points [14]. The phase-space analysis also provides a useful understanding of the unknown parameter p, describing the strength of the positive feedback between NF-κB and cytokines. We find that the positive feedback must have an appropriate intermediate strength in order to create oscillations. Too-weak positive feedback leads to almost no response, whereas too-strong positive feedback leads to a sustained strong amplification of cytokine concentration, a situation which can be related to chronic inflammatory response.

A spatial model is highly relevant for understanding possible spatial effects that might appear in nature and which are not captured in most laboratory experiments because of space-averaging or mixing. Our spatial model of the tissue naturally leads to the propagation of traveling waves of high cytokine concentration, because the system behaves as an excitable medium.

Excitable media are also observed in many other biological systems which share the need of sending information over many-cell distances, and the resulting traveling waves are in good agreement with the expected spatial form of a neutrophil directing signal.

We find that spatial effects play a large role in the model and contributes to the model's ability to propagate repetitive waves. By changing the parameters of the model, we observe qualitatively different spatial patterns and we see that a even very strong positive feedback leading to chronic inflammation gives rise to oscillations close to the absorbing

boundaries representing the blood stream. Hence the situation corresponding to chronic inflammation would also recruit neutrophils from the bloodstream, but they would not be able to orient themselves once in the tissue because there is no directed signal to guide them.

ACKNOWLEDGMENT

This work was supported by the Danish National Research Foundation through the Center for Models of Life, to whom we are grateful.

[1] A. L. Hodgkin and A. F. Huxley, Proc. R. Soc. London B **140**, 177 (1952).
[2] R. FitzHugh, Bull. Math. Biol. **17**, 257 (1955).
[3] G. Gerisch, Curr. Top. Dev. Biol. **3**, 157 (1968).
[4] F. Alcantara and M. Monk, J. Gen. Microbiol. **85**, 321 (1974).
[5] P. Yde, B. Mengel, M. H. Jensen, S. Krishna, and A. Trusina, BMC Systems Biol. **5**, 115 (2011).
[6] A. Hoffmann, A. Levchenko, M. L. Scott, and D. Baltimore, Science **298**, 1241 (2002).
[7] D. E. Nelson *et al*., Science **306**, 704 (2004).
[8] S. Ghosh, M. J. May, and E. B. Kopp, Annu. Rev. Immunol. **16**, 225 (1998).
[9] S. Krishna, M. H. Jensen, and K. Sneppen, Proc. Natl. Acad. Sci. USA **103**, 10840 (2006).
[10] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang, Cell **138**, 760 (2009).
[11] S. Tay, J. J. Hughey, T. K. Lee, T. Lipniacki, S. R. Quake, and M. W. Covert, Nature **466**, 267 (2010).
[12] D. A. Turner, P. Paszek, D. J. Woodcock, D. E. Nelson, C. A. Horton, Y. Wang, D. G. Spiller, D. A. Rand, M. R. H. White, and C. V. Harper, J. Cell. Sci. **123**, 2834 (2010).
[13] R. Cheong, A. Bergmann, S. L. Werner, J. Regal, A. Hoffmann, and A. Levchenko, J. Biol. Chem. **281**, 2945 (2006).
[14] E. Meron, Phys. Rep. **218**, 1 (1992).

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:

STATISTICAL PHYSICS, THERMODYNAMICS AND NONLINEAR DYNAMICS

INFORMATION THEORY AND COMPUTATION

MATHEMATICS AND COMPUTING

STATISTICS

Received
3 September 2012

Accepted
5 October 2012

Published
8 November 2012

Correspondence and requests for materials should be addressed to M.H.J. (mhjensen@nbi.dk)

Modular networks of word correlations on Twitter

Joachim Mathiesen, Pernille Yde & Mogens H. Jensen

Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen, Denmark.

Complex networks are important tools for analyzing the information flow in many aspects of nature and human society. Using data from the microblogging service Twitter, we study networks of correlations in the occurrence of words from three different categories, international brands, nouns and US major cities. We create networks where the strength of links is determined by a similarity measure based on the rate of co-occurrences of words. In comparison with the null model, where words are assumed to be uncorrelated, the heavy-tailed distribution of pair correlations is shown to be a consequence of groups of words representing similar entities.

Networks are elegant representations of interactions between individuals in large communities and organizations[1–3]. These networks are constantly changing according to demands, fashions and flow of ideas[4–6]. The worldwide popularity of social media such as Twitter[5–7] have made them a considerable component in research on social networks[8,9]. Twitter is a microblogging service that allows registered users to post short text-based announcements, limited to 140 characters in length, known as "tweets", to an online stream. The frequency by which users interact on a global scale on Twitter allows for a high-resolution real-time analysis of movements in the society.

From automatic queries to Twitter, we have estimated tweet rates of words from a given set M containing selected words from one of the three different categories, international brand names, nouns and US major city names. The rate is measured by the number of new tweets posted per hour. For each query submitted at time t about a specific word $a \in M$, Twitter returns a finite set of the $n_a(t)$ latest tweets $\{T_1, \ldots, T_{n_a(t)}\}$. In addition to the message text string s, each tweet contains the username of the author, the time t_i when the tweet was posted and further details that we have not used. A tweet T_i is therefore a list of information $T_i = (s, t_i, \ldots)$. The maximum number of tweets returned from each query is $n_a = 1500$.

The time signal of tweets mentioning a specific word a, $\eta_a(t)$, can be written on the form

$$\eta_a(t) = \sum_i \delta(t - t_i), \tag{1}$$

From the number of tweets and the timestamps we compute an averaged tweet rate of a word a,

$$\gamma_a(t) = \frac{1}{\tau}\int_{t_1}^{t_1+\tau} \eta_a(t)\mathrm{d}t = \frac{n_a(t)}{\tau}, \tag{2}$$

Similarly we define a rate by which words a and b co-occur in a tweet at the same time, $\gamma_{ab}(t) = n_{ab}(t)/\tau$.

Tweets containing words from the aforementioned categories were recorded over a period of 4 months November 2010 – February 2011 and a period of two months January 2012 – February 2012. In general the rate, at which new tweets appear containing words from each of the categories, is too high to count the total number of tweets. Our analysis is based on estimated tweet rates computed from Eq. (2) using $n_a = 100$–1500. When averaging over many queries, we did not see a significant difference in the results when using different values of n_a.

We analyse the correlation between individual words within the mentioned categories. For that purpose, we define a measure of similarity in terms of the co-occurrence rate of words. The measure is then used to construct networks where links represent the degree of similarity. The way that we consider correlation networks can be seen as an alternative to existing studies on semantic networks (see e.g.[10]).

Results

We define a similarity measure between two words a and b in terms of the rate γ_{ab} by which new tweets occur containing both a and b. For example, by considering queries to Twitter containing the terms "Google" and

"Microsoft", we get $\gamma_{Google} \approx 130000$ tweets per hour and $\gamma_{Microsoft} \approx 17000$ tweets per hour whereas $\gamma_{Google,Microsoft} \approx 700$ tweets per hour (January 2011). A normalized symmetric measure of similarity (the Jaccard index) is naturally defined by

$$\omega_{ab} = \frac{\gamma_{a\cap b}}{\gamma_{a\cup b}} = \frac{\gamma_{ab}}{\gamma_a + \gamma_b - \gamma_{ab}} \quad (3)$$

Alternatively one can use information theory to compute the similarity from the joint probability of observing two words in the same tweet[11]. This approach is in particular useful when we have access to the normalized probabilities of observing A and B. Here, because of limitations to the permissible sample rate of data we only have access

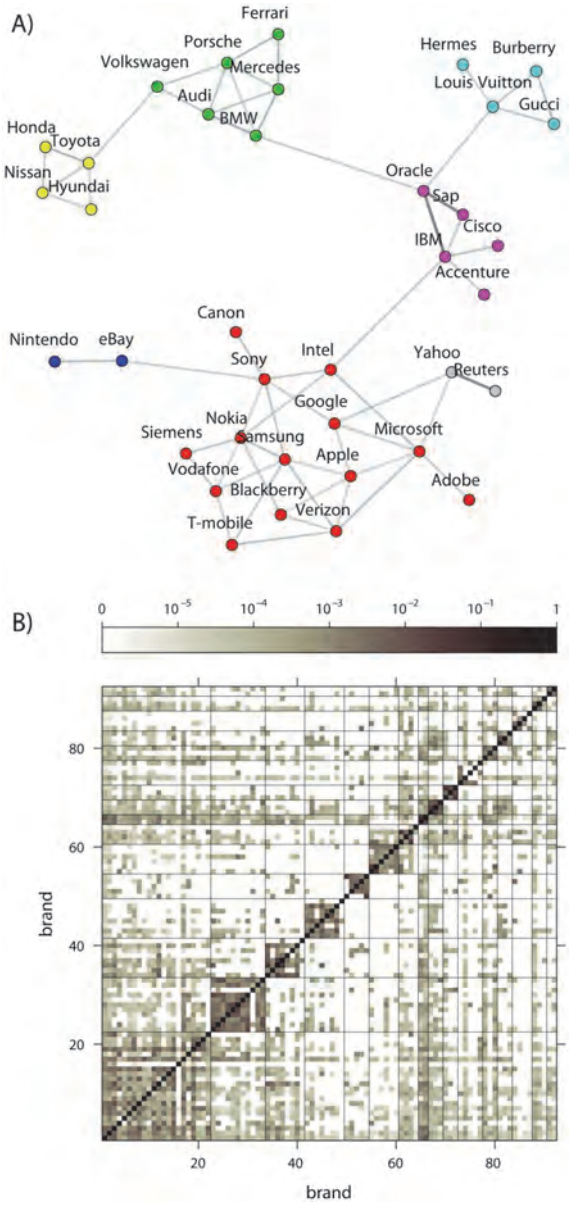


Figure 1 | Networks of correlations between international brands computed from the corresponding tweet rates on Twitter. A link in the networks represents the similarity measure computed using Eq. (3). In panel A, we show a network with links that have a strength larger than 0.004. The color of the nodes are modules found using community detection. Darker link colors mean stronger links. In Panel B, we show the adjacency matrix where the individual brands are ranked in modules. The colors represent the link strengths on a logarithmic scale. The block-structure is consistent with the clear modularity observed in panel A.

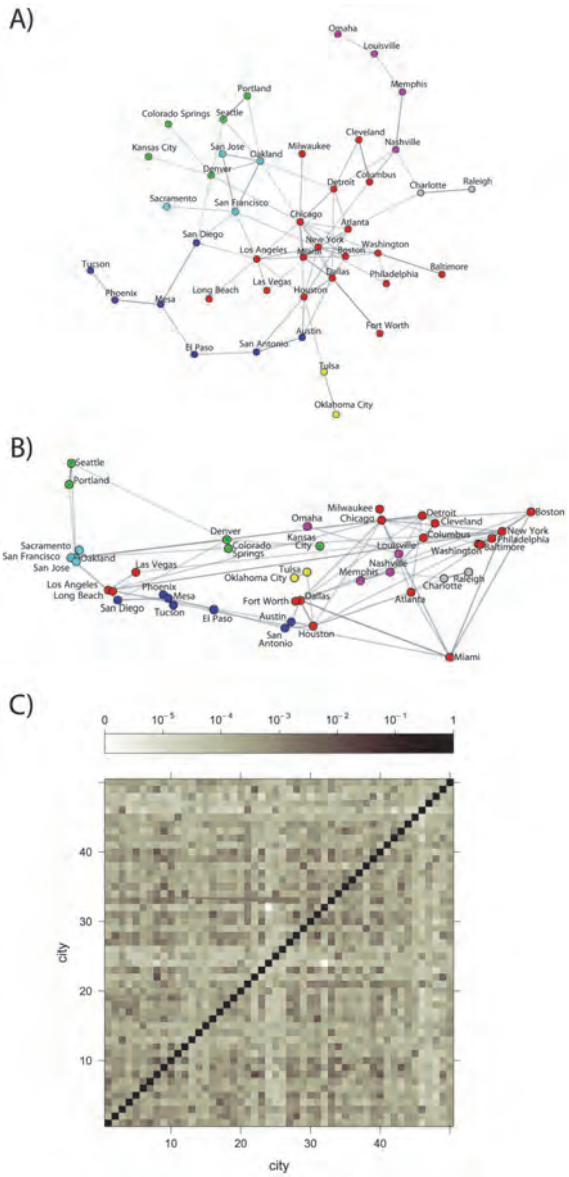


Figure 2 | Network of cities with high similarity. In panel A), we show a similarity network where nodes are located according to the algorithm of Fruchterman-Rheingold. In panel B), the corresponding network is shown where nodes are arranged according to the geographical location of the cities. In both panels only links with a strength larger than 0.004 are shown. In the network, darker link colors mean stronger links. In panel C), the network is shown in the corresponding matrix form.

to a fraction of the total number of posted tweets and can therefore at best estimate the relative probabilities.

In Fig. 1A we present a network of international brand names where the link strength is given by the measure Eq. (3). A threshold is introduced on the link strength in order to visualize primary structures, *i.e.* links between brand with a similarity $\omega_{AB} < 0.004$ are omitted. We observe that the network is strongly modular with groups of brands representing similar products or services. As an example one can observe distinct groups of European car brands, Asiatic car brands, consulting and IT companies, and fashion brands. The modules in the network are coloured according to the community detection algorithm introduced in[12]. Most of the connections inside the modules are rather obvious, whereas a few links connecting the modules represent less obvious relations between brands. In Fig. 1B we show the corresponding weighted adjacency matrix, where individual brands are ranked in modules. Note that the matrix contains information about brands that were not part of the largest connected component shown in Fig. 1A.

In Fig. 2A, a similarity network of US cities is shown. The network provides an alternative map where individual cities only to some extent are grouped according to their geographical location. The network is dominated by a central module consisting of New York, Chicago, Atlanta, Los Angeles and Boston. This is not surprising as these cities are hubs in the American society. We observe a module of Californian cities that connects naturally to cities like Denver and Seattle. We also detect a module of east-coast to mid-western cities connecting to a module of southern cities. Again the modules were detected by the algorithm presented in[12]. It is natural to ask how

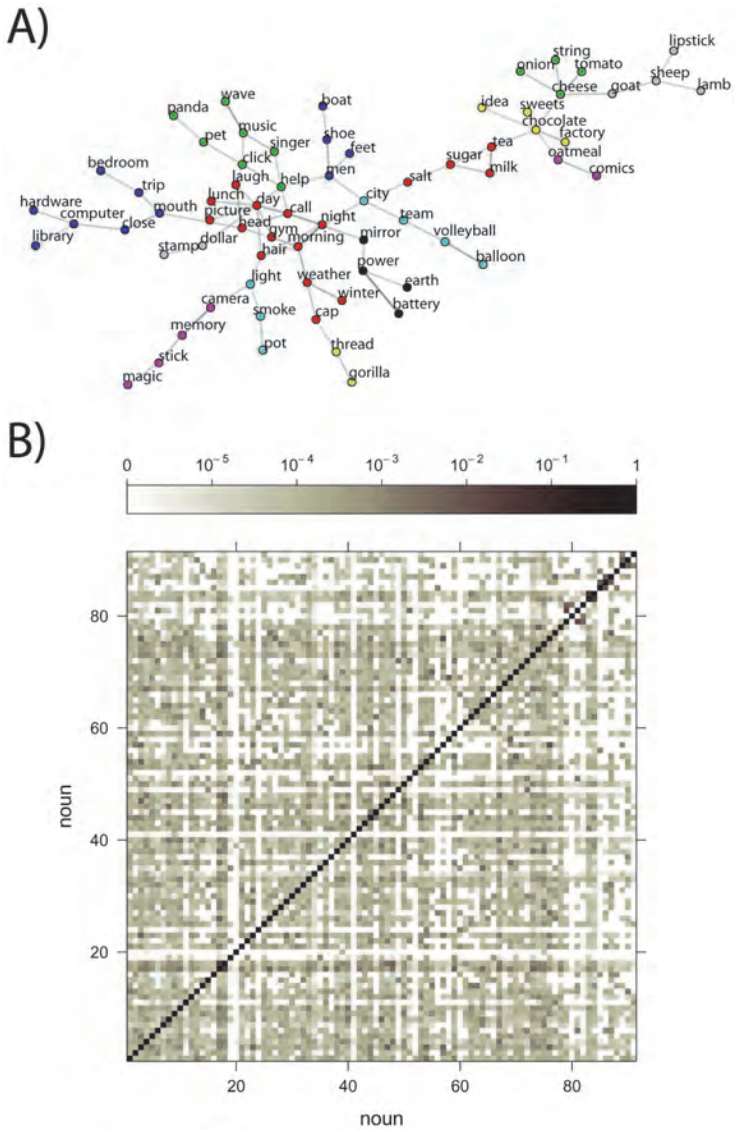


Figure 3 | **Network of nouns with high similarity.** Similarity network of 200 random nouns chosen from a list of the 2000 most common nouns. We only show the largest connected component for links with a strength larger than 0.04. The corresponding matrix form of the network including all nouns is shown in pnael B).

much of the similarity between cities is influenced by the geographical distance between them. To answer this question, we have compared tweet rates with the distance between cities as well as the size of the cities. It turns out that there is a weak to moderate correlation between the size of a city and the number of tweets referring to that city. The co-occurrence of two cities, however, has no clear correlation with their sizes and the distance between them. That said, when the nodes in the similarity network are arranged according to their geographical location it is evident that cities in same regions (states or neighbouring states) are better inter-connected and therefore often belong in the same module, see Fig. 2B.

As a final example of a similarity network, we present in Fig. 3 a network of nouns. From a list of 2000 common nouns in the English language, 200 nouns are randomly selected and the corresponding pairwise similarities are computed. Like the previous networks for brands and cities, the network of nouns also exhibits a pronounced modularity with modules e.g. representing similar food products.

We now consider further the data underlying the link strengths. As a main result, we obtain scale free distributions,

$$p(\gamma_{ab}) \sim \gamma_{ab}^{-\alpha}, \tag{4}$$

of the pairwise tweet rates γ_{ab} over six orders of magnitude using the brand names, nouns as well as major cities, see Fig. 4A. Surprisingly, the distributions all have the same scaling exponent $\alpha = 1.40 \pm 0.02$ (s.d.). The distribution of the tweet rates of individual search terms a, γ_a, does not follow a clear scale invariant distribution (see inset of Fig. 4). Moreover, the tweet rate of pairs γ_{ab} does not follow trivially from the rate of the individual brands, that is, the rate is not proportional to the product $\gamma_a\gamma_b$ which would be the case if a and b were uncorrelated. In particular we notice that if the distribution of the rates γ_x could be approximated by a scale invariant distribution $p(\gamma_x) \sim \gamma_x^{-\alpha}$ then the product $z = \gamma_a\gamma_b$ would follow a distribution

$$p(z) \sim z^{-\alpha} \log(z^2). \tag{5}$$

which follows from introducing the auxiliary variable $\nu = \gamma_a/\gamma_b$ and performing the integral

$$\int_{\epsilon^2/z}^{z/\epsilon^2} p(z, \nu)\mathrm{d}\nu = \int_{\epsilon^2/z}^{z/\epsilon^2} p(\gamma_a(z, \nu), \gamma_b(z, \nu)) \left|\frac{\partial(\gamma_a, \gamma_b)}{\partial(z, \nu)}\right| \mathrm{d}\nu, \tag{6}$$

where ϵ is a characteristic minimum tweet-rate that we observe.

The logarithmic correction to the scaling does not provide a statistically significant fit to the data presented in Fig. 4, that is a best fit has an exponent $\alpha \approx 2$ significantly larger than the tweet rate γ_x of individual search terms (see the inset of Fig. 4). A power-law distribution has also been observed for the co-occurrence of tags in social annotation systems[14] where users annotate online resources such as web pages by lists of words. The exponent of the distribution in the annotation systems ($\alpha > 2$) is larger than the one reported here and is close to the distribution of co-occurrence of nouns in sentences of novels considered below. The distribution of the similarity measure, Eq. (3), also has a scale invariant form. The value of α is in this case slightly larger, see Fig. 4B.

Discussion

For comparison, we have performed a similar analysis using search engines such as Google and Bing. The similarity between two words was computed from Eq. (3) by inserting the number of web pages that the search engines return containing the words. That is, instead of a rate we now use a fixed number. The distributions turn out to be significantly different (see Fig. 5A) and do not show a clear scaling behavior as in the case of Twitter. This may in part be explained by the fact that the search engines return results from web pages which are not restricted in size and they cover a wide range of media.

Finally, we compare the scaling behavior of word correlations observed on Twitter by considering the corresponding distribution of nouns in sentences of novels by Mark Twain (Huckleberry Finn) and Herman Melville (Moby-Dick). The sentences in these novels turn out to have a typical length comparable to the 140 character

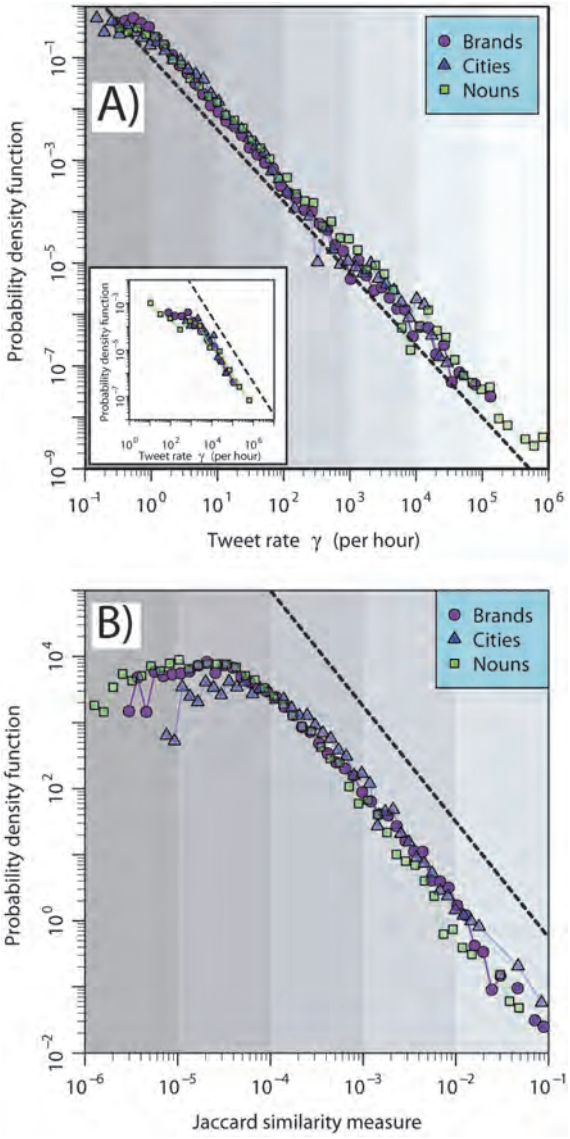


Figure 4 | Probability density function of tweet rates of pairs of international brands, major cities in the USA and common English nouns. The distributions include rates of individual search terms. The violet circles correspond to brand names, the blue triangles to cities and the green squares to nouns. Note that the rates of the cities have been multiplied by 20 to allow for a direct comparison. The distributions of the rates are scale invariant over more than six orders of magnitude and have the same exponent $\alpha = 1.40 \pm 0.02$ (s.d.). The dashed line corresponds to $\alpha = 1.4$. The inset shows distributions of tweet rates of single brands (purple circles), major US cities (blue triangles) and English nouns (green squares). For comparison we have inserted the same line as in the main panel and it is observed that the individual categories do not have the same scaling behavior. In panel B), we show the corresponding distribution for the similarity measure in Eq. (3).

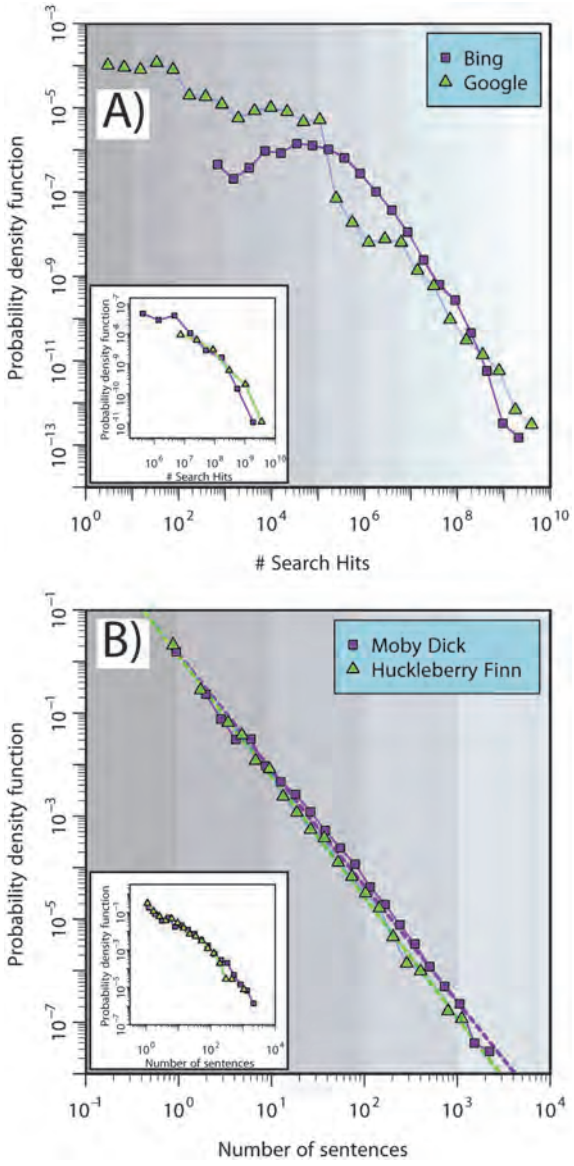


Figure 5 | Probability density functions of the number of search hits returned from Bing and Google and for the number of sentences in which two nouns co-occur in novels. In panel A) we performed pairwise queries on international brands to Bing and Google. In contrast to the result obtained from Twitter, we do not observe clear scale-free distributions. Inset: Probability density functions of search hits returned from queries on individual brands alone. Panel B) shows the number of sentences in which two nouns co-occur in the novels Huckleberry Finn (Mark Twain) and Moby-Dick (Herman Melville). The distributions are plotted on double-logarithmic scales and include the distributions of individual nouns. Dashed lines are best fit to a scale-free distribution and have exponents $\alpha = 2.34 \pm 0.04$(s.d.) (Huckleberry Finn) and $\alpha = 2.24 \pm 0.04$(s.d.) (Moby-Dick). Inset: Probability density function of the frequencies by which individual nouns occur in the same sentences.

limit of a tweet and do indeed lead to broad but significantly steeper distributions in the word correlations (see Fig. 5B). The novels are written by single authors and typically exhibits a more formal structure compared to the text messages. At the same time, the pair distribution of nouns are for the novels compatible with the null model where all words in the novels are randomized meaning that the correlated structures in the novels are rather weak. The distributions of individual words were considered for the same novels in[15]. Compared to the novels the distribution of the co-occurrence of words in tweets is less broad, which might be because the active vocabulary of the average user of Twitter is less diverse than that of the authors of the two novels.

Scale invariance is often described by Zipf's law[13] which states that the frequency of a word (for instance in a language) is inversely proportional to the rank in the frequency table. In its general formulation Zipf's law says that the frequency γ of a word is a power law in the rank $\gamma \sim r^{-\alpha}$. For the corresponding probability density functions we have $p(\gamma)d\gamma = p(r)dr \rightarrow p(\gamma) = p(r)\left|\frac{dr}{d\gamma}\right|$. Since $\frac{dk}{d\gamma} = \gamma^{-\frac{1+\alpha}{\alpha}}$ making the natural assumption that the PDF of the rank is a constant, we obtain the PDF of the frequency as

$$p(\gamma) \sim \gamma^{-\frac{1+\alpha}{\alpha}} \tag{7}$$

Empirically the value $\alpha \sim 1$ has been found for words in a corpus of a natural language where as for the population size of cities $\alpha \sim 1.1$. In Fig. 5 (inset) we observed a frequency distribution $p(\gamma) \sim \gamma^{-2}$ for words in the two novels leading to $\alpha \sim 1$ in good agreement with the 'established' Zipf result. For Twitter sentences on the other hand we found $p(\gamma) \sim \gamma^{-1.4}$ leading to a rank exponent of the order $\alpha = 2.5$ which is quite far from the usual Zipf exponent. We thus conclude, that texts from human communication on social media leads to a self-organized state that appears to have no resemblance with the structure of written texts.

Social media have become vital channels for advertising, dissemination of news and spreading of political opinions, therefore an understanding of the communication between users in social media provides important input not only to several branches of science but also for commercial purposes. For example, the value of a brand is determined by the consumer awareness and its apparent uniqueness. Companies put enormous efforts into positioning, i.e. to create the right image in the mind of potential customers. The modular structure of the brand network gives a first indication of the association between the various brands. For high-end fashion brands for instance, it might be preferable to be associated with similar brands instead of less valuable brands. At the same time the modular network can also be used to detect competing brands and as such provide invaluable information for commercial campaigns. In particular, the similarity measure could measure the correlation with 'up-coming' brands that might eventually turn into serious competitors. Likewise for cities, the network structure could provide a basis for urban strategies and business planning for travel-agencies.

1. Albert, R. & Barabasi, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
2. Borgatti, S. P., Mehra, A., Brass, D. J. & Labianca, G. Network analysis in the Social Sciences. *Science* **323,** *892–895 (2009).*
3. Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nature Physics* **6**, 888–893 (2010).
4. Ratkiewskicz, J., Fortunato, S., Flammini, F. & Vespignani, A. Characterizing and modeling the dynamics of online popularity, Phys. *Rev. Lett.* **105**, 158701 (2010).
5. Mandavilli, A. Peer review: Trial by Twitter. *Nature* **469**, 286–287 (2011).
6. Huberman, B. A., Romero, D. M. & Wu, F. Crowdsourcing, attention and productivity. *J. Inform. Sci.* **35**, 758–765 (2009).
7. Kwak, H., Lee, C., Park, H. & Moon, S. What is Twitter, a social network or a news media? Proceedings of the 19th international conference on World Wide Web, 591–600 (2010).
8. King, G. Ensuring the data-rich future of the social sciences. *Science* **331**, 719–721 (2011).

9. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
10. Steyvers, M. & Tenenbaum, J. B. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognit Sci* **29**, 4178 (2005).
11. Cilibrasi, R. L. & Vitanyi, P. M. B. The Google Similarity Distance, IEEE T. Knowl. *Data. En* **3**, 370383 (2007).
12. Rosvall, M. & & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1118–1123 (2008).
13. Zipf, G. K. Human behavior and the principle of least effort (Addison-Wesley, Cambridge, 1949).
14. Cattuto, C., Barrat, A., Baldassarro, A., Schehr, G. & Loreto V. Collective dynamics of social annotation. *Proc. Natl. Acad. Sci. U.S.A* **26**, 10511–10515 (2009).
15. Bernhardsson, S., Correa da Rocha, L. E. & Minnhagen, P. Size-dependent word frequencies and translational invariance of books, *Physica A* **389**, 330–341 (2010).

Acknowledgments

Suggestions and comments by Alex Hunziker and Pengfei Tian are gratefully acknowledged. This study was supported by the Danish National Research Foundation through the Center for Models of Life.

Author contributions

J.M., P.Y. and M.H.J. designed the research, performed the research, analyzed the data, and wrote the paper

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Mathiesen, J., Yde, P. & Jensen, M.H. Modular networks of word correlations on Twitter. *Sci. Rep.* **2**, 814; DOI:10.1038/srep00814 (2012).

BIBLIOGRAPHY

[1] Komla Ako, Dominique Durand, Taco Nicolai, and Lydiane Becu. Quantitative analysis of confocal laser scanning microscopy images of heat-set globular protein gels. *Food Hydrocolloids*, 23(4):1111 – 1119, 2009. ISSN 0268-005X. doi: http://dx.doi.org/10.1016/j.foodhyd.2008.09.003. URL http://www.sciencedirect.com/science/article/pii/S0268005X08002257. <ce:title>Food Colloids: Creating Structure, Delivering Functionality</ce:title>.

[2] E. Albrecht and H. R. Petty. Cellular memory: neutrophil orientation reverses during temporally decreasing chemoattractant concentrations. *Proc Natl Acad Sci U S A*, 95(9):5039–5044, Apr 1998.

[3] Tamara Aleksic, Bernd Baumann, Martin Wagner, Guido Adler, Thomas Wirth, and Christoph K Weber. Cellular immune reaction in the pancreas is induced by constitutively active ikappab kinase-2. *Gut*, 56(2):227–236, Feb 2007. doi: 10.1136/gut.2005.084665. URL http://dx.doi.org/10.1136/gut.2005.084665.

[4] Arno C Alting, Rob J Hamer, Cees G de Kruif, and Ronald W Visschers. Cold-set globular protein gels: interactions, structure and rheology as a function of protein concentration. *J Agric Food Chem*, 51(10):3150–3156, May 2003. doi: 10.1021/jf0209342. URL http://dx.doi.org/10.1021/jf0209342.

[5] Christian Beyschau Andersen, Hisashi Yagi, Mauro Manno, Vincenzo Martorana, Tadato Ban, Gunna Christiansen, Daniel Erik Otzen, Yuji Goto, and Christian Rischel. Branching in amyloid fibril growth. *Biophysical Journal*, 96(4):1529 – 1536, 2009. ISSN 0006-3495. doi: http://dx.doi.org/10.1016/j.bpj.2008.11.024. URL http://www.sciencedirect.com/science/article/pii/S0006349509000034.

[6] Andrew J Baldwin, Spencer J Anthony-Cahill, Tuomas P J Knowles, Guy Lippens, John Christodoulou, Paul D Barker, and Christopher M Dobson. Measurement of amyloid fibril length distributions by inclusion of rotational motion in solution nmr diffusion measurements. *Angew Chem Int Ed Engl*, 47(18):3385–3387, 2008. doi: 10.1002/anie.200703915. URL http://dx.doi.org/10.1002/anie.200703915.

[7] V?ronique Baud and Michael Karin. Signal transduction by tumor necrosis factor and its relatives. *Trends in Cell Biology*, 11(9):372–377, 2001. ISSN

0962-8924. doi: DOI:10.1016/S0962-8924(01)02064-5. URL http://www.sciencedirect.com/science/article/B6TCX-43RK8XD-K/2/233bb0952eb0afb21c6a7db67cef4314.

[8] M. F. Bishop and F. A. Ferrone. Kinetics of nucleation-controlled polymerization. a perturbation treatment for use with a secondary pathway. *Biophys J*, 46(5):631–644, Nov 1984. doi: 10.1016/S0006-3495(84)84062-X. URL http://dx.doi.org/10.1016/S0006-3495(84)84062-X.

[9] Domenico Bosco, Mathieu Armanet, Philippe Morel, Nadja Niclauss, Antonino Sgroi, Yannick D Muller, Laurianne Giovannoni, Géraldine Parnaud, and Thierry Berney. Unique arrangement of alpha- and beta-cells in human islets of langerhans. *Diabetes*, 59(5):1202–1210, May 2010. doi: 10.2337/db09-1177. URL http://dx.doi.org/10.2337/db09-1177.

[10] Cory M Bryant and D Julian McClements. Molecular basis of protein functionality with special consideration of cold-set gels derived from heat-denatured whey. *Trends in Food Science & Technology*, 9:143–151, 1998.

[11] S. Cairoli, S. Iametti, and F. Bonomi. Reversible and irreversible modifications of beta-lactoglobulin upon exposure to heat. *J Protein Chem*, 13(3):347–354, Apr 1994.

[12] Raymond Cheong, Adriel Bergmann, Shannon L Werner, Joshua Regal, Alexander Hoffmann, and Andre Levchenko. Transient ikappab kinase activity mediates temporal nf-kappab dynamics in response to a wide range of tumor necrosis factor-alpha doses. *J Biol Chem*, 281(5):2945–2950, Feb 2006. doi: 10.1074/jbc.M510085200. URL http://dx.doi.org/10.1074/jbc.M510085200.

[13] Fabrizio Chiti and Christopher M Dobson. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, 75:333–366, 2006. doi: 10.1146/annurev.biochem.75.101304.123901. URL http://dx.doi.org/10.1146/annurev.biochem.75.101304.123901.

[14] Samuel I A Cohen, Michele Vendruscolo, Christopher M Dobson, and Tuomas P J Knowles. Nucleated polymerization with secondary pathways. ii. determination of self-consistent solutions to growth processes described by non-linear master equations. *J Chem Phys*, 135(6):065106, Aug 2011. doi: 10.1063/1.3608917. URL http://dx.doi.org/10.1063/1.3608917.

[15] Samuel I A Cohen, Michele Vendruscolo, Christopher M Dobson, and Tuomas P J Knowles. Nucleated polymerization with secondary pathways. iii. equilibrium behavior and oligomer populations. *J Chem Phys*, 135(6):065107, Aug 2011. doi: 10.1063/1.3608918. URL http://dx.doi.org/10.1063/1.3608918.

[16] Samuel I A Cohen, Michele Vendruscolo, Mark E Welland, Christopher M Dobson, Eugene M Terentjev, and Tuomas P J Knowles. Nucleated polymerization with secondary pathways. i. time evolution of the principal moments. *J Chem Phys*, 135 (6):065105, Aug 2011. doi: 10.1063/1.3608916. URL http://dx.doi.org/10.1063/1.3608916.

[17] Charles A Dinarello, Marc Y Donath, and Thomas Mandrup-Poulsen. Role of il-1beta in type 2 diabetes. *Curr Opin Endocrinol Diabetes Obes*, 17(4):314–321, Aug 2010. doi: 10.1097/MED.0b013e32833bf6dc. URL http://dx.doi.org/10.1097/MED.0b013e32833bf6dc.

[18] Christopher M Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–890, Dec 2003. doi: 10.1038/nature02261. URL http://dx.doi.org/10.1038/nature02261.

[19] Marc Y Donath and Steven E Shoelson. Type 2 diabetes as an inflammatory disease. *Nat Rev Immunol*, 11(2):98–107, Feb 2011. doi: 10.1038/nri2925. URL http://dx.doi.org/10.1038/nri2925.

[20] Marc Y Donath, Marianne Böni-Schnetzler, Helga Ellingsgaard, Philippe A Halban, and Jan A Ehses. Cytokine production by islets in health and diabetes: cellular origin, regulation and function. *Trends Endocrinol Metab*, 21(5):261–267, May 2010. doi: 10.1016/j.tem.2009.12.010. URL http://dx.doi.org/10.1016/j.tem.2009.12.010.

[21] P. R. Ebrahimzadeh, C. Högfors, and M. Braide. Neutrophil chemotaxis in moving gradients of fmlp. *J Leukoc Biol*, 67(5): 651–661, May 2000.

[22] Jan A Ehses, Aurel Perren, Elisabeth Eppler, Pascale Ribaux, John A Pospisilik, Ranit Maor-Cahn, Xavier Gueripel, Helga Ellingsgaard, Marten K J Schneider, Gregoire Biollaz, Adriano Fontana, Manfred Reinecke, Francoise Homo-Delarche, and Marc Y Donath. Increased number of islet-associated macrophages in type 2 diabetes. *Diabetes*, 56(9):2356–2370, Sep 2007. doi: 10.2337/db06-1650. URL http://dx.doi.org/10.2337/db06-1650.

[23] A. A. Elayat, M. M. el Naggar, and M. Tahir. An immunocytochemical and morphometric study of the rat pancreatic islets. *J Anat*, 186 (Pt 3):629–637, Jun 1995.

[24] F. A. Ferrone, J. Hofrichter, H. R. Sunshine, and W. A. Eaton. Kinetic studies on photolysis-induced gelation of sickle cell hemoglobin suggest a new mechanism. *Biophys J*, 32(1):361–380, Oct 1980. doi: 10.1016/S0006-3495(80)84962-9. URL http://dx.doi.org/10.1016/S0006-3495(80)84962-9.

[25] F. A. Ferrone, J. Hofrichter, and W. A. Eaton. Kinetics of sickle hemoglobin polymerization. ii. a double nucleation mechanism. *J Mol Biol*, 183(4):611–631, Jun 1985.

[26] Flyvbjerg, Holy, and Leibler. Microtubule dynamics: Caps, catastrophes, and coupled hydrolysis. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, 54(5):5538–5560, Nov 1996.

[27] C. Frieden and D. W. Goddette. Polymerization of actin and actin-like systems: evaluation of the time course of polymerization in relation to the mechanism. *Biochemistry*, 22(25):5836–5843, Dec 1983.

[28] Jeremy Geiger, Deborah Wessels, and David R Soll. Human polymorphonuclear leukocytes respond to waves of chemoattractant, like dictyostelium. *Cell Motil Cytoskeleton*, 56(1):27–44, Sep 2003. doi: 10.1002/cm.10133. URL http://dx.doi.org/10.1002/cm.10133.

[29] G. J. Goodhill. Diffusion in axon guidance. *Eur J Neurosci*, 9(7):1414–1421, Jul 1997.

[30] Sundaram Gunasekaran, Sanghoon Ko, and Lan Xiao. Use of whey proteins for encapsulation and controlled delivery applications. *Journal of Food Engineering*, 83(1):31 – 40, 2007. ISSN 0260-8774. doi: http://dx.doi.org/10.1016/j.jfoodeng.2006.11.001. URL http://www.sciencedirect.com/science/article/pii/S0260877406006704. <ce:title>Future of Food Engineering</ce:title> <ce:subtitle>Selected Papers from the 2nd International Symposium of {CIGR} Section {VI} on Future of Food Engineering</ce:subtitle>.

[31] Su-Ji Han, Hyun-Mi Ko, Jung-Hwa Choi, Kook Heon Seo, Hyun-Suk Lee, Eun-Kyoung Choi, Il-Whan Choi, Hern-Ku Lee, and Suhn-Young Im. Molecular mechanisms for lipopolysaccharide-induced biphasic activation of nuclear factor-kappa b (nf-kappa b). *J Biol Chem*, 277(47):44715–44721, Nov 2002. doi: 10.1074/jbc.M202524200. URL http://dx.doi.org/10.1074/jbc.M202524200.

[32] Alexander Hoffmann and David Baltimore. Circuitry of nuclear factor κb signaling. *Immunological Reviews*, 210:171–186, 2006.

[33] Alexander Hoffmann, Andre Levchenko, Martin L Scott, and David Baltimore. The ikappab-nf-kappab signaling module: temporal control and selective gene activation. *Science*, 298(5596):1241–1245, Nov 2002. doi: 10.1126/science.1071914. URL http://dx.doi.org/10.1126/science.1071914.

[34] Thomas Holst-Hansen. The inflammatory response in islets of langerhans. Master's thesis, Univeristy of Copenhagen, 2013.

[35] Min W. Irwin, Susanna Mak, Douglas L. Mann, Rong Qu, Josef M. Penninger, Andrew Yan, Fayez Dawood, Wen-Hu Wen, Zhiping Shou, and Peter Liu. Tissue expression and immunolocalization of tumor necrosis factor-alpha in postinfarction dysfunctional myocardium. *Circulation*, 99(11):1492–1498, 1999. URL http://circ.ahajournals.org/cgi/content/abstract/99/11/1492.

[36] German Kilimnik, Billy Zhao, Junghyo Jo, Vipul Periwal, Piotr Witkowski, Ryosuke Misawa, and Manami Hara. Altered islet composition and disproportionate loss of large islets in patients with type 2 diabetes. *PLoS One*, 6(11):e27445, 2011. doi: 10.1371/journal.pone.0027445. URL http://dx.doi.org/10.1371/journal.pone.0027445.

[37] John E. Kinsella and Charles V. Morr. Milk proteins: Physicochemical and functional properties. *C R C Critical Reviews in Food Science and Nutrition*, 21(3):197–262, 1984. doi: 10.1080/10408398409527401. URL http://www.tandfonline.com/doi/abs/10.1080/10408398409527401. PMID: 6391823.

[38] Tuomas P J Knowles, Christopher A Waudby, Glyn L Devlin, Samuel I A Cohen, Adriano Aguzzi, Michele Vendruscolo, Eugene M Terentjev, Mark E Welland, and Christopher M Dobson. An analytical solution to the kinetics of breakable filament assembly. *Science*, 326(5959):1533–1537, Dec 2009. doi: 10.1126/science.1178250. URL http://dx.doi.org/10.1126/science.1178250.

[39] Sandeep Krishna, Mogens H Jensen, and Kim Sneppen. Minimal model of spiky oscillations in nf-kappab signaling. *Proc Natl Acad Sci U S A*, 103(29):10840–10845, Jul 2006. doi: 10.1073/pnas.0604085103. URL http://dx.doi.org/10.1073/pnas.0604085103.

[40] S. Kudo, K. Mizuno, Y. Hirai, and T. Shimizu. Clearance and tissue distribution of recombinant human interleukin 1 beta in rats. *Cancer Res*, 50(18):5751–5755, Sep 1990.

[41] J. D. Loike, J. el Khoury, L. Cao, C. P. Richards, H. Rascoff, J. T. Mandeville, F. R. Maxfield, and S. C. Silverstein. Fibrin regulates neutrophil migration in response to interleukin 8, leukotriene b4, tumor necrosis factor, and formyl-methionyl-leucyl-phenylalanine. *J Exp Med*, 181(5):1763–1772, May 1995.

[42] Kathrin Maedler, Gitanjali Dharmadhikari, Desiree M Schumann, and Joachim Størling. Interleukin-1 beta targeted therapy for type 2 diabetes. *Expert Opin Biol Ther*, 9(9):1177–1188, Sep 2009. doi: 10.1517/14712590903136688. URL http://dx.doi.org/10.1517/14712590903136688.

[43] Osvaldo Campanella Mary McSwiney, Harjinder Singh and Lawrence K. Creamer. Thermal gelation and denaturation of bovine -lactoglobulins a and b. *Jounal of Dairy Resarch*, 61:221–232, 1994.

[44] Joachim Mathiesen, Pernille Yde, and Mogens H Jensen. Modular networks of word correlations on twitter. *Sci Rep*, 2:814, 2012. doi: 10.1038/srep00814. URL http://dx.doi.org/10.1038/srep00814.

[45] Benedicte Mengel, Sandeep Krishna, Mogens H. Jensen, and Ala Trusina. Theoretical analyses predict a20 regulates period of nf-kb oscillation. *arxive[physics.bio-ph]*, 0911.0529, Nov 2009.

[46] Ehud Meron. Pattern formation in excitable media. *Physics Reports*, 218(1):1–66, 1992. ISSN 0370-1573. doi: DOI:10.1016/0370-1573(92)90098-K. URL http://www.sciencedirect.com/science/article/B6TVP-46SXPNW-7W/2/abaed74d9adfb66d4e4e6177ba4354d2.

[47] Andrea Moro, Carlos Gatti, and Néstor Delorenzi. Hydrophobicity of whey protein concentrates measured by fluorescence quenching and its relation with surface functional properties. *Journal of Agricultural and Food Chemistry*, 49(10):4784–4789, 2001. doi: 10.1021/jf001132e. URL http://pubs.acs.org/doi/abs/10.1021/jf001132e.

[48] D. E. Nelson, A. E C Ihekwaba, M. Elliott, J. R. Johnson, C. A. Gibney, B. E. Foreman, G. Nelson, V. See, C. A. Horton, D. G. Spiller, S. W. Edwards, H. P. McDowell, J. F. Unitt, E. Sullivan, R. Grimley, N. Benson, D. Broomhead, D. B. Kell, and M. R H White. Oscillations in nf-kappab signaling control the dynamics of gene expression. *Science*, 306(5696):704–708, Oct 2004. doi: 10.1126/science.1099962. URL http://dx.doi.org/10.1126/science.1099962.

[49] Min Nian, Paul Lee, Neelam Khaper, and Peter Liu. Inflammatory cytokines and postmyocardial infarction remodeling. *Circ Res*, 94(12):1543–1553, Jun 2004. doi: 10.1161/01.RES.0000130526.20854.fa. URL http://dx.doi.org/10.1161/01.RES.0000130526.20854.fa.

[50] Søren B. Nielsen, Kristina Wilhelm, Brian Vad, Jürgen Schleucher, Ludmilla A. Morozova-Roche, and Daniel Otzen. The interaction of equine lysozyme:oleic acid complexes with lipid membranes suggests a cargo off-loading mechanism. *Journal of Molecular Biology*, 398(2):351 – 361, 2010. ISSN 0022-2836. doi: http://dx.doi.org/10.1016/j.jmb.2010.03.012. URL http://www.sciencedirect.com/science/article/pii/S0022283610002500.

[51] Søren Bang Nielsen, Pernille Yde, Lise Giehm, Sabrina Sundbye, Gunna Christiansen, Joachim Mathiesen, Mogens Høgh Jensen, Poul Henning Jensen, and Daniel E Otzen. Multiple roles of heparin in the aggregation of p25α. *J Mol Biol*, 421 (4-5):601–615, Aug 2012. doi: 10.1016/j.jmb.2012.01.050. URL http://dx.doi.org/10.1016/j.jmb.2012.01.050.

[52] Koh Ono, Akira Matsumori, Tetsuo Shioi, Yutaka Furukawa, and Shigetake Sasayama. Cytokine gene expression after myocardial infarction in rat hearts : Possible implication in left ventricular remodeling. *Circulation*, 98(2):149–156, 1998. URL http://circ.ahajournals.org/cgi/content/abstract/98/2/149.

[53] F. Oosawa and S. Asakura. *Thermodynamics of the polymerisation of proteins*. Academic Press, 1975.

[54] F. OOSAWA and M. KASAI. A theory of linear and helical aggregations of macromolecules. *J Mol Biol*, 4:10–21, Jan 1962.

[55] Peter Parham. *The Immune System*. Garland Science, second edition, 2005.

[56] Tuan Phan-Xuan, Dominique Durand, Taco Nicolai, Laurence Donato, Christophe Schmitt, and Lionel Bovetto. On the crucial importance of the ph for the formation and self-stabilization of protein microgels and strands. *Langmuir*, 27(24):15092–15101, 2011. doi: 10.1021/la203357p. URL http://pubs.acs.org/doi/abs/10.1021/la203357p.

[57] Tuan Phan-Xuan, Dominique Durand, Taco Nicolai, Laurence Donato, Christophe Schmitt, and Lionel Bovetto. Heat induced formation of beta-lactoglobulin microgels driven by addition of calcium ions. *Food Hydrocolloids*, 34(0):227 – 235, 2012. ISSN 0268-005X. doi: http://dx.doi.org/10.1016/j.foodhyd.2012.09.008. URL http://www.sciencedirect.com/science/article/pii/S0268005X12002044. <ce:title>Food Colloids 2012: Creation and Breakdown of Structure</ce:title>.

[58] S. B. Prusiner. Molecular biology of prion diseases. *Science*, 252 (5012):1515–1522, Jun 1991.

[59] Nanik Purwanti, Mary Smiddy, Atze Jan van der Goot, Renko de Vries, Arno Alting, and Remko Boom. Modulation of rheological properties by heat-induced aggregation of whey protein solution. *Food Hydrocolloids*, 25(6):1482 – 1489, 2011. ISSN 0268-005X. doi: http://dx.doi.org/10.1016/j.foodhyd.2011.02.027. URL http://www.sciencedirect.com/science/article/pii/S0268005X11000737.

[60] X. L. Qi, S. Brownlow, C. Holt, and P. Sellers. Thermal denaturation of beta-lactoglobulin: effect of protein concentration at ph 6.75 and 8.05. *Biochim Biophys Acta*, 1248(1):43–49, Apr 1995.

[61] O. Rekdal, Z. Konopski, J. S. Svendsen, J. O. Winberg, T. Espevik, and B. Osterud. The tnf receptors p55 and p75 mediate chemotaxis of pmn induced by tnfalpha and a tnfalpha 36-62 peptide. *Mediators Inflamm*, 3(5):347–352, 1994. doi: 10.1155/S0962935194000487. URL http://dx.doi.org/10.1155/S0962935194000487.

[62] Amy M Ruschak and Andrew D Miranker. Fiber-dependent amyloid formation as catalysis of an existing reaction pathway. *Proc Natl Acad Sci U S A*, 104(30):12341–12346, Jul 2007. doi: 10.1073/pnas.0703306104. URL http://dx.doi.org/10.1073/pnas.0703306104.

[63] K.N. Ryan, B. Vardhanabhuti, D.P. Jaramillo, J.H. van Zanten, J.N. Coupland, and E.A. Foegeding. Stability and mechanism of whey protein soluble aggregates thermally treated with salts. *Food Hydrocolloids*, 27(2):411 – 420, 2012. ISSN 0268-005X. doi: http://dx.doi.org/10.1016/j.foodhyd.2011.11.006. URL http://www.sciencedirect.com/science/article/pii/S0268005X11003055.

[64] Christophe Schmitt, Claudine Bovay, Martine Rouvet, Sabrina Shojaei-Rami, and Eric Kolodziejczyk. Whey protein soluble aggregates from heating with nacl: physicochemical, interfacial, and foaming properties. *Langmuir*, 23(8):4155–4166, Apr 2007. doi: 10.1021/la0632575. URL http://dx.doi.org/10.1021/la0632575.

[65] Kate Schroder, Rongbin Zhou, and Jurg Tschopp. The nlrp3 inflammasome: a sensor for metabolic danger? *Science*, 327 (5963):296–300, Jan 2010. doi: 10.1126/science.1184003. URL http://dx.doi.org/10.1126/science.1184003.

[66] L. S. Tobacman and E. D. Korn. The kinetics of actin nucleation and polymerization. *J Biol Chem*, 258(5):3207–3214, Mar 1983.

[67] B. Vardhanabhuti and E. A. Foegeding. Rheological properties and characterization of polymerized whey protein isolates. *J Agric Food Chem*, 47(9):3649–3655, Sep 1999.

[68] Bongkosh Vardhanabhuti, E.Allen Foegeding, Matthew K McGuffey, Christopher R Daubert, and Harold E Swaisgood. Gelation properties of dispersions containing polymerized and native whey protein isolate. *Food Hydrocolloids*, 15(2):165 – 175, 2001. ISSN 0268-005X. doi: http://dx.doi.org/10.1016/S0268-005X(00)00062-X. URL http://www.sciencedirect.com/science/article/pii/S0268005X0000062X.

[69] M. G. Vicker. Ideal and non-ideal concentration gradient propagation in chemotaxis studies. *Exp Cell Res*, 136(1):91–100, Nov 1981.

[70] M. G. Vicker, J. M. Lackie, and W. Schill. Neutrophil leucocyte chemotaxis is not induced by a spatial gradient of chemoattractant. *J Cell Sci*, 84:263–280, Aug 1986.

[71] A. Wegner and J. Engel. Kinetics of the cooperative association of actin to actin filaments. *Biophys Chem*, 3(3):215–225, Jul 1975.

[72] Youling L. Xiong. Influence of ph and ionic environment on thermal aggregation of whey proteins. *Journal of Agricultural and Food Chemistry*, 40(3):380–384, 1992. doi: 10.1021/jf00015a004. URL http://pubs.acs.org/doi/abs/10.1021/jf00015a004.

[73] Wei-Feng Xue, Steve W Homans, and Sheena E Radford. Amyloid fibril length distribution quantified by atomic force microscopy single-particle image analysis. *Protein Eng Des Sel*, 22(8):489–496, Aug 2009. doi: 10.1093/protein/gzp026. URL http://dx.doi.org/10.1093/protein/gzp026.

[74] Pernille Yde. Inflammatory response and excitable media. Master's thesis, University of Copenhagen, 2010.

[75] Pernille Yde, Mogens Høgh Jensen, and Ala Trusina. Analyzing inflammatory response as excitable media. *Phys Rev E Stat Nonlin Soft Matter Phys*, 84(5 Pt 1):051913, Nov 2011.

[76] Pernille Yde, Benedicte Mengel, Mogens H Jensen, Sandeep Krishna, and Ala Trusina. Modeling the nf-κb mediated inflammatory response predicts cytokine waves in tissue. *BMC Systems Biology*, 5:115, 2011.

[77] Haiming Zhu and Srinivasan Damodaran. Heat-induced conformational changes in whey protein isolate and its relation to foaming properties. *Journal of Agricultural and Food Chemistry*, 42(4):846–855, 1994. doi: 10.1021/jf00040a002. URL http://pubs.acs.org/doi/abs/10.1021/jf00040a002.