Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 988

Intracellular Flows and Fluctuations

BY

JOHAN ELF

ACTA UNIVERSITATIS UPSALIENSIS
UPPSALA 2004

Dissertation presented at Uppsala University to be publicly examined in Room B41, Uppsala Biomedical Centre, Uppsala, Thursday, June 3, 2004 at 10:00 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

Elf, J. 2004. Intracellular Flows and Fluctuations. Acta Universitatis Upsaliensis. *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 988. 63 pp. Uppsala. ISBN 91-554-5988-9

Mathematical models are now gaining in importance for descriptions of biological processes. In this thesis, such models have been used to identify and analyze principles that govern bacterial protein synthesis under amino acid limitation. New techniques, that are generally applicable for analysis of intrinsic fluctuations in systems of chemical reactions, are also presented.

It is shown how multi-substrate reactions, such as protein synthesis, may display zero order kinetics below saturation, because an increase in one substrate pool is compensated by a decrease in another, so that the overall flow is unchanged. Under those conditions, metabolite pools display hyper sensitivity and large fluctuations, unless metabolite synthesis is carefully regulated. It is demonstrated that flow coupling in protein synthesis has consequences for transcriptional control of amino acid biosynthetic operons, accuracy of mRNA translation and the stringent response.

Flow coupling also determines the choices of synonymous codons in a number of cases. The reason is that tRNA isoacceptors, cognate to the same amino acid, often read different codons and become deacylated to very different degrees when their amino acid is limiting for protein synthesis. This was demonstrated theoretically and used to successfully predict the choices of control codons in ribosome mediated transcriptional attenuation and codon bias in stress response genes.

New tools for the analysis of internal fluctuations have been forged, most importantly, an efficient Monte Carlo algorithm for simulation of the Markov-process corresponding to the reaction-diffusion master equation. The algorithm makes it feasible to analyze stochastic kinetics in spatially extended systems. It was used to demonstrate that bi-stable chemical systems can display spontaneous domain separation also in three spatial dimensions. This analysis reveals geometrical constraints on biochemical memory circuits built from bistable systems. Further, biochemical applications of the Fokker-Planck equation and the Linear Noise Approximation have been explored.

Keywords: mesoscopic, reaction-diffusion, protein synthesis, amino acid, flow, fluctuations

Johan Elf, Department of Cell and Molecular Biology, Box 596, BMC, Uppsala University, SE-75124 Uppsala, Sweden

ISSN 1104-232X
ISBN 91-554-5988-9
urn:nbn:se:uu:diva-4291 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-4291)

Till min bättre Elft

ABBREVIATIONS

BST	Biochemical Systems Theory
FP	Fokker-Planck
LNA	Linear Noise Approximation
MC	Monte Carlo
MCA	Metabolic Control Analysis
ME	Master Equation
PDE	Partial Differential Equation
RDME	Reaction-Diffusion Master Equation

PAPERS

This thesis is based on the following papers, which are referred to in the text by the Roman numerals I-X:

I Elf J, Nilsson D, Tenson T and Ehrenberg M (2003) Selective charging of tRNA isoacceptors explains patterns of codon usage *Science* **300** 1718-1722.

II Elf J, Berg OG and Ehrenberg M (2001) Comparison of repressor and transcriptional attenuator systems for control of amino acid biosynthetic operons *Journal of Molecular Biology* **313** 941-954.

III Elf J, Paulsson J, Berg OG and Ehrenberg M (2003) Near-critical phenomena in intracellular metabolite pools *Biophysical Journal* **84** 154-170.

IV Elf J and Ehrenberg M (2003) Fast evaluation of fluctuations in biochemical networks with the linear noise approximation *Genome Research* **13** 2475-2484.

V Pedersen K, Zavialov A, Pavlov M, Elf J, Gerdes K and Ehrenberg M (2003) The bacterial toxin RelE displays codon specific cleavage of mRNAs in the ribosomal A-site *Cell* **112** 131-140.

VI Elf J, Doncic A & Ehrenberg M (2003) Mesoscopic reaction-diffusion in intracellular signaling *Fluctuations and noise in Biological,Biophysical and Biomedical Systems*. *SPIE proceedings series* **5110** 114-124.

VII Elf J, Lötstedt P, and Sjöberg P (2003) Problems of high dimension in molecular biology. *Proceedings of the 17th GAMM-Seminar* 21-30.

VIII Ehrenberg M, Elf J , Aurell E, Sandberg R and Tegnér J (2003) Systems Biology is Taking Off *Genome Research* **13** 2377-2380.

IX Elf J and Ehrenberg M (2004) Stochastic reaction-diffusion kinetics separates bi-stable biological systems into spatial domains of opposite phases. *Submitted manuscript*.

X Elf J and Ehrenberg, M (2004) Near-critical behavior of aminoacyl-tRNA pools at rate limiting supply of several types of amino acids in *E. coli*. *Manuscript*.

Reprints were made with permission from the journals.

CONTENTS

1. INTRODUCTION

The ambition of this thesis is to demonstrate how mathematical modeling can rationalize and clarify the thinking about a biological problem. Quantitative reasoning in Molecular Biology is increasingly important as the systems that are within experimental range are more and more complicated, and mere verbal reasoning often lead to unspecific conclusions that are hard to test.
The problem that I address is part of the old question (reviewed in [1]) how the molecular control circuits in the bacterium *Escherichia coli* are built to maintain a high rate of growth under varying nutritional conditions. This may seem to be a simple and sufficiently understood biological problem, as it is a major topic in standard textbooks on bacterial physiology [2, 3]. Many molecular biologists have therefore moved forward to more complicated problems in other organisms, leaving fundamental unresolved issues behind. These issues include global coordination of intracellular regulation and the dynamics of responses to changing environments, including major physiological changes under stress and starvation. Such issues are central to life in all organisms and it is likely that the principles that are clarified in *E. coli* later can be generalized to other organisms.
One particular systems of interest is the flow of amino acids into protein synthesis under conditions of amino acid limitation. The quantitative analysis of this system suggests a number of properties that hardly could have been deduced from verbal reasoning, but that are possible to test experimentally. For pedagogical reasons I can (post-) rationalize the modeling approach as an investigation of stationary and dynamic properties of simple kinetic motifs [4] (fig 1.1.) that are the metabolic building blocks of the protein synthesis machinery.

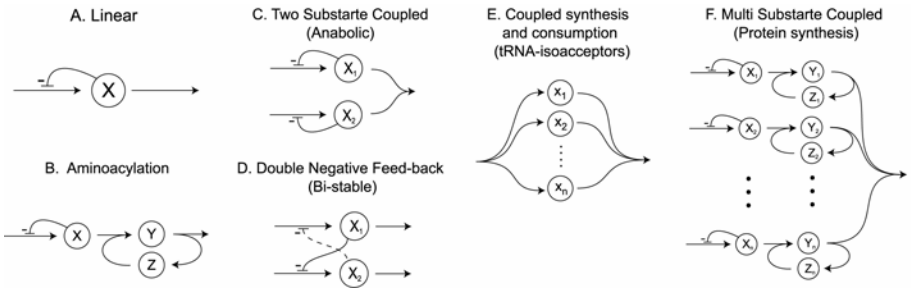


Figure 1.1 Kinetic motifs. The figure illustrates the building blocks of metabolic networks that are analyzed in this thesis. Macro- and mesoscopic properties of the different motifs are characterized under different flow constraints and consequences for the biological system in which they are embedded are discussed. A is described in paper III, B in II, C in III, IV,VII and section 3, D in IX and section 2.3, E in I and section 5, and F in X and section 4.

Metabolite flow in protein synthesis is clarified from general properties of these motifs. As the motifs, with some exceptions, are ubiquitous in metabolic networks, such general properties are of interest also outside the realm of protein synthesis. It should however be emphasized that it is the reason for evolution of a particular motif and its properties in a specific system, such as protein synthesis, that contribute to biological understanding.
The approach to characterize the motifs are principally similar to sensitivity analysis as described in biochemical systems theory (BST) [5-7] or metabolic control analysis (MCA) [8, 9]. Both MCA and BST do however have a strictly macroscopic perspective on chemical kinetics where internal fluctuations in the chemical reactions are disregarded. Such macroscopic models do generally not describe the average behavior correctly for systems of finite volume with non-linear kinetics. Therefore, the results of MCA or BST alone can be misleading and are here complemented with mesoscopic analysis based in the theory of stochastic processes for chemical reactions [10, 11]. Another concern is the spatial aspects of biological systems, which also will require extensions from MCA and BST, when appropriate.

1.1 Overview

Section 2 contains a review of the different levels of quantitative models that I have used in this thesis. The most interesting part of this section is the description of spatially dependant stochastic models, which is a practically unexplored, but potentially very important, domain of intracellular kinetics. In section 3 I describe and analyze a simple kinetics motif with coupled consumption of two metabolites in a bi-substrate reaction. This rather detailed analysis of a toy-system is motivated because the results have direct correspondences in the model for synthesis of amino acids under amino acid limitation and amino acid consumption in protein synthesis that is presented in section 4. Section 5 is an extension of a part of the model in section 4. It describes what can be expected to happen to the levels of aminoacylation of different transfer RNA (tRNA) isoacceptors under limitation of their cognate amino acid. Section 6 is another extension of the model in section 4. It contains an analysis of the different sources of sensitivity amplification in ribosome mediated transcriptional attenuation, which is a ubiquitous transcriptional control mechanism for amino acid biosynthetic operons. In section 7 I present some personal, non-scientific, reflections about the role of mathematical and computational models in molecular biology. Further the study of amino acid flow in protein synthesis is summarized.
Unpublished results that cannot be found in the main references I-X are indicated by * at the respective section title.
The thesis can be read from two different perspectives. The first perspective has the modeling tools as the main theme and the biological system for which they have been forged is only an example. The other perspective has the biological system and the principles that can be understood from it as the main focus, and the modeling tools are only methods. The choice of perspective is yours.

2. LEVELS OF KINETIC DESCRIPTION

Chemical reactions are discrete and stochastic events [10]. The reaction probabilities per unit time generally depend non-linearly on the molecule copy numbers. This invalidates macroscopic models as a correct description of the average behavior. The deviations from the macroscopic description are most pronounced when some component copy numbers are low or when the system is very sensitive [12], two conditions that are common in cells. In addition, intracellular diffusion is highly restricted [13] due to crowding and geometrical constraints. Features like these make intracellular chemistry qualitatively different from the typical test-tube experiment, where only few components of very high copy numbers are studied in isolation. The analysis must therefore be taken one step deeper in physical detail than the standard kinetic theory for macroscopic chemistry.
This section deals with the different levels of quantitative descriptions that have been used in this thesis, including some information about what conditions that has to be fulfilled in order to motivate a particular level of description. The chapter may therefore be too formal for some readers, that are recommended to move directly to section 4. The readers that lack rigor in this chapter will probably not be so interested in the rest of the thesis, and are recommended to too go directly to references [10, 12, 14, 15]. The intention is, however, to follow the narrow path on which a biologically interested physicist and a mathematically interested biologist can walk together if they have open minds.

2.1 Chemical reactions in the cell and the notion of state

Consider a biochemical system with volume Ω and N different components. The state of the system is defined by $\mathbf{n}=[n_1,..,n_N]$, where n_i is the number of molecules of species i. A state change takes place by any of R elementary reactions. The probability that the elementary reaction number j will occur in a small time interval δt is given by $\Omega\tilde{f}_j(\mathbf{n},\Omega)\delta t$, which is the transition rate for reaction j. By such an event the chemical component number i changes from n_i to $n_i + S_{ij}$ molecules where the integers S_{ij}, i=1,2…N; j=1,2,…,R are the elements of the $N\times R$ stoichiometric matrix $\mathbf{S}$ of the reaction network.
Elementary reactions mean that the transition rates $\tilde{f}_j(\mathbf{n},\Omega)$ should be nearly constant as long as the state $\mathbf{n}$ does not change. Another way to express this is that the system has no memory or that it is Markovian. Because the state description is contracted from a much larger state description, including for instance internal energy levels of the molecules, their spatial position and their momentum, we can only require that the transition rates are nearly constant. This means that we assume that all variables not included in the state description reach stationary distributions much faster than the chemical reactions that change in the state.
Typical transition rates that describe elementary reactions are derived from the law of mass action [16], *e.g.* the transition rate for the second order reaction $X_1+X_1 \longrightarrow X_2$, with a second order association rate constant k (ℓ)

is $\tilde{f}(n_1,\Omega) = kn_1(n_1-1)/\Omega^2$. The transition rate is in this way proportional to the number of different pairs of X_1 molecules.
A complex reaction, including many chemical intermediates can, under certain conditions, be considered elementary. The requirement is that the distribution of rates for the total reaction equilibrates on a faster timescale than state transitions that change the rates. The complex reaction can then be represented by its average rate and the chemical intermediates can safely be excluded from the state description. Care must however be taken not to forget that the molecules distributed in intermediate complexes can not participate in other reactions.
The assumption that the spatial distributions of the molecules are equilibrated on the timescale of the chemical reactions is a commonly used, but seldom motivated, assumption in intracellular modeling. In section 2.3 I will extend the state description to include local concentrations, but for now we settle with a well stirred and homogenous system.

2.2 Spatially homogenous systems

2.2.1 THE MASTER EQUATION

Given a discrete state space and memoryless probabilistic transition rates we can write the probability of being in a state $\mathbf{n}_m$ at time $t+dt$ as

$$\begin{aligned} &P(\mathbf{n}_m, t+dt) = \\ &P(\mathbf{n}_m, t) + dt \cdot \sum_{k\neq m} W(\mathbf{n}_k, \mathbf{n}_m) P(\mathbf{n}_k, t) - dt \cdot \sum_{k\neq m} W(\mathbf{n}_m, \mathbf{n}_k) P(\mathbf{n}_m, t) \end{aligned} \tag{2.1}$$

where $dt \cdot W(\mathbf{n}_k, \mathbf{n}_m)$ is the probability in a short time interval dt of a transition from state $\mathbf{n}_k$ to a state $\mathbf{n}_m$. Observe that the state transition rates W only depend on the state and not on t since the system is memoryless. When $P(\mathbf{n}_m, t)$ in Eq. (2.1) is moved to the left hand side, and all terms are divided by δt, one obtains the master equation [10] in the limit $dt \to 0$:

$$\frac{dP(\mathbf{n}_m, t)}{dt} = \sum_{k\neq m} W(\mathbf{n}_k, \mathbf{n}_m) P(\mathbf{n}_k, t) - \sum_{k\neq m} W(\mathbf{n}_m, \mathbf{n}_k) P(\mathbf{n}_m, t) \tag{2.2}$$

The master equation for the chemical system defined in section 2 is given by Eq. (2.2) if one for each reaction j, identifies the state transition rate $\Omega f_j(\mathbf{n},\Omega)$ for the transition from $\mathbf{n}$ to $\mathbf{n} + S_{\cdot j}$ [†] by $W(\mathbf{n}, \mathbf{n} + S_{\cdot j})$:

$$\frac{dP(\mathbf{n}, t)}{dt} = \Omega \sum_{j=1}^{R} \left(\prod_{i=1}^{N} \mathbb{E}_i^{-S_{ij}} - 1 \right) \tilde{f}_j(\mathbf{n}, \Omega) P(\mathbf{n}, t) . \tag{2.3}$$

$\mathbb{E}_i^m$ is a step operator, defined from $\mathbb{E}_i^m g(.,n_i,..) = g(..,n_i+m,..)$, where g is an arbitrary function of the state. The master equation Eq. (2.3) fully describes the time evolution of the homogeneous chemical systems defined in section 2.1 given an initial probability distribution $P(\mathbf{n},0)$.
Some examples of analytical solutions of master equations with one reactant are given in paper III. Master equations with more than one state variable and

[†] $S_{\cdot j} = \left[S_{1j} \cdots S_{Nj} \right]^T$

simple stoichiometry are rarely possible to solve explicitly. Valuable exact expressions for the moments of copy number distributions can however sometimes be obtained by moment generating function [10]. One example of this is given in section 3.6. Direct numerical solution of the master equation is usually prevented by the vastness of the state space. Therefore, the rest of this chapter will describe methods to approximate the solution of the master equation that have been used in this thesis. The early references to and some of the historical background of the use of the master equation for chemical reactions are given in [11].

2.2.2 Numerical simulation of the Markov Process corresponding to the Master Equation

One way to estimate the properties of the master equation is to simulate realizations of the Markov process that it describes using Monte Carlo (MC) methods. The probability that the next reaction will be of type j and occur between time $t+\tau$ and $t+\tau+\delta\tau$ when the system is in state $\mathbf{n}$ at time t is $\delta\tau \cdot p(\tau, j|\mathbf{n}, t)$, where

$$p(\tau, j|\mathbf{n}, t) = \underbrace{a(\mathbf{n})e^{-a(\mathbf{n})\tau}}_{I} \underbrace{\tilde{f}_j(\mathbf{n},\Omega)/a(\mathbf{n})}_{II} = \tilde{f}_j(\mathbf{n},\Omega)e^{-a(\mathbf{n})\tau},$$

$$\text{where } a(\mathbf{n}) = \Omega\sum_{j=1}^{R}\tilde{f}_j(\mathbf{n},\Omega) \tag{2.4}$$

Factor *I* is the probability density function for the time of any next reaction, and factor *II* is the probability that the reaction is of type *j*.
In year 1976 Gillespie [17] suggested two different ways to sample reactions from this distribution. Of these the *Direct Method* is most efficient. Here, the time to the next event is sampled from the exponential distribution (*I*) and the reaction type from (*II*). A practically equivalent but far less cited algorithm was published by Bortz, Kalos, and Lebowitz in 1975 [18]. In year 2000 Gibson and Bruck presented an algorithm, the *Next Reaction Method*, [19] that is more efficient, when the reaction network is sparse, such that a state change only affect a small number of transition rates.
In this thesis, the Direct Method has been used as a reference for other numerical and analytical methods and the Next Reaction Method has been used as a starting point for a new Monte Carlo algorithm for simulation of the Reaction Diffusion Master Equation (section 2.3).

2.2.3 The Fokker-Planck Approximation

The step operator $\Pi_i\mathbb{E}_i^{-S_{ij}}$ in the master equation Eq (2.3) implies that the function that follows, *i.e.* $\tilde{f}_j(\mathbf{n},\Omega)P(\mathbf{n},t)$, should be evaluated at a state displaced by $-S_{\cdot j}$ from $\mathbf{n}$. If the displacement is small and the function varies smoothly the displaced function can be approximated by a Taylor expansion. This is the idea behind the Kramers-Moyal approximation of the master equation (see [14] or [12]). For a general function $g_j(\mathbf{n})$ of the state vector $\mathbf{n}$ the Taylor expansion looks like

$$\prod_{i=1}^{N} \mathbb{E}_i^{-S_{ij}} g_j(\mathbf{n}) = g_j(\mathbf{n} + S_{\cdot j}) \approx$$
$$g_j(\mathbf{n}) + \sum_i S_{ij} \frac{\partial g_j(\mathbf{n})}{\partial n_i} + \frac{1}{2} \sum_{i,k} S_{ij} S_{kj} \frac{\partial^2 g_j(\mathbf{n})}{\partial n_i \partial n_k} + \ldots = \quad (2.5)$$
$$\left[1 + \sum_i S_{ij} \frac{\partial}{\partial n_i} + \frac{1}{2} \sum_{i,k} S_{ij} S_{kj} \frac{\partial^2}{\partial n_i \partial n_k} + \ldots \right] g_j(\mathbf{n})$$

It is seen that the step-operator is replaced by a differential operator. If we truncate the expansion after the second order term and use the differential approximation in instead of the step operator in the master equation (2.3) we get the Fokker-Planck approximation [20]:

$$\frac{dP(\mathbf{n},t)}{dt} = \Omega \sum_{j=1}^{R} \left(\sum_i S_{ij} \frac{\partial \tilde{f}_j(\mathbf{n},\Omega) P(\mathbf{n},t)}{\partial n_i} + \frac{1}{2} \sum_{i,k} S_{ij} S_{kj} \frac{\partial^2 \tilde{f}_j(\mathbf{n},\Omega) P(\mathbf{n},t)}{\partial n_i \partial n_k} \right) \quad (2.6)$$

The FP-equation Eq. (2.6) is a Partial Differential Equation (PDE) for the N-dimension probability density $P(\boldsymbol{n},t)$, which in this approximation is a continuous function of the state. This is unproblematic as long as it is smoothly varying at length scales corresponding to the small jumps in state space given by single reactions. The FP-equation uses the same initial condition as the master equation and additional boundary conditions to guarantee that the total probability remains constant.
For complicated reaction schemes the FP-equation is almost as complicated to work with as the original master equation. Its major advantage is that it can be solved numerically in many cases when the state space of the full master equation is too large. Numerical solution of the FP-PDE requires that the continuous state space is discretized on an artificial computational grid. When the probability varies smoothly over the state-space, the computational grid can be sparser than the state space discretization of the original master equation, which is based on molecule copy numbers. The major drawback of the FP-equation is that it is not exact and large deviations from the master equation can be expected when the approximation of $\tilde{f}_j(\mathbf{x},\Omega)$ and $P(\boldsymbol{x},t)$ as continuous smooth functions is poor. We are working on numerical methods to overcome these drawbacks in the FP-treatment of chemical fluctuations and to estimate the approximation error (Paper VII and Sjöberg et al. *in preparation*).
The primary solution is to use a discretization such that the state description is denser where the functions have large varaition. For instance, it is possible to choose a discretization corresponding to the real discretization of the master equation close to the lower boundaries, where a continuous description fails. In this way, the numerical treatment can seamlessly go from an exact master equation description for small numbers of molecules to a sparse Fokker-Planck description of large copy numbers of molecules.

2.2.4 The Linear Noise Approximation

The Fokker-Planck approximation would be exact if the jumps in state space were infinitely small. However, the sizes of the jumps in state space caused by chemical reactions are fixed in size, and there is no direct way to take the FP-approximation to a limit where it is exact. One could include more terms

from the Kramers-Moyal expansion but there is no systematic way to know when to truncate to get an approximation with sufficient accuracy. A solution to this potential and rather technical problem is the method known as size expansion [10]. This is a Taylor expansion in powers of $\Omega^{1/2}$ that can be systematically truncated when Ω, the volume of the well stirred system, gets large. If one truncates after the Ω^0-term the Linear Noise Approximation (LNA) of the master equation is obtained. The general LNA is derived in the supplementary material to paper IV. Similar treatments in one dimension can be found in references [21] and [10].

The LNA relies on that the transition rates can be accurately approximated by linear functions of the state variables over the whole region of the fluctuations. This is true in the limit of a large system far from critical points, where the internal fluctuations are small compared to the average value. However, as the biochemical systems addressed in this thesis are far from macroscopic, the LNA does not have more *a priory* validity than the Fokker-Planck approximation, Eq. (2.6). The LNA has, however, proved to be very useful in describing noise in intracellular systems, as it give simple analytic approximations for the sizes and correlations of fluctuations [22, 23] (and papers III, IV, X). I will recapitulate the most important results from the LNA, since they clarify the origin of internal fluctuations. More details are given in paper IV.

In the size expansion a new stochastic variable, ξ_i, is defined from the relation $n_i \equiv \Omega x_i + \Omega^{1/2}\xi_i$, where n_i is the copy number of component i, x_i is a deterministic function of time and Ω is the volume as before. The properties of x_i and ξ_i are both derived from the master equation using an expansion in powers of $\Omega^{1/2}$.

The first result from the LNA is that it derives the macroscopic rate equations from the master equation. These equations describe the kinetics in the limit of an infinitely large, well stirred, system. In this limit the stochastic fluctuations in $\mathbf{n}$ are negligible, the state can be fully represented by macroscopic average concentrations $\mathbf{x}=[x_1,\ldots,x_N]^T$ and $\tilde{f}_j(\mathbf{n},\Omega)$ will simplify to their macroscopic rate law counterparts $f_j(\mathbf{x})$. We define

$$\mathbf{x} = \lim_{\substack{\Omega\to\infty \\ \mathbf{n}\to\infty}} \Omega^{-1}\mathbf{n} \qquad \text{and} \qquad f_j(\mathbf{x}) = \lim_{\substack{\Omega\to\infty \\ \mathbf{n}\to\infty}} \tilde{f}_j(\mathbf{n},\Omega) \tag{2.7}$$

The time evolution of the macroscopic concentration variables $\mathbf{x}$ are governed by the deterministic rate equations

$$\dot{\mathbf{x}} = \mathbf{S}\mathbf{f}(\mathbf{x}), \tag{2.8}$$

where $\mathbf{S}$ is the stoichiometric matrix (see Eq. (2.3)) and $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x})\cdots f_R(\mathbf{x})]^T$. The macroscopic rate equation, Eq.(2.8), is commonly used as a staring point for modeling of intracellular kinetics without further motivation [9, 24].

The next order of the expansion gives the linear noise approximation of the fluctuations, $\Omega^{1/2}\xi_i$, around the macroscopic trajectory Ωx_i. The fluctuations are in this approximation characterized by the linear Fokker-Planck equation for the probability density function $\Pi(\boldsymbol{\xi},t)$:

$$\frac{\partial\Pi(\boldsymbol{\xi},t)}{\partial t} = -\sum_{i,k} A_{ik}\frac{\partial(\xi_k\Pi)}{\partial\xi_i} + \frac{1}{2}\sum_{i,k} D_{ik}\frac{\partial^2\Pi}{\partial\xi_i\partial\xi_k}, \tag{2.9}$$

where the **A** is the Jacobian and **D** is the diffusion matrix evaluated in the state **x**(t) as determined by Eq. (2.8). The elements of the matrices are

$$A_{ik} = \sum_{j=1}^{R} S_{ij} \frac{\partial f_j}{\partial x_k} \qquad \text{and} \qquad D_{ik} = \sum_{j=1}^{R} S_{ij} S_{kj} f_j(\mathbf{x}) \tag{2.10}$$

The stationary solution of the linear Fokker-Planck equation is a multivariate Gaussian. The average deviation from the macroscopic steady state is zero, *i.e.* $\langle \mathbf{n} \rangle \approx \Omega \bar{\mathbf{x}}$. The covariance matrix, **C**, of the fluctuations around the average is given by the Lyapunov equation

$$\mathbf{AC} + \mathbf{CA}^T + \Omega \mathbf{D} = \mathbf{0}, \tag{2.11}$$

where **C** is defined as

$$\mathbf{C} = \Omega \left\langle \boldsymbol{\delta\xi} \boldsymbol{\delta\xi}^T \right\rangle = \left\langle \left(\mathbf{n} - \langle \mathbf{n} \rangle \right) \left(\mathbf{n} - \langle \mathbf{n} \rangle \right)^T \right\rangle. \tag{2.12}$$

A and **D** are evaluated at the stationary state $\bar{\mathbf{x}}$ given by $\mathbf{Sf}(\bar{\mathbf{x}}) = 0$.
The general solution to Eq. (2.11) is given in paper IV. Here, I will focus on a few special cases.

2.2.4.1 LNA in single component systems
When there is only one chemical species, X, then Eq. (2.11) is scalar and the LNA of the stationary variance and Fano-factor [25] are

$$\sigma_X^2 = -\Omega \sum_j S_j^2 f_j \Big/ 2 \sum_j S_j f_j' \Big|_{x=\bar{x}} \quad \text{and} \quad \frac{\sigma_X^2}{\langle n \rangle} = -\sum_j S_j^2 f_j \Big/ 2\bar{x} \sum_j S_j f_j' \Big|_{x=\bar{x}} \tag{2.13}$$

Example 1: Flow and relaxation An important special case is when X is synthesized and consumed one molecule at the time by two different reactions with arbitrary rate laws $f_1(x)$ and $f_2(x)$. X can for instance be a metabolite in a biosynthetic pathway. At steady state the flow through the metabolite pool is $f_1(\bar{x}) = f_2(\bar{x}) = f$ and the Fano-factor evaluates to

$$\frac{\sigma_X^2}{\langle n_X \rangle} \approx -\frac{f}{\bar{x}(f_2' - f_1')}\Bigg|_{x=\bar{x}} = \frac{f/\bar{x}}{-\lambda}, \tag{2.14}$$

where $\lambda = f_1'(\bar{x}) - f_2'(\bar{x})$ is the rate of relaxation back to steady state $(\bar{x})$ after a small perturbation. The size of the fluctuations is therefore determined by the ratio between the rate of turnover of the pool, $f/\bar{x}$, and the macroscopic rate of relaxation back to steady state, λ.
We note that fluctuations in metabolite pools can have any size, since the reactions occur far from equilibrium. For instance, if synthesis of X in the example is constitutively synthesized by an enzyme that not is product inhibited, the Fano-factor will be 1 if the metabolite is consumed by an unsaturated Michaelis-Menten enzyme but it will increase to infinity as the flow increases and the consuming enzyme approaches saturation. In contrast, Fano-factor can be less then one if the consumption reaction is cooperative. In paper III and [26] this is described in more detail.

Example 2*: Equilibrium The Fano-factor is a measure of the size of internal fluctuations. It is equal to one in systems at thermodynamic equilibrium where fluctuations have Poisson statistics [14]. At the level of the LNA this can be seen if we assume a set of R reactions (j=1..R) $\mathrm{A}_j + m_{-j}\mathrm{X} \underset{k'_j}{\overset{k_j}{\rightleftharpoons}} m_j\mathrm{X}+\mathrm{B}_j$ that change the number of X molecules. A_j and B_j are in equilibrium independently of X, such that there are no conservation relations to consider. The rate laws and stoichiometries of the reactions are $f_j = k'_j a_j x^{m_{-j}} = k_j x^{m_{-j}}$ with $S_j = m_j - m_{-j}$ and $f_{-j} = k'_{-j} b_j x^{m_j} = k_{-j} x^{m_j}$ with $S_{-j} = -S_j$.

$$\frac{\sigma_X^2}{\langle X\rangle} \approx \left.\frac{-\sum_i (s_j)^2 k_j \bar{x}^{m_{-j}} + (-s_j)^2 k_{-j}\bar{x}^{m_j}}{2\bar{x}\sum_j (s_j) m_{-j} k_j \bar{x}^{m_{-j}-1} + (-s_j) m_j k_{-j}\bar{x}^{m_j-1}}\right|_{x=\bar{x}} = \left.\frac{-\sum_j (s_j)^2 k_j \bar{x}^{m_{-j}}}{\sum_j -(m_j - m_{-j})(s_j)_j k_j \bar{x}^{m_{-j}}}\right|_{x=\bar{x}} = 1 \quad , \tag{2.15}$$

where the last step follows from the equilibrium relation $k_j \bar{x}^{m_{-j}} = k_{-j}\bar{x}^{m_j}$.

Example 3*: Coupled fluctuations.
Two molecules X and Y are synthesized independently, but they are removed together from the system through a bimolecular reaction

$$\varnothing \xrightarrow{\tilde{f}_1(n_X,\Omega)} \mathrm{X} \quad \varnothing \xrightarrow{\tilde{f}_2(n_Y,\Omega)} \mathrm{Y} \quad \mathrm{X}+\mathrm{Y} \xrightarrow{\tilde{f}_3(n_X,n_Y,\Omega)} \varnothing \tag{2.16}$$

The transition rate vector $\tilde{\mathbf{f}}(\mathbf{n},\Omega)$ and the stoichiometric matrix $\mathbf{S}$ are

$$\tilde{\mathbf{f}}(\mathbf{x},\Omega) = \begin{bmatrix} \tilde{f}_1(n_X) & \tilde{f}_2(n_Y) & \tilde{f}_3(n_X)\end{bmatrix}^{\mathrm{T}}, \quad \mathbf{S} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1\end{pmatrix} \tag{2.17}$$

The master equation (2.3) is given by

$$\begin{aligned} dP(n_X,n_Y,t)/dt &= \Omega\tilde{f}_1(n_X-1)P(n_X-1,n_Y,t) \\ &+\Omega\tilde{f}_2(n_Y-1)P(n_X,n_Y-1,t) \\ &+\Omega\tilde{f}_3(n_X+1,n_Y+1)P(n_X+1,n_Y+1,t) \\ &-\Omega\left(\tilde{f}_1(n_X)+\tilde{f}_2(n_Y)+\tilde{f}_3(n_X,n_Y)\right)P(n_X,n_Y,t)\end{aligned} \tag{2.18}$$

The macroscopic equation is given by $\dot{\mathbf{x}} = \mathbf{S}\mathbf{f}(\mathbf{x})$ and its steady state, $\bar{\mathbf{x}} = \begin{bmatrix}\bar{x} & \bar{y}\end{bmatrix}$, is given by $\mathbf{S}\mathbf{f}(\bar{\mathbf{x}}) = 0$. In the LNA the average stationary copy numbers of the molecule $\langle\mathbf{n}\rangle \approx \Omega\bar{\mathbf{x}}$ and the covariance of the fluctuations, **C,** around this stationary value is given by the matrix equation

$$\mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{A}^T + \Omega\mathbf{B} = \mathbf{0}\,, \tag{2.19}$$

where the Jacobian matrix **A** and the diffusion matrix **B** are evaluated in the steady state. If we assume that the synthesis rates depend on their products in the same way, *i.e.* $\tilde{f}_1(n,\Omega) = \tilde{f}_2(n,\Omega)$, then $\bar{x} = \bar{y}$, $f_1(\bar{x}) = f_2(\bar{y}) = f_3(\bar{x},\bar{y})$ and $f'_{1x}(\bar{x}) = f'_{2y}(\bar{y})$

$$\mathbf{A} = \begin{bmatrix} f'_{1x} - f'_{3x} & -f'_{3y} \\ -f'_{3x} & f'_{1x} - f'_{3y} \end{bmatrix} \text{ and } \mathbf{B} = f_1 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \tag{2.20}$$

The covariance matrix **C** can now be solved explicitly, but the expressions for σ_X^2, σ_Y^2, and σ_{XY}^2 are complicated. The variance σ_{X-Y}^2 for the difference in copy number between X and Y is more informative and is given by

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}^2 = \frac{2\Omega f_1 \left(f'_{1x} - f'_{3y} \right)}{f'_{1x} \left(2f'_{1x} - f'_{3y} - f'_{3x} \right)} \tag{2.21}$$

If n_X and n_Y enter symmetrically in $f_3(n_X,n_Y)$, so that $f'_{3x} = f'_{3y}$, then Eq. (2.21) reduces to

$$\sigma_{X-Y}^2 \approx \Omega f_1 \big/ f'_1 \,. \tag{2.22}$$

This means that the variance of the difference between the numbers of X and Y molecules is approximately equal to the flow through the pools divided by how the supply rate responds to a perturbation in the pool from steady state. We will get back to this important result in sections 3 and 4.

2.2.4.2 Elimination of fast variables and LNA

A method for elimination of fast variables in LNA is also described in paper IV. The method can be used to make the LNA more accurate in describing fluctuations in non-linear systems. It works when it is possible to change variables such that the dynamics of the slowly relaxing variables is well approximated with linear rate laws. In these situations it is possible to do the LNA on two different time scales, where the fast variables with non-linear kinetics and small fluctuations are conditioned on the slow variables with large fluctuations.

2.3 Spatially dependant systems

When the spatial aspect of a system is part of its function [27], it is obvious that spatial models is necessary to describe relevant properties. Some examples of such systems are chemotaxis based on spatial sensing [28, 29], septum formation in bacterial cell division [30], signal processing in the dendritic tree of a neuron [31], or fly segmentation [32]. In other cases will the kinetics spatial dependence not be obvious for the function of the system by but still necessary for its correct description. One example is given in paper IX, where it is demonstrated that bi-stabile systems under some conditions behave qualitatively different when spatial aspects of their stochastic kinetics are considered.

2.3.1 The Reaction Diffusion Master Equation

In the derivation of the master equation (2.3) it was assumed that the spatial distribution of molecules equilibrates faster than the characteristic timescale of changes in the state variables. This was necessary in order to consider the transition rates $\tilde{f}_j(\mathbf{n},\Omega)$ as constant for a state **n**. If this condition is not ful-

filled, transition rates may depend on the time because the positions of previous reactions becomes important and the Markovian property is lost.
The strict condition for homogeneity by diffusion is that

$$\tau_i \gg L^2/D_i \qquad \text{for all } i\text{=}1..\text{N}, \tag{2.23}$$

τ_i is the characteristic time between two reactions for a molecule of compound i, D_i is its diffusion constant and L is the size in one dimension of the system. When Eq. (2.23) is satisfied, each molecule has an equal probability to participate in its next reaction anywhere. When Eq. (2.23) is not satisfied the master equation (2.3) is an approximation, which may be good or bad depending on if the transition rates are sensitive to local deviations in concentrations.
The transition rates averaged over space are not always equal to the transition rates of the concentration averaged over space, as the transition rates may be nonlinear.
In order to model the heterogeneity in space one needs to include local concentrations. Thus, the total volume Ω is divided into C artificial cubic sub-volumes of size $\Delta=\Omega/C$ and side length ℓ, so that Eq.(2.23) is satisfied, with L replaced by ℓ is satisfied. With this choice of sub-volume size the mean reaction free path is much longer than a sub-volume and each molecule has a higher probability to diffuse out of a sub-volume than to react in it.
At the same time, the sub-volumes should be much larger than the mean collision free path. This is necessary to describe the movement in the sub-volume as a diffusion process. The mean collision free path is very short in living cells due to the high concentration of non-reactive molecules. This makes the phase-space description [33, 34] that includes position and velocity of all molecules, unnecessary. More importantly for bio-molecular reactions are that ℓ is significantly larger than the reaction radius of all interactions, which is required for well defined association and dissociation rate constants within a sub-volume.
The extended state description is $\{\mathbf{n}_\lambda\} = [n_{11} \cdots n_{i\lambda} \cdots n_{\text{NC}}]$, where $n_{i\lambda}$ is the number of i-molecules in sub-volume λ [10]. The state-space is therefore huge (NC-dimensional), but this is what is required to model the system as Markovian. The state of the system is changed by (*i*) chemical reactions and (*ii*) diffusion events. The chemical reactions have different transition rates in different sub-volumes as they depend on the local concentrations of reactants. Diffusion is modeled as a memory-less random walk in discrete space [35], as implemented by a set of first order diffusion events:

$$[\cdots n_{i\gamma} \cdots n_{i\lambda} \cdots] \xrightarrow{n_{i\lambda} d_i^{\lambda\gamma}} [\cdots n_{i\gamma} + 1 \cdots n_{i\lambda} - 1 \cdots] \qquad \begin{matrix} i = 1,2,\ldots,N \\ \lambda = 1,2,\ldots,\text{C} \\ \gamma = 1,2,\ldots,\text{C} \end{matrix} \tag{2.24}$$

Here, an i-molecule diffuses from cell λ to cell γ. For the diffusion rate constants for i-molecules it is assumed that $d_i^{\lambda\mu} = d_i^{\mu\lambda} = D_i / \ell^2$ for neighboring cells and otherwise zero. This choice of $d_i^{\lambda\mu}$ gives the correct macroscopic diffusion behavior. Given the extended state description and a new set of state transition rates in the form of reaction and diffusion events, we can write down our reaction diffusion master equation [14, 36, 37]:

$$\frac{dP(\{\mathbf{n}_\lambda\},t)}{dt} = \Delta\sum_\lambda\sum_{j=1}^{R}\left(\prod_{i=1}^{N}\mathbb{E}_{i\lambda}^{-S_{ij}}-1\right)\tilde{f}_j(\mathbf{n}_\lambda,\Delta)P(\{\mathbf{n}_\lambda\},t)+ \sum_\lambda\sum_{\gamma\neq\lambda}\sum_i\left(\mathbb{E}_{i\lambda}^{1}\mathbb{E}_{i\gamma}^{-1}-1\right)d_i n_{\lambda i}\left(P\{\mathbf{n}_\lambda\},t\right) \quad (2.25)$$

The upper row contains the state transition rates that are due to reaction. These are exactly as before except that they have to be calculated for each sub-volume, λ. The step operator $\mathbb{E}_{i\lambda}^{m}$ means that the following function should be evaluated in a state where $n_{i\lambda}$ is changed to $n_{i\lambda}+m$. The lower row in Eq. (2.25) contains terms from diffusion between neighboring sub-volumes, *i.e.* Eq. (2.24). The reaction diffusion master equation Eq. (2.25) converges to the "ordinary" master equation Eq. (2.3) in the limit of fast diffusion.

2.3.2 The Spatial Next Reaction Algorithm

The reaction diffusion master equation (2.25) is prohibitively complicated for analytical approaches, especially if the system has "exotic" properties, such as bi-stability, ultra-sensitivity, oscillations, spatial pattern formation, all of which are common phenomena in living cells but not in the test-tube [38-41]. Direct simulation of the reaction diffusion Markov processes by, for instance, the Direct Method of Gillespie, is practically impossible as pointed out by him in the 1976 article [17]. This is still true even though CPU:s today are much faster than in 1976. The problem is that in each iteration one has to find the right reaction or diffusion event to execute, among all the C*R* reaction and C*N* diffusion events. The number of sub-volumes, C, is in the order of 10^5-10^7; *R*, the number of different reactions is in the order of 2-20; and *N*, the number of components, is in the order of 1-20. *R* and *N* are here limited to systems that are small enough to investigate with the hope of understanding how they work. The number of different events that have to be considered in order to know which one happened first is therefore huge, ($CR+CN>10^6$). With any of the two exact Gillespie algorithms one has to search through all these possible events in a linear fashion in each iteration. This implies that the algorithms scale linearly with the number of sub-volumes. The new algorithm I propose in paper VI also generates exact trajectories of the Markov process corresponding to the reaction diffusion master equation, but it scales logarithmically with the number of sub-volumes, which makes it approximately 10^4-10^5 times faster for a problem of realistic size. There are three improvements that are responsible for this remarkable speedup. These are briefly presented here, and in full details in the supplementary material of paper IX.

Firstly, it is only necessary to recalculate transition rates for events in sub-volumes where the state has changed. This implies that not more than $2(R+1)$ transition rates have to be recalculated after each event. An event is a reaction in a sub-volume or a diffusion out of a two sub-volumes.

Secondly, as long as the state in a sub-volume is constant, the distribution of the time for the next reaction in or diffusion out of a sub-volume will not change. Therefore, if the next event for a sub-volume is sampled to occur at time t_λ, it does not have to be resampled until after it actually occurs or if the state of the sub-volume changed because some molecule diffuses into it before time t_λ. Thus, one only has to keep track of in which sub-volume the next event occurs. When this reaction or diffusion event has occurred, one

must only recalculate the transition rates for the sub-volumes that were involved in this event. Given these transition rates the time of the next event in these sub-volumes can be sampled.
Finally, the event times of the sub-volumes are kept sorted in a priority queue, such that the next event always occurs in the sub-volume at the top of the queue. When new event times have been calculated for the sub-volumes that were involved in the last reaction, the queue is sorted, which can be done in $O(\log_2 C)$ time.
This algorithm is a combination of the Next Reaction Method by Gibson and Bruck [19] and the Direct Method by Gillespie [17]. The Next Reaction Method is used to keep track of in which sub-volume the next event occurs, and the direct method is used to sample which event and when the next event will occur in a sub-volume after that the state of the sub-volume has changed.

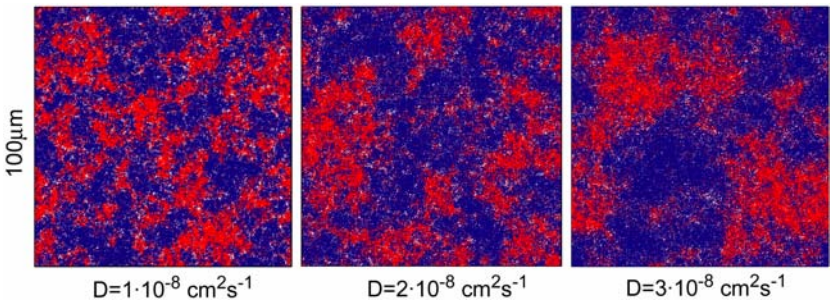


Figure 2.1 Domain separation. The mesoscopic reaction diffusion kinetics of a bi-stable chemical system (paper IX) has been simulated for different diffusion constants in a 0.3µm×100µm×100µm volume with periodic boundaries. The figure illustrates spontaneous separation of spatial domains in different attractors. The simulations with up to $3{\cdot}10^6$ sub-volumes were run on a standard PC.

A word of warning. It is very easy to make physically inconsistent simulations using this algorithm. The pitfalls that I have been struggling with the most are: (*i*) That the association and disassociation rate constants' dependence or the rate of diffusion must be considered [42]. Ideally, the reaction rates that are used in the model have been determined in conditions with intracellular diffusion rates, in which case there is no complication. However, if diffusion rates are varied in the model, the rate constants have to be adjusted accordingly and care has to be taken to ensure that the reaction radii of all molecular interactions are significantly smaller than the side-length ℓ of a sub-volume. (*ii*) Elimination of rapidly equilibrating reactions is complicated as diffusion out of a sub-volume should be fast compared to the rate of re-association after a dissociation event. (*iii*) Conservation relations of the total number of molecules can rarely be used to eliminate variables.

2.3.3 The Reaction-Diffusion Equation in the Macroscopic Limit

A spatially extended system can be modeled macroscopically in the limit that there is a *large* number of reaction partners within diffusion range of each molecule. In this limit the reaction diffusion master equation converges to the macroscopic reaction-diffusion equation [43]

$$\frac{dx_i(r,t)}{dt} = \sum_{j=1}^{R} S_{ij} f_j\left(\mathbf{x}(r,t)\right) + D_i \nabla^2 x_i(r,t) \qquad (2.26)$$

The definition of what is a sufficiently large number of molecules for a deterministic treatment will depend on the details of the system's kinetics, just as in the spatially homogeneous case.

The macroscopic limit of the mesoscopic description is illustrated in figure 2.2. This simulation of the reaction diffusion master equation with more than half a million sub-volumes and 8 reaction species has until now been out of reach, and demonstrates the efficiency of the new algorithm.

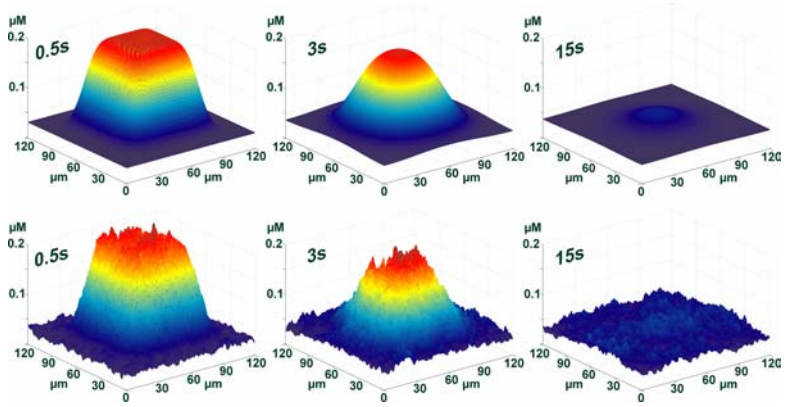


Figure 2.2. The macroscopic approximation. The figure shows three snap-shots (after 0.5s, 3s and 15s) of the concentrations of one of the reactants of the bi-stabile system. In the upper panel the system is described by macroscopic reaction-diffusion equations and in the lower by the reaction-diffusion master equation. The macroscopic approximation works because the rate of diffusion is high ($5 \cdot 10^{-7}$ $cm^2 s^{-1}$). More details are given in paper IX.

The macroscopic reaction-diffusion equation (2.26) has been successfully used to describe many biochemical phenomena [44] starting with Turing's seminal work on morphogenesis in 1952 [45]. In contrast, spatial models of intracellular systems considering internal fluctuations are very rare. One of the few example is given by [46]. The low number of spatial stochastic models of intercellular biochemistry is presumably not due lack of relevance, but because the experimental technology required to address spatial problems quantitatively is still in its infancy.

3. COUPLED IRREVERSIBVLE FLOWS

This section will elucidate properties of an idealized model for the coupled flows of two substrates that are irreversibly joined. The system exemplifies that very simple, yet biologically relevant, reaction schemes can give rise to kinetic properties that require mesoscopic analysis even though the average number of molecules is large.
In macroscopic terms, the system behavior is best described as an attenuated phase transition in metabolite pool dynamics, where the metabolite pools are ultra-sensitive to imbalances in their synthesis. The phase transition like behavior thus occurs at the point where the substrates are produced at equal rates. This is also the point, where the cell achieves a required flow at a minimal cost. In the phase transition region, the macroscopic description is inconclusive, as the mesoscopic analysis reveals that the metabolite fluctuations are very large in relation to the mean.
It is also demonstrated how adiabatic elimination of a fast variable in the phase transition regime makes it possible to solve the master equation. This is an example of the more general method suggested in paper IV. The analytical solution makes it possible to clarify the consequences of having feed-back systems or product inhibition regulating the synthesis of substrates. I show that these ubiquitous molecular control mechanisms efficiently attenuate the fluctuations in the system.
In section 4 these results are put in the context of coupled flows of amino acids into protein synthesis. It will be shown that a similar instability, as described here, is likely to reside in the levels of tRNA molecules charged with amino acids under balanced, but amino acid limited, growth. The instability or ultra-sensitivity of the metabolite pools may be important for the transcriptional control of gene expression, to which the metabolite pools often serve as control signals.
The results presented in this section are a simplified version of the results in paper III, where the reactions are treated as enzyme catalyzed reactions. A brief treatment is also given in paper IV as an example of LNA and in paper VII it is used as a test system for the numerical solution of the Fokker-Planck equation.

3.1 The reactions

A product molecule, C, is created by irreversible binding of two substrate molecules, X and Y, with rate constant k_2($s^{-1}M^{-1}$). The reaction is either spontaneous or catalyzed by an unsaturated enzyme in rapid equilibrium with both substrates. X and Y are synthesized with the rate constants k_X and k_Y ($s^{-1}M$), respectively, and are diluted through cell growth with rate constant μ (s^{-1}). The system is defined by the following reactions

$$\begin{array}{c} \varnothing \xrightarrow{k_X} \mathrm{X} \quad \varnothing \xrightarrow{k_Y} \mathrm{Y} \\ \mathrm{X} \xrightarrow{\mu} \varnothing \quad \mathrm{Y} \xrightarrow{\mu} \varnothing \quad \mathrm{X}+\mathrm{Y} \xrightarrow{k_2} \mathrm{C} \end{array} \tag{3.1}$$

3.2 Macroscopic analysis

The reactions in Scheme (3.1) are irreversible, as are energy driven intracellular biochemical pathways. The concentrations x and y of X and Y, respectively, obey the following macroscopic rate equations

$$dx/dt = k_X - k_2 xy - \mu x \qquad dy/dt = k_Y - k_2 xy - \mu y \tag{3.2}$$

As the system in Eq. (3.2) is symmetric, it is sufficient to study the parameter regime $k_X \geq k_Y$. The key assumption in the following analysis is that the turnover rate of the smallest pool, k_Y, is much higher than the rate of dilution, μy. In the limit $k_X = k_Y (\equiv k)$ this condition is equivalent to

$$\mu \ll \sqrt{kk_2} \tag{3.3}$$

Under this condition the stationary values of x and y are ultra sensitive to the balance between k_X and k_Y, as is illustrated in Figure 3.1.

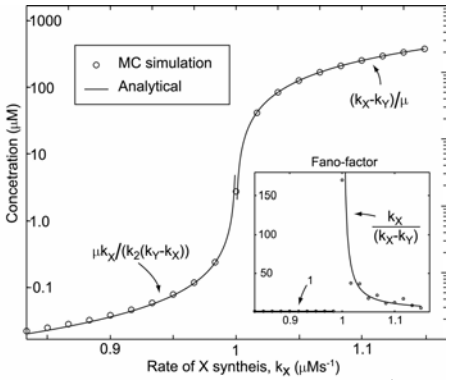


Figure 3.1 *Main figure* Concentration of X for different synthesis rates k_X. k_Y is constant at 1μMs^{-1}. The values indicated by circles are calculated from MC (Gillespie) simulations of the reactions. The lines represent analytical approximations of the macroscopic stationary concentrations, as given by the expressions in the figure. *Insert* The Fano-factor $(=\sigma_X^2/\langle n_X \rangle)$ as estimated from MC simulations (circles) and analytical approximations from LNA (lines). The association rate constant is $k_2 = 6 \cdot 10^4$ M^{-1}s^{-1}, the growth rate $\mu = 4 \cdot 10^{-4} s^{-1}$ and the cell volume Ω is $10^{-15}\ell$

A useful measure of the system sensitivity at steady state is the relative change in the pool concentration x, normalized to the relative change in its rate of synthesis k_X, *i.e.* the sensitivity amplification a_{xk} [38]. The sensitivity amplification peaks at $k_X = k_Y (\equiv k)$, where it evaluates to:

$$a_{xk} = \left(\frac{dx}{dk_X} \frac{k_X}{x} \right)_{k_X = k_Y = k_X} \approx \frac{\sqrt{kk_2}}{2\mu} \tag{3.4}$$

It follows from condition in Eq. (3.3) that $|a_{xk}| \gg 1$. The high sensitivity caused by the flow coupling of substrate pools generalizes the classical zero-order ultra-sensitivity concept [38, 47, 48] from one to two dimensions.
When the relative difference between k_X and k_Y is larger than a_{xk}^{-1}, the dilution term μy in Eq. (3.2) is negligible compared to the $k_2 xy$ term and the stationary concentrations are given by $x^s = (k_X - k_Y)/\mu$ and $y^s = \mu k_Y / [k_2 (k_X - k_Y)]$ (see Figure 3.1). The sensitivities in this case evaluate to $a_{yk_Y} = k_Y/(k_X - k_Y)$, for the limiting substrate Y, and $a_{xk_X} = k_X/(k_X - k_Y)$, for the non-limiting substrate X.

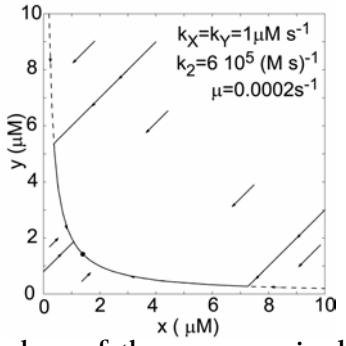


Figure 3.2 Phase-plane of the macroscopic description. Black lines are numerically evaluated trajectories. The dashed line is the curve $k_X = k_Y = k_2 xy$. Arrows indicate the direction of the trajectories. The filled circle represents the macroscopic steady state.

Introducing the variable change $w = x - y$ and $u = x + y$ in (3.2) gives

$$dw/dt = k_X - k_Y - \mu w \qquad du/dt = k_X + k_Y - k_2 \left(u^2 - w^2\right)/2 - \mu u \qquad (3.5)$$

In the phase transition region, where $k_X \approx k_Y$, a given investment in the enzymes that produce X and Y gives the largest rate of production of C. In this region, *u(t)* rapidly adjusts to a quasi-steady state, $u^s(w)$, corresponding to zero right hand side of the second equation (3.5), before there is a significant change in *w*. Separation of time scales can therefore be used to approximate the global behavior of the system by adiabatic elimination of the fast variable [14, 15]:

$$w(t) = (k_X - k_Y)/\mu + c_1 e^{-\mu t} \qquad u(t) = \frac{\lambda}{k_2}\left(\frac{1 - e^{-\lambda(t+c_2)}}{1 + e^{-\lambda(t+c_2)}}\right) - \frac{\mu}{k_2}, \qquad (3.6)$$

with $\lambda = \sqrt{2k_2(k_X + k_Y) + (k_2 w)^2 + \mu^2}$. In the slow time scale, the expression for $u^s(w)$ simplifies to

$$u^s(w) = \sqrt{\frac{2(k_X + k_Y)}{k_2} + w(t)^2 + \frac{\mu^2}{k_2^2}} - \frac{\mu}{k_2} \approx \sqrt{\frac{2(k_X + k_Y)}{k_2} + w(t)^2} \qquad (3.7)$$

In the phase plane [49] (Figure 3.2), the separation of time scales makes all trajectories initially move diagonally towards the nearly parallel nullclines at $xy \approx k/k_2$, which they follow at a rate set by μ to the stationary state at $w = (k_X - k_Y)/\mu$.

3.3 Mesoscopic analysis*

When the turnover rate of the metabolite pools is higher than the rate of relaxation towards the steady state, the macroscopic description is inconclusive and sometimes misleading due to the emergence of very large molecule-number fluctuations (see examples in section 2.2.4). Analysis of intracellular biosynthetic reactions, as described by Eq.(3.2), therefore requires mesoscopic considerations based on the master equation (section 2.2). Monte Carlo simulations [50] for the scheme (3.1) with biologically relevant parameter values illustrate the erratic behavior of the X- and Y-pools when their rates of synthesis are equal (Fig. 3.3). The stationary probability density function for the numbers n_X and n_Y of molecules in the X- and Y-pools, respectively, estimated from extensive simulations like the one in Figure 3.3, is shown in Figure 3.4 a-b. The master equation of scheme (3.1) is given by

$$\begin{aligned} dP/dt &= k_X\Omega(\mathbb{E}_X^{-1}-1)P + k_Y\Omega(\mathbb{E}_Y^{-1}-1)P + \left(k_2/\Omega\right)(\mathbb{E}_X^{1}\mathbb{E}_Y^{1}-1)n_X n_Y P \\ &+\mu(\mathbb{E}_X^{1}-1)n_X P + \mu(\mathbb{E}_Y^{1}-1)n_Y P \end{aligned} \tag{3.8}$$

P is the probability $P(n_X, n_Y, t)$ that there are n_X molecules of type X and n_Y molecules of type Y in the system at time t. Ω is the system (cell) volume and $\mathbb{E}$ is a step operator defined from $\mathbb{E}_X^i f(n_X) = f(n_X + i)$. Dilution of X- and Y-concentrations due to growth and cell division is in Eq. (3.8) approximated by a first order degradation rate constant μ [23].

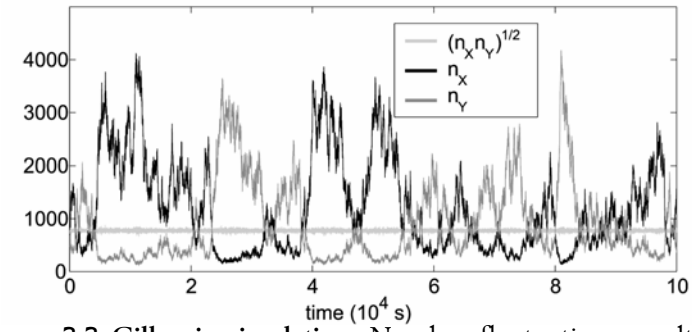


Figure 3.3 Gillespie simulation. Number fluctuations resulting from a balanced production of $k_X = k_Y = 1\mu Ms^{-1}$. The association rate constant is $k_2 = 6\cdot10^5 (Ms)^{-1}$, the growth rate $\mu = 2\cdot10^{-4}s^{-1}$ and the cell volume is $\Omega = 10^{-15}\,\ell$

As in the macroscopic analysis the kinetics of the two substrate pools is effectively decoupled when $k_X > k_Y$. In this case, the rate of consumption of the limiting pool, Y, is proportional to its concentration. The stationary probability density function is therefore approximated by a Poissonian distribution with an average of $\langle n_Y \rangle = \Omega k_Y / \left(k_2 x^s\right) = \Omega k_Y \mu / \left[k_2 (k_X - k_Y)\right]$. The quotient between the variance and the average value equals one for the Poisson distribution. For the non-limiting pool, X, the rate of consumption through the bimolecular reaction is set by k_Y and the one-dimensional

stochastic process is therefore linear and the stationary probability function is given by

$$P(n_X) = P(0)\left(k_X/\mu\right)^{n_X} \frac{\Gamma\left(1+k_Y/\mu\right)}{\Gamma\left(1+n_X+k_Y/\mu\right)} \tag{3.9}$$

The Fano-factor in this case is very well approximated by the sensitivity, $a_{xk_X} = k_X/(k_X - k_Y)$ as the variance of the linear one-step process is equal to the flow (Ωk_X) divided by the rate of relaxation, μ [10]. The Fano-factor takes very high values close to the transition point, $k_X = k_Y$. In the insert of figure 3.1, the analytical approximation of the Fano-factor is compared to the estimate calculated from MC simulations of the full system, Eq. (3.8).

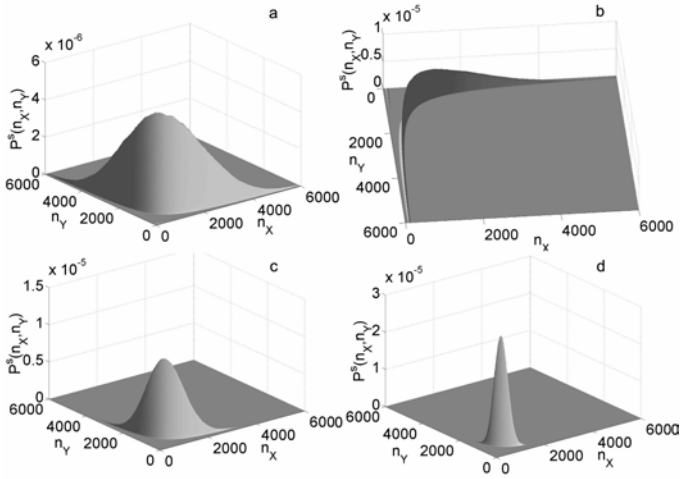


Figure 3.4 Stationary probabilities for different states (n_X,n_Y) as estimated from MC simulations using the Direct Method. a and b are without feed-back inhibition $\mu = 2 \cdot 10^{-4}\, s^{-1}$. c is feed-back inhibited with $K = 1.7$ mM and d with $K = 0.17$ mM. Other parameters are as in Figure 3.3.

Close to the transition point, *i.e.* when $k_X \approx k_Y$, the fluctuations in the pools are coupled and the full master equation Eq. (3.8) has to be considered. Exact solutions can not be obtained, but evaluation of the moment generating function (see section 3.6) gives an exact expression for the variance σ_w^2 of n_W=n_X-n_Y, when $k_X = k_Y (\equiv k)$:

$$\sigma_w^2 = k\Omega/\mu + \langle n_X \rangle \tag{3.10}$$

The variance can take very large values for high flow rates, k, and slow relaxation, μ. For large association rate constants, k_2, the standard deviation is even larger than the number of molecules in the steady state of the macroscopic model ($\approx \Omega\sqrt{k/k_2}$).
A change in variables and adiabatic elimination of the fast variable [14] can be used also mesoscopically to obtain an accurate approximation to the solution of the master equation Eq. (3.8). When $k_X \approx k_Y$, the probability distribution for the variable n_U=n_X+n_Y rapidly adjusts to a quasi steady state, conditional on n_W, which will fluctuate slowly with n_U as slave [14].

The slow variable n_W increases by one step when an X-molecule is synthesized and decreases by one step when a Y-molecule is synthesized. Conversely, n_W decreases when an X-molecule is degraded and increases when a Y-molecule is degraded. n_W is unaffected by the rate of product formation. This random walk is described by the birth and death scheme

$$\{n_W\} \underset{\Omega k_Y + \mu n_X(n_{W+1})}{\overset{\Omega k_X - \mu n_Y(n_W)}{\rightleftarrows}} \{n_{W+1}\} \quad \rightarrow \quad \{n_W\} \underset{\Omega k_Y + \mu n_{W+1}/2}{\overset{\Omega k_X - \mu n_W/2}{\rightleftarrows}} \{n_{W+1}\} \quad (3.11)$$

$n_Y = (n_U - n_W)/2$ and $n_X = (n_U + n_W)/2$. In the simplification that is indicated by the arrow we have used condition (3.3), *i.e.* $k_X \gg u\mu$ and $k_Y \gg u\mu$ when $k_X \approx k_Y$.

In the continuous description this is an Ornstein-Uhlenbeck process [10], for which the stationary distribution is Normal with an average of $\langle n_W \rangle = \Omega(k_X - k_Y)/\mu$ and a variance of $\sigma_w^2 \approx \Omega\left((k_X + k_Y)/(2\mu)\right)$. This approximate variance is close to the exact expression in Eq. (3.10) for $k_X = k_Y$. The autocorrelation [10] for n_W is given by

$$g_W(\tau) = \langle n_W(t), n_W(t+\tau) \rangle - \langle n_W \rangle^2 = \sigma_w^2 e^{-\mu\tau} \quad (3.12)$$

In Fig. 3.5, the autocorrelation for n_W in Eq. (3.12) is compared with an estimate from MC simulations of the full system.

To justify that the fluctuations in $n_U(n_W)$ are small and fast, we perform a Linear Noise Approximation (see 2.2.4). This is the way to formulate a linear Fokker-Planck equation for the distribution of fluctuations around the macroscopic value of n_U. The stationary solution (paper IV) of this equation is a Normal distribution $P(n_U) = N\left(\Omega u^s(w), \sigma_U^2\right)$, with

$$u^s(w) \approx \sqrt{2(k_X + k_Y)/k_2 + w^2}\,,$$

$$\sigma_U^2(w) \approx 3\Omega(k_X + k_Y)/\left(2k_2 u^s(w)\right). \quad (3.13)$$

Here $w = n_W/\Omega$. (The same notations are used for the discrete variables n_X, n_Y, n_W, n_U and their continuous counterparts).

The maximal variance, reached for $w = 0$, is approximately 3/4 of the mean value. This deviation from Poisson statistics arises because n_U decreases in steps of 2 through product formation between X and Y (see [10], pp. 246-247 for a discussion of this point). The relaxation rate for the autocorrelation of n_U is $\left(k_2 u^s(w) + \mu\right)$. This corresponds to the macroscopic relaxation rate λ in Eq. (3.6).

The separation of time scales allows for analytical approximations of Eq. (3.8) that are valid for biologically relevant parameter values. This is because $P(n_W, n_U, t) = P(n_W, t)P(n_U | n_W, t)$, where approximations of $P(n_U | n_W, t)$ and $P(n_W, t)$ are known. The probability distribution, $P(n_X, n_Y, t)$, for n_X and n_Y follows after a change of variables. For instance, the stationary probability distribution $P^s(n_X, n_Y)$ is given by

$$P^s(n_X, n_Y) \approx \mathcal{N} e^{-\frac{(n_X - n_Y - \Omega(k_X - k_Y)/\mu)^2}{2\sigma_w^2}} \cdot e^{-\frac{(n_X + n_Y - \Omega u^s(w))^2}{2\sigma_u^2}} \quad (3.14)$$

where $\mathcal{N}$ is a normalization constant. There is no visible difference between the approximate probability distribution in Eq. (3.14) and the estimate for

the exact steady state probability distribution obtained by Gillespie simulation and shown in Fig. 3.4. See also the comparisons in paper III.

3.4 The effects of product inhibition

A common feature of intracellular pathways is product inhibition or feed-back inhibition. These are control mechanisms, where the activity of enzymes that synthesize intracellular metabolites is inhibited by large concentrations of their own products, or of product molecules that appear further downstream along the pathway [2, 51]. To illustrate the effects of such inhibition on the stochastic properties of metabolite pools, we replace the rate constants k_X and k_Y in Scheme (3.1) by $k_X(x)$ and $k_Y(y)$, which depend on concentrations according to

$$k_X(x) = k_X/(1+x/K) \qquad k_Y(y) = k_Y/(1+y/K) \tag{3.15}$$

where K is an inhibition constant. When Eq. (3.15) can be Taylor expanded around the point ($x/K=0$, $y/K=0$) and $k_X = k_Y = k$, the linear noise description for n_W in Eqs. (3.12) and (3.14) remains on the same form, except that the growth rate μ is now replaced by $\mu' = \mu + k/K$. The variance for n_W becomes $\sigma_w^2 \approx \Omega(k/\mu') \approx \Omega K$, showing that even very moderate product inhibition reduces drastically both the size and correlation time ($1/\mu'$ rather than $1/\mu$) of pool fluctuations. Figs. 3.4c-d show how increasing strength of feedback inhibition in the synthesis reactions reduces fluctuations in the substrate pools, and Fig. 3.5 illustrates how the decay of the autocorrelation function for n_W is speeded up by feed-back inhibition. The figure also shows that there is no significant difference between autocorrelation estimates based on Eq. (3.12) and Monte-Carlo simulations based on Eq. (3.8).
These results suggest that product/feedback inhibition of enzymes, in addition to its previously identified system properties [51], is essential for control of gene expression by attenuating the large and slowly decaying molecule number fluctuations that otherwise spontaneously emerge in substrate pools that are coupled by a common exit flow. However, because the output per enzyme is reduced by feedback inhibition, its advantages will at some point be offset by reduced enzyme efficiency.

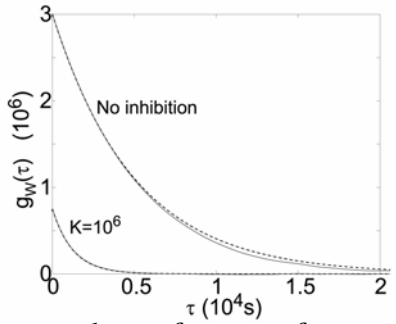


Figure 3.5 Auto correlation functions for $n_X - n_Y$. Estimates from Gillespie simulations (solid) are compared with analytical approximations (dashed) for non inhibited and product/feed-back inhibited synthesis. Other parameters are as in Fig. 3.

3.5 Summary: Coupled irreversible flows

It has been demonstrated that large fluctuations arise when the substrates in bimolecular reactions are produced at approximately equal rates. This is an inherent property of all irreversible multi-substrate reactions where the rates of synthesis of substrates are balanced to the consumption and the reaction is the dominant consumer of the substrates. The origin of the fluctuations is a zero order consumption rate in several dimensions, meaning that many different combinations of substrate concentrations give the same rate of the multi-substrate reaction. Therefore, an increase in one substrate can be compensated for by a decrease in another such that the overall consumption rate is unchanged. Notable examples of irreversible multi-substrate reactions of this kind are translation, transcription and replication. The inherent instability in these reactions imposes constraints on biochemical network wiring, which can be taken care of by for example feedback systems.
When the rates of substrate synthesis are unbalanced, the supply of one substrate limits the consumption of all other substrates. The pool of non-limiting substrates increases until their synthesis is inhibited to match the limited rate of consumption.
The system properties in both the balanced and the unbalanced regimes will be important for the analysis of turn-over of amino acids in protein synthesis described in the next section.

3.6 Appendix: The moment generating function*

It is possible to derive an exact expression for the stationary variance in $n_A - n_B$ by using moment-generating functions [10]. This gives a point of reference for the numerical and approximate solutions. The moment generating function

$$G(x,y,t) = \sum_{n_X,n_Y=0}^{\infty} x^{n_X} y^{n_X} P(n_X,n_Y,t) \tag{3.16}$$

satisfies the partial differential equation

$$\frac{\partial G}{\partial t} = k_0\Omega(x-1)G + k_0\Omega(y-1)G + \mu(1-x)\frac{\partial G}{\partial x} + \mu(1-y)\frac{\partial G}{\partial y} + \frac{k_2}{\Omega}(1-xy)\frac{\partial^2 G}{\partial x\partial y} \tag{3.17}$$

when $P(n_X,n_Y,t)$ is given by the master equation in Eq.(3.8). Eq. (3.17)is obtained by multiplying each term in Eq. (3.8) by $x^{n_X}y^{n_Y}$ and summing over all n_X and n_Y, *e.g.*

$$\begin{aligned}&\sum_{n_X,n_Y=0}^{\infty} x^{n_X}y^{n_Y}(n_X+1)(n_Y+1)P(n_X+1,n_Y+1,t) = \\ &\sum_{n_X,n_Y=0}^{\infty} x^{n_X-1}y^{n_Y-1}(n_X)(n_Y)P(n_X,n_Y,t) = \frac{\partial^2 G}{\partial x\partial y}\end{aligned} \tag{3.18}$$

Evaluating Eq. (3.17) in the appropriate limits gives

$$\frac{d\bar{n}_X}{dt} = \frac{\partial}{\partial t}\frac{\partial G}{\partial x}\bigg|_{x,y=1} = k_0\Omega - \mu\bar{n}_X - \frac{k_2}{\Omega}\overline{n_X n_Y} \tag{3.19}$$

$$\frac{d\bar{n}_Y}{dt} = \frac{\partial}{\partial t}\frac{\partial G}{\partial y}\bigg|_{x,y=1} = k_0\Omega - \mu\bar{n}_Y - \frac{k_2}{\Omega}\overline{n_X n_Y} \tag{3.20}$$

$$\frac{d\overline{n_X^2}}{dt} - \frac{d\bar{n}_X}{dt} = \frac{\partial}{\partial t}\frac{\partial^2 G}{\partial x^2}\bigg|_{x,y=1} = 2k_0\Omega\bar{n}_X - 2\mu\left(\overline{n_X^2} - \bar{n}_X\right) - 2\frac{k_2}{\Omega}\left(\overline{n_X^2 n_Y} - \overline{n_X n_Y}\right) \tag{3.21}$$

$$\frac{d(\overline{n_X n_Y})}{dt} = \frac{\partial}{\partial t}\frac{\partial^2 G}{\partial x\partial y}\bigg|_{x,y=1} = (\bar{n}_X + \bar{n}_Y)k_0\Omega - 2\mu\left(\overline{n_X n_Y}\right) - \frac{k_2}{\Omega}\left(\overline{n_X^2 n_Y} - \overline{n_X n_Y} + \overline{n_Y^2 n_X}\right) \tag{3.22}$$

$$\frac{d\sigma_X^2}{dt} = 2k_0\Omega - \frac{d\bar{n}_X}{dt} - 2\mu\sigma_X^2 - \frac{2k_2}{\Omega}\left(\overline{n_X^2 n_Y} - \bar{n}_X\overline{n_X n_Y}\right) \tag{3.23}$$

$$\frac{d\sigma_Y^2}{dt} = 2k_0\Omega - \frac{d\bar{n}_Y}{dt} - 2\mu\sigma_Y^2 - \frac{2k_2}{\Omega}\left(\overline{n_Y^2 n_X} - \bar{n}_Y\overline{n_X n_Y}\right) \tag{3.24}$$

$$\frac{d\sigma_{XY}^2}{dt} = -2\mu\sigma_{XY}^2 - \frac{k_2}{\Omega}\left(\overline{n_X^2 n_Y} - \bar{n}_X\overline{n_X n_Y} - \overline{n_X n_Y} + \overline{n_Y^2 n_X} - \bar{n}_Y\overline{n_X n_Y}\right) \tag{3.25}$$

where $\sigma_{XY}^2 = \overline{n_X n_Y} - \bar{n}_X\bar{n}_Y$ and $\overline{\text{bar}}$ indicates average. Thus, the variance in $n_X - n_Y$, σ_{X-Y}^2, can be expressed as

$$\sigma_{X-Y}^2 = \overline{((n_X - n_Y) - (\bar{n}_X - \bar{n}_Y))^2} = \left(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}^2\right) \tag{3.26}$$

From Eqs. (3.23), (3.24) and (3.25) we get

$$\frac{d}{dt}\left(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}^2\right) = 2k_0\Omega - 2\mu\left(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}^2\right) + \mu\left(\bar{n}_X + \bar{n}_Y\right) \tag{3.27}$$

At the stationary state, the distribution must be symmetric and $\bar{n}_X = \bar{n}_Y$, $\sigma_X^2 = \sigma_Y^2$, and $\overline{n_X^2 n_Y} = n_X n_Y^2$. This gives

$$\sigma_{X-Y}^2 = 2(\sigma_X^2 - \sigma_{XY}^2) = \frac{k_0\Omega}{\mu} + \bar{n}_X \tag{3.28}$$

4. COUPLED CONSUMPTION OF AMINOACYL-tRNA

The following sections (4, 5 and 6) will deal with how constraints for coupled metabolite flows can be used to better understand how amino acid synthesis is regulated and what consequences this has for protein synthesis under conditions of amino acid limitation. In section 4 I will investigate consequences of the common assumption that aminoacyl tRNAs are partially charged with amino acids in poor media. In section 5 the analysis is focused on what happens when charging of one tRNA isoacceptor family is limiting the rate of translation. In section 6 I will focus on a transcriptional control system that responds to changes in the concentration of aminoacylated tRNA. All these models are closely related, but they are valid under different experimental conditions and they can therefore be tested separately.
The aim is to capture the critical aspects of the system dynamics that governs how amino acid flows are balanced between different biosynthetic pathways. Some aspects of the model described below are therefore crude, whereas other aspects of general system properties are considered in more detail.

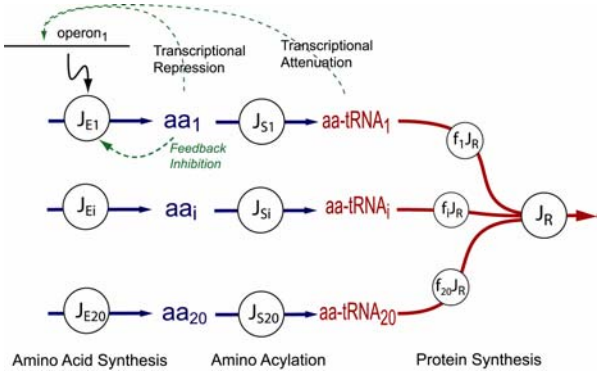


Figure 4.1 System overview. 20 amino acids (aa_i) are synthesized at rates J_{Ei}. The amino acids are used in aminoacylation of tRNA at rates J_{Si}. The aminoacylated tRNA in ternary complexes ($aa\text{-}tRNA_i$) are consumed at different frequencies (f_i) in protein synthesis. The synthesis enzymes are feedback inhibited by respective free amino acid. In *E. coli* is expression of biosynthetic enzymes under control of repressor or attenuation mechanisms, sensing the amino acid concentration and the rate of translation respectively.

4.1 The flow of amino acids into protein synthesis

The intracellular kinetic motif with twenty regulated pathways for amino acid synthesis and a common outflow via protein synthesis is summarized in figure. 4.1. The figure also summarizes the most important regulatory systems.

The genes for the amino acid biosynthetic enzymes are often organized in operon structures, such that enzymes that function in the same biosynthetic pathway are translated from the same transcript. This basic regulatory structure coordinates expression of enzymes that must be simultaneously present in well-defined stoichiometries [2].

Transcriptional regulation of the operons for the amino acid synthetic enzymes is mediated by repressor systems (arg, lys, met, asp), by ribosome dependent attenuation systems (leu, his, val, ile, thr, phe), or by the combined action of both types of systems (trp). References to studies of these control systems are given in paper II.

Repressors are activated for DNA binding and repression of transcription by an allosteric transition that occurs when it forms complex with the amino acid that is produced by the enzymes that are expressed from the regulated operon [52]. Accordingly, a high intracellular concentration of a particular free amino acid signals that expression from the operon that encodes the enzymes making this amino acid should be shut down.

A ribosome-mediated transcriptional attenuation system responds, in contrast, to the rate of translation of "own" codons in the leader sequence of the transcript from the amino acid biosynthetic operon [53]. This implies that the rate of peptide elongarion must be slowed down to activate the expression from attenuation controlled operons. This mechanism is described in detail in section 6.

In addition to the transcriptional regulation of the concentrations of the amino acid synthetic enzymes, the flows through their metabolic pathways are modulated by feedback control mechanisms (Fig. 4.1). These sense metabolite concentrations late in the pathways, and adjust the enzymatic activities at an earlier stage. Such feed-back inhibition helps to balance and stabilize the metabolite flows in the different pathways on a short time scale when enzyme concentrations are approximately constant [5, 51, 54]. The biosynthetic enzymes will in many cases also be subject to direct inhibition by their own products. However, at the level of simplification chosen in this model, no distinction will be made between feed-back inhibition in the beginning of the pathway and direct product inhibition late in the pathway.

The model below is most closely related to the model by Marr [55], with the major difference that we consider parallel pathways for supply of different amino acids, and that we neglect global responses mediated by ppGpp. Such global feed-fack systems are, however, natural extensions of the present model.

4.2 Model for amino acid synthesis, aminoacylation and protein synthesis

The following differential equations are used for the macroscopic analysis of the system described in figure 4.1

$$\begin{aligned} \frac{dx_i}{dt} &= J_{Ei}(x_i) - J_{Si}(x_i, y_i) \\ \frac{dy_i}{dt} &= J_{Si}(x_i, y_i) - f_i J_R(y_1, .., y_{20}) \end{aligned} \quad ; \text{i=1,2,..,20} \qquad (4.1)$$

x_i is the concentration of amino acid i, y_i is the concentration of aminoacylated-tRNA i. J_{Ei} is the rate of synthesis of amino acid i, J_{Si} is the rate of aminoacylation of the tRNAs for amino acid i and J_R is the rate of consumption of all amino acids in protein synthesis. f_i is the usage frequency of amino acid i in protein synthesis. At this level of simplification I neglect that there are several tRNA isoacceptors for most of the amino acids, a system which will be described in more detail in section 5. The rate laws for the different reactions in Eq. (4.1) are described below.
First, I will only discuss the scenario in which transcriptional regulation is disregarded and the capacities of the amino acid biosynthetic pathways to supply amino acids are held constant. This analysis applies to cases when the relaxation rate of enzyme concentrations, set by the rate of growth, is slow compared to the metabolite dynamics; an assumption frequently made in metabolic modeling [9, 24]. However, it will become evident that this assumption is not always justified because metabolite pools may have very slow relaxation rates.

4.2.1 Amino acid synthesis

The rate law used for amino acid biosynthesis is:

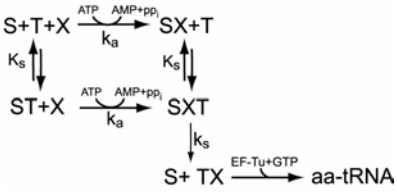
$$J_{Ei}(x_i) = \frac{k_{Ei}}{1 + x_i / K_I} \qquad (4.2)$$

This simple model of the amino acid supply module reflects two basic properties. Firstly, the rate is proportional to k_{Ei}, which is the capacity of the amino acid biosynthetic enzymes. Capacity is defined as the rate by which uninhibited enzymes produce amino acid. The capacity is proportional to the concentration of the enzymes that are active in the pathway, since the availability of substrates for amino acid synthesis is assumed to be constant in a specific environment. Secondly, the enzymes are feedback inhibited by free amino acid, with K_I as the inhibition constant [56].

4.2.2 Aminoacylation

For sixteen out of the twenty aminoacyl-tRNA synthetases in *E. coli*, the amino acid is activated to aminoacyl-adenylate (aa-AMP) independently of the presence of a transfer RNA molecule on the synthetase [57]. When the level of pyrophosphate in the cell is low, the energy rich aa-AMP molecule is bound to the synthetase in a stable complex until a transfer RNA molecule binds and aminoacyl-tRNA is formed and released [58]. When the binding of

aa-AMP is irreversible and deacylated transfer RNA equilibrates rapidly with the synthetase, the aminoacylation reactions scheme takes the form:



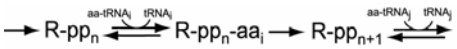
In the scheme, S is the synthetase, X is the amino acid, T is the tRNA and aa-tRNA is aminoacylated-tRNA in the ternary complex with EF-Tu and GTP. k_S is the maximal turnover rate of the enzyme when it is saturated with both tRNA and amino acid, K_S is the dissociation constant for the binding of tRNA to the enzyme and k_a is the association rate constant for the binding of amino acid to the enzyme multiplied with the probability that an aa-AMP complex is formed. The rate equation for aminoacylation derived from this scheme (see paper II) and used in Eq.(4.1), is

$$J_{Si}(x_i, y_i) = \frac{[S_i]k_S}{1 + (k_S/k_a)/x_i + K_S/(t_{0i} - y_i)} \tag{4.3}$$

Here, t_{0i} is the total concentration of tRNA and $t_i = t_{0i} - y_i$ is the concentration of deacylated-tRNA corresponding amino acid i. Alternatively, the concentrations of aminoacylated and deacylated tRNA will be written $\alpha_i t_{0i} = y_i$ and $(1-\alpha_i)t_{0i} = t_i$, where α_i is its fraction of tRNA that is aminoacylated.

4.2.3 Protein Synthesis

Protein synthesis is a multi-step mechanism, where incorporation of a tRNA-bound amino acid in a nascent polypeptide (pp_n) on a ribosome (R) is an irreversible step.



The average rate of protein elongation on ribosomes depends on the concentrations of all different ternary complexes in the cell. It is the inverse of the mean time (τ) to incorporate an amino acid. τ is an average over the mean times (τ_i ; i=1,2,..20) for incorporation of amino acid i, weighted by the codon usage frequencies f_i. By assuming Michaelis-Menten kinetics for each step, the rate law takes the form [59][I]:

$$J_R(y_1,..,y_{20}) = r\tau^{-1} = r\left(\sum_i f_i\tau_i\right)^{-1} = r\left(\sum_i f_i \frac{1 + K_R/y_i}{k_R}\right)^{-1} \tag{4.4}$$

r is the concentration of ribosomes in elongation phase. K_R and k_R are the K_m- and k_{cat}-values, respectively, for incorporation of amino acid from ternary complexes reading a cognate codon. The association rate constant for the binding of a ternary complex to the ribosome multiplied with the probability that binding is followed by peptidyl-transfer is the ratio k_R/K_R. The inverse of

k_R is the mean time it takes to hydrolyse GTP on EF-Tu, execute peptidyl-transfer,
translocate peptidyl-tRNA from A- to P-site and dissociate elongation factor EF-G from the ribosome [60]. K_R and k_R are here assumed to be the same for all amino acids.

4.2.5 THE PARAMETERS

Representative *in vivo* values of the rate constants are given in table 4.1. The references to these model parameters are given in paper X. As always, *in vivo* parameters are uncertain, and results that depend on finely tuned parameter sets can in general not be trusted. Analytical expressions are more useful as they can characterize system behaviors over large parameter regions.

Table 4.1

Parameter	Value
k_s	$100\ s^{-1}$
k_R	$20\ s^{-1}$
k_a	$10^6\ M^{-1}s^{-1}$
K_s	10^{-6} M
K_I	10^{-4} M
K_R	10^{-6} M
t_i	10^{-5} M
$[S_i]$	10^{-6} M
r	$1.7 \cdot 10^{-5}$ M
f_i	0.05

4.3 Flows and fluctuations in amino acid and aminoacyl-tRNA pools

When the rate of supply of at least one amino acid limits the rate of protein synthesis, the model in Eq. (4.1) has properties similar to those in the idealized system described in section 3. In both cases there are several irreversible synthesis pathways for substrates that are joined in a multi-substrate reaction. The multi-substrate reaction is in this case protein synthesis mediated by the ribosome and its substrates are the aminoacyl-tRNA in ternary complex with EF-Tu and GTP. There are, however, three important features that distinguishes the system in Eq. (4.1) from that in Eq. (3.2). In protein synthesis (*i*) there are twenty substrates and not only two; (*ii*) there are upper limits to the substrate pool sizes, set by the concentrations of tRNA; and most importantly (*iii*) aminoacyl-tRNA, cannot directly inhibit their own synthesis since they are sequestered in ternary complexes and feedback inhibition of amino acid synthesis is mediated by free amino acids and not aminoacyl-tRNAs.
The fact that there are many, rather than two, substrates does however not qualitatively change the system properties. A limiting substrate supply rate will still determine the consumption of all other substrates. To determine which amino acid supply that is limiting, the synthesis rates k_{Ei} are normalized to the maximal rate of amino acid consumption in protein synthesis $s_i = k_{Ei}/(f_i r k_R)$. For example when s_i=0.7 it implies that amino acid *i* is supplied at a rate that is 70% of the maximal capacity of protein synthesis. The amino acid supply with the lowest $s_i = s_{\min}$ will limit the rate of protein synthesis.
The major difference between the idealized system in section 3 and protein synthesis is the aminoacylation reaction. This reaction separates the amino acid pools, that are mediating feedback inhibition of the amino acid supply, from the aminoacyl-tRNA pools, that are the substrates in the multi-substrate

reaction. Separation of the feedback from the substrate functions by aminoacylation give the tRNA charged levels a switch-like behavior, so that the charged level of tRNA in the limiting pathway is very low while the charged level of tRNA in a non-limiting pathway will be high. This is illustrated in figure 4.2, where the numerically determined stationary concentrations of amino acids 1 and 2 and the corresponding aminoacyl-tRNAs are plotted for the parameters given in table 4.1. The concentrations are plotted as functions of the normalized supply rate of amino acid 1, s_1, whereas s_2=0.7 and $s_{3\text{-}20}$=1. The corresponding analytical approximations of the concentrations of amino acid 1 and aminoacyl-tRNA 1 described in paper X are also indicated in the figure.

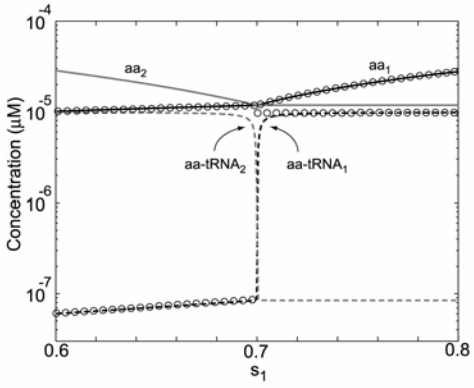


Figure 4.2. Amino acid and aminoacyl-tRNA concentrations for different synthesis capacities (s_1). The steady state concentrations for aa_1, aa_2, aa-$tRNA_1$ and aa-$tRNA_2$ at different synthesis capacities for aa_1 (s_1) when s_2=0.7 and $s_{3\text{-}20}$>0.7. The analytical approximations (paper X) are indicated with circles.

The reason why tRNAs in non-limiting pathways are fully charged with amino acid is the high affinity binding of deacylated-tRNAs to their synthetases. As a result, near 100% charging of a tRNA must be reached before access to deacylated-tRNA starts to limit the aminoacylation reaction. Thus, only at close to full tRNA charging will the amino acid pool start increasing in response to limited consumption over the synthetase. The amino acid pool increases until its synthesis is feedback inhibited to match the supply of the limiting amino acid.

At the same time the tRNA in the limiting pathway has a very low charged level which is required to reduce the average translation rate of the ribosome to match the limiting supply flow. Because the K_R-value for interaction of cognate ternary complexes with ribosomes is low compared to the total tRNA concentration and since only a small fraction of the codons will be cognate for the limiting amino acid, tRNA charging level in the limiting pathway has to drop very low.

At a point where two amino acid supply rates are equally limiting, $s_1=s_2=s_{min}$, there will be a dramatic switch in tRNA charging. An approximation of the sensitivity amplification in the concentration of aminoacylated-tRNA, to a change in amino acid supply at the switch point is given by

$$\frac{dy_1}{ds_1}\frac{s_1}{y_1} = \beta\underbrace{\left(\frac{dy_1}{dt_1}\frac{t_1}{y_1}\right)}_{a_{y1t1}}\underbrace{\left(\frac{dt_1}{dx}\frac{x}{t_1}\right)}_{a_{t1x}}\underbrace{\left(\frac{dx}{ds_1}\frac{s_1}{x}\right)}_{a_{xs}} \qquad (4.5)$$
$$= \beta\frac{t_{01}(1-s_1)}{(2s_1 fK_R)}\frac{k_S t_{01}[S_1]}{k_{E1}K_S}\frac{K_I k_a[S_1]}{k_{E1}}$$

This expression is derived in the appendix of paper X. It is seen that the sensitivity is a product of three factors (*i*) the sensitivity amplification, a_{y1t1}, in concentration of ternary complexes to a change in deacylated-tRNA, (*ii*) the sensitivity, a_{t1x}, in deacylated-tRNA to a change in free amino acid concentration, x, (*iii*) the sensitivity , a_{xs}, in free amino acid concentration to a change in supply. β is a constant that can be restricted to the interval [0.5,1], when the two limiting pathways are identical (paper X).
The factor a_{y1t1} is numerically large when the concentration of ternary complex is small in relation to the concentration of deacylated tRNA, which is the case if the balance point is significantly below saturation of the ribosome and if $fK_R << t_0$. The factor a_{t1x1} is large if the dissociation constant of deacylated tRNA is low compared to the total tRNA concentration ($K_S << t_0$) and the synthetase works below saturation $k_S[S] > k_1$. Finally, the factor a_{x1s1}, is large when x_1 is small compared to the feed-back inhibition constant K_I.
For the parameters in table 4.1 the sensitivity at the balance point ($s_1=s_2=0.7$) evaluates to >10000.

4.4 Near critical fluctuations in aminoacyl-tRNA concentrations

Figure 4.3 shows a Monte Carlo simulation of stochastic fluctuations in the two amino acid and aminoacyl-tRNA pools below, at and above the switch, described macroscopically in section 4.3. The assumption made when doing the simulation is that the metabolites equilibrate rapidly with their respective enzymes such that each reaction described in section 4.2 can be considered to occur in a single step.
The most striking feature of figure 4.3 is the huge and very slow fluctuations in tRNA charging level when $s_1=s_2=s$, corresponding to the switch point where both amino acid supply rates are equally limiting for protein synthesis. This corresponds to the fluctuations with balanced rates of synthesis of the two substrates in section 3.3. The fluctuations in aminoacyl-tRNA levels are correlated such that the flow into protein synthesis is nearly constant at all times, *i.e.* so that

$$J_{E1}/f_1 = J_{E2}/f_2 \approx J_R(y_1, y_2, y_{3-20} = 1)\,. \qquad (4.6)$$

The rate of consumption of one aminoacyl-tRNA is zero order on this curve, *i.e.* an increase in one charged level can be compensated for by a decrease in another such that the overall flow is unchanged. Therefore, any combination of y_1 and y_2 that satisfies Eq. (4.6) would have same probability unless the rate of aminoacylation is reduced when the charged level takes high values.

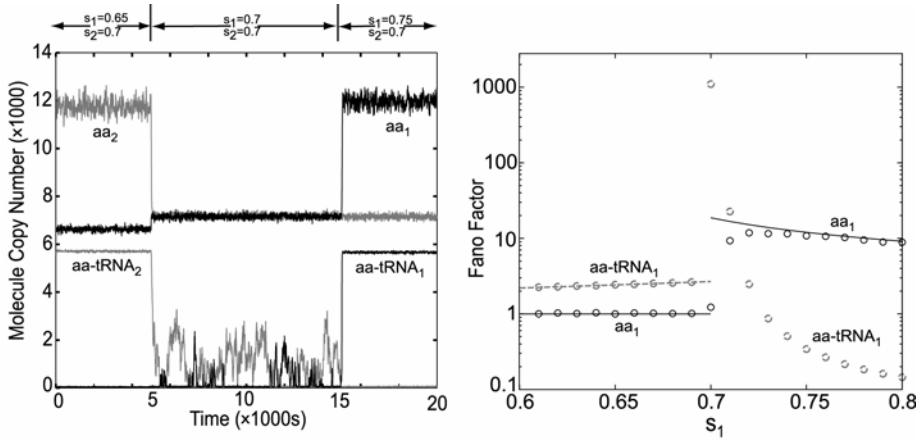


Figure 4.3 Near-critical fluctuations in pools of amino acid and aminoacylated tRNA. *left*: Stochastic trajectories of the system as simulated with the Gillespie algorithm. The value of s_1 and s_2 in different regions is seen in top of the figure and $s_{3\text{-}20}$=1. The fluctuations in tRNA changing levels huge and anti-correlated due to stoichiometric flow coupling at the critical point (s_1=s_2). *right*. The circles are the Fano-factors (variance/mean) for amino acid 1 and ternary complex 1 as estimated by Gillespie simulation. The corresponding LNA estimates are given by solid lines. For $s_1 < s_2$ the LNA estimates are equal to the sensitivity. For $s_1 > s_2$ the fluctuations in the ternary complex pool are very small and the fluctuations in amino acid pool are approximately twice the sensitivity. The reason is explained in paper X. At the balance point the fluctuations are very large and direct application of LNA fails.

In section 2.2.4 (example 3) an expression for the variance in the difference between two substrates at the point of balanced synthesis was derived with the LNA. It was found that the variance is approximately the flow through one of the pools divided by the first derivative in its supply rate with respect to the pool size. For the concentration difference in between two limiting ternary complexes this would give

$$\frac{\sigma^2_{y_1-y_2}}{\Omega y_1} \approx \frac{k_{E1}}{y_1\left(\frac{dJ_{E1}}{dy_1}\right)} = \frac{k_{E1}}{y_1}\frac{dy_1}{dk_{E1}} = \frac{s_1}{y_1}\frac{dy_1}{ds_1}. \qquad (4.7)$$

Eq.(4.7) is equal to the sensitivity amplification given in Eq. (4.5), except for the marginal effect of the β-factor. This suggests that the fluctuations are huge and, more importantly, that they depend on the same parameters as the sensitivity in Eq. (4.5). The LNA-estimate in Eq. (4.7) is however based on a local linearization that over-estimative the fluctuations since these are globally limited by the total concentrations of tRNA.

As shown in figure 4.3, the ternary complex concentrations are ultra-sensitive at the point of balanced, but limiting supply, of amino acids. This is due to (*i*) zero order coupled consumption in protein synthesis and (*ii*) close to zero order aminoacylation kinetics below full charging of tRNA. The zero order

behavior implies very slow relaxation back to steady state when the pools of aminoacyl-tRNA are perturbed. The relaxation rate constant is essentially given by dJ_{Ei}/dy_i, which is in the order of the rate of dilution through cell growth that also determines the timescale for changes in the enzyme concentrations.
As a consequence, the assumption of well separated timescales between metabolite and enzyme dynamics may not be justified and the transcriptional regulation of amino acid biosynthetic genes must be considered. This is complicated for several reasons. The major problems from a modeling point of view are the time delays between sensing the metabolite signal and the synthesis of new enzymes, and the stochastic aspects of gene-expression [22, 61]. However, an attempt to investigate stationary properties of the transcriptional control systems was made in paper II, and an attempt to address the dynamics of transcriptional control of amino acid biosynthetic operons are given in the end of paper X. Full understanding of the dynamical aspects of transcriptional coordination of amino acid biosynthetic operons is however lacking. The clarification will require more experimental work, probably at the level of single cells.

4.5 Robustness in coordination of transcriptional control*

In this section I will outline a principle for robust transcriptional regulation of enzymes that synthesize substrates for multi-substrate reactions. Let us consider a simplified scenario where a number of enzymes make products that are substrates for an irreversible unsaturated multi-substrate reaction. Further, the capacity to make the product increases with increasing concentrations of the enzymes, but not necessarily linearly. We will temporarily disregard the inhibition of the enzymes by product inhibition. In this case, the rate of synthesis by the enzyme with the lowest capacity, in relation to stoichiometry in the multi-substrate reaction of its products, will determine the rate of the multi-substrate reaction.

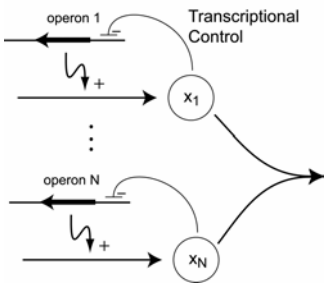


Figure 4.4 Robust control for optimal enzyme investment. Stoichiometrically coupled exit-flows cause transcriptional down regulation of all pathways, except the one making the limiting substrate. This is because the non-limiting metabolite pools operate as integrators of the difference between their supply rates and the limiting supply rate.

If we now want to minimize the total investment in enzymes for a certain rate of the multi-substrate reaction, we must choose the concentrations of the enzymes such that the enzymes' capacities are in proportion to the stoichiometry of their product in the multi-substrate reaction. This can be seen by the following argument: When the rates of synthesis of substrates are unbalanced, the concentrations of the enzymes responsible for the excess capacity can be reduced to the balance point at an unsaturated rate of the multi-substrate reaction The opposite is also true, *i.e.* for a fixed total concentration of enzymes the rate of the multi-substrate reaction is maximized by adjusting the individual concentrations such that the capacity per stoichiometry ratio is equal for all substrate producing enzymes.

Now the following question arises: how can transcriptional regulation of the different enzymes adjust the enzyme concentrations to their optimal levels without knowledge of the concentrations of all other enzymes, their different relations between concentration and synthesis capacity, and the stoichiometry of all substrates in the multi-substrate reaction? There is a very simple and extremely robust way to achieve optimal global balance between enzyme concentrations and it is built on the principle of integral feed-back [62, 63]. The trick is to let transcriptional regulation for individual biosynthetic enzymes respond to the concentrations of their products in such a way that transcription is down-regulated if the product concentration increases. The shape of the transcriptional response function is irrelevant as long as it is monotonically decreasing with increasing product concentration.

This simple magic works because the concentrations of non-limiting substrates build up over time, such that any unbalance in synthesis capacity per stoichiometry ratio, no-matter how small, will create strong signals to down regulate the non-limiting synthesis after some time. Only the case when the enzymes capacities are balanced to the stoichiometry in the multi-substrate results in non-increasing control signals.

The control structure where the enzyme's products control the transcription of their own biosynthetic enzymes is ubiquitous. That this structure robustly maximizes the flow over a multi-substrate reaction per enzyme investment has not to my knowledge been described before.

The product inhibition of the biosynthetic enzymes does however make the picture more complicated, and it can in the worst case destroy all the good properties of the integral feedback control. For instance if the enzymes are inhibited at a lower concentration of product than the one that down regulates transcription this could result in the very unfortunate situation that the cell operates with severely inhibited enzymes. On the other hand, as long as transcription is down-regulated for lower product concentrations than what it takes to get strong product inhibition, all the positive coordination effects of integral feedback control still remain, particularly the robustness to variations in the kinetic details of transcriptional control, substrate synthesis and consumption in the multi-substrate reaction.

5. SELECTIVE CHARGING OF tRNA ISOACCEPTORS

The fact that most amino acids have several tRNA isoacceptors that read partly overlapping codons is not included in the model outlined in section 4 above. In this section, the theory is extended to cases when the rate of supply of one type of amino acid limits the charging of several isoacceptors that read different codons.

5.1 What about the isoacceptors?

When different tRNA isoacceptors have the same kinetics in aminoacylation, the isoacceptors are charged in proportion to their deacylated concentrations. This is formalized as

$$J_{Sij} = J_{Si}\frac{(1-\alpha_{ij})t_{ij}}{\sum_j (1-\alpha_{ij})t_{ij}} \tag{5.1}$$

The subscript i identifies the amino acid and subscript $j=1..n$ the isoacceptor. J_{Sij} is the rate of aminoacylation of isoacceptors ij. t_{ij} is the total concentration of this isoacceptor and α_{ij} is its level of charging with the amino acid. J_{Si} is, as before, the total rate of aminoacylation by amino acid i.
A similar relation can be written for the deacylation flow into the ribosome, where the aminoacyl-tRNAs compete during translation of their cognate codons:

$$J_{Rij} = \sum_k f_{ik}\frac{k_{jk}\alpha_{ij}t_{ij}}{\sum_j k_{jk}\alpha_{ij}t_{ij}} j_R \tag{5.2}$$

The parameter k_{jk} is k_{cat}/K_m for translating a codon of type k in the ribosomal A-site by a ternary complex with tRNA isoacceptor j. k_{jk} is assumed to be zero for non-cognate interactions. Codon k of amino acid i has codon usage f_{ik}. In stationary state the aminoacylation flow and the deacylation flow must be equal for each isoacceptor, *i.e.* $J_{Sij} = J_{Rij}$ for all j. This results in n-1 independent pair-wise relations between different isoacceptors (j, m).

$$\frac{(1-\alpha_{ij})t_{ij}}{(1-\alpha_{im})t_{im}} = \frac{\sum_k \left(f_{ik}k_{jk}\alpha_{ij}t_{ij} \Big/ \sum_j k_{jk}\alpha_{ij}t_{ij} \right)}{\sum_k \left(f_{ik}k_{mk}\alpha_{im}t_{im} \Big/ \sum_m k_{mk}\alpha_{ij}t_{ij} \right)} \tag{5.3}$$

When each isoacceptor only reads one codon, Eq. (5.3) reduces to

$$\frac{(1-\alpha_{ij})t_{ij}}{(1-\alpha_{im})t_{im}} = \frac{f_{ij}}{f_{im}} \tag{5.4}$$

In Eq. (5.4) isoacceptors j and m read codons with codon usage f_{ij} and f_{im}, respectively.
Eq. (5.3) or (5.4) defines flow balance constraints on the relations between charged levels of different isoacceptors imposed by codon usage in protein synthesis and the total concentrations of the different isoacceptors. The principle is illustrated by the cartoon in figure 5.1.

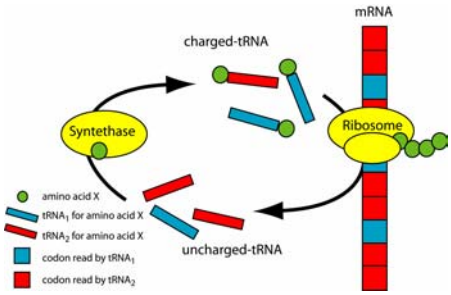


Figure 5.1 Flow balances and the isoacceptors. C*harging*: tRNAs that read synonymous codons are charged with the same amino acid in proportion to the concentrations of uncharged tRNAs. *Protein synthesis*: Charged tRNAs are consumed in protein synthesis in proportion to the usage of their codons. Flow balance requires that the ratio between uncharged tRNAs that read synonymous codons is equal to the ratio between the frequencies of the codons they read. C*onsequences*: The rates of reading of some codons are sensitive to starvation of the amino acid, while the rates of reading other codons for the same amino acid are insensitive. In the example shown in the figure are blue codons read at a higher rate than red codons.

When there are n different isoacceptors Eq. (5.3) gives n-1 independent equations. To determine the absolute values of the charged levels one additional equation is needed. This is supplied by Eq. (4.4) that determines the total flow of the limiting amino acid into protein synthesis

$$s_i = \frac{J_{Ei}}{f_i r k_R} = \left(1 + \sum_{\substack{q=1..20 \\ q \neq i}} \left(f_q K_R / t_q \right) + \sum_k \left(f_{ik} K_R \Big/ \sum_{j_k} \alpha_{ij} t_{ij} \right) \right)^{-1} \tag{5.5}$$

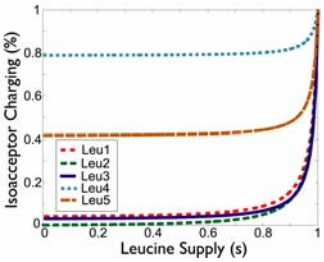
The flow of the limiting amino acid i is divided by the maximal capacity of protein synthesis, which defines the degree of saturation s_i[‡]. Eq. (5.5) includes the additional parameter K_R, which determines how far the charged levels drop at a particular level of amino acid limitation.

[‡] With this definition of s, it can never reach s=1, as the total concentrations of tRNA limits the maximal translation rate to $k_R/(1+K_R/t_0)$.

5.2 Selective charging in the Leu-family*

When the concentrations of isoacceptors and the codon usage frequencies are known, the flow balance equations (5.3) and (5.5) can be used to predict how the charging of individual isoacceptors responds to limiting supply of their cognate amino acid. For *E. coli* these parameters have been estimated for several different growth conditions [64]. I will use data for the leucine family as an example, because there are as many as six different Leu codons, which are read by five different isoacceptors. The estimated tRNA concentrations, codon frequencies and the predicted charged levels at different degrees of starvation are given in figure 5.2.

tRNA	Genes	Anti-codon	Codon-recognition	Copy Numbers (t_{ij})
Leu1	*leuPQTV*	CAG	CUG	4470
Leu2	*leuU*	GAG	CUC, CUU	943
Leu3	*leuW*	UAG	CUA, CUG	666
Leu4	*leuX*	CAA	UUG	1913
Leu5	*leuZ*	UAA	UUG, UUA	1031



Codon	Codon usage (f_{ij})
UUG	.0063
UUA	.0061
CUG	.0601
CUC	.0062
CUU	.0057
CUA	.0022

Figure 5.2 Selective charging of leu-isoacceptors. Predictions of selective charging of leucine tRNA-isoaccptors at different leucine supply rates. The amount of the different isoacceptors are estimated from *E. coli* K12 grown at 0.4 doublings h^{-1} [64] and the codon recognition [65] is given in the table together with the estimated codon usage at the same growth rate.

It can be seen that the Leu4 and Leu5 isoacceptors keep high residual charging even when translation is completely stalled due to leucine starvation. The reason is in this case that the concentrations of the Leu4 and leu5 isoacceptors are high compared to the frequencies of the codons that they read. This means that they can be recharged by the synthetase although Leu1, Leu2 and Leu3 have lower relative charging. To put it in an other way, the average rate of translation is limited by the supply of amino acid leucine; the same average translation rate can be achieved by stalling on different leucine codons; what codons is given by the flow balance equations Eq.(5.3).
As a consequence of selective charging the cell can evolve different codon usage in different positions in the genome depending on if translation should stall under amino acid limitation or not. In paper I it is described how this may have shaped codon usage in genes for amino acid biosynthetic enzymes, where it is bad to stall during amino acid limitation and in the leader peptide

coding sequence used for ribosome mediated transcriptional attenuation, where it is essential that the ribosome stalls under amino acid limitation.
The possibility for selective charging of isoacceptors under amino acid limitation, adds a dynamic aspect to the genetic codon, since "synonymous" codons that are translated at the same rate in one medium can be translated at different rates in another medium. This would imply that the genetic code is not as redundant or degenerate as has been suggested [66] from the fact that 20 amino acid residues are encoded by 61 sense codons. Judging from the examples described in reference I, the extra information seems to be used in regulatory responses to amino acid limitation.
The theory for selective charging outlined in section 5.1 includes a large number of assumptions concerning how the processes of aminoacylation and translation works, and simplifications such that tRNA bound to ribosomes and synthetases are omitted and that ternary complexes form very rapidly after aminoacylation. The predictions from the theory are however clear and the theory can therefore be tested by experiments.
The major prediction is that charged levels of different isoacceptors will respond differently to starvation of the cognate amino acid and that the differences depend on what codons they read, the codon usage of these codons, and the concentrations of the isoacceptors. The most direct way to test these predictions is to measure charging of individual isoacceptors under amino acid limitation. It should be possible to detect selective residual charging of some isoacceptors also at severe starvation. Such selective charging has not been investigated before. Secondly, if selective charging is observed, it should be possible to change the charging pattern in a predictable way by adjusting isoacceptor concentration or codon usage.
These experiments are not easy, but with the expertise of Michael Sörensen, Copenhagen, it has been possible to investigate the isoacceptor charging patterns of the leu-family in response to starvation. The results will be presented elsewhere (Sörensen at al., *in preparation*), but they strongly support the theory.

5.3 Coupled fluctuations in charging of isoacceptors*

What is the distribution of tRNA charging levels given a fixed amino acid supply rate? In section 4 we have seen that is can be very broad if the system operates close to balanced synthesis of two limiting pathways. However, to take the concept of coupled zero order fluctuations one step further I will introduce another phenomenon that can cause large fluctuations in the charging levels of the limiting tRNA even if the supply of different amino acids are unbalanced. The phenomenon arises if the concentrations of several isoacceptors are well balanced to their codon usage frequencies, such that they lose their charging simultaneously during limitation of their cognate amino acid. Under such conditions the deacylated forms of the isoacceptors will dominate, which makes the aminoacylation reaction close to zero order with respect to isoacceptor charging level. This implies that the rate of aminoacylation will be insensitive to large relative variations in the low concentrations of the aminoacylated isoacceptors. As there are many different combinations of charging levels of the different isoacceptors that result in the same reduction in translation rate, we get another example of coupled zero order fluctuations.

For example, assume that there are two isoacceptors for the limiting amino acid that can read one codon each. If they have approximately equal concentration per codon usage, the charging levels of the isoacceptors can vary freely on the curve given by Eq.(5.5), for which the translation rate equals the limiting supply. The resulting very broad distribution of isoacceptor charging levels is illustrated in figure 5.3. The peculiar "bi-phased" shape of these probability distributions are due to that the isoacceptor is the one limiting in the lower range, whereas the other isoacceptor is limiting in the higher range.

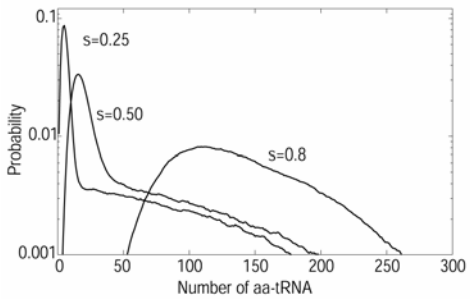


Figure 5.3. The stationary distribution function for the copy number of one of two aminoacylated isoacceptors is shown for different supply rates, s, of their cognate amino acid. The probability distributions were determined by Monte Carlo simulations using Gillespie's direct method [17]. The parameters are as described in section 4.2.5, and the codon usage of the starved codons are 0.05.

These fluctuations implies that balanced aminoacylation of different isoacceptors can not be achieved by fine tuning of tRNA concentrations and codon usage, as may be expected from the macroscopic analysis in paper I. In this situation the starvation, can not be directed to selected codons, and it is therefore likely to have a less pronounced effect on codon usage. In the next section we will see that the fluctuations in isoacceptors charging levels may instead contribute to increased sensitivity in the mechanism of transcriptional attenuation by the principle of stochastic focusing [67].

6. RIBOSOME MEDIATED ATTENUATION OF TRANSCRIPTION

I will use the name attenuation for the specific mechanism of ribosome mediated transcriptional attenuation that is described below in section 6.1. This mechanism should not be confused with any of the other control mechanisms that operate by very different principles but also are called attenuation [68].
Attenuation is a ubiquitous mechanism for transcriptional control of amino acid biosynthetic operons [69]. The question addressed here is how attenuation works in its intracellular context described in chapters 4 and 5. The focus is on what makes attenuation sensitive to insufficient supply of amino acid. It turns out that the attenuation mechanism is a beautiful example of how the sensitivity amplification in a molecular control system can depend on a combination of several different mechanisms.

6.1 Introduction to attenuation

Several amino acid biosynthetic operons are under control of attenuation. The mechanism makes the probability that the RNA polymerase (RNAP) synthesizes a full length transcript of the operon dependent on the rate of codon translation. More specifically, the outcome of an initiated transcription depends on a race between the RNAP transcribing the leader of the regulated operon and a ribosome translating this leader transcript. The leader contains a number of "own" codons for the amino acid that is synthesised by the enzymes encoded by the mRNA [53]. If the supply of the amino acid is insufficient to meet the demand from protein synthesis, the ribosome will be slowed down on these codons and transcription will continue into the coding sequences of the operon. If, in contrast, the amino acid supply is in excess, the ribosome will move fast over the own codons and this will allow formation of a secondary structure in the leader that signals termination of transcription (attenuation). The ribosome mediated transcriptional attenuation was first found in the *trp*- and *his*-operons of *Escherichia coli* [53, 70] and *Salmonella Typhimurium* [71, 72] respectively. The attenuation mechanism has later been found to control the expression of the *leu*, *thr*, *ilvGMEDA*, *ilvBN* and *pheA* operons of *E. coli* and *Salmonella* [69].
A scheme for attenuation control of the *trp*-operon, as unveiled by Yanofski and co-workers [69, 73], is shown in figure 6.1. The mechanism has been extensively studied for the *trp*-operon, but is believed to be very similar for the other operons. Our model follows this scheme closely. The leader sequence contains, starting from the 5´-end, the start signal AUG for mRNA translation followed by region *I* in which there are *m* "own" codons for the amino acid that is synthesised by the enzymes in the controlled operon as well as other codons. Region *I* is followed by region *II* of the leader and then by a strong pause site for the RNA polymerase (RNAP). Further downstream, there are *n* transcriptional steps subdivided in regions *III* and *IV* of the leader.

When the RNA polymerase has reached the pause site, it stops and remains there until a ribosome starts melting the hairpin structure formed by regions *I* and *II* [74-76]. The polymerase is released and resumes transcription, moving forward in synchrony with the ribosome. If the ribosome is slow in translating codons of region *I* due to deficient amino acid supply, it will remain in the "control region" *I* of the leader when the RNAP finishes transcription of region IV (Figure 6.1). In this case the *II:III*, but not the *III:IV*, hairpin is formed and the RNA polymerase will continue into the open reading frames of the operon. When, in contrast, the amino acid synthetic activity of the enzymes encoded by the operon is sufficient or in excess, the ribosome will move fast over the codons in region *I* and prevent formation of the anti-terminator loop *II:III*, when the RNAP finishes transcribing region *IV*. In this case the hairpin *III:IV*, which signals rho-independent termination of transcription, will be formed so that transcription is aborted (attenuated) before the RNAP starts transcribing the coding sequences of the operon.
If the ribosome reaches the stop-codon and releases the leader peptide and RNA before the RNAP has escaped termination, the anti-terminator may form although translation is fast over the control codons. The probability of this event determines the basal expression level of an attenuation controlled initiation [77]. Here, I will disregard this mechanism for forming the anti-terminator and focus on the principles that make attenuation sensitive under amino acid limitation. It is, however, essential to include the possibility of read-though when the ribosome leaves the stop codon to correctly describe the basal expression of some attenuation mechanisms (J. Elf *manuscript in preparation*).

6.2 The model of attenuation

The probability, Q, that transcription is attenuated, *i.e.* that transcription of the operon is aborted, is given by

$$Q = \int_0^\infty R(t,m)P(t,n)dt \qquad (6.1)$$

Time zero is when the RNAP resumes transcription from its pausing state under the influence of an approaching ribosome. $R(t,m)$ is the probability that that the ribosome has left the control region with m codons at time t. $dt \cdot P(t,n)$ is the probability that the polymerase leaves the n:th base counted from the pause site and therefore escapes attenuation between time t and $t+dt$. The number n is operationally defined as the difference in the number of bases between the attenuated transcript and transcripts that are terminated at the pause site. The probability that transcription successfully continues into the structural genes is given by 1-Q.

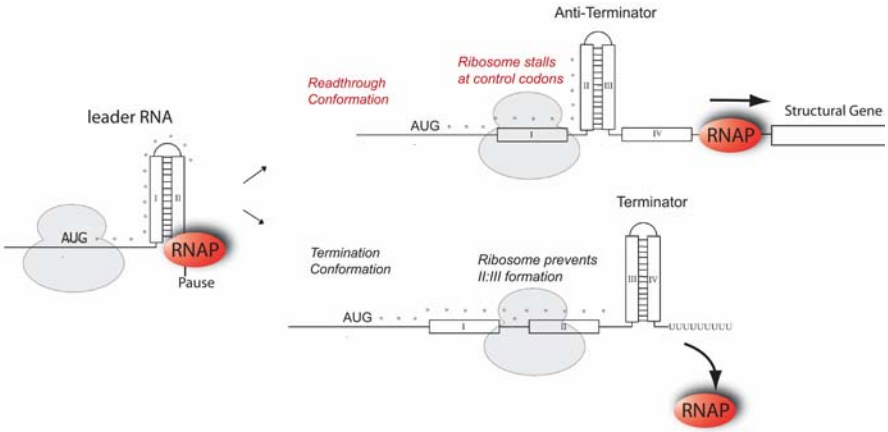


Figure 6.1 The leader-transcript of the *trp*-operon in *E. coli.* Attenuated transcription results in a 141nt leader RNA. Aborted transcription of the paused RNAP results in a 91 nt transcript. The number of nucleotides that have to be transcribed after pausing to escape attenuation are set to be the difference between these values (n=141-91=50). The transcript includes an open reading frame of 15 codons, encoding a very short lived 14 residue leader peptide. The RNAP is believed to be released from the pause site when the fifth codon is read. Two of the three codons in the control region are trp-codons (no 10, 11). Also ribosome stalling on the arg–codon (no 12) prevent *I:II*-hairpin formation and attenuation. After reaching the stop-codon (UGA) the ribosome disassociate in about 1s. The *II:III* and *III:IV* conformations are approximately equally stabile and the attenuator forms in 50% of the cases after release of the ribosome. This determines the basal read trough level of 10-15%. Under severe tryptophane starvation attenuation is totally relieved. This results in a dynamic range of 6-8 fold of the attenuation system. The range of the repressor system that also controls the trp-operon is 70-fold. In total the range of transcriptional control is about 500-fold.

In order to focus on principles and not on detail, I will assume that each of the m codons in the control region is translated at rate k, which only depends on the availability of aminoacylated-tRNA cognate to the controlled amino acid. I assume further that translation of a codon can be characterised by an exponentially distributed time with average $1/k$, which makes the ribosome stepping through the control region a Poisson process and R is given by

$$R(t,m) = 1 - \sum_{i=0}^{m-1} \mathrm{Po}(i,kt), \tag{6.2}$$

where $\mathrm{Po}(i,\lambda) = \lambda^i e^{-\lambda}/i!$. Similarly, transcription of the n bases in regions *III* and *IV* will be modelled in n identical steps, each characterised by an exponentially distributed time with average $1/q$. Exponentially distributed step-times for ribosomes and RNA polymerases correspond to assuming that both translation of a codon and transcription of a base have one rate limiting

step, which can be described by a single rate constant, k or q, respectively. With these simplifications, the probability $P(t)$ is given by

$$P(t,n)=q\cdot \mathrm{Po}(n-1,qt) \tag{6.3}$$

This model has the same basic structure as the model presented by Manabe in 1981 [78], except that the pause site, that synchronizes the polymerase and ribosome, has been included. The pause site was not known at the time of Manabe's model, but it makes a fundamental difference for the sensitivity of the mechanism, as the ribosome jump process otherwise would be dominated by a single step corresponding to waiting for the ribosome to initiate translation at the AUG codon. There have been other attempts to include the RNAP pause site into the Manabe model [79, 80], but these missed the crucial point that RNAP actually waits for the ribosome, which is critical for the synchronization between ribosome and RNAP and therefore for the whole mechanism to work properly.

6.3 What makes the transcriptional attenuation mechanism sensitive?*

There seems to be a major drawback of the whole attenuation mechanism, in the sense that it can not possibly respond to a limited rate of supply of an amino acid before peptide elongation is slowed down. In the book [81] by Ingraham, Maaloe and Neidhardt from 1983 they describe this as a paradox. The question they ask is how an operon, solely controlled by attenuation, can be sufficiently expressed to ensure a high growth rate. The problem is that high expression requires slow ribosomes, which is not compatible with high growth rate. The solution, they suggest, is that attenuation may be very sensitive, such that the protein synthesis only needs to be slowed down marginally to activate expression of attenuation controlled operons. And, indeed, as we will see in this section, there are several features in the attenuation mechanism that can make it a hyper-sensitivity control system if used correctly. However, I believe that the potential for extreme sensitivity in the attenuation mechanism is only part of the solution to the "Ingraham-Maaloe-Neidhardt paradox". The other part being, that attenuation controlled operons may be expressed in bursts such that the rate of protein elongation is not limited of amino acid most of the time. This is a hypothesis that must be addressed experimentally.

6.3.1 Sensitivity Amplification

Using the theory outlined in sections 4 and 5 we can now ask what will happen to the probability Q of attenuation when the rate of supply of the controlled amino acid decreases. This is the relevant biological control signal for this control system since it is the amino acid supply provided by the biosynthetic enzymes that is regulated.

I will discuss the factors which make the attenuation mechanism sensitive to changes in the rate of supply of the amino acid. The sensitivity is quantified in terms of the sensitivity amplification a_{Qs} [5, 38] (compare Eq. (3.4)), which is the relative change in Q in response to a relative change in the normalized rate s ($=k_E/fk_Rr$) of amino acid synthesis at steady state. That is

$$a_{Qs} = \frac{dQ/Q}{ds/s} = \frac{\ln Q}{\ln s} \tag{6.4}$$

This can be interpreted as the percent reduction of transcription of the operon when the rate of amino acid supply increases by one percent.
I will describe four independent factors that contribute to high sensitivity amplification of the attenuation mechanism.

6.3.2 DOUBLE MULTISTEP SENSITIVITY AMPLIFICATION

The double multistep sensitivity amplification that is intrinsic for attenuation will be presented it in the context of ordinary multistep sensitivity amplification [82]. Consider first a process where transcription is attenuated if the ribosome reads one codon at rate k' before the polymerase reads one nucleotide at rate q'. The probability of attenuation is in this case simply

$$Q = \frac{k'}{k'+q'} \tag{6.5}$$

The sensitivity to a change in the rate of translation is thus

$$a_{Qk} = \frac{dQ}{dk}\frac{k}{Q} = \frac{q'}{q'+k'} < 1, \tag{6.6}$$

which asymptotically approaches is largest value $a_{Qk} = 1$ when $k' \ll q'$, *i.e.* when Q is very small. To get to an ordinary multi-step scenario, it is required that the ribosome translates m codons in a row at a rate of $k = k' \cdot m$ instead of one codon at rate k'. The probability for attenuation of transcription is in this case

$$Q = \left(\frac{k}{k+q'}\right)^m \tag{6.7}$$

The sensitivity is

$$a_{Qk} = \frac{mq'}{q'+k} = \frac{mq'}{q'+mk'} \tag{6.8}$$

In the limit that $m \to \infty$ there will be a well defined time delay ($1/k'$) before the ribosome has finished reading the codons. The probability of attenuation will then only depend on if the single polymerase reaction occurs before or after this time delay, *i.e.* $Q = e^{-q'/k'}$ and $a_{Qk} = q'/k'$. An ordinary multi step mechanism can therefore get high sensitivity ($a \gg 1$) if the number of steps, m, is high and Q is low ($k' \ll q'$).
In the attenuation mechanism, also the polymerase movement is a multi-step process. If the number of steps in the polymerase process, n, increases together with the rate of each step $q = q' \cdot n$, the distribution, Eq.(6.3), for the time it takes to transcribe the n bases gets more and more narrow. As a consequence, the time over which $R(t,m)$ is averaged in Eq. (6.1) becomes shorter and shorter. Therefore, when R makes a sharp transition from 0 to 1, due to many translation steps m, the sensitivity is unbounded also at high Q. In the limiting case where both $m \to \infty$ and $n \to \infty$ there is no uncertainty in the time it takes to translate the control codons or to transcribe the bases. In

this case, Q(*k*) is a step function with infinite sensitivity for all Q. In real attenuation systems in *E. coli* n=50-100 and m=2-16. An explicit expression for the sensitivity amplification are in this case too complicated to be informative. Figure 6.2 does, however, illustrate system behavior.

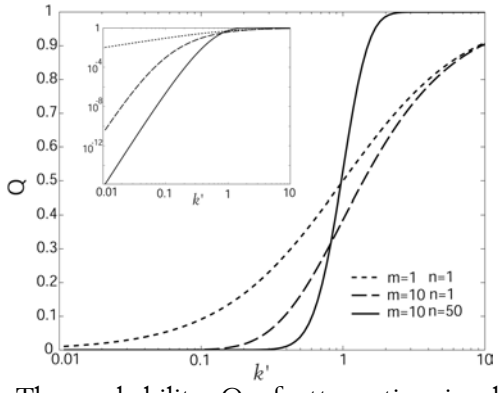


Figure 6.2 The probability Q of attenuation is plotted as a function of $k'=1/\tau$, where τ is the average time for translation of all the *m* codons. The translation rate of the individual codons is $k=k'\cdot m$. The average time to transcribe the *n* nucleotides is kept constant at 1/q'=1s. The transcription rate for an individual nucleotide is $q=q'\cdot n$. For *m*=1, *n*=1 the response function is given by Eq.(6.5), for *m*=10, *n*=1 the response function is given by Eq.(6.7). These curves should be compared to the much more sensitive response that is reached with two multi-step processes, as in attenuation (*e.g. m*=10, *n*=50). The *insert* shows the same curves as the main figure but in a log-log scale, where the slopes correspond directly to the sensitivity amplification.

6.3.3 Sensitivity amplification caused by stoichiometrically coupled flows

In order to estimate a_{Qs} one also has to consider how sensitive the rates of translation of an individual codon *k* is to changes in the rate of amino acid supply $s_i=k_E/fk_Rr$; that is a_{ks}. The sensitivity in *k* to a change in *s* can then simply be multiplied by the sensitivity in Q to changes in *k* to get the sensitivity $a_{Qs} = a_{Qk}a_{ks}$.

The first thing to notice is that the rate *k* of translation is constant as long as the tRNA is fully charged. Therefore, attenuation will only be sensitive when tRNA charging decreases, as when the rate of supply of one amino acid is limiting, *i.e.* s<1. Furthermore, the analysis in chapter 4 shows that the rate of supply with the lowest s (=s_{min}) determines the rate of protein synthesis, whereas tRNAs for other amino acids are nearly completely charged. The whole reduction in translation rate must therefore depend on ribosomes stalling at a small fraction *f* of the codons, as given by:

$$s_{\min} = \frac{1}{k_R\left(\frac{f}{k} + \frac{(1-f)}{k_{\max}}\right)} \qquad (6.9)$$

Eq.(6.9) follows directly from Eq.(4.4), when the translation rate of the codons read by fully charged tRNAs is $k_{max}=k_R(1+K_R/t_0)$. The translation rate of codons for the limiting amino acid is k and their frequency is f. This gives transcriptional attenuation an additional sensitivity amplification that evaluates to

$$a_{ks} = 1 + \frac{k}{k_{\max}} \frac{(1-f)}{f} \tag{6.10}$$

Because f is small [0.02-0.1], this effect contributes a factor between 10-50 to the sensitivity amplification, when the rate limiting pathway is close to, but below, saturation.

6.3.4 Sensitivity amplification caused by selective charging of tRNA isoacceptors

The codon bias in the attenuation leader sequences has long been suggested to be related to the sensitivity of the mechanism [69, 79]. The striking fact that the leu-operon leader sequence in *E. coli* uses four rare CUA codons suggested rare codons to be preferable, and it is a common misconception that this is the case. An equally striking counter example is the *thr*-operon, that uses 8 major ACC codons.

From section 5 about selective charging of tRNA isoacceptors it is, however, clear that only codons, read by tRNA isoacceptors that lose their charging during amino acid limitation, should be used. Right choice of synonymous codon can contribute additional sensitivity amplification whereas the wrong choice can eliminate all sensitivity.

Until now k has been defined as the average rate of translation of synonymous codons for the limiting amino acid, irrespective of which one is used in the leader sequence. To analyze the consequence of differences between synonymous codons, Eq. (6.9) must be replaced with the more complete Eq. (5.5), which in combination with Eq. (5.3) will give the charging levels of individual isoacceptors for different levels of starvation. Next, the rate of translation of the codons in the attenuation leader is evaluated as

$$k = k_R \Big/ \left(1 + K_R \Big/ \sum_j \alpha_j t_j \right), \tag{6.11}$$

where the sum is taken over all isoacceptors that can read the codon. When the amino acid has several isoacceptors, some of which read the control codons, this gives a fairly complicated expression for a_{ks} that replaces Eq.(6.10).

To estimate the maximal contribution that selective tRNA charging can give to the sensitivity amplification one can consider the limit where there is high abundance of all isoacceptors other than the one that reads the control codon. In this limit, only the isoacceptor that reads the control codon loses its charging and Eq. (6.10) is correct if the frequency f is taken to be the frequency of codons read by that specific isoacceptor. This will increase the sensitivity by a factor of f_i/f_{ij}, where f_i is the total codon usage of the amino acid and f_{ij} is the codon usage of the attenuation codon. It should be noted that this sensitivity amplification is associated with a cost of making large

amounts of tRNAs that read other synonymous codons. The lower f_{ij}, the higher tRNA concentrations must be maintained.

6.3.5 STOCHASTIC FOCUSING

Finally, an additional sensitivity amplification can be obtained by stochastic focusing [67]. If the translation rate of the control codons, k, varies due to fluctuations in the charged cognate tRNA, the probability of attenuation will be different for each transcription initiation attempt also at a fixed levels of amino acid supply. If we assume these fluctuations to be slow enough so that the rate of translation is constant during translation of the leader, but fast enough to be uncorrelated at different initiation attempts, the probability of attenuation, Q, at a certain amino acid supply, s, evaluates to

$$Q(s) = \int_k Q(k)p(k|s)dk \tag{6.12}$$

$p(k|s)$ is the probability density of the translation rate of the control codons conditional on the normalized rate of amino acid supply. In the simplest attenuation scenario, described by Eq.(6.5), this means that

$$Q(s) = \int_k \frac{k}{k+q} p(k|s)dk = \int_\alpha \frac{k(\alpha)}{k(\alpha)+q} p(\alpha|s)d\alpha \tag{6.13}$$

When only one tRNA isoacceptor with charged level α reads the control codon, then $k(\alpha) = k_R/(1+K_R/\alpha t_0)$.

The next issue is to evaluate $p(\alpha|s)$, *i.e.* the probability density of the charging level α, conditional on s. For example, we may consider the scenario described in section 5.3, where the charging distributions for the case when there are two isoacceptors reading one codon each was calculated. Examples of $p(\alpha|s)$ are given in fig. 5.3 (also inserted in fig. 6.3). Given these distributions, the attenuation response function, Q(s), can be calculated from Eq.(6.13). The results for q=0.1 and q=0.01 are shown in fig. 6.3. For q=0.01 (the upper curve in figure 6.3) it can be seen that the response function is more sensitive when the fluctuations in tRNA charging are included (circles) compared to when they are neglected (solid line). This is because the probability of the low charging levels contributes disproportionally to the total probability of attenuation.

Figure 6.3 shows that the response in this case can be more sensitive in a mesoscopic analysis than what is seen in the macroscopic description. Superficially, this phenomenon could be described by saying that addition of noise makes the mechanism more sensitive. However, such a statement is erroneous and misleading. It is wrong because noise was not added; it was there all the time. Its effects were just neglected in the macroscopic description. It is misleading because it suggests that a macroscopic average value of a quantity has a special position compared to any other value in the distribution function, which does not have to be the case for non-linear systems with finite numbers of molecules. The macroscopic average is generally not equal to the mesoscopic average value and sometimes not even close to the most likely value.

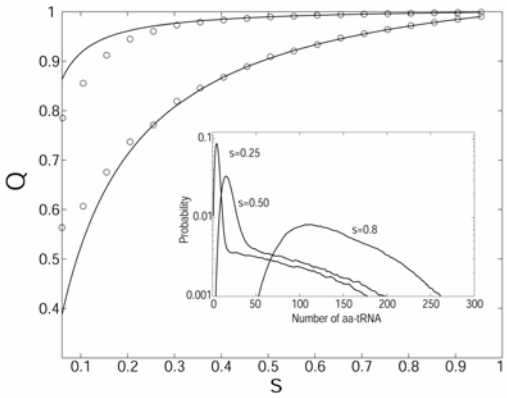


Figure 6.3 Stochastic focusing in attenuation. The probability of attenuation, Q, is plotted as a function of the normalized amino acid supply, s. The very insensitive attenuation mechanism from Eq. (6.5) is used with $q=0.01s^{-1}$ (upper curves) or $q=0.1s^{-1}$ lower curves. The solid lines corresponds to a deterministic description of the translation rate as a function of s, where as the circles corresponds to a mesoscopic description. The *insert* is from fig. 5.3. It shows the copy number distribution of the limiting aminoacyl-tRNA for different amino acid supply rates.

The physiological importance of stochastic focusing in attenuation is probably not great, as high sensitivity follows naturally from the multi-step movement of RNAP and the ribosome.

6.4 Summary: Attenuation

A model for transcriptional attenuation in amino acid biosynthetic operons has been outlined and it has been demonstrated that a number of new mechanisms can contribute to make that attenuation response very sensitive to the supply of a limiting amino acid. The most important contribution to sensitivity amplification originates in a double multi-step mechanism, where the ribosome and RNAP competes in a translation-transcription race. The sensitivity of this mechanism depends critically on RNAP pausing until the ribosome approaches. The sensitivity is further increased by ribosome stalling on a small fraction of all mRNA codons.

7. CONCLUDING REMARKS

This section contains some personal, non-scientific, remakes on the role of quantitative modeling in molecular biology and a brief summary of the results from the analysis of amino acid flows in protein synthesis.

7.1 Quantitative modeling of intracellular kinetics

7.1.1 MODELING MEANS MAKING ASSUMPTIONS

Modeling a biological system necessarily means making assumptions, focusing on what one believes is important and disregarding what one not believes is important. This is true both for models that are formalized in mathematical terms and those that are not. Our quantitative understanding of intracellular processes is essentially an extrapolation from what has been observed in the test-tube, to what is believed to be intracellular conditions. Like in all models of physics and biology such assumptions can only be justified *a posteriori*, by the observations that the model explains and the strength of its predictions. The more unexpected observations the model can explain, the more confidence can be put in the validity of the assumptions.
Successful modeling is equivalent to making the correct assumptions. Correct means that the simplified description, *i.e.* the model, still captures the relevant aspects in a way that clarifies how the system works.

7.1.2 THE PURPOSE OF MATHEMATICS IN BIOLOGICAL MODELS

One purpose of formulating a biological model in terms of mathematics is to rationalize thinking in such a way that logical consequences of the assumption can be deduced and tested. Another is to help pinpoint the critical features of a system that make it work. Yet another is to rule out models that can not possibly work as suggested by verbal arguments, even when they are interpreted in the most generous way.
All steps in the mathematical modeling includes introduction of assumptions that is omitted in a verbal description. Introducing these additional constraints explicitly makes the model more precise and therefore more vulnerable to criticism and falsification. This is very good as it makes it possible identify and correct erroneous assumptions. A less well defined model that does not make unambiguous predictions is always worse as it is too vague to be tested and improved.
In practice, the additional assumptions may get the character of predictions as only some of the assumptions will give models which are compatible to previous observations. For instance, the mathematical modeling can result in conditions on reaction and diffusion rate constants or reaction mechanism, that are unspecified in the verbal model, but that are necessary to get the observed system properties, *i.e.* stability, sensitivity, oscillations, speed of response *etc.* In this way mathematical models can be very powerful in suggesting the mechanism of biological action required to get an observed behavior. Beautiful examples are given by the conditions for robust formation

of the BMP-morphogen gradient in *Drosophila* [32] and the perfect adaptation of the chemotactic response in *E. coli* [62].

7.1.3 Contractions

There are simplifications of a different type than the assumptions that define the model. These simplifications can be deduced within the model and I will call them 'contractions' (following [15]). Typical contractions that often are used in biological modeling are to represent the copy number distribution of a molecule by its macroscopic average value, to assume that some reactions are rapidly equilibrated or that diffusion is sufficiently fast to allow disregarding spatial coordinates. Common to contractions is that a larger state description is contracted to a few variables, such that the degrees of freedom of the model are reduced. In contrast to the assumptions, that only can be justified *a posteriori*, contractions can be validated using analytical or numerical methods and the limits of their validity can be found.

A model may have qualitatively different properties at different levels of contractions. The most contracted description that correctly retains the properties of interest is usually the one to choose. This simply means that it is preferable to work with a simple model if possible.

Often the modeling of intracellular processes starts out at a high level of contraction, *i.e.* the macroscopic well stirred approach adapted from test-tube chemistry. This is of course easily motivated by that the experimental data rarely support a more detailed model. However, it should be remembered that there must always exist at least some physically consistent model at all lower contraction levels that is compatible with the highly contracted model, in order to make the highly contracted model valid. In more simple terms, biochemical models must not break physical laws. For instance, one may ask if the dynamics of concentrations, as described a macroscopic model, can be reproduced as average values in any reasonable mesoscopic model of the same system, or as an average over space when realistic diffusion rates are considered. This kind of physical constraint from lower levels of contractions can be very useful as it reduces the number of possible models.

7.2 What has been learned about amino acid limitation?

In this thesis I have scratched on the surface of how prokaryotic life may respond to amino acid limitation. Simplistic quantitative kinetic motifs taken from protein synthesis have been analyzed meso- and macroscopically. Some things have been learnt about how dramatic dynamics in tRNA charging levels can be expected if ribosomes operate sub-saturated in poor media. These near critical properties originate in coupled zero-order consumption and the switch characteristics of the aminoacylation reaction. The sensitivity in the charging of tRNA and the concentrations of free amino acids to the balance in amino acid supply have consequences for transcriptional control of amino acid biosynthetic operons. As a first prediction, it seems that only one attenuation controlled operon can be expressed at the time. The test of this prediction requires quantitative measurements on single cells.

Further, we have seen that tRNA isoacceptors are not uniformly deacylated under amino acid limitation. This has implications for how the genetic code is used, especially in genes and control systems that are involved in the response

to amino acid limitation. Another response to amino acid limitation is the one that is described in paper V. In this paper the mechanism of action the endogenous bacterial toxin RelE is characterized by biochemical methods. RelE is normally inhibited by its anti-toxin partner RelB, which in contrast to RelA is rapidly turned over by proteolysis. Therefore, when protein synthesis is reduced due to amino acid limitation, the concentration of RelB drops and RelE is activated. In paper V it is shown how RelE cleaves mRNA codon-specifically in the ribosome. The physiological role of RelE is presumably to stop translation of mRNAs that were transcribed before amino acid starvation occurred, such that the available amino acids are used for more urgently needed proteins translated from newly made RNA transcripts. The codon specificity of RelE action makes it possible for the cell to use codons that are not cleaved by RelE in genes that should be expressed under amino acid limitation. For instance, the UAG stop codon, which is efficiently cleaved by RelE, is significantly under represented in amino acid biosynthetic operons. Together with the principle of selective charging of isoacceptors the cleavage pattern of RelE can be used to define a starvation codon adaptation index for classification of genes that are expressed under amino acid limitation (T. Tenson el al. *in preparation*).

In paper X we suggest a mechanism for RelA action in the stringent response [83] which makes the rate of ppGpp synthesis depend directly on the rate of translation and not only on the concentrations of deacylated tRNA and ribosomes with an open A-site. We suggest that RelA is removed from the ribosome in the petidyltransfer or translocation event, which would restrict the stringent response to situations where the rate of protein synthesis is slow.

Finally, the reasons for high sensitivity in transcriptional attenuation have been clarified. High sensitivity in this control system is critically important for the expression of amino acid biosynthetic enzymes, as the mechanism only can respond after amino acid limitation already has started affecting the rate of protein synthesis.

The modeling is still one step away from the global response to amino acid limitation, including the stringent response and changes in stabile RNA composition. The framework introduced in this thesis is however a starting point for quantitative modeling of bacterial growth, where the regulated investments in protein synthesis is a central part.

8. SUMMARY IN SWEDISH

I det här avsnittet har jag valt att kortfattat beskriva avhandlingens viktigaste upptäckter, med ambitionen att göra dem tillgängliga för en bred läsekrets.

DYNAMISK ANVÄNDNING AV DEN GENETISKA KODEN

Den genetiska koden använder 61 kodord för att koda för 20 aminosyror. Kodorden översätts till aminosyror med hjälp av tRNA molekyler. tRNA molekyler som läser synonyma kodord (olika kodord för samma aminosyra) kallas isoacceptorer. Koncentrationerna av de olika isoacceptorerna är inte helt balanserade mot frekvensen av de kodord de läser. I artikel I visas att detta leder till att isoacceptorerna kan förväntas tappa sin aminosyraladding i olika hög grad när det blir brist på aminosyran. Eftersom olika isoacceptorer ofta läser olika synonyma kodord så påverkas hastigheten i översättningen av de olika kodorden olika mycket vid aminosyrabrist. Den genetiska koden är således inte så redundant som man kan tro om man bara ser till att det är 61 kodord för 20 aminosyror. Cellerna tycks utnyttja överskottsinformationen till att klara växlingarna mellan rika och fattiga miljöer på ett bra sätt. Till exempel så används kodord som läses av tRNA molekyler som inte tappar sin laddning helt i ökad utsträckning för proteiner som behövs vid svält.
Ett annat exempel på hur synonyma kodord kan användas för regulatoriska syften ges i artikel V. Där beskrivs hur proteinet RelE, som aktiveras vid svält, klipper sönder mRNA i ribosomen på ett kodordsspecifikt sätt.

KOPPLADE METABOLA FLÖDEN ORSAKAR KÄNSLIGHET OCH FLUKTUATIONER

Ett annat tema i avhandlingen är vad som händer i metabolitpooler med stökiometriskt kopplade utflöden. Analysen är särskilt gjord för att förstå proteinsyntesen vid brist på en eller flera aminosyror, men resultaten går lätt att generalisera till andra flersubstratreaktioner. Det mest intressanta beteendet uppkommer då syntesen av metaboliter (aminosyror) är balanserad mot sina stökiometrier i den gemensamma konsumtionsreaktionen. I denna biologiskt relevanta balanspunkt bryter den makroskopiska beskrivningen samman och man måste beakta slumpmässigheten i de kemiska reaktionerna. Nära balanspunkten blir medelkoncentrationerna av metaboliter mycket känsliga för obalansen i deras syntesflöden vilket kan utnyttjas av cellen för effektiv koordination av transkriptionell reglering (artikel III och X).

NYA METODER FÖR ANALYS SPATIELLA FLUKTUATIONER

I avhandlingen beskrivs en ny algoritm som gör det praktiskt möjligt att simulera kemiska processer så som de beskrivs med reaktion-diffusions masterekvationen. Algoritmen gör att det går att undersöka vad som kan förväntas hända om man tar hänsyn till att kemiska reaktioner både är diskreta slumpmässiga händelser och att diffusionen är så långsam att komponenterna i en cell inte hinner blandas ordentligt.
I uppsats IX används algoritmen för att demonstrera hur egenskaparna för kemiska system som utgör grunden för biokemist minne är beroende av reaktionsvolymens geometri.

ACKNOWLEDGEMENTS

Discussions with Måns Ehrenberg have led to many of the results presented in this thesis. Our *fundamental* results concerning black holes and the spatio-temporal properties of photons will be published elsewhere or, most likely, not at all. I am very grateful that he generously has shared his devotion to science, his creativity, and his humor with me. Måns has been a fantastic supervisor and a good friend.
I want to thank Johan Paulsson for his work on the relation between sensitivity and fluctuations, which has been very important for this thesis. Johan has also carefully and openheartedly commented several of my manuscripts, which have improved the published articles significantly. I want to thank Martin Lovmar for many things but particularly for helpful discussions regarding experiments and lots of proofreading. Otto Berg has generously contributed with his expertise in stochastic processes and reactions-diffusion kinetics, as well as codon usage. The results in section 3.6 were originally derived by Otto. I also want to thank Paul Sjöberg for clarifying discussions regarding mesoscopic reaction-diffusion problems and the Fokker-Planck equation. Finally, I want this thank my parents for encouraging my curiosity and creativity as a child and Kristin for encouragement, support and cozy mornings.
The project has been supported by the National Graduate School in Scientific Computing, NGSSC.

REFERENCES

1. Neidhardt, F.C. (1999), Bacterial growth: constant obsession with dN/dt. *J Bacteriol*, **181**:7405-8.
2. Neidhardt, F.C., R. Curtiss, J. Ingraham, E. Lin, K. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter, and H. Umbarger (1996), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. Washington, D.C., ASM Press.
3. Neidhardt, F.C., J.L. Ingraham, and M. Schaechter (1990), *Physiology of the bacterial cell*. Sunderland, Sinauer.
4. Shen-Orr, S.S., R. Milo, S. Mangan, and U. Alon (2002), Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet*, **31**:64-8.
5. Savageau, M.A. (1976), *Biochemical systems analysis: a study of function and design in molecular biology*. Reading, Addison-Wesley.
6. Savageau, M.A. (1969), Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J Theor Biol*, **25**:365-9.
7. Savageau, M.A. (1971), Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature*, **229**:542-4.
8. Kacser, H. and J.A. Burns (1973), The control of flux. *Symp Soc Exp Biol*, **27**:65-104.
9. Heinrich, R. and S. Schuster (1996), *The regulation of cellular systems*. New York, Chapman and Hall.
10. Van Kampen, N.G. (1997), *Stochastic processes in physics and chemistry, second edition*. second ed. Amsterdam, Elsevier.
11. McQuarrie, D.A. (1967), Stochastic approach to chemical kinetics. *J Appl Prob*, **4**:413-478.
12. Érdi, P. and J. Tóth (1989), *Mathematical models of chemical reactions*. Princeton, Princeton University Press.
13. Elowitz, M.B., M.G. Surette, P.E. Wolf, J.B. Stock, and S. Leibler (1999), Protein mobility in the cytoplasm of Escherichia coli. *J Bacteriol*, **181**:197-203.
14. Gardiner, C. (1985), *Handbook of stochastic Methods*. second edition ed. Berlin, Springer-Verlag.
15. Keizer, J. (1987), *Statistical Thermodynamics of Nonequlibrium Processes*. Berlin, Springer-Verlag.
16. Guldberg, C. and P. Waage (1864), Forhandlingar i Videnskabs-Selskabet i Christiana. **35-40**:111-120.
17. Gillespie, D. (1976), A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, **22**:403-434.

18. Bortz, A., M. Kalos, and J. Lebowitz (1975), A New Algorithm for Monte Carlo Simulation of Ising Spin Systems. *J. Comp. Phys.*, **17**:10-8.
19. Gibson, M. and J. Bruck (2000), Efficient exact stochastic simulation of chemical systems with many species and channels. *J Phys. Chem. A*, **104**:1876-1889.
20. Risken, H. (1984), *The Fokker-Planck equation*. Berlin, Spinger Verlag.
21. Van Kampen, N.G. (1961), *Can. J. Phys.*, **39**:551.
22. Paulsson, J. (2004), Summing up the noise in gene networks. *Nature*, **427**:415-8.
23. Paulsson, J. and M. Ehrenberg (2001), Noise in a minimal regulatory network: plasmid copy number control. *Q Rev Biophys*, **34**:1-59.
24. Fell, D.A. (1996), *Understanding the Control of Metabolism*. London, Portland Press.
25. Fano, U. (1947), Ionization yield of rations. II. the fluctuations of the number of ions. *Phys Rev*, **72**:26-29.
26. Paulsson, J. (2000), *The stochastic nature of intracellular control circuits*. Uppsala, Acta Universitatis Upsaliensis.
27. Shapiro, L. and R. Losick (2000), Dynamic spatial regulation of the bacterial cell. *Cell*, **100**:89-98.
28. Levchenko, A. and P.A. Iglesias (2002), Models of Eukaryotic Gradient Sensing: Application to Chemotaxis of Amoebae and Neutrophils. *Biophys. J.*, **82**:50-63.
29. Thar, R. and M. Kuhl (2003), Bacteria are not too small for spatial sensing of chemical gradients: an experimental evidence. *Proc Natl Acad Sci U S A*, **100**:5748-53.
30. Howard, M., A.D. Rutenberg, and S. De Vet (2001), Dynamic compartmentalization of bacteria: accurate division in E. coli. *Phys Rev Lett*, **87**:278102.
31. Finch, E.A. and G.J. Augustine (1998), Local calcium signaling by inositol-1,4,5-trisphosphate in Purkinje cell dendrites. *Nature*, **396**:753-6.
32. Eldar, A., R. Dorfman, D. Weiss, H. Ashe, B.Z. Shilo, and N. Barkai (2002), Robustness of the BMP morphogen gradient in Drosophila embryonic patterning. *Nature*, **419**:304-8.
33. Nicolis, G. and I. Prigogine (1971), Fluctuations in nonequilibrium systems. *Proc Natl Acad Sci U S A*, **68**:2102-2107.
34. Nicolis, G. (1972), Fluctuations around nonequilibrium states in open nonlinear systems. *J Stat Phys*, **6**:195-222.
35. Stundzia, A.B. and C.J. Lumsden (1996), Stochastic simulation of coupled reaction-diffusion processes. *J Comp Phys*, **127**:196-207.
36. Baras, F. and M.M. Mansour (1997), Microscopic simulation of chemical instabilities. *Advances in Chemical Physics*, **100**:393-475.
37. Nicolis, G. and I. Prigogine (1977), *Self-organization in nonequilibrium systems*. New York, John Wiley & sons.
38. Goldbeter, A. and D.E. Koshland, Jr. (1982), Sensitivity amplification in biochemical systems. *Q Rev Biophys*, **15**:555-91.

39. Prigogine, I., R. Lefever, A. Goldbeter, and M. Herschkowitz-Kaufman (1969), Symmetry breaking instabilities in biological systems. *Nature*, **223**:913-6.
40. Goldbeter, A. (2002), Computational approaches to cellular rhythms. *Nature*, **420**:238-45.
41. Bhalla, U.S. and R. Iyengar (1999), Emergent properties of networks of biological signaling pathways. *Science*, **283**:381-7.
42. Berg, O.G. (1978), On diffusion-controlled dissociation. *Chemical Physics*, **31**:47-57.
43. Arnold, L. and M. Theodosopulu (1980), Deterministic limit of the stochastic model of chemical reactions with diffusion. *Adv. Applied Prob.*, **12**:367-379.
44. Murray, J.D. (2003), *Mathematical Bioogy II: Spatial Models and Biomedical Applications*. Berlin, Springer-Verlag.
45. Turing, A.M. (1952), The chemical basis of morphogenesis. *Philosophical transactions of the Royal society of London*, **237**:37-72.
46. Howard, M. and A.D. Rutenberg (2003), Pattern formation inside bacteria: fluctuations due to the low copy number of proteins. *Phys Rev Lett*, **90**:128102.
47. Koshland, D.E., Jr., A. Goldbeter, and J.B. Stock (1982), Amplification and adaptation in regulatory and sensory systems. *Science*, **217**:220-5.
48. Berg, O.G., J. Paulsson, and M. Ehrenberg (2000), Fluctuations and quality of control in biological cells: zero-order ultrasensitivity reinvestigated. *Biophys J*, **79**:1228-36.
49. Strogatz, S.H. (1994), *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Cambridge MA, Addison-Wesley.
50. Gillespie, D. (1977), Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**:2340-2361.
51. Alves, R. and M.A. Savageau (2000), Effect of overall feedback inhibition in unbranched biosynthetic pathways. *Biophys J*, **79**:2290-304.
52. Jacob, F. and J.J. Monod (1961), Genetic Regulatory Mechanisms in the Synthesis of Proteins. *J Mol. Biol.*:3,318-356.
53. Yanofsky, C. (1981), Attenuation in the control of expression of bacterial operons. *Nature*, **289**:751-8.
54. Bliss, R.D., P.R. Painter, and A.G. Marr (1982), Role of feedback inhibition in stabilizing the classical operon. *J Theor Biol*, **97**:177-93.
55. Marr, A.G. (1991), Growth rate of Escherichia coli. *Microbiol Rev*, **55**:316-33.
56. Fersht, A.R. (1999), *Structure and mechanism in protein science*. New York, W.H. Freeman and company.
57. Ibba, M. and D. Söll (2000), Aminoacyl-tRNA synthesis. *Annu Rev Biochem*, **69**:617-650.
58. Fersht, A.R. and M.M. Kaethner (1976), Mechanism of aminoacylation of tRNA. Proof of the aminoacyl adenylate pathway

for the isoleucyl- and tyrosyl-tRNA synthetases from Escherichia coli K12. *Biochemistry*, **15**:818-23.
59. Ehrenberg, M. and C.G. Kurland (1984), Costs of accuracy determined by a maximal growth rate constraint. Q *Rev Biophys*, **17**:45-82.
60. Kaziro, Y. (1978), The role of guanosine 5'-triphosphate in polypeptide chain elongation. *Biochim Biophys Acta*, **505**:95-127.
61. Berg, O.G. (1978), A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol*, **71**:587-603.
62. Barkai, N. and S. Leibler (1997), Robustness in simple biochemical networks. *Nature*, **387**:913-7.
63. Yi, T.M., Y. Huang, M.I. Simon, and J. Doyle (2000), Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci U S A*, **97**:4649-53.
64. Dong, H., L. Nilsson, and C.G. Kurland (1996), Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J Mol Biol*, **260**:649-63.
65. Björk, G. (Stable RNA modification, in *ref. 16*. 861-886.
66. Crick, F.H., L. Barnett, S. Brenner, and R.J. Watts-Tobin (1961), General nature of the genetic code for proteins. *Nature*, **192**:1227-1232.
67. Paulsson, J., O.G. Berg, and M. Ehrenberg (2000), Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc Natl Acad Sci U S A*, **97**:7148-53.
68. Yanofsky, C. (2000), Transcription attenuation: once viewed as a novel regulatory strategy. *J Bacteriol*, **182**:1-8.
69. Landick, R. and C. Yanofsky (1987), Transcription attenuation, in *In Escherichia coli and Salmonella Typhmurium: Cellular and Molecular Biology*, N.F.C.e. al., Editor. ASM Press Washington D.C. 276-1301.
70. Jackson, E.N. and C. Yanofsky (1973), Thr region between the operator and first structural gene of the tryptophan operon of Escherichia coli may have a regulatory function. *J Mol Biol*, **76**:89-101.
71. Lewis, J.A. and B.N. Ames (1972), Histidine regulation in Salmonella typhimurium. XI. The percentage of transfer RNA His charged in vivo and its relation to the repression of the histidine operon. *J Mol Biol*, **66**:131-42.
72. Kasai, T. (1974), Regulation of the expression of the histidine operon in Salmonella typhimurium. *Nature*, **249**:523-7.
73. Landick, R. and J.C.L. Turnbough (1992), Transcriptional Attenuation, in *Transcriptional Regulation*, S.L. McKnight and K.R. Yamamoto, Editors. Cold Spring Harbour Laboratory Press. 407-447.
74. Landick, R., J. Carey, and C. Yanofsky (1985), Translation activates the paused transcription complex and restores transcription of the trp operon leader region. *Proc Natl Acad Sci U S A*, **82**:4663-7.

75. Winkler, M.E. and C. Yanofsky (1981), Pausing of RNA polymerase during in vitro transcription of the tryptophan operon leader region. *Biochemistry*, **20**:3738-44.
76. Landick, R., J. Carey, and C. Yanofsky (1987), Detection of transcription-pausing in vivo in the trp operon leader region. *Proc Natl Acad Sci U S A*, **84**:1507-11.
77. Roesser, J.R., Y. Nakamura, and C. Yanofsky (1989), Regulation of basal level expression of the tryptophan operon of Escherichia coli. *J Biol Chem*, **264**:12284-8.
78. Manabe, T. (1981), Theory of regulation by the attenuation mechanism: stochastic model for the attenuation fo the Escherichia coli tryptophan operon. *J Theor Biol*, **91**:527-44.
79. Von Heijne, G. (1982), A theoretical study of the attenuation control mechanism. *J Theor Biol*, **97**:227-238.
80. Suzuki, H., T. Kunisawa, and J. Otsuka (1986), Theoretical evaluation of transcriptional pausing effect on the attenuation in trp leader sequence. *Biophys J*, **49**:425-35.
81. Ingraham, J.L., O. Maaloe, and F.C. Neidhardt (1983), *Growth of the Bacterial Cell*. Sunderland, Mass, Sinauer Associates, Inc.
82. Ehrenberg, M. (1996), Hypothesis: hypersensitive plasmid copy number control for ColE1. *Biophys J*, **70**:135-45.
83. Cashel, M., D.R. Gentry, V.J. Hernandez, and D. Vinella (1996), The stringent response, in *Escherichia coli and Salmonella cellular and molecular biology*, F.C. Neidhardt, Editor. ASM Press Washington DC.

Acta Universitatis Upsaliensis

Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology*. (Prior to October, 1993, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science".)

Distribution:
Uppsala University Library
Box 510, SE-751 20 Uppsala, Sweden
www.uu.se, acta@ub.uu.se

ISSN 1104-232X
ISBN 91-554-5988-9