Kinetic Modelling of Gene Expression

From Linear Genome Sequence to Nonlinear Cellular Dynamics

Kinetische Modellierung der Genexpression

Von linearer Genomsequenz zu nichtlinearer zellulärer Dynamik

Von der Fakultät Maschinenbau der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Ingenieurwissenschaften (Dr.-Ing.) genehmigte Abhandlung

vorgelegt von Sabine Arnold
aus Stuttgart

Hauptberichter:	Prof. Dr.-Ing. Matthias Reuss
Mitberichter:	Prof. Dr. Jay D. Keasling
Tag der mündlichen Prüfung:	8. November 2002

Institut für Bioverfahrenstechnik
Universität Stuttgart

2003

Acknowledgements

I am indebted to Prof. Dr.-Ing. Matthias Reuss for academic guidance and encouraging support throughout this work. His valuable contributions over the course of this work are gratefully acknowledged.

I am much obliged to Prof. Dr. Jay Keasling for his interest in this research topic and his participation in the examination committee.

I thank Dr. rer. nat. Martin Siemann-Herzberg for kindly providing various experimental data and his invaluable patience in interpreting experimental and simulation results to improve scientific understanding.

Many thanks are addressed to Sandra Baumann and Kai Scharnweber for a wonderful atmosphere in our group and their thorough experimental work, which made modelling worthwhile.

Further thanks are granted to the students contributing substantially to this study: Steffen Knapp, Joachim Schmid, Andreas Schwarz, Claudia Thiele, Markus Werner, Nikolaus Rahaus.

Financial support by the German Ministry of Research (ZSP project A3.10U) and by the "Deutsche Forschungsgemeinschaft" (project RE 632/8-1) is gratefully acknowledged. This project was associated also with a joint project "Cell-free protein biosynthesis reactor", (project FKZ 0 311 302). I thank our collaboration partners for stimulating discussions, in particular with Prof. Volker Erdmann (Institute of Biochemistry, FU Berlin, Germany), Prof. Alexander Spirin (Institute for Protein Research, Pushchino, Russia), Dr. Herbert Stadler (Institute for Bioanalytics, Göttingen) and our industrial collaboration partner Roche Diagnostics Ltd. (Penzberg, Germany) represented by Dr. Albert Röder.

I also thank Prof. Dr. rer. nat. Christoph Syldatk for a stimulating research environment, PD. Dr. rer. nat./sc. nat. Stefan Schuster for fruitful discussions on network analysis, and PD. Dr. rer. nat. Markus Pietzsch - his delicious *Zwiebelkuchen* and Wine parties are remembered for long.

I thank all my co-workers at present and in the past for the friendly and cheerful working environment, especially Naruemol Noisommit-Rizzi, Klaus Mauch, Harald Teves, Andrea Seipel, Sabine Pietruszka, Hans-Jürgen Christian, Kerstin Ragnitz, Thomas Waniek, Oliver Vielhauer, Kerstin Schierholz, Stefan Buziol, Ester Guerrero, Luciano Aguilera, Franz Macherhammer, Dirk Mueller, Gerhard Stadler, Susanne Zibek, Achim Hauck, Andreas Freund, Hans-Joachim Daniel, Ralf Otto, Michael Dauner, Claudia Dössereck, Alexander Schwarz, Sven Schmalzriedt, Marc Jenne, Alexei Lapin, Ulrich Peckmann, Hiltrud Schweitzer, and Renate Moser.

To my family for their love and on-going support.

Stuttgart, Germany Sabine Arnold
November 2002

Contents

Nomenclature

Symbols

a_i	number of codons representing a particular amino acid i
A	number of naturally occurring amino acids
c	codon usage
C	metabolite concentration (μM)
d	spacing between ribosomes and degradosomes, and between SD-sequence and translational start codon
D	promoter contained on DNA template
E	enzyme
f	fraction of single-stranded bases within the 23 bases subsequent to the SD-sequence
$f_{j,i}$	relative portion of base j contained in transcript i (%)
G	free energy (kJ/mol)
J	number of base tripletts of a mRNA
k_i	respective rate constant
K	last codon of a coding region
K_a	association constant
K_d	dissociation constant
K_I	inhibition constant for respective metabolite (μM)
K_M	Michaelis-Menten constant for respective substrate, (μM)
L_j	physical diameter of a ribosome and degradosome, respectively
m	mass (g)
m_i	ratio of RNA species i to total measured RNA (g/g)
$m_{i,j}$	element of matrix $\underline{\underline{M}}$
m_j	reference state of a ribosome and degradosome, respectively
M	mRNA
M	number of mRNA molecules
$\underline{\underline{M}}$	mRNA matrix
n	number
n_i	transcript length for RNA species i (kb)
n_{cod}	number of base tripletts to denote a state
N	number of ribonucleic bases
N_A	Avogadro number
r	regression coefficient
R	number of RNA species synthesized from a given DNA template
S	number of segments
t	time (min)
T	number of tRNA species

T	temperature (K)
T	time (s)
V	reaction rate (μM/min)
V	volume (μl)
V_P	relative protein expression rate (%)
X	measured radioactivity (dpm/μL)
z	position of endonucleolytic cleavage site
Z	number of fragments of a mRNA obtained by endonucleolytic cleavage

Greek letters

α	reduced scintillation measurement due to sample immobilization (%)
γ	specific activity of radionuclide (Bq/mmol)
δ	dilution factor of nonradiolabelled and labelled metabolite
η	fractional codon usage
μ	specific growth rate (h^{-1})
Φ	efficiency factor
ϕ	T7 transcription terminator
$\phi 10$	T7 promoter
φ	energy charge

Indices

aq	aqueous
avg	average
bla	β-lactamase
cell	referring to a single cell
CR	catabolite repression
d	degradation
D	refers to promoter sequence of a DNA
D0	refers to a degradosome association site
eff	effective
eq	thermodynamic equilibrium
exp	experimentally determined
f	folded
f	formyl-
f	forward reaction
i	count index
in	entering equilibrium computation
I	induction
j	count index
k	count index
m	methionine
NTP	nucleoside triphosphate

out	outcome of equilibrium computation
qss	quasi-stationary state
r	reverse reaction
R0	refers to a ribosome binding site
s	count index
sim	predicted from simulation
t	denotes total concentration
u	unfolded
un	unbound
X	biomass

Superscript

$\prime$	refers to new codon grid representation
0	initial condition
0	standard condition
A	refers to the A-site of a ribosome
D	degradosome
M	mRNA
M	methionine
max	maximum value
P	refers to the P-site of a ribosome
R	ribosome
$R*$	ribosome bound to the initiation codon prior to IF2-dissociation

Abbreviations

30S	small prokaryotic ribosomal subunit
30SIC	30S initiation complex
50S	large prokaryotic ribosomal subunit
70S	free, undissociated prokaryotic ribosome
70SIC	70S initiation complex
A	adenine
aa	amino acid(s)
aa-tRNA	aminoacyl-tRNA
Ac	acetate
Ack	acetate kinase
AcP	acetyl phosphate
ACSL	Advanced Continuous Simulation Language
Adk	adenylate kinase
ADP	adenosine diphosphate
Ala	alanine
AMP	adenosine monophosphate
Arg	arginine

ARS	aminoacyl-tRNA-synthetase
Asn	asparagine
Asp	aspartic acid
ass	association
ATP	adenosine triphosphate
AUG	translational start codon
bp	base pairs
BSA	bovine serum albumin
C	cytosine
car	carbamoylase
CDP	cytosine diphosphate
CFCF	continuous-flow cell-free
CMP	cytosine monophosphate
CTP	cytosine triphosphate
Cys	cystein
DNA	desoxy-ribonucleic acid
dsRNA	double-stranded RNA
DTT	dithiothreitol
E	enzyme
EC	Enzyme Commission
EF	translational elongation factor
EMBL	European Molecular Biology Laboratory
endo	endonucleolytic
exo	exonucleolytic
F	folded conformation of the ribosome binding site
fMet-tRNA$_f^M$	*N*-formylmethionyl-tRNA
Frag	mRNA fragment
FRET	fluorescence resonance energy transfer
fTHF	Formyl-H_4-Folate
G	guanine
GDP	guanosine diphosphate
GFP	green fluorescent protein
Gln	glutamine
Glu	glutamic acid
Gly	glycine
GMP	guanosine monophosphate
GTP	guanosine triphosphate
h	hour
His	histidine
hyd	hydantoinase
IC	initiation complex

IF	translational initiation factor
IF2D	IF2-dependent GTP hydrolysis
Ile	isoleucine
IOD	integrated optical density
K	Kelvin
kb	kilo bases
kDa	kilo Dalton (1 Da $\hat{=}$ 1 g/mol)
kJ	kilo Joule
Leu	leucine
Lys	lysine
malE	maltose binding protein E
Met	methionine
MFA	metabolic flux analysis
Mg	magnesium
min	minute
mRNA	messenger RNA
mv	degradosome movement
n.d.	not determined
Ndk	nucleoside diphosphate kinase
NDP	nucleoside diphosphate
Nmk	nucleoside monophosphate kinase
NMP	nucleoside monophosphate
NMR	nuclear magnetic resonance
nt	nucleotide(s)
NTP	nucleoside triphosphate
ORF	open reading frame
P	promoter
PAGE	polyacryl amide gel electrophoresis
PAP I	poly-adenylate phosphorylase
pelB	pelB leader sequence
pET	*E. coli* expression system
Phe	phenylalanine
Pi	inorganic phosphate
PNPase	polynucleotide phosphorylase
PPi	inorganic pyrophosphate
PPK	polyphosphate kinase
Pro	proline
RBS	ribosome binding site
rDNA	recombinant DNA
RF	translational termination factor
RFH	a particular translational termination factor

RNA	ribonucleic acid
RNAP	DNA-dependent RNA polymerase
RNase	ribonuclease
RP	ribosomal protein
RRF	ribosome release factor
rRNA	ribosomal RNA
s	second
S1	ribosomal protein S1 (contained in 30S ribosomal subunit)
scFv	single chain Fv-fragment
SD	Shine-Dalgarno
Ser	serine
SNP	single-nucleotide polymorphism
ssRNA	single-stranded RNA
T	terminator
T	thymine
T	tRNA
T3	ternary complex (consists of one copy of EFTu, GTP, and aa-tRNA)
TC	transcription
TCA	tricarboxylic acid
TCE	transcription elongation
TCI	transcription initiation
TCT	transcription termination
TE	termination efficiency
THF	H_4-Folate
Thr	threonine
TL	translation
TLE	translation elongation
TLI	translation initiation
TLT	translation termination
tmRNA	transfer-messenger RNA
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
Trp	tryptophan
Tyr	tyrosine
U	unit
U	uracil
U	unfolded conformation of the ribosome binding site
UDP	uracil diphosphate
UMP	uracil monophosphate
UTP	uracil triphosphate
Val	valine

Summary

In this study, a dynamic model of prokaryotic gene expression was developed that heavily makes use of gene sequence information. The main contribution arises from the fact that the combined gene expression model allows to assess mechanistically the impact of nucleotide sequence alteration on the dynamics of gene expression rates. The high level of detail of the mathematical model enables to provide very detailed insight into the various steps of the protein expression process.

Modelling required the development of a valid model structure for template-bound biopolymerization processes within a continuous analysis method. In contrast to a discrete model, or a combination of both approaches (i.e., hybrid modelling), the continuous model presented is a mechanism-based deterministic description of system states in terms of differential and algebraic sets of equations. Characteristically, a codon-specific representation of state variables was chosen for this model.

Transcription kinetics were described mathematically at the example of T7 RNA polymerase. Parametrization of the transcription model was carried out for selected model constants, i.e., for the rate constants of initiation, elongation, and termination, as well as for the maximum rate of transcription. According to enzyme kinetics, most influential parameters determining transcription rate are T7 RNA polymerase concentration, and promoter concentration at typical reaction conditions of simultaneous *in vitro* transcription/translation.

The process of mRNA degradation was modelled allowing for a distinction between endonucleolytic and exonucleolytic reaction steps. The effects of increased translational efficiency greatly improving mRNA stability, as observed experimentally, were correctly demonstrated by the model. On the basis of simulating *lacZ* mRNA degradation, it was possible to identify the parameters contained in the degradation model.

Because mRNA can constitute a significant sink for nucleoside triphosphates, transcription rate was suggested to be kept at moderate levels, in particular in batch systems. Otherwise, the resulting nucleotide concentrations may drop to limiting thresholds, as they are incorporated into mRNA molecules. Model-assisted simulations can help in identifying an appropriate counterbalance between mRNA degradation rate and a suitable adjustment of transcription rate.

The translation model presented covers the mechanisms of protein synthesis initiation, elongation, and termination, at the same time considering the particular mechanistic role of key translation factors. An earlier approach for describing sterical interference among template-bound catalysts (MacDonald et al., 1968) was extended in this study, in order to cover also a situation where two different types of catalysts (i.e., ribosomes and degradosome) can be bound in multiple copies to a same template.

To enhance the applicability of the model for large expression systems, a model reduction was introduced. In the suggested procedure, the number of state variables could be significantly diminished by merging groups of base tripletts together, while at the same time taking into account the implications on reaction kinetics and material balancing.

The current status of the combined model allows to reveal several causes for production limitation, whether they are due to substrate depletion or inactivation processes, or else caused by unfavourable initial catalyst concentrations and their stoichiometric relations. An application of the combined gene expression model to simulating cell-free protein synthesis dynamics demonstrated limited volumetric productivitites to be caused by unfavourably low translation factor levels that are typical for these dilute *in vitro* systems. Equilibrium binding calculations suggested a requirement for at minimum equal molar ratios of initation factors IF1, IF2, and IF3, with respect to total concentration of unbound ribosomes. When these conditions are met, about 85 % of all freely dissolved 30S ribosomal subunits are predicted to prevail in their activated form, i.e., complexed with all of these three initiation factors. By appropriately raising the concentrations of both translation initiation and elongation factors, a 4-fold improvement of volumetric protein synthesis rate and a 5-fold higher final product yield are predicted in comparison to a non-optimized reference batch process.

From a stand-point of reduced model complexity, it may be beneficial to use the overall model for estimation of mechanism-related parameters or decay constants of a gene expression model, prior to applying these parameters within a whole-systems modelling frame. The immediate value of such models arises from their applicability for describing the expression of individual genes or a few genes at a time, which are typical for recombinant protein production.

Gene sequence information enters the overall model at the following stages: (a) within the transcription process, by assigning different rate constants for initiation and termination of mRNA synthesis, respectively, (b) the endo- and exonuclease activities in the ordered process of $5'$ to $3'$-degradation of messenger RNA, (c) during translation by distinguishing codon-specific elongation rates and effects related to sterical interactions among translating ribosomes.

An attempt was made also to investigate the impact of predicted mRNA secondary structure within the ribosome binding site on measured translation rate. The results of this analysis suggest improved protein synthesis rates for mRNA conformations that favour a high degree of single-strandedness within a region of approximately 20 nucleotides downstream of the Shine-Dalgarno sequence. Coincidently, this mRNA region is covered by the ribosome during the event of ribosome binding.

Altogether, the mathematical gene expression model presented in this study provides a comprehensive frame for a thorough analysis of sequence-related effects on the stages of mRNA synthesis, mRNA degradation, and ribosomal translation, as well as their nonlinear interconnectedness, and may as such serve useful for rational design of recombinant bacterial protein synthesis systems.

Zusammenfassung

In dieser Studie wurde ein mathematisches Modell zur dynamischen Beschreibung der prokaryotischen Genexpression ausgearbeitet, das sich in hohem Maße die in der Gensequenz enthaltenen Informationen zunutze macht. Ein wesentlicher Beitrag dieser Arbeit besteht darin, dass das vollständige Genexpressionsmodell einen mechanistischen Zusammenhang zwischen der Nukleotidsequenz und der Dynamik der Expressionsraten herzustellen vermag. Der hohe Detailliertheitsgrad des Modells erlaubt einen tiefen Einblick in die verschiedenen Stufen des Proteinsyntheseprozesses.

Im Vordergund der Modellbildung stand die Entwicklung einer geeigneten Modellstruktur zur Abbildung von Biopolymerisationsprozessen, bei denen der Biokatalysator an ein Templat gebunden ist. Im Gegensatz zu diskreten Modellen oder Hybridmodellen, einer Kombination aus diskreter und kontinuerlicher Systembeschreibung, basiert der gewählte, ausschließlich kontinuierliche Modellansatz auf einer deterministischen und mechanistischen Beschreibung der Zustandsgrößen in Form von Differenzial-algebraischen Gleichungssystemen. Ein Charakteristikum des Modells ist die codon-spezifische Darstellung des Zustandsvektors.

Die mathematische Beschreibung der Transkription erfolgte am Beispiel der Phage T7 RNA Polymerase. Eine Parametrierung des Transkriptionsmodells wurde für die Reaktionsgeschwindigkeitskonstanten der Initiation, Elongation und Termination, sowie für die Maximalrate der Transkription durchgeführt. Anhand der ermittelten Enzymkinetik konnten die Konzentrationen der T7 RNA Polymerase und des Promoters als wichtigste Einflußfaktoren unter typischen Reaktionsbedingungen der gleichzeitigen Transkription/Translation identifiziert werden.

Die Modellierung des mRNA-Abbaus erlaubt eine Unterscheidung zwischen endo- und exonukleolytischen Reaktionsschritten. Die allgemeinhin bekannte stabilisierende Wirkung erhöhter Translationseffizienz auf die Lebensdauer der mRNA konnte korrekter Weise durch das Modell wiedergegeben werden. Anhand des Abbaus von *lacZ* mRNA wurden die Parameter des Degradationsmodells bestimmt. Da die mRNA eine wesentliche Senke für Nukleosidtriphosphate verkörpert, ist es insbesondere in absatzweisen Prozessen notwendig, die erforderliche Transkriptionsrate sinnvoll auszuwählen. Modellgestützte Simulationen können wertvolle Vorhersagen liefern, um die gegenläufigen Reaktionen der Synthese und des Abbaus von mRNA adäquat aufeinander abzustimmen.

Das vorgestellte Translationsmodell umfaßt den Mechanismus des Kettenstarts, der Kettenverlängerung und des Abbruchs der Proteinsynthese unter gleichzeitiger Berücksichtigung der katalytischen Beeinflussung durch Translationsfaktoren. Ein früherer Ansatz zur Beschreibung sterischer Wechselwirkungen eines Katalysators, der an verschiedenen Positionen eines Templats gebunden ist (MacDonald et al., 1968), wurde in dieser Arbeit auf die Interaktionen zweier verschiedener Katalysatoren (z.B. Ribosom und Degradosom) erweitert.

Um das Anwendungsfeld des Gesamtmodells auf umfassendere Expressionssysteme ausdehnen zu können, wurde eine Modellreduktion eingeführt. Gemäß der vorgeschlagenen Vorgehensweise kann die Zahl der Zustandsvariablen durch ein Zusammenfassen benachbarter Basen-

tripletts deutlich reduziert werden. Die sich daraus ergebenden Implikationen wurden bei der Formulierung von Reaktionskinetiken und bei der Bilanzierung berücksichtigt.

Das vorgestellte Gesamtmodell der Genexpression ist in der Lage, Ratenlimitierungen aufgrund von Substratmangel, Inaktivierungsprozessen, oder auch ungünstig gewählten Katalysatorkonzentrationen und deren stöchiometrischen Beziehungen aufzudecken. Bei der Anwendung des umfassenden Genexpressionsmodells auf die dynamische Simulation der zellfreien Proteinbiosynthese konnten Limitierungen der Raum-Zeit-Ausbeute an Zielprotein eindeutig auf die in diesen Systemen üblicherweise vorherrschenden geringen Konzentrationen an Translationsfaktoren zurückgeführt werden. Die Ergebnisse thermodynamischer Gleichgewichtsberechnungen innerhalb der Translationsinitiation zeigen die Notwendigkeit für ein jeweils zumindest äquimolares stöchiometrisches Verhältnis der Initiationsfaktoren IF1, IF2 und IF3 im Vergleich zur Konzentration an nicht-translatierenden Ribosomen. Wenn diese Bedingung erfüllt ist, liegen laut Modellvorhersage ca. 85 % aller frei in Lösung befindlicher 30S ribosomaler Untereinheiten komplexiert mit den genannten Initiationsfaktoren vor. Dieser Komplex geht linear in die Rate der Translationsinitiation ein. Durch eine geeignete Anhebung der Konzentrationen an Initiations- und Elongationsfaktoren konnten in der Simulation um einen Faktor 4 gesteigerte Raum-Zeit-Ausbeuten und fünffach höhere Endkonzentrationen an Zielprotein im Vergleich zu einem nicht-optimierten absatzweisen Prozess erreicht werden.

Aus Sicht einer Reduzierung der Modellkomplexität ist es vorteilhaft, kinetische Parameter und Abbaukonstanten zunächst mit Hilfe eines detaillierten Modells der Genexpression abzuschätzen, bevor diese Parameter im Rahmen ganzheitlicher Modelle eingesetzt werden. Ein Hauptanwendungsgebiet des erarbeiteten Gesamtmodells ist die rekombinante Proteinsynthese, bei der üblicherweise einzelne oder wenige Gene gleichzeitig überexprimiert werden.

Gensequenzinformation wird im Gesamtmodell in folgenden Stufen verarbeitet: (a) innerhalb der Transkription durch eine Unterscheidung der Bindungsstärke regulatorischer Erkennungssequenzen des Initiations- bzw. Terminationsbereichs, (b) bei der zeitlichen Abfolge von endo- und exonukleolytischen Reaktionsschritten im gerichteten Prozeß des mRNA-Abbaus vom 5′-Ende zum 3′-Ende der mRNA, (c) während der Translation durch eine Unterscheidung Codon-spezifischer Effekte aufgrund sterischer Wechselwirkungen unter den translatierenden Ribosomen.

Desweiteren wurde der Einfluß von Vorhersagen der mRNA-Sekundärstruktur innerhalb der Ribosomenbindestelle (RBS) auf die gemessene Translationsrate untersucht. Ergebnisse dieser Analysen lassen einen Zusammenhang zwischen Einzelsträngigkeit des RBS im Bereich unterhalb der Shine-Dalgarno Sequenz und erhöhten Proteinsyntheseraten erkennen.

Insgesamt erlaubt das in dieser Arbeit entwickelte mathematische Modell der bakteriellen Genexpression detaillierte Analysen der durch die Nukleotidsequenz hervorgerufenen Effekte innerhalb der Transkription, des mRNA-Abbaus, der ribosomalen Translation und deren nichtlinearen Kopplung und stellt somit ein geeignetes Werkzeug zur rationalen Prozeßentwicklung der rekombinanten Proteinsynthese dar.

1 Introduction

Rapid advances in genomics research with improved molecular biological, analytical and computational technologies have created a massive explosion of bioinformatic databases. Owing to the development of high-throughput desoxyribonucleic acid (DNA) sequencing methods, complete genomes are now available for as diverse organisms as, for example, the bacterium *Escherichia coli*, the yeast *Saccharomyces cerevisiae*, the roundworm *Caenorhabditis elegans*, and the weed plant *Arabidopsis thaliana*. With the full deciphering of the human genome recently, in June 2000, a 'milestone in the history of mankind' has been reached, which was celebrated as the biomedical equivalent to the technological achievements of the moon-landing or the invention of the wheel.

The primary reason for the noted exhileration is the fact that the genome of an organism contains in its most condensed form all the information necessary for the explanation of life. It is the particular order of the nucleotides, of which genomic DNA is composed, that specifies the uniqueness of an organism. A comprehensive analysis and interpretation of genomic data constitutes the essential key to a profound understanding of the complex molecular processes underlying biological systems.

There are many ways by which genomic data can be analyzed and interpreted. The methods applied in functional genomics primarily aim at a prediction of protein structure and function on the basis of the coding sequence. Genomes are scanned for known consensus sequences (transcription factor binding sites, promoters, attenuators, transcription terminators, ribosomal binding sites, etc.) within open reading frames (ORFs) to pinpoint coding regions and their expression regulation on the genetic level. Over the past few years, enormous improvements have been made in the development of efficient search routines for fulfilling this task (as reviewed by Bork et al. (1998)). Through multiple sequence alignment of known protein-encoding genes, regulatory elements and homologous genes with similar functional properties can be identified that have been substantially conserved during evolution.

While the study of genomic sequences is generally agreed to be useful in tackling health-related issues, the emerging opportunities for the design and optimization of bioproduction processes have not been fully recognized. The rate and ambitiousness of genomic data collection exceed by far the rate at which these data are reflected by modern biotechnological application. An apparent lack of suitable engineering tools, that are able to make full use of biomolecular sequencing information, was pointed out recently (Bailey, 1998; Evans, 2000; Lee and Lee, 2000). Understanding the relationship between genome sequence and protein expression levels is an essential prerequisite for suggesting suitable strategies of genetic modification in the optimum design of gene expression systems.

Gene expression is carried out via two consecutive processes: the transcription of DNA into ribonucleic acid (RNA), and the subsequent translation of RNA into protein. This transfer of one-dimensional genome sequence information into three-dimensional function constitutes the central dogma of molecular biology. Implicit in the genomic nucleotide sequence is the amino acid sequence of encoded proteins. Protein sequences can therefore be inferred according to the

rules of the genetic code (units of three nucleotides characteristically code for one amino acid).

Since both RNA and proteins themselves are participating as numerous catalysts and effectors at all stages of overall protein synthesis, the reactions involved therein are extremely complex and highly coupled. While the genome of a cell remains essentially static, different genes are found to be expressed at different rates. The surrounding conditions (e.g., substrate availability, levels of inhibitors or effectors like alarmons or pheromones, and conditions like temperature and pH value) determine the type of genes which are needed to be expressed for a cell to survive. Yet, protein expression levels are not simply correlated with RNA levels (Gygi et al., 1999). Changes in the rates of transcription and translation reveal the existence of significant regulation. The possibility to fine-tune these regulatory mechanisms is a necessary basis for the cell to be able to adapt to genetic or environmental perturbation. For defined environmental conditions, the physiological state of an organism is essentially dependent on the genetic background of the organism. It should thus, in principle, be possible to derive the entire spectrum of cellular functionality and phenomenological display on the basis of genomic sequence information (Schilling et al., 1999).

The extent of protein expression is in many ways critically influenced by the encoded gene sequence. Regulatory elements at the initiation and termination sites of both the transcription and translation process are known to affect overall protein expression rate. The causes for differential mRNA degradation can mainly be attributed to nucleotide sequence variation (Coburn and Mackie, 1999). Translation rate varies notably with the coding sequence (Chaney and Morris, 1979). The functionality of a fully translated protein is dependent on amino acid sequence. Already the substitution of single amino acids (point mutations) can lead to a significant alteration of functional activity (Shen et al., 1999). Protein degradation rate is also a matter of amino acid sequence, which provides the recognition sites for proteolytic attack.

A holistic assessment of gene expression rates and the global cellular state is possible employing high-throughput bioanalytical technologies. The routine use of DNA chip technology to analyze mRNA levels (transcriptomics), the achievements made in the development of HR-2D gel electrophoresis for monitoring protein expression patterns (proteomics), and the progress in measuring *in vivo* metabolite concentrations on a cellular scale (metabolomics), have generated huge libraries of detailed biomolecular data obtained under various physiological conditions. These databases are often on-line accessible, linked with each other, and continuously updated as data collection progresses. A combination of data extracted from these libraries offers new opportunities of interpretation.

Mathematical modelling provides a quantitative frame for integrating data obtained from the various "omics" technologies (Evans, 2000). The intention of this study is to exploit, in particular, linear nucleotide sequence information for mathematically describing the dynamics involved in bacterial gene expression. At the same time, the developed model structure allows to integrate annotated genomic data with information extracted from transcriptome, proteome, and metabolome libraries. Simulation studies using this model will facilitate the *in silico* design of genomic sequences, such that improved expression rates may be resultant. The ultimate goal pursued in this context envisions the analysis and tailored design of cellular dynamics of whole

organisms within a cell population. The modelling approach taken is fully purpose-driven, as there are many applications in the field of metabolic engineering of both *in vivo* and *in vitro* recombinant gene expression systems, that could substantiate on model-based design strategies:

- **Production of RNA and homologous and heterologous proteins.** The ability to control gene expression rates is important for the large-scale production of RNA (Kern and Davis, 1999), and both homologous and heterologous protein products (Weickert et al., 1996). Achieving maximum functional activity of the correctly folded product (both protein and RNA) determines the efficacy of the synthesis process. An example for large-scale application is the production of technical enzymes such as, e.g., tissue plasminogen activator used as a dissolving agent in blood-clotting (Christou et al., 2002).
- **Modulation of enzyme activities in pathway design for metabolite production.** Pathway engineering requires full control of the expression levels of enzymes, in order to redirect the metabolic flux pattern in a predefined optimum way. Such desired flux patterns pursue the goals of metabolite overproduction and - simultaneously - minimum byproduct formation, to further improve yields of a desired product (e.g., an amino acid or vitamin to be used in food and feed industries (Moritz et al., 2000)). Alternatively, when inserting new synthesis pathways into an organism, e.g., in order to catabolize new substrate sources, it is of great interest to fine tune the relative expression levels of the enzymes connected in this pathway. An excessive overproduction of proteins constitutes a tremendous energetic effort on the cellular metabolism, which may lead to reduced growth rates (Stephanopoulos et al., 1998). In order to lower the cellular burden, an alternative approach may be taken by keeping the total cellular protein concentration essentially static and modulating the relative enzyme expression levels appropriately (Mauch et al., 2002; Visser et al., 2002).
- ***In vitro* protein biosynthesis.** *In vitro* protein biosynthesis systems allow to study particular aspects of transcription and translation, such as, e.g., the dynamic behaviour and regulation sensitivity in response to system perturbation, as well as their high-throughput application for selection and evolution of functional proteins (Hanes and Plückthun, 1997). A model-based analysis and design of these systems, e.g., for the production of toxic rDNA proteins and proteins derived from unnatural amino acids, can help to identify the causes for production limitation and to suggest suitable counter measures.
- **DNA-based and RNA-based vaccination.** Furthermore, new challenges arise in the endeavour of vaccination with genetic material (both DNA and RNA). In particular a sufficient stability of heterologous RNAs, in order to ensure a sufficient expression level and the biological functionality of the vaccine, as well as their application for target identification are important issues in this context (Saenz-Badillos et al., 2001).

Model predictions are useful to study the regulatory performance of gene networks and the differential expression of synthesis pathways in response to environmental or genetic change.

For some applications, like e.g. bulk mRNA production, it may be necessary to investigate only particular aspects of overall protein biosynthesis. To allow also for isolated studies, the gene expression model is in this study partitioned into several modules. That is, the rates of transcription, mRNA degradation, and protein production are initially treated separately. The kinetics of these reactions are modelled in terms of gene sequence and its genetic context, and moreover consider codon usage and codon adaptation. Reaction kinetics further account for a mechanistic dependence on substrate, product, and effector concentrations. In a later stage of this study, the modular building blocks of the model are assembled in a combined gene expression model to be tested in comparison to experimental observations. While the focus of this study was clearly placed on the mathematical modelling aspects, model building was supported by experimental investigations that were kindly provided by the people of the "Molecular Biotechnology Lab" (lab head Dr. Martin Siemann, Institute of Biochemical Engineering, University of Stuttgart).

Due to the nature of this study are the upcoming chapters virtually self-contained. Chapter 2 revisits the fundamental aspects of modelling the biosynthesis of macromolecules. Chapter 3 provides a kinetic model for the polymerization process of RNA synthesis by phage T7 RNA polymerase. In chapter 4, a mathematical description of prokaryotic mRNA degradation is derived from common understanding of a major bacterial mRNA degradation mechanism. A parametrization of this model is carried out in chapter 5 at the example of *lacZ* mRNA. In chapter 6, the translational apparatus for bacterial protein synthesis is modelled kinetically on the basis of genomic sequence information. In chapter 7, the translation model is applied for an optimization of catalyst stoichiometries, mainly focussing on an improvement of protein synthesis initiation. Furthermore, effects related to codon-specificity are examined in this chapter. Next, chapter 8 combines the modelling schemes of the previous chapters. In order to assess the predictability of the combined gene expression model, simulation results are compared to experiments conducted with bacterial cell-free protein synthesis systems. The impact of predicted mRNA secondary structure on protein synthesis rate is investigated in chapter 9. Finally, in chapter 10, the results of this work are critically viewed within the context of future-related studies.

2 General Aspects of Modelling Gene Expression

In order to set the basis for model selection, a cryptic definition of various model types is reviewed in the following. **Unstructured models** (black-box models) treat cellular systems as a whole, in contrast to **structured models** capturing the interaction among subcellular components. Models that track the fate of each individual species within a population (**segregated models**) distinguish themselves from **unsegregated models**, which represent the averaged behaviour of a population. Knowledge about reaction kinetics is required for **dynamic models**, which may then simplify under steady state conditions to give **stationary models**.

Additionally, mathematical models can be distinguished according to their representation of system states. Table 2.1 provides an overview of optional model structures and their main characteristics. These model types are reviewed in the following sections.

Table 2.1: Comparison of modelling structures.

Discrete Models	Continuous Models	Hybrid Models
• rule-based, probabilistic	• deterministic	• probabilistic and deterministic
• discrete states of variables	• continuous variables (must fulfill the law of large numbers)	• both discrete and continuous variables
• computationally efficient	• computationally costly	• computationally very costly (problem: synchronization of continuous and discrete modelling domains)
• Boolean models, neural networks, Monte-Carlo simulations	• differential equations	• both differential equations and, e.g., Monte-Carlo simulations

2.1 Discrete modelling

Discrete models are rule-based, where a stochastic event either takes place or not, according to the probability for this event to occur. Their computational efficiency makes these models attractive particularly for application to large systems. On the other hand, a main drawback arises from the fact that only finite changes from one discrete state to another can be monitored. Discretely modelled systems are found to respond highly sensitively to discontinuous parameter variation (D'haeseleer et al., 2000), which may not always reflect biochemical reality.

Discrete modelling is used to describe genetic networks, where variables can take the two states of being switched ON or OFF (Boolean model). Selected examples of such models applied for biochemical reaction networks are Petri Nets (Hofestädt and Thelen, 1998), neural networks (Kamimura et al., 1996), and expert systems (Brutlag et al., 1991). Moreover, discrete models were extensively applied for the purpose of mathematically describing protein biosyn-

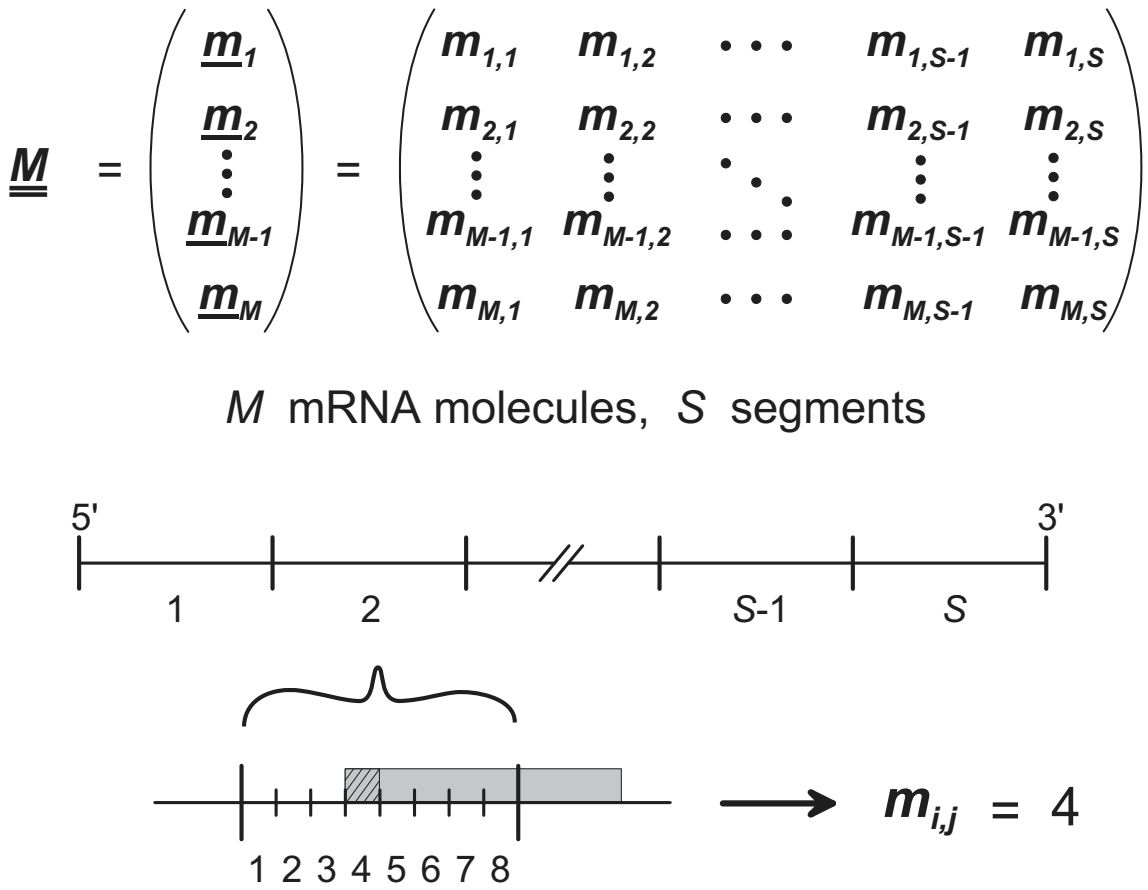

Figure 2.1: Discrete modelling of ribosome states. Matrix element $m_{i,j}$ denotes the position of a ribosome (grey-shaded rectangular) bound to segment j of mRNA i.

thesis. Gordon (1969) modelled the states of ribosomes bound to a single mRNA in terms of vector notation and computed polysomal size-distributions for various parameter sets. In this model, conditional probabilities for each discrete event, such as e.g., translation initiation, elongation, and termination, were chosen arbitrarily using Monte-Carlo simulations. Vassart et al. (1971) extended the earlier approach to cover ribosome dynamics for a fixed number of mRNA molecules by using a matrix representation (cf. Figure 2.1).

Rows denote mRNA molecules, columns indicate mRNA segments. The number given in each matrix element indicates the position (relative to each segment) that is covered by a ribosome. The model was later refined (Bergmann and Lodish, 1979; Liljenström and Blomberg, 1987) and used for investigation of various aspects of ribosomal translation: Harley et al. (1981) simulated protein synthesis under severe amino acid limitations. Menninger (1983) considered the impact of an erroneous tRNA-selection. Liljenström and von Heijne (1987) accounted for variable elongation rate, as well as Bagnoli and Liò (1995), who differentiated between codons and tRNA diversity.

A similar discrete model than the one by Vassart et al. (1971) was developed by Li et al. (1972). However, these authors achieved a deterministic model by assigning fixed time intervals to the different states a system variable can take. Singh (1969) developed a stochastic model to simulate the size distribution of polyribosomes and mRNA degradation. Much later,

the same author combined his earlier model with a Markov model (Singh, 1996), which provides the necessary probabilities for state transitions.

Another discrete modelling approach was taken by Gouy and Grantham (1980). These authors derived a probabilistic model of the tRNA cycle that simulates the behaviour of single molecules. Such an approach makes it necessary to consider the spatial three-dimensional distribution of state variables. Although computationally expensive, these models are valuable, in particular, for systems that contain state variables in very small numbers. On the other hand, for systems where each quantity balanced obeys the law of large numbers, continuum modelling applies. This approach is described in the next section.

2.2 Continuous modelling

Continuous models exhibit the form of (nonlinear) differential and algebraic equations and thereby allow to trace the continuous changes of system variables including their intermediate states. These models provide the most detailed information with the highest predictive precision. Since the computational cost is very high for large systems (often associated with many model parameters), these types of models are restricted only to moderately-sized systems. Differential equations are routinely used for a quantitative description of biological problems. The software for numerical integration, sensitivity analysis, and parameter estimation is straightforward and well-established. The computational tools frequently come already with suitable visualization packages.

A continuous framework, however, does not always apply, even though this is usually tacitly assumed. In general, it is necessary to check whether system variables exist in sufficiently large numbers of molecules to justify a continuity. If this premise is violated (e.g., in the case of only a few copy numbers of a gene), the approach is not readily applicable, unless further model assumptions are made. A continuous model may, however, still apply in the given example, if the total number of copies of the entire cell population adds up to a reasonably large number to then suffice to count as a continuum.

Models based on the continuum assumption have been formulated treating the rates of transcription, translation, and mRNA degradation in a black-box approach. In these models, state variables, like e.g., the concentrations of genes and mRNA, enter the kinetic expression in a linear fashion. First-order reaction rates are thus obtained with respect to these state variables (see Figure 2.2). Black-box models are widely used where there is only a limited amount of knowledge available about a particular reaction. When the main emphasis of an investigation is placed primarily on the model structure (i.e., the connecting links between the state variables), it may be worthwhile to accept a reduced level of detail in the description of reaction kinetics. In this context, black-box models were considered as part of structured gene expression systems (Lee and Bailey, 1984a; Biblia and Flickinger, 1992), and also for stability analysis (Hargrove and Schmidt, 1989; Hatzimanikatis and Lee, 1999). Black-box models are further attractive for large reaction networks, such as e.g., the study of pharmacokinetics in gene therapy (Ledley and Ledley, 1994).

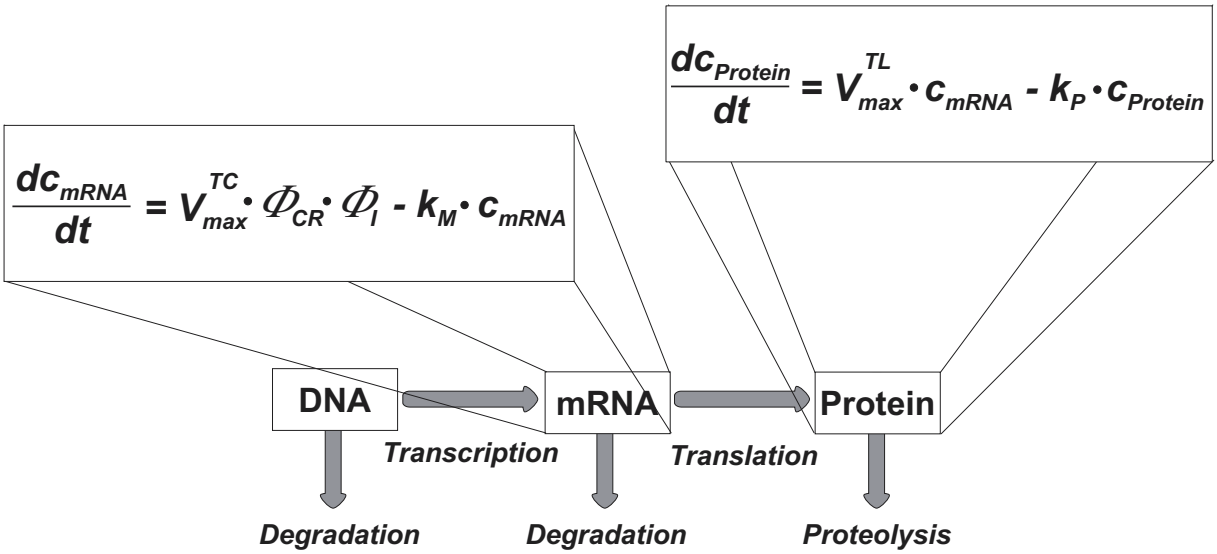

Figure 2.2: Example of unstructured modelling for representing gene expression. Material balance equations are provided for both concentrations of mRNA and protein. Symbol V_{max} denotes the maximum rate of both transcription (TC) and translation (TL), respectively. The efficiency factors for catabolite repression (Φ_{CR}) and induction (Φ_I) may themselves represent functional dependencies on the concentrations of repressor and operator region. Constants k_M and k_P are first-order degradation constants.

Probably the most compelling advantage of unstructured models is their simplicity. Not rarely, an analytical solution exists to these models, making numerical integrations obsolete. Only a single parameter is needed per each first-order reaction to fully describe the kinetics. This benefit constitutes at the same time the most severe limitation of unstructured models, because further rate-determining factors are neglected. In particular, for gene expression models based on the black-box assumption, this means that they miss out on the impact of cellular regulation denoted by the variety of synthesis rates and degradation rates observed. Model parameters thus need to be estimated experimentally and for each protein product separately, which puts great constraints on the predictive capacity of such models.

To tackle this problem and with more knowledge becoming available about a reaction mechanism, unstructured gene expression kinetics may be refined appropriately. The initial idea goes back to a formalism provided in the 1970s by Aiba and co-workers (Aiba et al., 1973), who derived an efficiency factor for both transcription and translation. These factors express a functional dependency on the concentration of regulatory components and may be multiplied with the respective maximum rate to modulate the conversion rate (cf. Figure 2.2). Model expansions leading to genetically structured models were given by Bailey and co-workers (Lee and Bailey, 1984b; Chen et al., 1991).

More sophisticated continuous models were developed for simulating DNA replication by distinguishing among site-specific states of a template (Simha et al., 1963; Zimmerman and Simha, 1965). In a later study, this modelling approach was extended to account for neighbouring effects among the template-bound catalysts (Silberberg and Simha, 1968). Gerst and Levine (1965) developed a deterministic model on the basis of differential equations for de-

scribing the dynamics of polyribosomes. However, these authors omitted the impact of sterical interactions among translating ribosomes. In a steady-state analysis, Godefroy-Colburn and Thach (1981) investigated the effect of mRNA competition on regulating translation rates. These authors further considered the case where translation initiation is blocked by ribosomes that are already bound within the initiation site.

A continuous model for reversible polymerization processes on a template was developed by the working group of Gibbs (Pipkin and Gibbs, 1966; MacDonald et al., 1968; MacDonald and Gibbs, 1969). Characteristic to their approach is the step-wise travel of a catalyst along the template, whereby at each step, a monomer is linked to a nascent product chain. The biopolymer synthesis considered an analogy to the physical problem of cooperative diffusion along a one-dimensional lattice (Pipkin and Gibbs, 1966). Mass transfer rates for successive monomer addition were derived on the basis of the fractional loading of each template site (MacDonald et al., 1968). The same model structure, was later extended to describe the impact of mRNA secondary structure on the overall translation rate (von Heijne et al., 1977; 1978). Under simplifying assumptions regarding the original model, it was moreover possible to reduce the number of differential equations to a single one (Heinrich and Rapoport, 1980). This model reduction holds only for the special situation if translating ribosomes are uniformly distributed over the length of a mRNA (including the termination site), and when they all propagate at the same specific rate.

Heinrich and Rapoport (1980) performed a transition from fractions to molarities and included a balance for total ribosomes. These authors were the first to provide time-dependent solutions to a translation model. They also treated a system of two competing mRNAs, which differed in their rate constant for translation initiation.

Apart from the above continuous models, gene expression has been modelled as an autocatalytic relaxation process (Chela-Flores et al., 1988). Mahaffy (1993) lumped all steps involved in both transcription and translation together to form a time delay until the full-length protein is assembled. In order to study the effects of clustering of low-usage codons (rare codons) as a function of their position along the mRNA and their impact on protein production rate, Zhang et al. (1994) developed a prokaryotic translation model consisting of algebraic equations. Their model illustrates the position of ribosomes on a mRNA and their residence times at different codons. The model is also capable of including interactions among polyribosomes. Götz and Reuss (1997) modelled time delays in microbial growth by considering the polymerization reaction of ribosome synthesis. In a recent study by Drew (2001), prokaryotic protein synthesis was modelled on the basis that transcription initiation rate is modulated by various states the polymerase binding site can take (e.g., being activated or repressed). Probabilities for the different states of DNA were represented by a Markov model, their time evolution was given by a contiuous black-box model. However, no polyribosomes and hence no queueing were considered.

2.3 Hybrid modelling

Hybrid modelling constitutes a combination of both continuously modelled variables and discrete state transitions. Conceivable for protein expression is a combination where the steps of the polymerization processes are treated in a discrete model, while the reaction rates are described in a continuum model, giving their functional dependence on the concentrations of substrates, products and effectors (Schmid, 1999). It is first problematic to interpret finite changes of mass transfer (discrete model) in terms of continuously modelled kinetic rates. And in the reverse direction, the question of converting kinetic rates into probabilities is an unsolved problem.

Apart from these interpretational obstacles, also serious technical difficulties need to be solved for such an approach to become computationally tractable. The exchange of information between the discrete model, which typically uses a constant finite step-time, and the numerical integrator needs to take place at certain (fixed) intervals of computed time. This imposes hard constraints on the nature of the numerical integrator. In particular, a multi-step method with variable step-size may have difficulties to match these pre-set instants of time. Alternatively, a constant integrator step-size may be used throughout, however, at the expense of choosing a smaller step-size associated with low computational efficiency. In this case, it may also be particularly difficult to guarantee numerical stability. Hybrid models have been applied for describing automated transportation, e.g., in developing software to avoid aircraft collisions during the flight (Livadas et al., 1999).

Hybrid modelling evidently requires a particular expertise in understanding the core processes involved in numerical integration. It may after all be necessary to develop a new numerical algorithm that is specifically capable of handling these problems.

2.4 Classification of polymerization processes

All polymerization processes are characterized by a reaction initiation providing activated reactants, a stage of propagation with the successive increase in the degree of polymerization, and the termination reaction completing polymer growth. Two classes of polymerizations can be distinguished: **chain polymerization** and **step polymerization**. Chain polymerization requires an activating initiator (e.g., a free radical or ionic species) to start polymer synthesis. At step polymerization, monomers with two different functional groups are needed in order for continued monomer addition to occur. The stoichiometry of the initial feed in terms of monomer ratios determines the degree of polymerization. For synthetic polymers, a population of unbranched, branched, and cross-linked polymers is typically obtained with different number average degrees of polymerization. Populations are generally represented by molecular weight distributions.

In that sense, **template-driven** and **enzyme-catalyzed polymerization** processes, like e.g., DNA, RNA, and polypeptide synthesis may be viewed as special, simplified cases of free-radical polymerization, where the diffusion problem reduces from a three-dimensional to a one-dimensional system (Pipkin and Gibbs, 1966). The role of the template is to provide the

Table 2.2: Comparative juxtaposition of the characteristic features of general polymerization versus enzymatic polymerization on templates.

General polymerization	**Enzymatic polymerization on templates**
• 3D-diffusion problem	• 1D-diffusion problem
• random order of monomer addition	• controlled order of monomer addition (cooperativity)
• unbranched, branched, and cross-linked products possible	• unbranched product
• molecular weight distribution of products	• fully completed polymers are identical
• initial stoichiometry of reactants determines the product spectrum	• initial stoichiometry determines the final amount of the single polymer species
• no correction mechanisms	• enzymatic correction mechanisms (proofreading) for minimizing error rate

order of monomers to be added to the growing unbranched polymer. Due to the enzyme involvement, each step is cooperatively linked to previous polymerization steps. Random monomer incorporation is excluded, as far as an erroneous selection of monomers can be ruled out. All polymer molecules completed exhibit the same monomer sequence and thus have identical physical and chemical properties (e.g., they are unbranched and have the same molecular weight and functionality). Yet further processing of the product, like e.g., post-transcriptional or post-translational modification, can take place subsequent to the polymerization process, thereby increasing product variety and altering their activity status. Enzymatically catalyzed polymerization processes are highly controllable and permit proofreading steps to minimize error frequencies.

According to the noted differences between synthetic polymerization and enzymatic polymerization on templates (see also the list provided in Table 2.2), typical polymerization models used in chemical engineering do not readily apply for this kind of biopolymerization.

2.5 RNA secondary structure

Among the primary determinants for governing mRNA lifetime are the stabilizing and destabilizing single-stranded and double-stranded regions, which in their entirety characterize the **secondary structure** of mRNA. Dependent on the nucleotide sequence, these structural properties of mRNA are formed over the course of RNA folding due to hydrogen bond formation and stacking of adjacent nucleic bases. Some examples for RNA secondary structures are shown in Figure 2.3. Binding between Guanine (G) and Cytosine (C), a triple-hydrogen bond, is usually stronger than the double-hydrogen bond between Adenine (A) and Uracil (U). Further types of interactions between bases have been identified (Nagaswamy et al., 2000; and the references therein), among which the wobble pair G-U is the most common. The strength of hydrogen bond formation is extremely context-dependent (Mathews et al., 1999; Guerra et al., 2000).

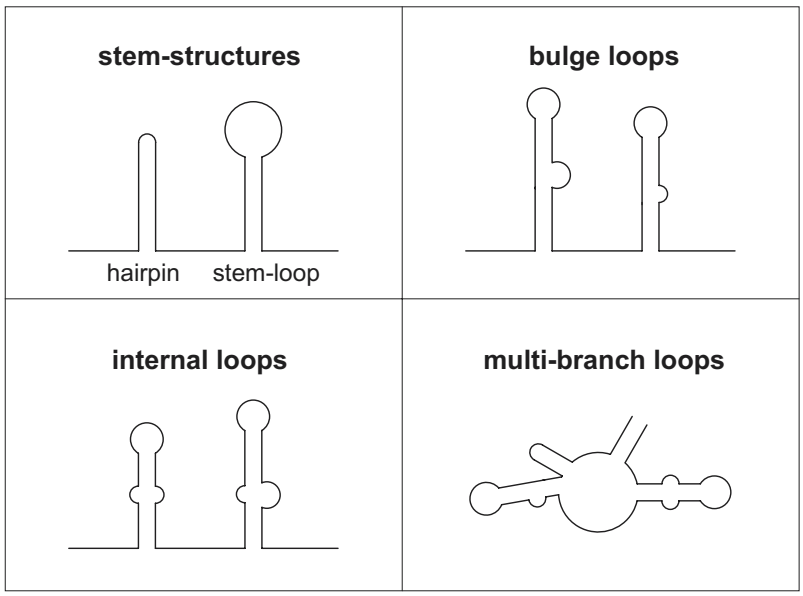

Figure 2.3: Secondary structural elements of RNA.

Secondary structure prediction. Powerful computer programs have been developed to perform the predictions of RNA secondary structure automatically. Some examples are the Vienna RNA Package[(1)] (Hofacker et al., 1994), ESSA[(2)] (Chetouani et al., 1997), X2s[(3)] (Juan and Wilson, 1999), and the *mfold* package[(4)] (Mathews et al., 1999; Zuker et al., 1999). These tools have been developed on the basis of nucleotide sequence and by summation of individual base-pair interactions. Standard free energies of formation, ΔG_f°, denote the binding strength of base pairing. Thermodynamic parameters used in the computation have been tabulated in the context of their nucleotide sequence (as reviewed by Zuker (2000)). Prediction methods typically give the RNA conformation with the lowest free energy. Secondary structure predictions may be further improved in combination with comparative sequence alignment methods.

mRNA secondary structures are a simplification of the three-dimensional folding of an RNA molecule. Some 3D-structures, like e.g., the pseudoknot, which results from base-pairing to further distant RNA sections, have been predicted with a dynamic programming algorithm (Rivas and Eddy, 1999).

There have been many efforts in creating suitable visualization tools for displaying predicted RNA secondary structures (as reviewed by Zuker (2000)). One outstanding example is the graphics package RNA movies[(5)] (Evers and Giegerich, 1999). This package is capable of dynamically illustrating chain growth or transient folding pathways, that arise when switching between different folding conformations.

In many cases, predicted RNA secondary structures have been verified experimentally. Commonly used experimental methods for secondary structure detection are X-ray crystallography (Baeyens et al., 1996) or NMR spectroscopy (Cheong et al., 1996).

[(1)] Vienna RNA Package at `http://www.tbi.univie.ac.at/~ivo/RNA/`

[(2)] ESSA at `http://www-bia.inra.fr/T/essa/Doc/essa_home.html`

[(3)] X2s at `http://wyrant.ucsc.edu/software.html`

[(4)] *mfold* at `http://www.ibv.wustl.edu/~zuker/rna/mfold-3.1.html`

[(5)] RNA movies at `http://BiBiServ.TechFak.Uni-Bielefeld.DE/rnamovies/`

Figure 2.4: The unfolded (U) ribosome binding site is involved in the two competing reactions of attaining a folded (F) conformation, and its association with the 30S subunit (modified figure taken from de Smit and van Duin (1990b)). $\Delta G^0_{\rm f}$ denotes the free energy of secondary structure formation. $\Delta G^0_{30\rm S}$ represents the free energy of 30S binding to the unfolded RBS.

Equilibrium between folded and unfolded structure. One of the key mechanisms in controlling gene expression is message-specific translational control. Since ribosomal interaction with the Shine-Dalgarno sequence and initiator tRNA association involve base-pairings, the mRNA needs to be single-stranded during the process of translation initiation (de Smit and van Duin, 1990a). Structural properties of the ribosome binding site (RBS) have a profound effect on the location of the thermodynamic equilibrium between the folded and unfolded conformation of this site (cf. Figure 2.4). With higher strengths of secondary structures in the RBS is this equilibrium shifted towards the folded conformation (de Smit and van Duin, 1994a; 1994b). On the other hand, sequence mutations within the RBS can alter the quality of Shine-Dalgarno interaction during the binding of the 30S subunit to mRNA and can thus affect the binding affinity.

In a series of studies, de Smit and van Duin provided a quantitative relationship between translational efficiency and the strength of a mRNA secondary structure located particularly within the initiation region (de Smit and van Duin, 1990a; 1990b; 1994a; 1994b). According to their findings, the efficiency of ribosomal binding is correlated with the fraction of mRNA, $f_{\rm u}$, exhibiting unfolded ribosome binding sites (de Smit and van Duin, 1990a; 1990b; Draper, 1993). $f_{\rm u}$ was given by

$$f_{\rm u} \;=\; \frac{1}{K_{\rm f}+1}. \tag{2.1}$$

The equilibrium constant, $K_{\rm f} = C^{\rm f}_{\rm RBS}/C^{\rm u}_{\rm RBS}$, for the equilibrium between the folded and unfolded RBS can be calculated from the $\Delta G^0_{\rm f}$, the free energy of secondary structure formation, according to (de Smit and van Duin, 1990b)

$$\Delta G^0_{\rm f} \;=\; -RT\,\ln K_{\rm f}. \tag{2.2}$$

Equation (2.2) is also known as the Van't Hoff equation. R is the universal gas constant (8.314 kJ/kmol·K), T is the absolute temperature (K). $\Delta G^0_{\rm f}$ serves as a measure for the stability of the secondary structure. A stability increase of a RBS secondary structure by -5.9 kJ/mol reduces translational efficiency (indicated by $f_{\rm u}$) by about one order of magnitude (de Smit

and van Duin, 1994a). Structural elements of mRNA which had a stability of less than about -25.1 kJ/mol were found to cause no adverse impact on translational initiation (de Smit and van Duin, 1994a). Solving equation (2.2), parameter K_f may thus be written as

$$K_\mathrm{f} = \begin{cases} e^{\left[-\dfrac{\Delta G^0_\mathrm{f}}{RT}\right]} & \text{if } \Delta G^0_\mathrm{f} \leq \text{-25.1 kJ/mol,} \\ 0 & \text{if } \Delta G^0_\mathrm{f} > \text{-25.1 kJ/mol.} \end{cases} \tag{2.3}$$

Considering equations (2.1) and (2.3), the fraction of unfolded mRNA becomes

$$f_\mathrm{u} = \begin{cases} \left(e^{\left[-\dfrac{\Delta G^0_\mathrm{f}}{RT}\right]} + 1 \right)^{-1} & \text{if } \Delta G^0_\mathrm{f} \leq \text{-25.1 kJ/mol,} \\ 1 & \text{if } \Delta G^0_\mathrm{f} > \text{-25.1 kJ/mol.} \end{cases} \tag{2.4}$$

2.6 Conclusions

The selection of model structure is guided by the application purpose of the model and the required level of biochemical detail. In order to be able to derive a functional relationship between genomic sequence information and gene expression rates, both on the mRNA and protein level, a highly detailed model is needed. The necessity to dynamically represent codon-related effects, as well as to handle mechanism-based kinetic rate expressions, impose significant constraints on the applicability of potential model structures. It is trivial to note that a black-box modelling approach is insufficient in pursuing this task.

Discrete models of protein synthesis offer an illustrative representation of codon-specific states. However, the handling of functional dependencies of reaction rates is problematic. For discrete models of protein synthesis, as far as the published works are concerned, the rates of transcription, translation, and mRNA degradation are required to be model entries. These rates are typically kept constant, even under dynamically changing conditions. In an attempt to predict these rates on the basis of nucleotide sequence information, a discrete modelling approach is thus not applicable in this study.

Hybrid modelling, which combines discretely modelled state variables with continuously modelled ones, is in principle feasible, albeit not practical from a modelling standpoint. It could allow for both codon-specific effects and the necessary kinetic detail. The excluding factors arise, however, mainly because of the associated disadvantages in data transfer between both modelling domains.

In contrast, a continuous modelling approach based on differential and algebraic equations appears particularly favourable to match the pre-defined model requirements. A continuous model formulation offers a straightforward method to dynamically simulate kinetically interacting state variables, also down to a codon-specific resolution. Further, the consideration

of mRNA secondary structure seems in principle compatible with a continuous modelling approach, due to the possibility to derive model parameters (like e.g., equilibrium constants) from hairpin stabilities. The ready possibility to link modular building blocks, simply by increasing the number of material balance equations, is a necessary premise for a holistic model concept, in which the individual modelling aspects are combined. Based on the above reasoning, a continuous modelling approach is chosen in this study to describe the key processes involved in bacterial gene expression.

3 Transcription

3.1 Introduction

Transcription is the enzymatically catalyzed process of RNA formation. The enzyme carrying out this biopolymerization reaction is a DNA-dependent RNA polymerase (RNAP). Due to its relatively simple catalytic mechanism in comparison to other RNA polymerases, bacteriophage T7 RNAP (EC 2.7.7.6) was chosen here as a model system for investigating transcription kinetics. This enzyme consists of a monomeric subunit of 98 kDa (Stahl and Zinn, 1981) and performs transcription termination factor-independently. The biotechnological importance of T7 RNAP is evident, since the enzyme is frequently employed for both *in vivo* recombinant protein production (Weickert et al., 1996) and cell-free gene expression systems (Stiege and Erdmann, 1995; Jermutus et al., 1998).

A general scheme of the transcription process is provided in Figure 3.1. Initial GTP binding to T7 RNAP was suggested to infer promoter specificity (Basu and Maitra, 1986; Sen and Dasgupta, 1993), or else to occur subsequent to promoter association to the enzyme (Martin and Coleman, 1987; Jia and Patel, 1997). Moreover, GMP and guanosine were reported to enable transcript initiation (Martin and Coleman, 1989). Inorganic pyrophosphate was shown to inhibit RNA synthesis rate (Cunningham and Ofengand, 1990; Guajardo and Sousa, 1997). In the absence of substrate nucleotides other than GTP, transcript slippage might take place within the initiation region and when the transcript begins with GGG (Martin et al., 1988). Transcript abortion, the premature termination of RNA synthesis, is a phenomenon commonly observed for T7 RNAP (Sousa et al., 1992). Abortive transcription leads to oligonucleotides that are typically less than 14 nucleotides in size. At termination, a switch from a processive to an abortive transcription stage occurs, which is accompanied with the release of a fully synthesized RNA molecule. This RNA liberation is thought to be mediated through hairpins impairing the interactions between the nascent RNA and the nucleic acid binding site (Job et al., 1988; Sousa et al., 1992).

The stoichiometry of the transcription reaction leading to the synthesis of a RNA of length n is given by:

$$n_{\mathrm{A}}\mathrm{ATP} + n_{\mathrm{C}}\mathrm{CTP} + n_{\mathrm{G}}\mathrm{GTP} + n_{\mathrm{U}}\mathrm{UTP} + \mathrm{DNA} \longrightarrow \mathrm{RNA}_n + (n-1)\ \mathrm{PP}_i + \mathrm{DNA} \qquad (3.1)$$

For each nucleoside triphosphate (NTP) other than the initiator nucleotide that is incorporated into the growing RNA chain, one molecule of inorganic pyrophosphate (PP_i) is released. n_{A}, n_{C}, n_{G}, and n_{U} denote the number of adenine, cytosine, guanine, and uracil bases, respectively, of which the RNA product is assembled. Their sum equals total RNA length n.

Modelling the dynamics of RNA polymerization is important for obtaining an improved system understanding as a basis for application in both *in vitro* and *in vivo* gene expression processes. Transcription models are moreover useful for the optimization of large-scale *in vitro* RNA production (Kern and Davis, 1999), and for the design of RNA-based tumor vaccination (Saenz-Badillos et al., 2001).

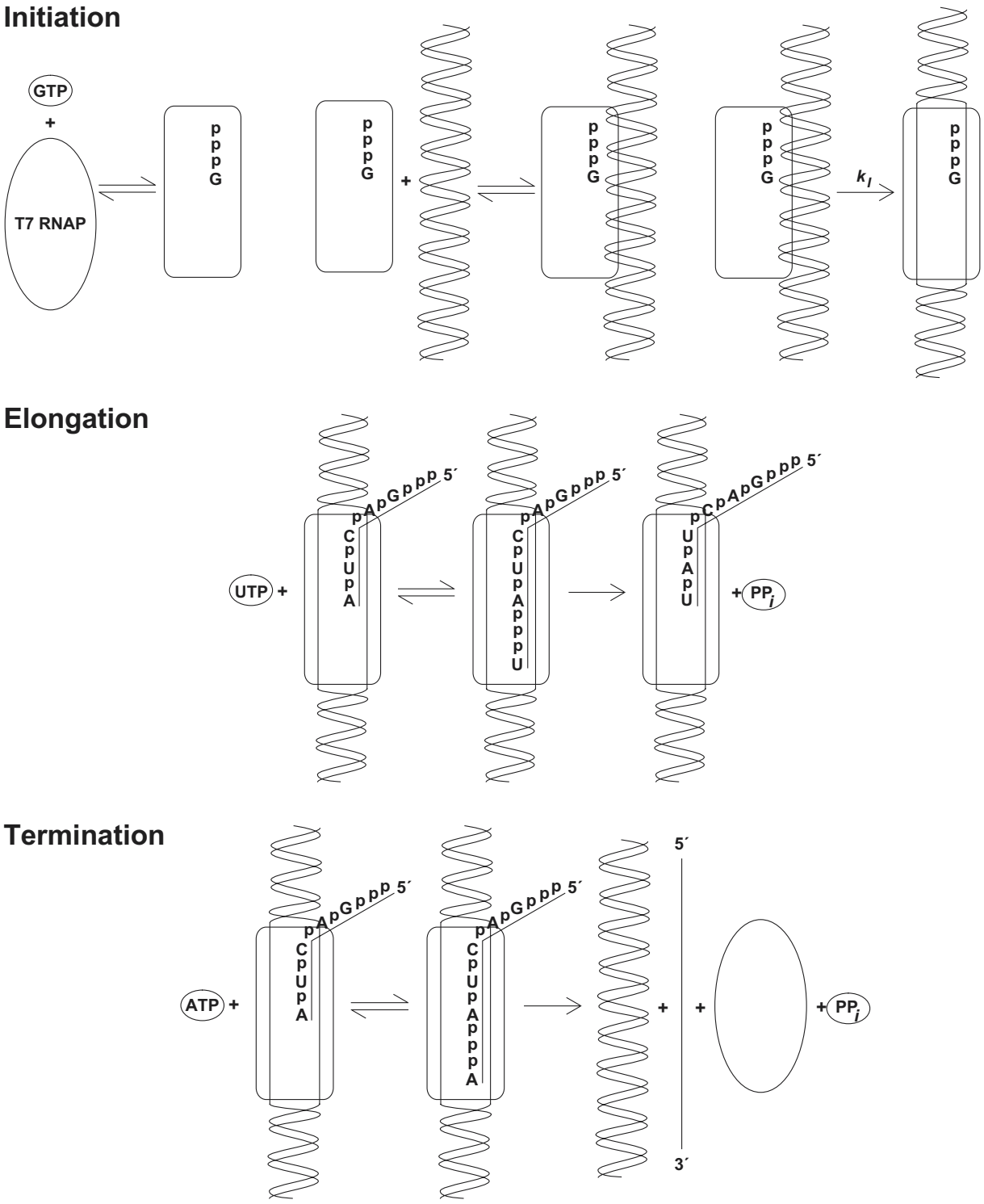

Figure 3.1: Schematic of transcription by T7 RNA polymerase.

In previous modelling, particular emphasis has been placed on elucidating the kinetics of transcription initiation (Martin and Coleman, 1987; Maslak and Martin, 1994; Újvári and Martin, 1996; Kuzmine and Martin, 2001), which is commonly believed to be the rate-limiting step of the overall RNA synthesis process. Transcription elongation has been modelled as a ping-pong bi-bi mechanism for *Escherichia coli* RNA polymerase and considering competition among substrate nucleotides (Rhodes and Chamberlin, 1974). Cooperative transcription kinetics as a result of enzyme isomerization were derived for wheat-germ RNA polymerase II (Job et al., 1988). The effects of pH and free Mg^{2+} ion concentration on RNA synthesis rate were accounted for short oligonucleotides by Young et al. (1997). Transcript polymerization kinetics distinguishing initiation, elongation, and termination arrived at a Michaelis-Menten-type equation for each limiting substrate nucleotide separately (Pozhitkov et al., 1998).

Despite the number of mathematical models developed for describing transcription kinetics, only a few studies have actually made use of these models for simulation purposes. Young et al. (1997) simulated transcription dynamics for optimization of an initial system composition to obtain maximum RNA yields. In studying the mRNA degradation mechanism embedded in prokaryotic gene expression, Carrier and Keasling (1997) derived a probabilistic model for transcription. In their model, the dynamics of RNAP population were represented by a series of discrete, conditional events to reflect the steric interactions among the transcribing enzymes. Discrete models are very illustrative modelling approaches, and they are useful to readily account for erroneous substrate incorporation (Menninger, 1983). However, when reaction rates, which are in the discrete modelling approach expressed as model constants, are found to vary over the course of the synthesis process, it is unclear how their functional dependencies on pool concentrations should be included in a stochastic model. Essentially, it is an unresolved question as to how the probabilities for a particular event to take place may be derived from the corresponding reaction rates. This is a problem generally encountered whenever deterministic and probabilistic models need to be combined (Hofestädt and Thelen, 1998).

The spectacular advances in the biosciences require predictive modelling tools to describe RNA production and protein synthesis (Bailey, 1998). An attempt is made in the following to derive - on the basis of genomic sequence information - a mechanistic model for T7 RNA polymerase kinetics that incorporates, in particular, template-specific properties and additionally reflects the impact of key reactant concentrations. Some essential features of this model have been published elsewhere (Arnold et al., 2001).

3.2 Materials and methods

3.2.1 Plasmids

Three vector systems, namely pT3/T7luc, pEThyd, and pETK411BscFv (see Table 3.1), have been employed for investigation of transcription kinetics. Each of these plasmids contains a T7 promoter and three transcription terminators, respectively, whose recognition sequences are specified in Table 3.2. Promotor $P_{pT3/T7luc}$ is contained on plasmid pT3/T7luc. Plas-

Table 3.1: Vector characteristics.

Plasmid	Structural genes	Source
pT3/T7luc	firefly luciferase	Clontech Inc. (Palo Alto, USA)
pEThyd	L-hydantoinase from *Arthrobacter aurescens*	May et al. (1998)
pETK411BscFv	recombinant single chain Fv-fragment (scFv) of antibody K411B against atrazine	kind gift by Prof. Rolf Schmid of the Institute of Technical Biochemistry (University of Stuttgart, Germany)

Table 3.2: Recognition sequences of promoters (P) and terminators (T) investigated.

Site	Gene sequence
$P_{pT3/T7luc}$	TAATACGACTCACTATAGGGCGA
$P_{\phi 10}$	TAATACGACTCACTATAGGGAGA
T_{ϕ}	CAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAG
T_{bla}	AAACCACCGCTGGTAGCGGTGGTTT
$T_{pT3/T7luc,1}$	n.d.
$T_{pT3/T7luc,2}$	n.d.
$T_{pET,1}$	n.d.

mids pEThyd and pETK411BscFv were derived from the same vector (pET *E. coli* expression system) and thus have the same T7 promotor, $P_{\phi 10}$, and termination signals.

3.2.2 *In vitro* transcription buffer

In all transcription experiments, the reaction conditions contained 6 mM $MgCl_2$, 2 mM spermidine, 10 mM DTT, 40 mM TrisHCl (pH 8.0). Unless otherwise indicated, 84 nM of the respective plasmid were included in the reaction assay. The following components were added at varying initial concentrations: nucleoside 5′-triphosphates, inorganic pyrophosphate, and yeast inorganic pyrophosphatase. Generally, a reaction temperature of 37°C was chosen. Transcription dynamics were started through the addition of T7 RNAP (Roche Diagnostics), as reported elsewhere (Werner, 1997). The specific activity of the enzyme was measured and gave 320,000 U/mg (Siemann, 1998). When the transcription assay was tested for DNase, RNase, and NTPase activity, no such degradation reactions were measurable (data not shown).

3.2.3 Radioactive labelling of RNA

The two radionuclides applied for isotope labelling were [5,6-^{3}H]-UTP and [8-^{14}C]-ATP, both purchased from Amersham Pharmacia Biotech (Buckinghamshire, UK) and had a specific activity, γ, of 1.67 TBq/mmol and 1.92 GBq/mmol, respectively, as noted by the supplier. The isotopes were diluted appropriately (dilution factor δ) with their nonlabelled analogues.

Samples taken from the transcription assay including [^{3}H]-UTP were precipitated with TCA solution, applied onto a Whatman filter, and finally analyzed in a Tri-Carb Liquid Scin-

Table 3.3: Effective rate constants for synthesis of the various RNA species obtained from the applied vector systems. Transcript length, respective number of uracil bases, and relative amounts of individual RNAs synthesized are also given.

Plasmid	Promoter	Terminator	RNA (kB)	n_{U}	m (% IOD)	$k_{TC,\mathrm{eff}}$ (s^{-1})
pT3/T7luc	$\mathrm{P}_{\mathrm{pT3/T7luc}}$	$\mathrm{T}_{\mathrm{pT3/T7luc,1}}$	1.890	504	9	8.8
	$\mathrm{P}_{\mathrm{pT3/T7luc}}$	$\mathrm{T}_{\mathrm{pT3/T7luc,2}}$	2.410	634	53	52.1
	$\mathrm{P}_{\mathrm{pT3/T7luc}}$	$\mathrm{T}_{\mathrm{bla}}$	2.951	749	38	37.5
pEThyd	$\mathrm{P}_{\phi 10}$	T_{ϕ}	1.512	292	64	73.8
	$\mathrm{P}_{\phi 10}$	$\mathrm{T}_{\mathrm{pET,1}}$	2.406	570	6	6.9
	$\mathrm{P}_{\phi 10}$	$\mathrm{T}_{\mathrm{bla}}$	3.057	708	30	34.6
pETK411BscFv	$\mathrm{P}_{\phi 10}$	T_{ϕ}	0.996	231	68	67.6
	$\mathrm{P}_{\phi 10}$	$\mathrm{T}_{\mathrm{pET,1}}$	2.175	593	4	4.0
	$\mathrm{P}_{\phi 10}$	$\mathrm{T}_{\mathrm{bla}}$	2.744	718	28	27.8

tillation Analyzer (Model 1900TR, Packard Instrument Co.), as indicated elsewhere (Arnold et al., 2001). The radioactivity signal, X (measured in dpm/μL), was corrected for the background obtained from reactions without any plasmid.

Size determination of the synthesized RNA products was carried out in denaturing 5 % PAGE containing 7 M urea at 65°C. In this analysis, the standard reaction assay contained additionally [^{14}C]-ATP for transcript labelling. Each sample of the transcription assay was treated as described previously (Sambrook et al., 1989) and applied onto the gel. The finally obtained fluorograms were quantified using the ImageMaster software (Amersham Pharmacia). The results of the gel analysis (see Figure 3.2) yielded the lengths, n_i, of all RNA transcripts synthesized and their relative amounts, m_i, given in % integrated optical density (IOD). These data are summarized in Table 3.3. With this knowledge, the concentration of each individual RNA species produced by a vector system can be calculated from scintillation data using the subsequent relationship:

$$C_{\mathrm{RNA},i} \;=\; \frac{1}{\gamma}\cdot\frac{\delta}{\alpha}\cdot\frac{m_i\,n_i}{n_{\mathrm{U},i}}\cdot X \tag{3.2}$$

Parameter α was determined to be 0.3 in the experimental set-up and for [^{3}H]-UTP. $n_{\mathrm{U},i}$ is the number of uracil residues of transcript i. Other constants depicted in equation (3.2) were introduced in the above. Within this chapter, RNA concentration is indicated in nucleotide moles per unit of reaction volume. The total RNA concentration, C_{RNA}, follows from

$$C_{\mathrm{RNA}} \;=\; \sum_{i=1}^{R} C_{\mathrm{RNA},i} \tag{3.3}$$

with R being the number of RNA species generated from each plasmid.

3.2.4 Ion complex formation

Both Kern and Davis (1999) and this study stress the necessity to distinguish between total magnesium concentration added to a transcript synthesis system and the resultant free Mg^{2+}

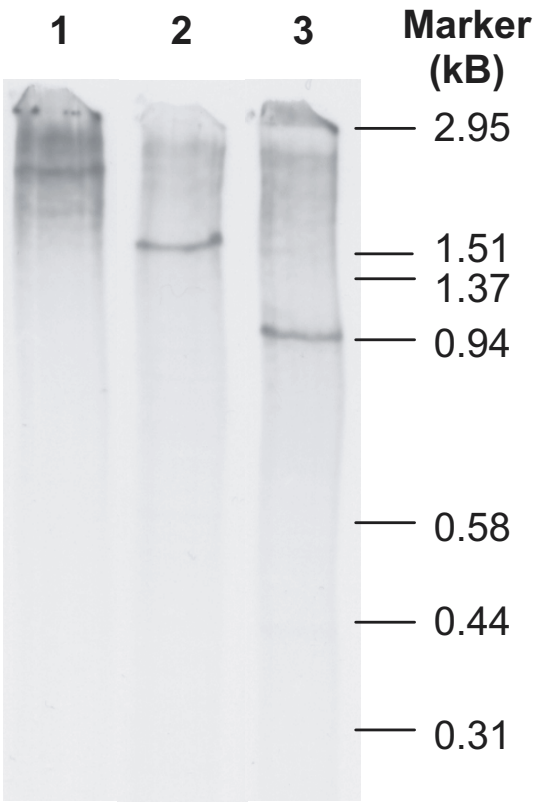

Figure 3.2: RNA length distribution investigated with acrylamide gel electrophoresis for templates (1) pT3/T7luc, (2) pEThyd, and (3) pETK411BscFv. Samples were taken after 90 min of a batch transcription performed under standard reaction conditions containing the respective plasmid. Independent transcription assays were made under standard reaction conditions adding the respective plasmid. The evaluated data using the ImageMaster software (Amersham Pharmacia Biotech) are summarized in Table 3.3.

ion concentration. The latter is determined by the concentrations of Mg-binding reactants and their binding affinities. Therefore, the free Mg^{2+} ion concentrations were estimated in this study. The iterative estimation procedure relied on an earlier suggested method of material balancing (Storer and Cornish-Bowden, 1976). Additionally, Magnesium binding to DNA and RNA was considered by assuming in average 1 molecule of Mg^{2+} to be bound per each 2 phosphate groups of nucleic acid (Record et al., 1976). Mg-binding to Tris buffer was shown earlier to be negligible (Good et al., 1966). Ion complex formation could then be calculated using the dissociation constants listed in Table 3.4. Calculus details of this method are provided in the Appendix B. The free Mg^{2+} ion concentrations indicated in this study represent the geometric mean between the conditions at the beginning and termination of each experiment.

According to our estimation of ionic species concentrations, there was only a minor change in the resulting concentration of free Mg^{2+} during an experiment. The relative difference in free Mg^{2+} concentration calculated between the start and termination of an experiment was maximally 8 %. While in our study, there was thus no immediate need to include the impact of changing free Mg^{2+} concentration, the study of Young et al. (1997) required to compute the free Mg^{2+} concentration with progressing reaction time. This was due to the high initial NTP concentration applied by these authors and the resultant dynamics of (in particular)

Table 3.4: Applied dissociation constants for complex formation between dissolved ions and low molecular weight components.

Dissociation constant	Unit	Value	Source
K_{HNTP}	(M)	$1.26 \cdot 10^{-7}$	Langer et al. (1977)
$K_{\mathrm{H_2NTP}}$	(M)	$8.71 \cdot 10^{-5}$	dto.
$K_{\mathrm{Mg_2NTP}}$	(M)	$5.01 \cdot 10^{-2}$	dto.
$K_{\mathrm{MgNTP_2}}$	(M)	$3.98 \cdot 10^{-3}$	dto.
K_{MgNTP}	(M)	$1.15 \cdot 10^{-5}$	O'Sullivan and Perrin (1964)
K_{MgHNTP}	(M)	$1.41 \cdot 10^{-3}$	dto.
K_{MgPi}	(M)	$1.32 \cdot 10^{-2}$	Smith and Alberty (1956)
K_{MgPPi}	(M)	$1.70 \cdot 10^{-5}$	Käpylä et al. (1995)
$K_{\mathrm{Mg_2PPi}}$	(M)	$2.14 \cdot 10^{-3}$	dto.

NTP concentration, total RNA and precipitated $Mg^{2+}PP_i$ during NTP conversion. The time-dependent variation of these compounds brought about also a profound change in free Mg^{2+} concentration.

3.3 Dynamic modelling

3.3.1 Model assumptions

From the common understanding of T7 RNAP mechanism, the reaction scheme displayed in Figure 3.3 has been derived. Model assumptions inherently made in this representation are:

1. **Initiation.** GTP is the initiator nucleotide. A random order of binding of T7 RNAP to the promoter, D, and GTP is possible. T7 RNAP is highly specific to its promoter, with the binding constants given for promoter association to be $1.0 \cdot 10^8$ M^{-1} versus non-promoter association to be $2.1 \cdot 10^4$ M^{-1} (Gunderson et al., 1987). Nonspecific binding to DNA is neglected.

2. **Elongation.** Nucleotide association to the transcription complex of T7 RNAP, DNA, and RNA_j (complex termed $E \cdot D \cdot M_j$ in Figure 3.3) is independent of neighbouring nucleotides of the DNA sequence. The rate constant, k_{TCE}, denotes an irreversible translocation step, at which one molecule of inorganic pyrophosphate is released.

3. **Competitive Inhibition.** Nucleotides and inorganic pyrophosphate competing with the binding of cognate substrate nucleotide are allowed to bind to freely dissolved T7 RNAP, to the enzyme-promoter complex, and to the elongating enzyme. The error frequency of transcription is negligible with a reported probability of 10^{-5} (Blank et al., 1986).

4. **Termination.** The processes involved in transcription termination are combined into one irreversible reaction step, at which the fully synthesized RNA product is released.

The kinetic model developed inherently assumes that the system has settled in a pseudo-steady state. While the validity of this assumption has not been deliberately tested in this study, there

Initiation

$$\mathrm{E} + \mathrm{GTP} \overset{K_\mathrm{G}^\mathrm{I}}{\rightleftharpoons} \mathrm{E{\cdot}GTP} \qquad \mathrm{E} + \mathrm{D} \overset{K_\mathrm{D}}{\rightleftharpoons} \mathrm{E{\cdot}D}$$

$$\mathrm{E{\cdot}GTP} + \mathrm{D} \overset{K_\mathrm{D}}{\rightleftharpoons} \mathrm{E{\cdot}D{\cdot}GTP} \qquad \mathrm{E{\cdot}D} + \mathrm{GTP} \overset{K_\mathrm{G}^\mathrm{I}}{\rightleftharpoons} \mathrm{E{\cdot}D{\cdot}GTP}$$

$$\mathrm{E{\cdot}D{\cdot}GTP} \overset{k_{TCI}}{\rightarrow} \mathrm{E{\cdot}D{\cdot}M_1}$$

Elongation

$$\mathrm{E{\cdot}D{\cdot}M}_j + \mathrm{ATP} \overset{K_\mathrm{A}}{\rightleftharpoons} \mathrm{E{\cdot}D{\cdot}M}_j\mathrm{{\cdot}ATP} \overset{k_{TCE}}{\rightarrow} \mathrm{E{\cdot}D{\cdot}M}_{j+1} + \mathrm{PP}_i$$

$$\mathrm{E{\cdot}D{\cdot}M}_j + \mathrm{CTP} \overset{K_\mathrm{C}}{\rightleftharpoons} \mathrm{E{\cdot}D{\cdot}M}_j\mathrm{{\cdot}CTP} \overset{k_{TCE}}{\rightarrow} \mathrm{E{\cdot}D{\cdot}M}_{j+1} + \mathrm{PP}_i$$

$$\mathrm{E{\cdot}D{\cdot}M}_j + \mathrm{GTP} \overset{K_\mathrm{G}}{\rightleftharpoons} \mathrm{E{\cdot}D{\cdot}M}_j\mathrm{{\cdot}GTP} \overset{k_{TCE}}{\rightarrow} \mathrm{E{\cdot}D{\cdot}M}_{j+1} + \mathrm{PP}_i$$

$$\mathrm{E{\cdot}D{\cdot}M}_j + \mathrm{UTP} \overset{K_\mathrm{U}}{\rightleftharpoons} \mathrm{E{\cdot}D{\cdot}M}_j\mathrm{{\cdot}UTP} \overset{k_{TCE}}{\rightarrow} \mathrm{E{\cdot}D{\cdot}M}_{j+1} + \mathrm{PP}_i$$

Competitive Inhibition

$$\mathrm{E} + \mathrm{NTP} \overset{K_{I,\,NTP}}{\rightleftharpoons} \mathrm{E{\cdot}NTP} \qquad \mathrm{E{\cdot}D} + \mathrm{NTP} \overset{K_{I,\,NTP}}{\rightleftharpoons} \mathrm{E{\cdot}D{\cdot}NTP}$$

$$\mathrm{E} + \mathrm{PP}_i \overset{K_{I,\,PPi}}{\rightleftharpoons} \mathrm{E{\cdot}PP}_i \qquad \mathrm{E{\cdot}D} + \mathrm{PP}_i \overset{K_{I,\,PPi}}{\rightleftharpoons} \mathrm{E{\cdot}D{\cdot}PP}_i$$

$$\mathrm{E{\cdot}D{\cdot}M}_j + \mathrm{NTP} \overset{K_{I,\,NTP}}{\rightleftharpoons} \mathrm{E{\cdot}D{\cdot}M}_j\mathrm{{\cdot}NTP}$$

$$\mathrm{E{\cdot}D{\cdot}M}_j + \mathrm{PP}_i \overset{K_{I,\,PPi}}{\rightleftharpoons} \mathrm{E{\cdot}D{\cdot}M}_j\mathrm{{\cdot}PP}_i$$

Termination

$$\mathrm{E{\cdot}D{\cdot}M}_n \overset{k_{TCT}}{\rightarrow} \mathrm{E} + \mathrm{D} + \mathrm{M}_n$$

Figure 3.3: Reaction scheme for RNA polymerization by phage T7 RNA polymerase.

is some support to be found in the literature. Guajardo et al. (1998) observed a simultaneous linear increase of the concentrations of different RNA species (run-off, fall-off, and abortive transcripts). This increase continued on proportionately above nonliminting substrate levels. These results provide strong evidence that steady-state synthesis was indeed achieved within the short time frame of a few seconds. Thus, the period of pre-steady state kinetics appears to be negligible when this model is applied for simulating several minutes of process time.

3.3.2 Reaction kinetics

On the basis of Figure 3.3, the rate of total RNA synthesis, V_{TC}, by T7 RNAP under *in vitro* conditions has been derived symbolically (see Appendix A) to give the following functional

dependence on the concentrations of NTP, total promoter (C_{D}), and inhibitory byproduct PP_i:

$$V_{TC} = \frac{V_{TC}^{max}}{1+\sum_{j=1}^{N}\frac{K_{M,\mathrm{NTP},j}}{C_{\mathrm{NTP},j}}\left(1+\frac{C_{\mathrm{PP}_i}}{K_{I,\mathrm{PPi}}}+\sum_{\substack{i=1\\ i\neq j}}^{N}\frac{C_{\mathrm{NTP},i}}{K_{I,\mathrm{NTP},i}}\right)+\frac{K_{M,\mathrm{D}}}{C_{\mathrm{D}}}\left[1+\frac{K_{\mathrm{G}}^{\mathrm{I}}}{C_{\mathrm{GTP}}}\left(1+\frac{C_{\mathrm{PP}_i}}{K_{I,\mathrm{PPi}}}+\sum_{i=1}^{N\text{-}1}\frac{C_{\mathrm{NTP},i}}{K_{I,\mathrm{NTP},i}}\right)\right]} \quad (3.4)$$

Model parameters used in this rate equation are themselves composed of rate constants for elementary reaction steps and association constants for substrate binding. Their mathematical expressions are shown in Table 3.5. Importantly, the derived transcription kinetics include genomic sequence information in terms of transcript length, transcript composition, and the rate constants for initiation, elongation, and termination of RNA polymerization. These rate constants are vector-specific and vary with the consensus sequence of regulatory elements like the sites of promoter binding and transcription termination.

Table 3.5: Estimated kinetic parameters for *in vitro* transcription by T7 RNA polymerase using plasmid pT3/T7luc. Comparison to literature data.

Parameter			Unit	This study	Chamberlin and Ring (1973)	Ikeda and Richardson (1987)
V_{TC}^{max}	=	$k_{TC,\mathrm{eff}}C_{\mathrm{E},t}$	(μM/min)	188 ± 100		
$K_{M,\mathrm{D}}$	=	$\frac{k_{TC,\mathrm{eff}}}{k_{TCI}}K_{\mathrm{D}}$	(nM)	6.3 ± 2.6		
$K_{M,\mathrm{ATP}}$	=	$n_{\mathrm{A}}\frac{k_{TC,\mathrm{eff}}}{k_{TCE}}K_{\mathrm{A}}$	(μM)	76 ± 22	47	31
$K_{M,\mathrm{CTP}}$	=	$n_{\mathrm{C}}\frac{k_{TC,\mathrm{eff}}}{k_{TCE}}K_{\mathrm{C}}$	(μM)	34 ± 7	81	23
$K_{M,\mathrm{GTP}}$	=	$(n_{\mathrm{G}}-1)\frac{k_{TC,\mathrm{eff}}}{k_{\mathrm{E}}}K_{\mathrm{G}}+\frac{k_{TC,\mathrm{eff}}}{k_{TCI}}K_{\mathrm{G}}^{\mathrm{I}}$	(μM)	76 ± 12	160	190
$K_{M,\mathrm{UTP}}$	=	$n_{\mathrm{U}}\frac{k_{TC,\mathrm{eff}}}{k_{TCE}}K_{\mathrm{U}}$	(μM)	33 ± 6	60	40
$K_{I,\mathrm{PPi}}$			(μM)	200 ± 45		
$k_{d,TC}$			(min^{-1})	0.014		

Parameter $K_{\mathrm{G}}^{\mathrm{I}}$, the dissociation constant for initial GTP binding was taken to be 25 nM according to Sen and Dasgupta (1993). A value of 4.8 mM was reported for $K_{I,\mathrm{NTP},i}$ in the case of ATP competing versus GTP binding and using poly(dC)-DNA together with T7 RNAP (Guajardo and Sousa, 1997). Thus, neglecting substrate competition leads to a simplification of equation (3.4) to yield

$$V_{TC} = \frac{V_{TC}^{max}}{1+\sum_{j=1}^{N}\frac{K_{M,\mathrm{NTP},j}}{C_{\mathrm{NTP},j}}\left(1+\frac{C_{\mathrm{PP}_i}}{K_{I,\mathrm{PPi}}}\right)+\frac{K_{M,\mathrm{D}}}{C_{\mathrm{D}}}\left[1+\frac{K_{\mathrm{G}}^{\mathrm{I}}}{C_{\mathrm{GTP}}}\left(1+\frac{C_{\mathrm{PP}_i}}{K_{I,\mathrm{PPi}}}\right)\right]}. \quad (3.5)$$

In the case of only one limiting substrate nucleotide j and simultaneous promoter saturation, the kinetic rate equation reads

$$V_{TC} = \frac{V_{TC}^{max}\, C_{\mathrm{NTP},j}}{K_{M,\mathrm{NTP},j}\left(1+\frac{C_{\mathrm{PP}_i}}{K_{I,\mathrm{PPi}}}\right)+C_{\mathrm{NTP},j}}. \quad (3.6)$$

Equation (3.6) is further reduced when a sufficient amount of inorganic pyrophosphatase is present in solution, giving:

$$V_{TC} = \frac{V_{TC}^{max}\, C_{\mathrm{NTP},j}}{K_{M,\mathrm{NTP},j} + C_{\mathrm{NTP},j}} \tag{3.7}$$

At excess nucleotide concentrations, transcription rate depends on promoter concentration (C_{D}) according to

$$V_{TC} = \frac{V_{TC}^{max}\, C_{\mathrm{D}}}{K_{M,\mathrm{D}} + C_{\mathrm{D}}}. \tag{3.8}$$

3.3.3 Material balances

Material balances of a batch-wise transcription employing T7 RNAP may be formulated for total RNA concentration, all substrate nucleotides individually, and for inorganic pyrophosphate to achieve:

$$\frac{dC_{\mathrm{RNA}}}{dt} = \sum_{i=1}^{R} V_i \tag{3.9}$$

$$\frac{dC_{\mathrm{NTP},j}}{dt} = -\sum_{i=1}^{R} f_{j,i}\, n_i\, V_i \qquad \text{for} \quad j = 1 \text{ to } N \tag{3.10}$$

$$\frac{dC_{\mathrm{PP_i}}}{dt} = \sum_{i=1}^{R} \frac{n_i - 1}{n_i}\, V_i \tag{3.11}$$

Parameter $f_{j,i}$ indicates the molar fraction of base j contained in transcript i. The reaction rates, V_i, were chosen from equations (3.5) to (3.7), respectively, to match the applied system conditions. Numerical integration of equations (3.9) to (3.11) was performed applying the Advanced Continuous Simulation Language (ACSL) by Mitchell & Gauthier Associates Inc. Simulations provided the time courses of substrate and product concentrations involved in transcription.

3.4 Parameter identification

3.4.1 Maximum transcription rate

In initial rate studies with nonlimiting substrate levels, the dependency of total RNA formation rate on varying concentrations of plasmid pT3/T7luc was measured. The outcome of this analysis is shown in Figure 3.4. From these experiments and using the Michaelis-Menten kinetics given by equation (3.8), the maximum rate of total RNA synthesis, V_{TC}^{max}, was estimated to be 188±100 μM/min (corresponding with 1 U/μl T7 RNAP), and $K_{M,\mathrm{D}}$ equalled 6.3±2.6 nM. 95 % confidence intervals for both parameters were calculated from Student's t-distribution. An effective rate constant for overall transcription, $k_{TC,\mathrm{eff}}$ = 97±52 s^{-1}, could be calculated from V_{TC}^{max} using the relation expressed in Table 3.5. The particular meaning of parameter $k_{TC,\mathrm{eff}}$ is that total RNA synthesis proceeded in average by 97 nucleotides per second and per molecule of T7 RNAP.

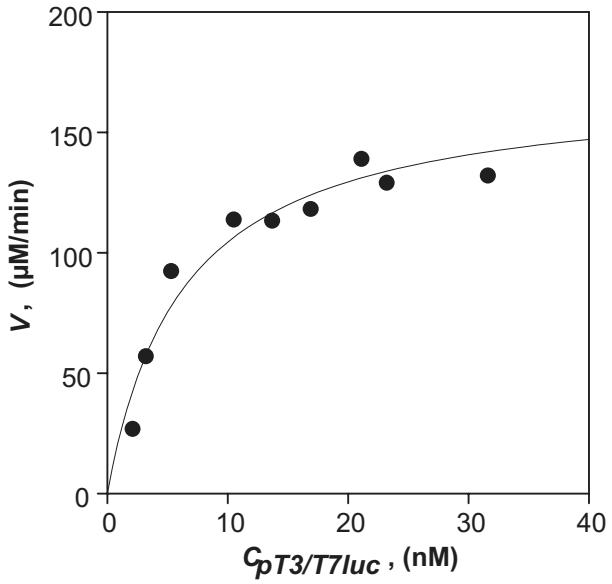

Figure 3.4: Transcription rate versus pT3/T7luc concentration at 3.7 mM free Mg^{2+} ions. Comparison between simulation and experiment. One of three independent experiments is displayed, all of which formed the basis for nonlinear regression analysis.

Since the noted affinity constant for promoter binding is located well within the range of Michaelis-Menten constants of 6 to 8 nM previously reported for promoter ϕ10, the value identified in this study was also applied for the two vectors pEThyd and pETK411BscFv, which contain this promoter sequence. The maximum transcription rates in systems containing pEThyd and pETK411BscFv were estimated to be 222±17 μM/min and 192±23 μM/min, respectively, for 1 U/μl T7 RNAP (Figure 3.5). These maximum rates correspond with effective rate constants of 115±9 s^{-1} and 99±7 s^{-1}, respectively. The results of this parameter identification are summarized in Table 3.6. Thus, the effective rate constants of overall transcription estimated in this study are located in the order of magnitude of 100 s^{-1} for all applied vector systems. This value compares to an average reported T7 RNAP progression of 60 s^{-1} on DNA template (Ikeda and Richardson, 1987), a rate constant of chain growth of 100 nucleotides per second (Chamberlin and Ring, 1973), and an elongation rate constant of 230 nucleotides per second for T7 RNAP travelling on T7 DNA (Golomb and Chamberlin, 1974).

The observed variation among the rate constants received in this study and also in the literature suggests some variability probably due to the usage of different DNA templates. More precisely, different types of regulatory regions like the initiation and termination sites, and the spectrum of RNA products of different lengths may be causally related to the noted range of transcript synthesis rates.

3.4.2 Rate constants for initiation, elongation, and termination

In order to try and pinpoint the individual contributions of initiation, elongation, and termination on maximum transcription rate, the experimental data of the different vector systems

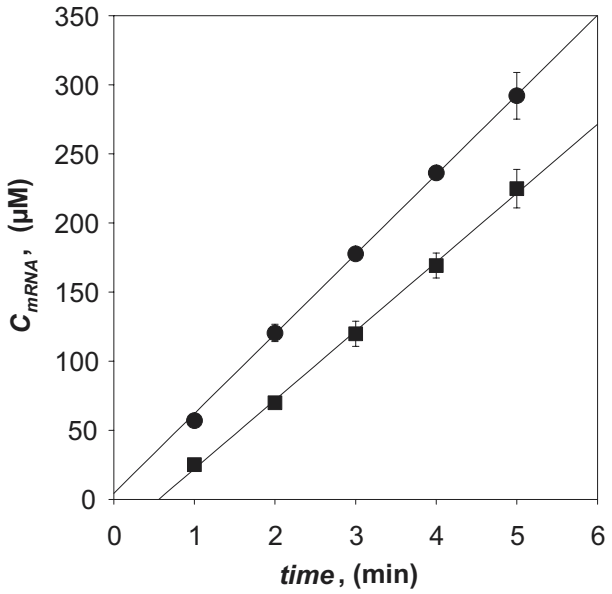

Figure 3.5: Total RNA concentration versus the time using 41 nM of the respective vector (pEThyd (circles) and pETK411BscFv (squares)) and for a free Mg^{2+} ion concentration of 3.7 mM. Comparison between simulation and experiment. Straight lines represent the average of three data sets. Error bars denote the standard deviation among the respective measurements.

Table 3.6: Comparison of maximum transcription rates of total RNA synthesis (indicated with respect to 1 U/μl T7 RNAP) and corresponding effective rate constants.

Plasmid	V_{TC}^{max} [μM/min]	$k_{TC,\text{eff}}$ [s^{-1}]
pT3/T7luc	188 ± 100	97 ± 52
pEThyd	222 ± 17	115 ± 9
pETK411BscFv	192 ± 23	99 ± 7

were further evaluated. V_{TC}^{max} may be written as

$$V_{TC}^{max} \;=\; \sum_{i=1}^{R} V_i^{max} \;=\; \sum_{i=1}^{R} k_{TC,\text{eff},i}\, C_{\text{E},t} \;=\; k_{TC,\text{eff}}\, C_{\text{E},t} \tag{3.12}$$

Parameter $k_{TC,\text{eff}}$ was defined from rate derivation according to

$$k_{TC,\text{eff},i} \;=\; \frac{1}{\dfrac{1}{k_{TCI}} + \dfrac{n_i - 1}{k_{TCE}} + \dfrac{1}{k_{TCT,\,i}}}. \tag{3.13}$$

k_{TCI} and k_{TCE} are the rate constants for transcription initiation and elongation, respectively. Parameter $k_{TCT,\,i}$ represents the rate constant for transcription termination at terminator site i. Using the relation

$$k_{TC,\text{eff},i} \;=\; m_i\; k_{TC,\text{eff}} \tag{3.14}$$

and the values of m_i determined from PAGE analysis (see section 3.2), parameter $k_{TC,\text{eff}}$ can be broken down into the contributions for synthesis of each RNA product received from this vector.

Table 3.7: Estimated parameters for transcription initiation, elongation, and termination using pT3/T7luc, pEThyd, and pETK411BscFv.

Parameter	Unit	Value
$k_{TCI,\text{pT3/T7luc}}$	(s^{-1})	146
$k_{TCI,\,\phi 10}$	(s^{-1})	79
k_{TCE}	(s^{-1})	$5.8{\cdot}10^7$
$k_{TCT,\,\phi}$	(s^{-1})	630
$k_{TCT,\text{bla}}$	(s^{-1})	50
$k_{TCT,\text{pT3/T7luc,1}}$	(s^{-1})	80
$k_{TCT,\text{pT3/T7luc,2}}$	(s^{-1})	9
$k_{TCT,\text{pET,1}}$	(s^{-1})	5

With this procedure, the effective rate constant for RNA synthesis could be determined for each of the nine transcripts generated from the investigated three vector systems (last column in Table 3.3). The substitution of these $k_{TC,\text{eff},i}$ values into equation (3.13) delivers a set of nine nonlinear algebraic equations, from which the eight unknown rate constants were estimated by minimization of the sum of squares of relative errors (Simulated annealing algorithm of OptdesX, Design Synthesis, Inc.).

The parameters characterizing the pET-derived vectors were identified first: the rate constant for initiation at promoter $\phi 10$ was found to be $k_{TCI,\phi 10} = 79\ \text{s}^{-1}$; the elongation rate constant was $k_{TCE} = 5.8 \cdot 10^7\ \text{s}^{-1}$; the termination rate constants were $k_{TCT,\phi} = 630\ \text{s}^{-1}$ for T7 terminator ϕ; $k_{TCT,\text{bla}} = 50\ \text{s}^{-1}$ for β-lactamase terminator; and $k_{TCT,\text{pET,1}} = 5\ \text{s}^{-1}$ for the unidentified terminator sequence contained on this type of vector. In a next step, while keeping these parameters constant, the rate constants for plasmid pT3/T7luc were estimated to be $k_{TCI,\text{pT3/T7luc}} = 146\ \text{s}^{-1}$, $k_{TCT,\text{pT3/T7luc,1}} = 80\ \text{s}^{-1}$, and $k_{TCT,\text{pT3/T7luc,2}} = 9\ \text{s}^{-1}$. Table 3.7 provides a summary of this parameter estimation. At close inspection of the obtained data, it is striking that the rate constant for elongation was estimated to be much larger, by several orders of magnitude, than in comparison to the other rate constants in this table. Thus, the elongation rate itself appears to be of minor importance in determining transcription rate. Only at extreme transcript lengths, elongation rate may become a decisive factor for mRNA synthesis rate.

In general, the rate constants for initiation and termination exhibit values within approximately one order of the corresponding effective rate constant, and they are thus indicating a pronounced impact on overall transcription rate. Their influence on transcription rate is congruent with the fact that there are multiple enzymic isomerization steps involved when T7 RNAP interacts with the regulatory sequences at the initiation and termination sites (Lyakhov et al., 1998; Kuzmine and Martin, 2001).

3.4.3 Termination efficiency

Commonly, the termination efficiency (TE) of a particular termination site is defined as the fraction of the termination product corresponding to this site with respect to the sum of all termination products synthesized from the investigated DNA template (von Hippel and Yager,

1991; Lyakhov et al., 1998). The TE for terminator j may thus be written as

$$\mathrm{TE}_j \;=\; \frac{C_{\mathrm{RNA},j}}{\sum_{i=1}^{R} C_{\mathrm{RNA},i}}. \tag{3.15}$$

Termination efficiency can thus be understood as the percentage of termination events that take place versus continuing on, whenever the transcribing RNA polymerase reaches this site. Therefore, the TE denotes the relative frequency of termination events leading to a particular RNA species.

There exists a connection between termination efficiency and parameter k_{TCT} in the derived model, as is detailed in the following. The above definition of termination efficiency shows a dependency on the RNA concentrations. Since the RNA concentrations themselves are determined by their respective formation rates, TE may be expressed in terms of transcript synthesis rates (von Hippel and Yager, 1991),

$$\mathrm{TE}_j \;=\; \frac{V_j}{\sum_{i=1}^{R} V_i}, \tag{3.16}$$

Thus, a connection is made between termination efficiency and the kinetic rate constants, among which k_{TCT} is one of them. Although such an approach will quickly lead to rather large, nonlinear mathematical formulae for TE, in particular under rate-limiting conditions, it is nevertheless suitable to provide an estimate for the termination efficiency of a certain transcription terminator. For any one of the investigated plasmids (say, e.g., pEThyd) and its RNA products, the same promoter applies and thus the same rate constant for initiation. Therefore, the differences in synthesis rates of the three RNA species obtained from this plasmid are mostly caused by the different rate constants for termination.

The termination rate constant, which governs the rate of release of the corresponding RNA product, is influenced by both (a) a temporary hold-up of RNA polymerase during RNA release and (b) the competing reaction of forward progression of RNA polymerase continuing the polymerization process. The model developed does not distinguish between both phenomena (a) and (b), but provides a - what might be called - overall rate constant for termination at a particular site, k_{TCT}. Importantly, small values given for k_{TCT} do thus not necessarily imply a slow termination process with delayed progression of T7 RNAP within this terminator region. They are indicative of a low termination efficiency, in the sense that the RNA polymerase may simply not recognize this termination signal as such and would thus continue transcript polymerization along the DNA. In summary, a small termination rate constant corresponds with a *low* (but not necessarily *slow*) rate of release of the corresponding RNA species, where the low rate may be caused by either one of the above items (a) or (b), or by both.

To give an example, the termination efficiency is subsequently calculated for terminator T_ϕ with the transcription data taken from the pEThyd experiment. For simplification, V_i may

be assumed to equal V_i^{max}. From equations (3.12), (3.13) and (3.16) it follows:

$$\begin{aligned} \mathrm{TE}_\phi &= \frac{k_{TCT,\mathrm{eff},\phi}}{k_{TCT,\mathrm{eff},\phi} + k_{TCT,\mathrm{eff,pET1}} + k_{TCT,\mathrm{eff,bla}}} \\ &= \left(1 + \frac{\left(\frac{1}{k_{TCI,\,\phi 10}} + \frac{n_{TCT,\,\phi}-1}{k_{TCE}} + \frac{1}{k_{TCT,\,\phi}}\right)}{\left(\frac{1}{k_{TCI,\,\phi 10}} + \frac{n_{TCT,\mathrm{pET1}}-1}{k_{TCE}} + \frac{1}{k_{TCT,\mathrm{pET1}}}\right) + \left(\frac{1}{k_{TCI,\,\phi 10}} + \frac{n_{TCT,\mathrm{bla}}-1}{k_{TCE}} + \frac{1}{k_{TCT,\mathrm{bla}}}\right)}\right)^{-1} \end{aligned} \tag{3.17}$$

Substituting the transcript lengths and rate constants from Tables 3.3 and 3.7, the termination efficiency is computed to be $\mathrm{TE}_\phi = 0.67$. Taking into account the simplifying assumption of maximum transcription rate, this value compares quite well with a reported termination efficiency for this terminator of 0.80 (Lyakhov et al., 1998). With the introduction of abbreviation x given by the following

$$x = \sum_{\substack{i=1 \\ i \neq j}}^{R} \left(\frac{1}{k_{TCI}} + \frac{n_{TCT,\,i}-1}{k_{TCE}} + \frac{1}{k_{TCT,\,i}}\right)^{-1}, \tag{3.18}$$

termination efficiency can be expressed as

$$\mathrm{TE}_j = \frac{1}{1 + x\left(\frac{1}{k_{TCI}} + \frac{n_{TCT,\,j}-1}{k_{TCE}} + \frac{1}{k_{TCT,\,j}}\right)}. \tag{3.19}$$

A general case study of equation (3.19) shows that small values of $k_{TCT,j}$ are indicative of a low termination efficiency at this site, whereas increased termination constants correspond with more efficient terminations.

3.4.4 Enzyme inactivation

The enzyme T7 RNA polymerase is known to be catalytically unstable in dilute reaction systems (Chamberlin and Ring, 1973; Maslak and Martin, 1994). Its catalytic activity may be maintained in the presence of high concentrations of bovine serum albumin (BSA). From an investigative point of view, however, it is desirable to avoid BSA addition, whose inherent RNase, DNase, and protease activities would make it an impossible task to conclusively evaluate transcription data. Therefore, the process stability of T7 RNAP was investigated by incubating the enzyme under static conditions in a buffered continuous-flow membrane reactor (8MC Micro-Ultrafiltration System, Amicon, Inc., Beverly, MA). The applied reaction conditions can be taken from Arnold et al. (2001). Some selected results of this analysis are briefly stated here.

The gradual loss of activity observed for T7 RNAP catalysis (see Figure 3.6) was assumed to obey a first-order kinetic rate law according to

$$V_{TC}^{max}(t) = V_{TC}^{max,0}\, e^{(-k_{d,\,TC}\,t)}. \tag{3.20}$$

Linear regression analysis of data gathered in the experiment gave an inactivation constant, $k_{d,TC} = 0.014$ min^{-1}, which corresponds with a half-life of T7 RNAP of 50 min. Whatever

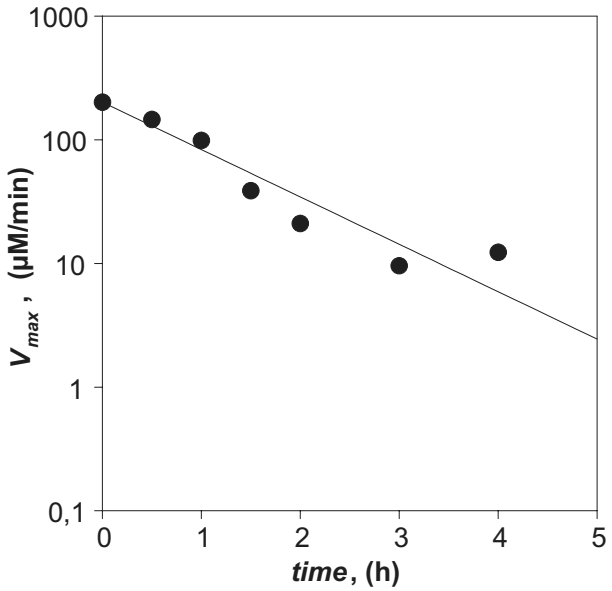

Figure 3.6: Time dependence of T7 RNAP activity under stationary reaction conditions. Comparison between simulation and experiment. The free Mg^{2+} concentration was 4.9 mM inside the reactor and 1.7 mM in the initial rate measurements.

the physical or chemical causes for the observed time-dependent enzyme denaturation might have been, their identification was not further investigated in this study. The ability, however, to quantify the noted rate of inactivation was an important prerequisite for the subsequent dynamic experiments. Because these experiments were conducted for a time-period of up to two hours, catalytic inactivation of T7 RNAP could not be omitted.

3.4.5 Substrate affinity and product inhibition

Apparent affinity constants for T7 RNAP towards substrate nucleotides were determined from dynamic experiments, as can be viewed from Figure 3.7. In these systems, a single NTP species was limiting, while the other NTP were kept at excess levels throughout the experiment duration. The Michaelis-Menten constants estimated including their 95 % confidence intervals were $K_{M,\mathrm{ATP}} = 76\pm22$ μM, $K_{M,\mathrm{CTP}} = 34\pm7$ μM, $K_{M,\mathrm{GTP}} = 76\pm12$ μM, and $K_{M,\mathrm{UTP}} = 33\pm6$ μM. The parameters are found to be very similar, including the K_M value for GTP, which had been suggested earlier to be much higher than for the other substrate nucleotides (see Table 3.5).

The inhibitory impact of inorganic pyrophosphate is reflected by the inhibition constant $K_{I,\mathrm{PPi}}$, which was determined to be equal to 200 ± 45 μM. This value compares favourably to the 830 μM reported earlier for the inhibition constant (Guajardo and Sousa, 1997). Therefore, PP_i accumulation appears to be significant mostly under *in vitro* conditions using the isolated enzyme. In cell-free synthesis systems and under *in vivo* conditions, where PP_i levels of 0.5 mM have been measured for *E. coli* (Kukko-Kalske et al., 1989), PP_i inhibition should not be a ruling factor for transcription rate.

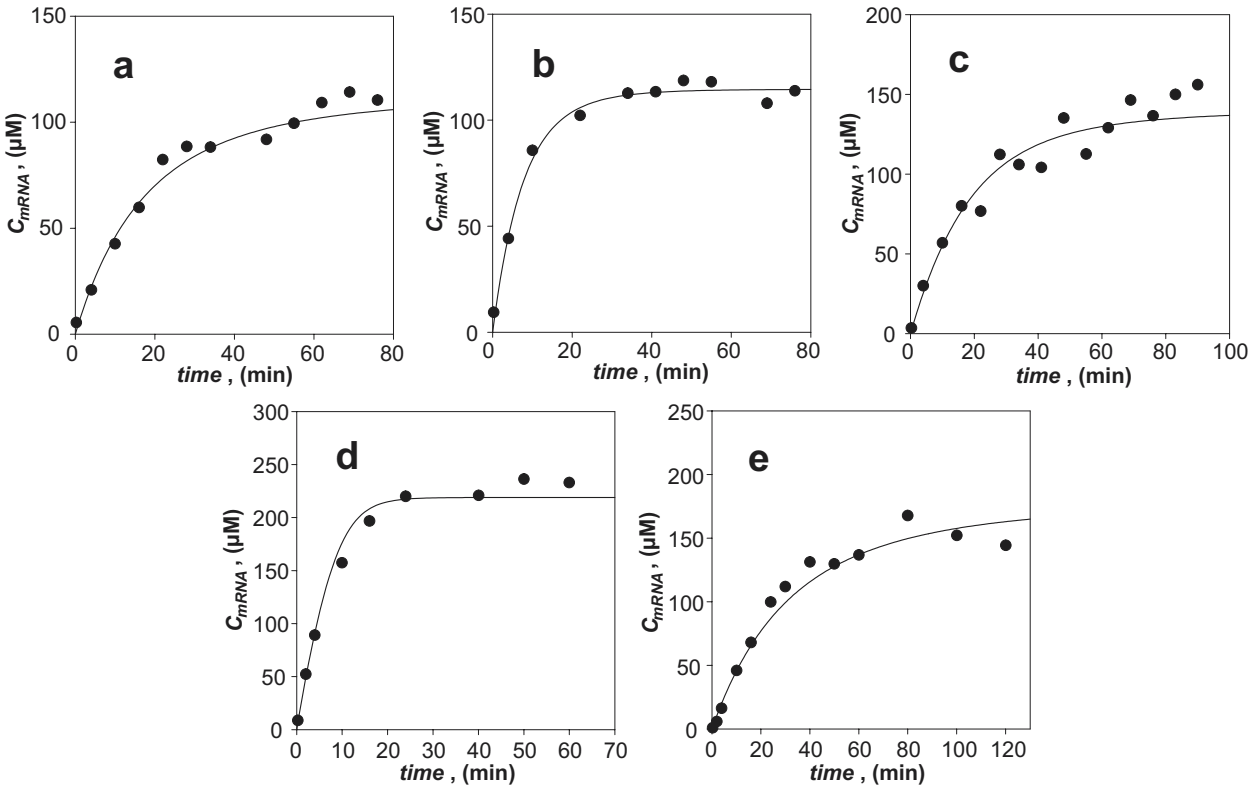

Figure 3.7: Batch experiments of transcription by T7 RNAP for estimation of nucleotide affinities and byproduct inhibition. Measured and simulated concentration courses of total RNA are shown for systems employing a single limiting substrate concentration. The free Mg^{2+} concentration was 2.8, 3.2, 2.7, 2.7, and 0.2 mM in systems (a), (b), (c), (d), and (e), respectively. (a) ATP limitation with 30 μM initial ATP. (b) CTP limitation with 30 μM initial CTP. (c) GTP limitation with 50 μM initial GTP. (d) UTP limitation plus inorganic pyrophosphatase addition. (e) UTP limitation (30 μM initial UTP) and supplementation of 3 mM PP_i at reaction start.

3.5 Conclusions

Although other kinetic models have been developed in the past to describe transcription by T7 RNA polymerase, apparently none of these models has placed enough emphasis on a systematic mechanistic model derivation, which could have ultimately led to express transcription rate in terms of specific DNA characteristics. The particular novelty of this chapter arises from the fact that the developed transcription model is trying to make use of genomic sequence data and their annotated information for the prediction of transcript synthesis rate. Sequence data incorporated in the model are (a) the explicit location of initiation and termination sites contained on this vector, and (b) the nucleotide sequence in between these sites, in order to obtain the sequences of the resulting RNA transcripts. From these two pieces of information, the lengths of RNA transcripts to be synthesized and their nucleotide composition are readily calculated. When additionally the specific recognition sequences of initiation and termination sites located on this vector are known also from plasmid sequence, and when they have been tabulated earlier together with their corresponding rate constants, then these parameters can be conveniently selected from such a library and used for simulation of transcription rate, importantly for a specifically chosen vector system. A large collection of transcription factor recognition sites and the annotated information concerning their binding properties is accessible in such databases, like e.g., TRRD (Kolchanov et al., 1999) and TRANSFAC (Wingender et al., 2000).

The general formulation of lumped model constants in terms of sequence-oriented parameters allows to enter the respective information of each investigated system and thus greatly improves the range of applicability of this model. From among the model parameters, the maximum transcription rate, V_{TC}^{max} was selected to undergo a more detailed examination with respect to its particular influencing by the genomic sequence. Other model parameters like the affinity constants for substrate nucleotides have been estimated for the particular vector pT3/T7luc. In order to discern among effects related to nucleotide sequence among these latter parameters, further experimental work is necessary, which is beyond the scope of this study.

The model developed may be used for dynamic simulation of mRNA synthesis rate as part of (both *in vivo* and *in vitro*) recombinant protein production systems employing T7 RNA polymerase and the investigated transcription initiation and termination sites. In combination with a mathematical model of mRNA degradation, the transcription model could serve as a basis for system design. In order to establish desirable transcript levels over the course of mRNA synthesis, process parameters of interest (like for example, the initial concentrations of substrates and RNA polymerase in a batch process) could be estimated in an optimization procedure using this model.

The structural similarities identified between nucleic acid polymerases (Sousa, 1996) may also provide an indication for mechanistic similarities between these enzymes. It would thus be interesting to test the transferability of this model to describe mRNA synthesis rate by a RNA polymerase other than from bacteriophage T7. In such an approach, obviously the respective kinetic parameters specific to this particular RNA polymerase need to be known. E.g., the

endogenous concentration of *E. coli* RNAP is about 5 μM (Bremer and Dennis, 1996). RNA chain growth proceeds in average with a rate of 40 to 55 nucleotides per second for mRNA, and 85 nucleotides per second for stable RNA (i.e., rRNA and tRNA). Additional kinetic features, such as for example the involvement of transcription factors, are at present not included in this model. With the current model formulation, however, it should in principle be possible to add further mechanistic properties. In this context, knowledge about binding constants for transcription factor binding is necessary. Modelling would then greatly benefit from studies providing these binding constants, either obtained from experimental detection, or alternatively from theoretic derivation on the basis of thermodynamic constraints (Kolchanov et al., 1999).

4 Prokaryotic mRNA Degradation

4.1 Introduction

Messenger RNA (mRNA) plays a central role in gene expression regulation since this molecule constitutes the connecting link between genetic information and ribosomal protein synthesis. In general, protein expression rates are correlated with the levels and the translatability of mRNA transcripts. The effective mRNA concentration results from a superposition of transcript synthesis and the degradation through ribonucleolysis. While the synthesis of mRNA has been treated in the previous chapter, the main factors influencing mRNA degradation rate are described mathematically in this chapter.

Transcript decay rates vary from gene to gene according to their sequence-specific properties. mRNA is highly unstable, in particular in prokaryotic systems. Functional half-lives of mRNA typically range within 1 to 5 min in *E. coli* (Blundell et al., 1972; Pedersen et al., 1978), with reduced temperatures generally giving higher mRNA stabilities. Eukaryotic transcripts show a comparatively higher stability. mRNA half-lives of up to 25 min are reached for yeast, and up to 16 hours for mammalian cell cultures (Court, 1993; Coburn and Mackie, 1999; Rauhut and Klug, 1999). While a fast mRNA turnover is a vital requirement for the cell to quickly adapt to environmental changes, a sufficient mRNA stability is on the other hand necessary for the successful application of recombinant DNA technologies.

Among the primary determinants for governing mRNA lifetime are the stabilizing and destabilizing single-stranded and double-stranded regions, which in their entirety characterize the **secondary structure** of mRNA (see section 2.5). Differently structured mRNAs are subject to alternate modes of decay, with several ribonucleases (RNases) participating in the degradation process. Table 4.1 provides an overview of some major known prokaryotic RNases together with their specific function. In general, these enzymes are distinguished between endonucleases and exonucleases. Endonucleases perform RNA cleavages at specific internal cleavage sites. Exonucleases catalyze the nucleotide-wise digestion of RNA. In prokaryotes, their direction is fixed from the 3′-end to the 5′-end of mRNA.

The majority of degradation reactions in *E. coli* is believed to involve the so-called **degradosome**. This aggregate of multiple enzymes (see Figure 4.1) contains both endonucleases and exonucleases, and is moreover capable of unwinding mRNA secondary structures (Carpousis et al., 1994; Py et al., 1996; Miczak et al., 1996). **RNase E**, a main component of the degradosome, selectively recognizes endonucleolytic cleavage sites that are characterized by their richness in adenine (A) and uracil (U) (see also Figure 4.5). A recent study by McDowall et al. (1994) suggested these sites to be determined rather by the A/U-content than with respect to a particular order of nucleotides. RNase E was shown to associate to the 5′-end of the mRNA to initiate the degradation process (Bouvet and Belasco, 1992). RNA secondary structural elements like stem-loops at the 5′-terminus constitute sterical obstacles in the association of the degradosome. Stem-loop structures may also affect degradosomal migration along the mRNA in the search for endonucleolytic cleavage sites and may further impair the catalytic step of

Table 4.1: Classification of enzymes that are catalytically involved in bacterial mRNA degradation (Coburn and Mackie, 1999; Rauhut and Klug, 1999)). ssRNA = single-stranded RNA, dsRNA = double-stranded RNA.

Enzyme	Reaction
Endonucleases	
RNase E	cleaves A/U-rich ssRNA in $5' \rightarrow 3'$-direction
RNase III	cleaves dsRNA in $5' \rightarrow 3'$-direction
Exonucleases	
RNase II	stepwise hydrolysis of ssRNA in $3' \rightarrow 5'$-direction
PNPase	stepwise phosphorolysis of ssRNA in $3' \rightarrow 5'$-direction
OligoRNase	hydrolysis of short oligonucleotides
Further enzymes	
PAP I	oligoadenylation at $3'$-end of RNA
RhlB	RNA helicase for unwinding of dsRNA

endonucleolytic cleavage itself.

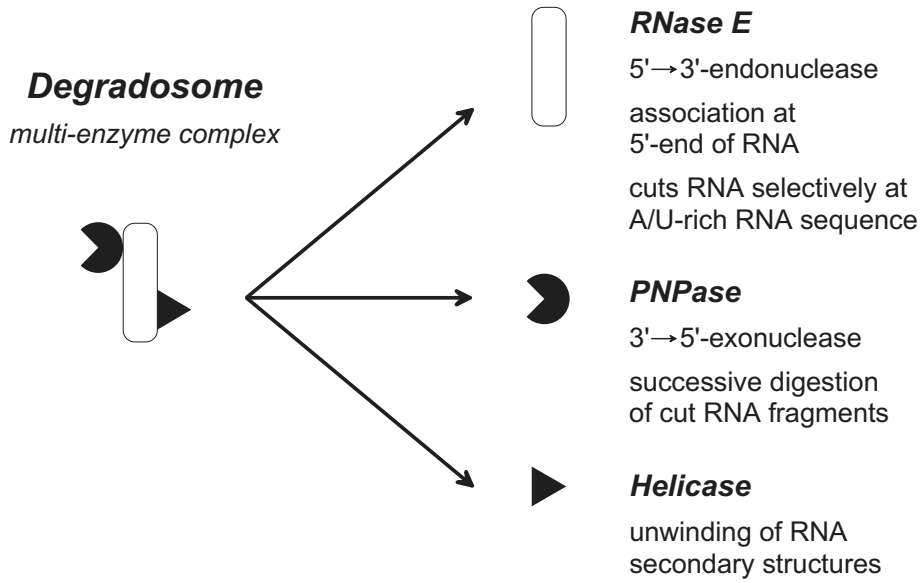

Figure 4.1: Main compounds of the bacterial degradosome.

The exonuclease **polynucleotide phosphorylase** (PNPase) contained also in the degradosome accomplishes the decay of RNA fragments resultant from endonucleolytic cleavage. According to common belief, PNPase remains attached to the mRNA molecule until the latter is fully digested (Régnier and Arraiano, 2000).

The function of RNA-**helicase RhlB** is at least two-fold. On the one hand, the enzyme is involved in an unwinding of RNA stem-loop structures that prevent access to endonucleolytic cleavage sites. On the other hand, this enzyme is to remove RNA secondary structures that impede the propagation of exonucleases (Rauhut and Klug, 1999). The unwinding process is

highly ATP-dependent. The hydrolysis of one ATP molecule per base was suggested for DNA helicase PcrA in a recent review about helicase mechanism (Soultanas and Wigley, 2001).

The importance of the degradosome as a key player in bacterial mRNA degradation is further emphasized as new enzymes are evidenced to participate in degradosome action. Additionally, the enzymes enolase, polyphosphate kinase (PPK), together with chaperones DnaK and GroEL needed for protein folding, were found to co-purify with RNase E (Liou et al., 2001). The particular function of these enzymes during the degradation process is not in all cases known.

Bacterial mRNA often encodes for multiple genes (polycistronic mRNA). Although these genes are controlled coordinately on the transcription level, not necessarily an equimolar synthesis of the encoded proteins may be observed (Schirmer and Hillen, 1998). Endonucleolytic cleavages of polycistronic mRNA into a number of monocistronic RNA fragments with different structural properties may lead to a variation in mRNA stability (Newbury et al., 1987; Klug, 1993). This process is commonly referred to as **differential mRNA degradation**.

The pivotal regulatory role of message stability needs to be viewed within the network of gene expression control reactions. A tight coupling exists between the mRNA degradation process, transcriptional rates, and the efficiency of translation, as was previously illustrated at the example of *lacZ* gene expression (Yarchuk et al., 1992). Weak ribosome binding sites, which were generated in that study by point mutation analysis, were found to greatly diminish mRNA expression levels. A reduced efficiency of translation initiation was shown to be associated with a premature termination of transcription and raised mRNA degradation rates. Furthermore, temporal uncoupling of transcription and translation was demonstrated to lead to labile mRNAs (Iost and Dreyfus, 1995). Such a non-synchronous situation emerges, e.g., when in recombinant protein expression, bacteriophage T7 RNA polymerase is applied instead of the host polymerase. T7 RNA polymerase transcribes approximately 5-fold faster than *E. coli* RNA polymerase (cf. chapter 3). mRNA stability could be restored by using a mutant T7 RNA polymerase with a reduced transcription rate (Makarova et al., 1995).

The broad spectrum of mRNA stabilities together with the numerous enzymes catalytically involved in mRNA degradation are likewise reflected in a diversity of degradation mechanisms. Some of the major degradation pathways are outlined subsequently.

4.1.1 Degradation mechanisms

5′-Binding model. — The common mechanism for mRNA degradation is conceived to proceed from 5′ to 3′ of the mRNA. After the initial degradosome binding to the mRNA at its 5′-terminus (Bouvet and Belasco, 1992), an alternating sequence of degradosome propagation, scanning the mRNA for endonucleolytic cleavage sites, and endonucleolytic cleavage followed by exonucleolytic digestion leads to the successive degradation of the mRNA molecule. The degradosome remains attached to the mRNA throughout this entire degradation process (see also Figure 4.5). The degradosome travel has been viewed as a sliding along the mRNA following translating ribosomes (Belasco and Higgins, 1988). Alternatively, degradosomes bound to 5′-tails of mRNA were considered to stochastically loop inwards and thus scan the mRNA for

putative endonucleolytic cleavage sites (Carrier and Keasling, 1997).

Polyadenylation. — Stable stem-loop structures at the 3′-end of mRNA may be made accessible to exonucleolytic degradation by adding a sequence of adenyl-residues to the terminal stem-loop structure. This polymerization reaction is catalyzed by the enzyme poly-adenylate phosphorylase (PAP I). Polyadenylation competes with the counteractive reaction of exonucleases constantly removing the growing poly(A)-tail. The trade-off between these reactions was suggested to constitute a further method of regulating mRNA degradation (Coburn and Mackie, 1999; Rauhut and Klug, 1999).

3′-Binding model. — This degradation mechanism involves the initial polyadenylation of the 3′-end of mRNA by PAP I to overcome stable stem-loop structures in this site. Subsequently, exonucleolytic degradation of mRNA proceeds in a $3' \rightarrow 5'$ direction until the mRNA is fully degraded.

Ribosome protection mechanism. — mRNA degradation rate is in many ways modulated by ribosomal **translation**. Binding of the 30S ribosomal subunit to the Shine-Dalgarno sequence in the vicinity of the 5′-terminal mRNA was capable of stabilizing *lacZ* mRNA (Wagner et al., 1994). Ribosomes bound to a mRNA may physically block degradosomes from entering the sites of nucleolytic cleavage (Petersen, 1993). Further, amino acid starvation was found to delay the degradation of *trp* mRNA (Morse and Guertin, 1971; Kennell and Simmons, 1972). All of these examples share in common a modulation of ribosome densities along the mRNA. Thus, the spacing of translating ribosomes can be taken as an indicator for the level of mRNA protection (Lopez et al., 1998; Coburn and Mackie, 1999).

mRNA stabilizers or destabilizers. — In the above listed mechanisms, 5′-terminal segments can enhance the stability of otherwise labile messages through impaired RNase E association. For the same reasoning, 3′-terminal and intercistronic stem-loop structures can impede degradosome initiation and the propagation of 3′-exonuclease digestion.

4.1.2 Previous modelling

Mathematical modelling of RNA degradation is typically performed employing unstructured models (Hargrove and Schmidt, 1989). In such an approach, the rate of mRNA degradation is reflected by a first-order rate. The kinetics consider a linear dependency on total mRNA concentration. A single parameter, the degradation rate constant, characterizes the decay rate, according to

$$\frac{dC_{\mathrm{mRNA}}}{dt} = k_{d,\mathrm{mRNA}}\, C_{\mathrm{mRNA}} \tag{4.1}$$

An assumption inherently made in equation (4.1) is that the multitude of reactions involved in the degradation process may be ascribed to a single rate-limiting step. In the light of the diversity of mRNA structures and the intricate regulation schemes involved in determining mRNA levels, this is clearly an oversimplification of the real picture. Since the rate constant of mRNA degradation, $k_{d,mRNA}$, needs to be determined for each system separately, the predictability of unstructured models is severely limited. Black-box models are, however, still useful when investigating the interconnectedness of mRNA degradation within gene networks.

In contrast to the unstructured modelling approach, other mathematical models of mRNA degradation have been developed that treated the decay as a multi-step process. The stochastic model by Singh (1969) envisions a random inactivation of the 5′-terminal mRNA by exonuclease activity, which is followed by a sequential mRNA degradation towards the 3′-end of mRNA. In a similar modelling approach, Rigney (1979) considered a modulation of degradation rate by the reaction of ribosome binding to the messenger. The concept of unidirectional mRNA degradation was further applied also by Kessling et al. (1976). These authors developed a deterministic model of mRNA degradation, which consists essentially of a sequence of first-order reactions. The idea behind was to theoretically reconstruct the base sequence of a RNA from the measuring signal obtained in an enzymatic sequencing analysis. While the assumption of a 5′-exonuclease may apply for eukaryotic mRNA degradation, it is, however, not applicable to prokaryotes, where no such enzyme has been verified to this date. Further attention in modelling mRNA degradation has been paid to mathematically describe the size distribution of a decaying mRNA population (Lim and Kennell, 1979). Moreover, in an attempt to discern between individual contributions to the overall observed chemical decay rate, Liang et al. (1999) developed a deterministic model with two model parameters, one of which related to endonucleolytic cleavage and the other to exonucleolytic digestion.

Carrier and Keasling (1997) provided a remarkably detailed example of mechanistically describing prokaryotic mRNA degradation. Their modelling approach took into account both degradosome binding and ribosome protection, which were embedded within the context of both mRNA and protein synthesis. Three types of decay theories were distinguished in that study: (*i*) ribosome protection mechanism, (*ii*) 5′-binding, and (*iii*) a hybrid mechanism combining features of both (*i*) and (*ii*). The modelling frame is based on the stochastic model by Vassart et al. (1971) (see section 2.1), where characteristically, the rates of polymerization steps (i.e., initiation, elongation, and termination of both transcription and translation, respectively) are taken to be model constants. The occurence of each stochastic event was determined from Monte-Carlo simulation. Among the degradation mechanisms tested, the hybrid model was found overall to best comply with the experimental observations of (a) increased mRNA stabilities at higher ribosome loadings and (b) improved mRNA life-times with reduced rates of translation elongation, and (c) the known direction of decay from 5′ to the 3′-end of mRNA.

While the model by Carrier and Keasling (1997) was very valuable in discriminating against degradation mechanisms, such a non-deterministic model is limited in its capacity to predict mRNA decay rates. For improved general applicability, ideally to cover universal mRNA products, a functional dependence of mRNA degradation rate on the specific transcript properties, which are in their condensed form represented by nucleotide sequence, is essential.

In this study, a first modelling approach is taken to include nucleotide sequence information for representation of mRNA degradation kinetics. The model aims in particular to account for both endonucleolytic and exonucleolytic reaction steps encountered during the decay process, as well as to describe mechanistically the interactions of mRNA degradation and ribosomal translation. Furthermore, a connection between degradation rate and the structural features of mRNA is discussed. The details of model derivation are provided in the next section.

4.2 Dynamic modelling

In the sequel, a mathematical model of mRNA degradation is developed based on the 5′-binding concept (cf. section 4.1.1). First, some definitions are given, before the degradation kinetics are derived. Mass balancing and possible means of model simplification complete this section.

4.2.1 Model assumptions

Nomenclature. According to Figure 4.2, mRNA base tripletts are consecutively numbered in 5′ to 3′-direction from $j = 1$ to J. The coding region stretches from the translational start site ($j = j_{R0}$) to codon $j = K$, just prior to the translational stop codon. It is assumed that $K \leq J$.

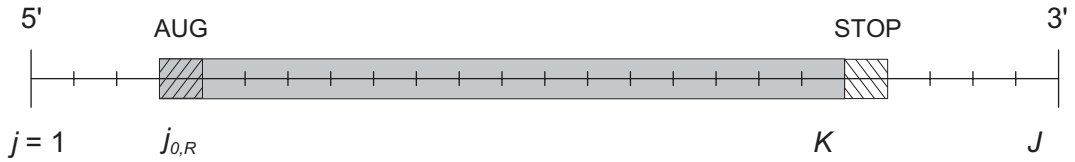

Figure 4.2: mRNA with coding region (grey-shaded).

Bound to a mRNA, a degradosome covers L_D base tripletts at a time. A ribosome extends over L_R codons simultaneously (cf. Figure 4.3). The catalytic center of bound degradosomes is located at m_D (with $1 \leq m_D \leq L_D$). The active center for protein synthesis is situated at position m_R of the ribosome (with $1 \leq m_R \leq L_R$).

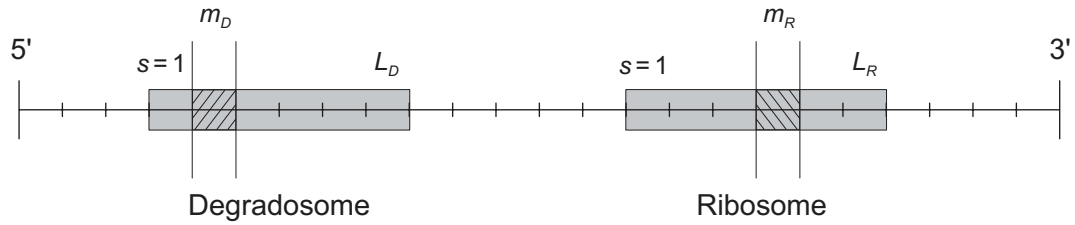

Figure 4.3: Nomenclature for degradosomes and ribosomes bound to mRNA.

Figure 4.4 sketches Z selected sites of an arbitrary mRNA molecule. Position $z_1 = 1$ denotes the 5′-terminal base triplett of this mRNA. Base tripletts j with $j \in \{z_2, \ldots, z_{Z-1}\}$ are characterized by an A/U-richness among their neighbouring bases. In order to ensure full mRNA degradation, an additional cleavage site was introduced arbitrarily at the 3′-terminal base triplett ($j = J$).

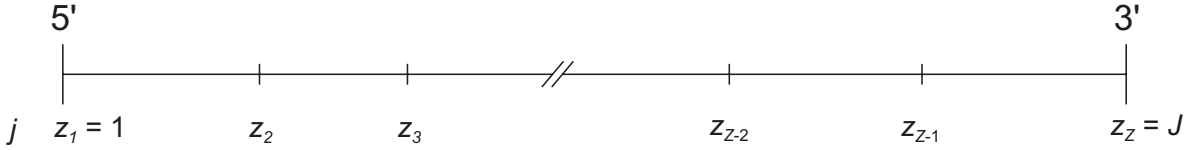

Figure 4.4: mRNA with endonucleolytic cleavage sites.

Reaction scheme. The mechanism of mRNA degradation considered is conform with a typically observed 5′ to 3′-directed mRNA decay (Figure 4.5). Ribosomes are assumed to be stripped off the mRNA before degradation. The ordered series of reactions starts out with

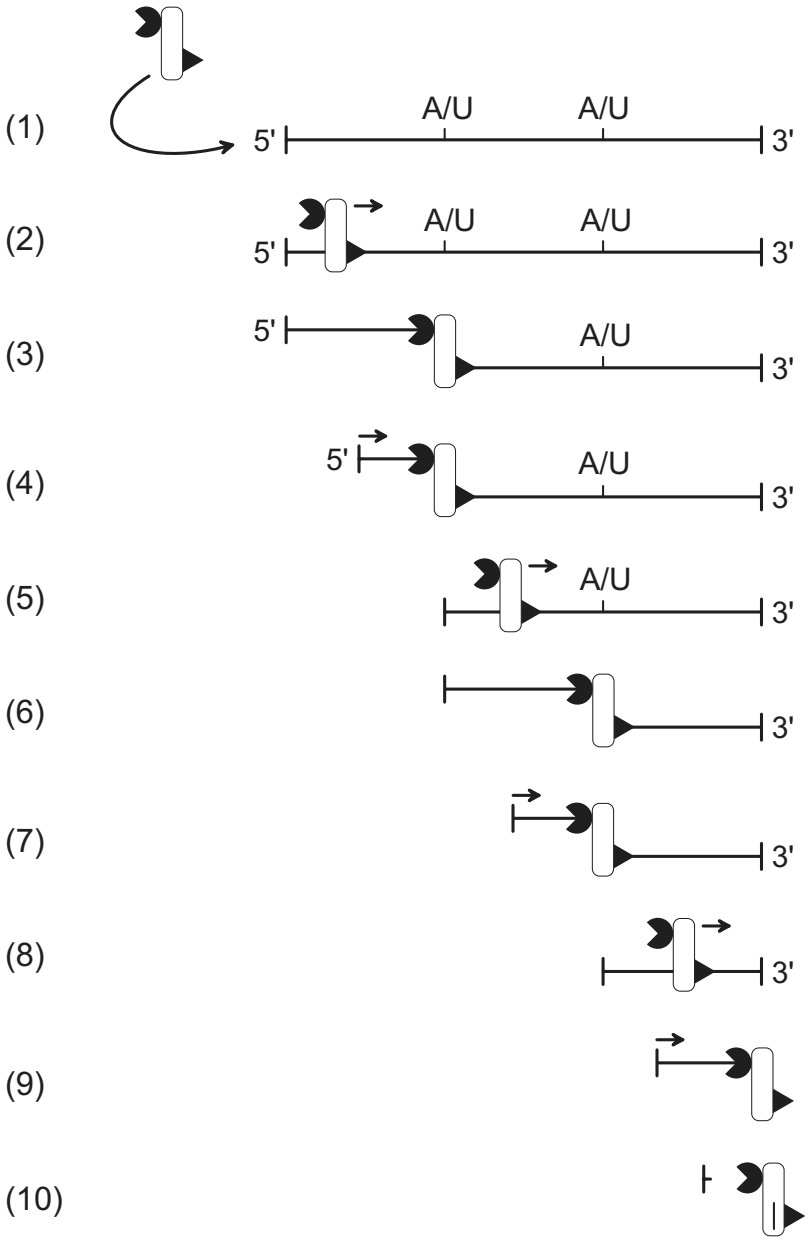

Figure 4.5: Mechanism of $5'$ to $3'$-directional mRNA degradation.

degradosome association to the $5'$-end of substrate mRNA (step (1)). The degradosome travels along the mRNA until an A/U-rich stretch is recognized as an endonucleolytic cleavage site (step (2)). In this position, the degradosome will pause and endonucleolytically cut the mRNA. The newly generated mRNA fragment is then transferred to the catalytic center of exonuclease activity (step (3)). Here, the fragment is successively degraded (step (4)). When this reaction is completed, the degradosome will continue its journey along the mRNA strand (step (5)) and will repeatedly undergo the stages of endonucleolytic and exonucleolytic digestion (steps (6) to (8)). The degradosome arrives eventually at the $3'$-terminal end of the mRNA, and the remaining mRNA fragment is exonucleolytically degraded (step (9)). The decay process is terminated with the release of the degradosome (step (10)), which can subsequently reenter another degradation cycle. In the model derivation, the $3'$-end was assumed to be chemically inert and thus not degraded by other reactions than those involved in the displayed mechanism.

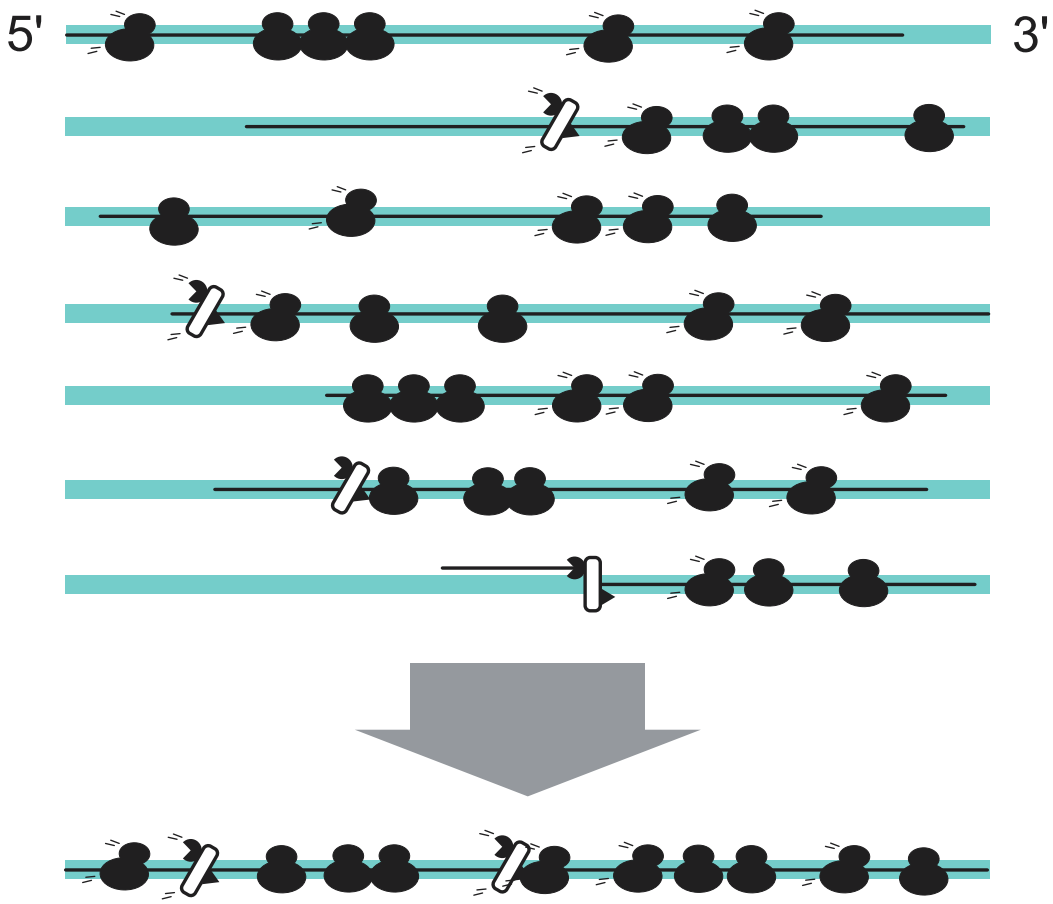

Figure 4.6: Projection of a mRNA population obtained from the same gene sequence onto one copy of full-length mRNA.

4.2.1.1 Choice of state variables

In the living cell, as well as under *in vitro* conditions, it is difficult to imagine mRNA as a single type of species. This holds also for the case of monocistronic mRNA, that is the product of a single gene. The totality of mRNA stemming from this one gene appears rather as a population of intermediates, as some mRNA molecules are in the process of being generated, while others are getting decomposed. The majority of these mRNAs thus differs in their $5'$ and $3'$-ending. Additional variation accrues from the stochastic nature of translational reaction steps. Altogether this leads to an array full of mRNAs with different lengths and divergent ribosome patterns. Such a situation is visualized in the upper part of Figure 4.6. In this figure, ribosomes followed by degradosomes are bound to mRNAs in different places, and they all travel in $5'$ to $3'$-direction.

From a modelling standpoint, this level of system complexity causes severe problems, in particular with increasing size of gene sequences. It appears impossible to track the fate of individual mRNA species by means of population balancing, unless further assumptions are made. In this case, each new state would have to be explicitly represented by a variable. In fact, the number of states could then be as high as the number of mRNA molecules, and the total number of ribosomes and degradosomes.

To arrive at a more practical model formulation, a site-specific state representation is

chosen in this study. A reduction of system complexity is achieved through a projection of the entire mRNA population onto a single species of full-length mRNA. This procedure is illustrated in the lower section of Figure 4.6. Material balance equations can now be derived for codon-specific state variables. That is, the total concentration of each base triplett j, (C_j^M) with $1 \leq j \leq J$, and the concentrations of degradosomes (C_j^D) and ribosomes (C_j^R) situated in j. These concentrations express averaged states with respect to the entire pool of each base triplett j. As a direct consequence of the projection, the model describes also situations where degradosomes are bound downstream of ribosomes, which is in contrast to the real system. Nevertheless, degradosomes and ribosomes bound to a particular codon j upstream of an endonucleolytic cleavage site, do not get lost in the instant of cleavage. Instead, they are - inherently in the model - redistributed within the remaining pool of base triplett j. Moreover, a reasonably sized set of state variables (maximally $3 \times J$) is obtained to characterize the concentrations of mRNA, and bound ribosomes and degradosomes, respectively. The state vector is thus expected to be computationally more inexpensive than in comparison to a system of population balances. On the other hand, the projection procedure is clearly accompanied by a loss of information. In particular, conclusions upon the loading pattern of individual mRNA molecules, their characteristic lengths, or presence and integrity of their native $5'$ and $3'$-termini cannot be drawn with this model.

4.2.1.2 Concentration of base tripletts

mRNA synthesis may be efficiently stopped, e.g., through the addition of antibiotic rifampicin, which inhibits transcription initiation. In such a system, mRNA concentration is then influenced only by the reactions involved in the ribonucleolysis process. One essential feature of the model regarding material balancing follows as a consequence of the fixed order of reaction steps that each degradosome needs to undergo in a degradation cycle. Namely, the pool of each base triplett j is affected in a direct manner only by the rates of endonucleolytic cleavage (given that transcription is ceased in this case). This means in particular that the concentration of base tripletts can temporarily remain unaltered, even though having been traversed by a degradosome. In this case, apparently the $(z_{i-1} - z_i)$ base tripletts in between two consecutive cleavage sites, z_{i-1} and z_i, change their states in parallel. This is shown in Figure 4.7 for an arbitrary mRNA. In order to describe the time-dependent decrease of all J base tripletts of a decaying transcript, it is thus sufficient to derive material balances for only Z selected base tripletts (i.e., one for each mRNA fragment upstream of an endonucleolytic cleavage site, plus one balance for the $3'$-terminal base triplett). The other concentrations of base tripletts, C_j^M, can then be represented in terms of these reference states. For example,

$$C_j^M \;=\; C_{z_{i-1}}^M \qquad \text{for} \;\; 1 \leq j < J-1 \;\; \text{and} \;\; z_{i-1} \leq j < z_i \tag{4.2}$$

4.2.2 Reaction kinetics

Degradosome association was reflected by the rate expression

$$V_{D,ass} \;=\; k_{D,ass}\, q_{j_{D0}}^{D0}\, C_{j_{D0}}^M \tag{4.3}$$

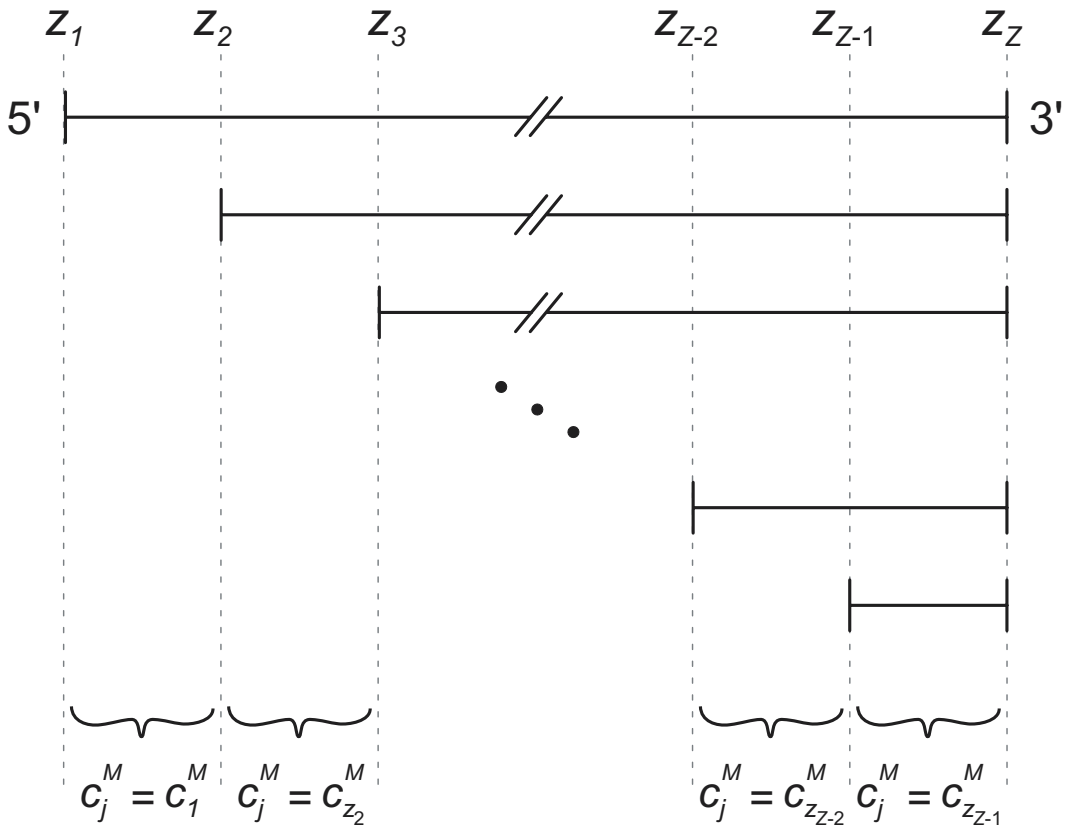

Figure 4.7: When mRNA synthesis is prevented, the concentrations of base tripletts change concomittantly for a region in between two successive endonucleolytic cleavage sites.

In equation (4.3), symbol $C^M_{j_{D0}}$ defines the total concentration of base triplett $j = j_{D0}$, at which degradosome association takes place. It is assumed that this position is located very closely to the 5′-terminus of mRNA. Parameter $q^{D0}_{j_{D0}}$ denotes the probability for the 5′-binding site to be in an unoccupied state. This parameter is explained in section 4.2.3. The rate constant for degradosome association, $k_{D,ass}$, may express a linear dependency on the concentration of freely dissolved degradosomes, which is assumed to be inherent in this parameter. Secondary structural features encountered in the binding region will also render $k_{D,ass}$.

The stepwise one-directional diffusion of degradosomes along the mRNA requires to take into account sterical blocking by catalysts bound further downstream. The rate of degradosome movement from base triplett j with $j_{D0} \leq j < J$ to position $j + 1$ is thus described by

$$V_{D,mv,j} \;=\; k_{D,mv}\, q^D_j\, C^D_j \tag{4.4}$$

Parameter q^D_j written in equation (4.4) denotes the probability of forward movement and is defined in section 4.2.3. Degradosome movement takes place until one of the endonucleolytic cleavage sites j is reached, with $j = z_i$ and $2 \leq i \leq Z$ (see definitions given on page 40). At these particular sites j, the degradosome will pause and adopt a state, here denoted by C^{D*}_j. In this state, an endonucleolytic cleavage reaction is considered to occur right upstream of codon j, which generates a mRNA fragment of $(z_{i-1} - z_i)$ bases in length. In Figure 4.5, such a stage is visualized by reaction steps (3) and (6). The kinetics for this cleavage reaction are represented

by a first-order rate according to

$$V_{D,endo,j} \;=\; k_{D,endo,j}\, C_j^{D*} \qquad \text{for} \;\; j \in \{z_2, z_3, \ldots, z_{Z-1}, z_Z\} \tag{4.5}$$

The rate constants, $k_{D,endo,j}$, may vary across all endonucleolytic cleavage sites. This study, however, treats all endonucleolytic cleavage sites the same, thus assigning the same parameter $k_{D,endo}$ to any of such sites. While the degradosome remains bound to the endonucleolytic cleavage site, the newly produced mRNA fragment is successively degraded by an exonuclease contained in the degradosome. The totality of all exonucleolytic steps can be summarized as to

$$V_{D,exo,j,i} \;=\; \sum_{s=z_{i-1}}^{z_i} k_{D,exo,s}\, C_j^{D*Frag} \qquad \text{for} \;\; j \in \{z_2, z_3, \ldots, z_{Z-1}, z_Z\} \;\text{ and }\; 2 \le i \le Z \tag{4.6}$$

The rate constant for exonuclease activity, $k_{D,exo,s}$ may differ with the type of base to be cleaved. It could be influenced additionally by sequence context. For example, each of the mRNA fragments may exhibit a unique secondary structural conformation. The unwinding of this structure, which is necessary during the process of an exonuclease reaction, would then lead to diverse rates of cleavage for each individual base in the exonuclease reaction. Although the model in its general form accounts for such differences, the rate constants for individual exonucleolytic cleavage steps will, for the most cases, be unknown. For practical reasons, it is assumed further on that this parameter remains invariant with nucleotide sequence.

The termination rate of mRNA degradation, which occurs at the final base triplett, $j = J$, is assumed to obey a first-order rate law, according to

$$V_{D,T} \;=\; k_{D,T}\, C_J^D \tag{4.7}$$

4.2.3 Queueing factor

4.2.3.1 mRNA degradation

Parameters q_{jD0}^{D0} and q_j^D used in equations (4.3) and (4.4) take into account the physical blocking of degradosomes both at their association to mRNA and during their journey along the template (as outlined in Figure 4.8). The probability for an unoccupied next site has been

Figure 4.8: Interference between degradosomes travelling along a mRNA.

reported previously by Gibbs and co-workers (Pipkin and Gibbs, 1966; MacDonald et al., 1968; MacDonald and Gibbs, 1969), for a situation where one type of catalyst (e.g., ribosome or degradosome) is bound in various positions of a template. This probability is further on termed the 'queueing factor' or 'queueing parameter'. When the fractional notation given in the original study is substituted by the concentrations of state variables involved in mRNA degradation,

the following set of equations can be obtained:

$$q_{j_{D0}}^{D0} = 1 - \sum_{s=1}^{L_D} \frac{\sum^{i} C_{i,j_{D0}+s-1}^{D}}{C_{j_{D0}+s-1}^{M}} \qquad \text{for} \quad j_{D0} = m_D \tag{4.8}$$

$$q_j^D = \frac{1 - \sum_{s=1}^{L_D} \frac{\sum^{i} C_{i,j+s}^{D}}{C_{j+s}^{M}}}{1 - \sum_{s=1}^{L_D-1} \frac{\sum^{i} C_{i,j+s}^{D}}{C_{j+s}^{M}}} \qquad \text{for} \quad j_{D0} \leq j \leq J - L_D \tag{4.9}$$

$$q_j^D = 1 \qquad \text{for} \quad J - L_D < j \leq J \tag{4.10}$$

j_{D0} is the base triplett, at which degradosome association takes place. Parameters L_D, m_D, and J have been defined earlier (on page 40). C_j^M denotes the total concentration of the jth base triplett of the mRNA. As equations (4.8) and (4.9) demonstrate, queueing factors are by no means to be understood as model constants. Instead, they change dynamically, as the binding states of base tripletts vary with the time. According to their definition, queueing factors can take values between 0 and 1. The summation over index i used in equations (4.8) and (4.9) denotes the sum of degradosomes in different conformations bound to a codon j, according to

$$\sum^{i} C_{i,j}^{D} = C_j^{D*} + C_j^{D*Frag} + C_j^{D} \tag{4.11}$$

4.2.3.2 mRNA degradation and translation

In the case where mRNA degradation and ribosomal translation take place simultaneously, both degradosomes and ribosomes are typically attached at the same time to a same mRNA molecule. Under these conditions, a mutual interaction among both sorts of catalysts exists. Such a scenerio is illustrated in Figure 4.9. The probability of forward movement provided earlier, now needs to be modified to account additionally for the inter-species interactions.

Figure 4.9: Interference between ribosomes and degradosomes travelling along a mRNA.

A mathematical relation for the probability, which then applies, has been developed in this study, as is detailed in Appendix E. The final results of this derivation (equations (E.21) to (E.24)) are redisplayed here. For degradosome association, which occurs at base triplett $j_{D0} = m_D$, the probability for this site to be unblocked depends on both concentrations of degradosomes and ribosomes bound to this site. $q_{j_{D0}}^{D0}$ is thus expressed by

$$q_{j_{D0}}^{D0} = 1 - \sum_{s=1}^{L_D} \frac{\sum^{i} C_{i,j_{D0}+s-1}^{D}}{C_{j_{D0}+s-1}^{M}} - \sum_{s=1}^{L_D+L_R-1} \frac{C_{j_{D0}+s-m_D-L_R+m_R}^{R}}{C_{j_{D0}+s-m_D-L_R+m_R}^{M}} \tag{4.12}$$

Degradosome movement along a mRNA is influenced by both degradosomes and ribosomes bound to proximate sites downstream of a base triplett j. The probability for site $j+1$ to be empty is given by

$$q_j^D = \frac{1 - \sum_{s=1}^{L_D} \frac{\sum^{i} C^D_{i,j+s}}{C^M_{j+s}} - \sum_{s=1}^{L_D} \frac{C^R_{j+s-m_D+m_R}}{C^M_{j+s-m_D+m_R}}}{1 - \sum_{s=1}^{L_D-1} \frac{\sum^{i} C^D_{i,j+s}}{C^M_{j+s}} - \sum_{s=1}^{L_D-1} \frac{C^R_{j+s-m_D+m_R}}{C^M_{j+s-m_D+m_R}}} \qquad \text{for} \quad j_{D0} \le j \le J - L_D \tag{4.13}$$

and by

$$q_j^D = \frac{1 - \sum_{s=1}^{L_D} \frac{C^R_{j+s-m_D+m_R}}{C^M_{j+s-m_D+m_R}}}{1 - \sum_{s=1}^{L_D-1} \frac{C^R_{j+s-m_D+m_R}}{C^M_{j+s-m_D+m_R}}} \qquad \text{for} \quad J - L_D < j \le J \tag{4.14}$$

Analogously, the queueing factor for ribosome association at the initiation codon $j = j_{R0}$ is affected by both ribosomes and degradosomes covering this site. That is,

$$q^{R0}_{j_{R0}} = 1 - \sum_{s=1}^{L_D+L_R-1} \frac{\sum^{i} C^D_{i,j_{R0}+s-m_R-L_D+m_D}}{C^M_{j_{R0}+s-m_R-L_D+m_D}} - \sum_{s=1}^{L_R} \frac{C^R_{j_{R0}+s}}{C^M_{j_{R0}+s}} \tag{4.15}$$

The queueing factor for ribosomal elongation, q_j^R, describes a dependency on both concentrations of neighbouring degradosomes and ribosomes, according to

$$q_j^R = \frac{1 - \sum_{s=1}^{L_R} \frac{\sum^{i} C^D_{i,j+s-m_R+m_D}}{C^M_{j+s-m_R+m_D}} - \sum_{s=1}^{L_R} \frac{C^R_{j+s}}{C^M_{j+s}}}{1 - \sum_{s=1}^{L_R-1} \frac{\sum^{i} C^D_{i,j+s-m_R+m_D}}{C^M_{j+s-m_R+m_D}} - \sum_{s=1}^{L_R-1} \frac{C^R_{j+s}}{C^M_{j+s}}} \qquad \text{for} \quad j_{R0} \le j \le K \tag{4.16}$$

Given the finite dimensions of a degradosome, degradosome binding to base tripletts upstream of j_{D0} is excluded, thus

$$C_j^D = 0 \qquad \text{for} \quad j < j_{D0} \tag{4.17}$$

Further, ribosome binding within non-coding regions is neglected. This yields,

$$C_j^R = 0 \qquad \text{for} \quad j < j_{R0}, \quad \text{and} \quad K < j \le J \tag{4.18}$$

4.2.4 Material balancing

4.2.4.1 mRNA degradation

From the given mechanism of mRNA degradation, the following material balances have been derived for a system in which transcription initiation and translation initiation have been

efficiently switched off. The model considers the time-dependent concentration changes of various degradosome states. The pool of freely dissolved degradosomes is considered to remain constant under *in vivo* conditions.

$$\frac{dC^D_{j_{D0}}}{dt} = V_{D,ass} - V_{D,mv,j_{D0}} \tag{4.19}$$

$$\frac{dC^D_j}{dt} = V_{D,mv,j-1} - V_{D,mv,j} \qquad \text{for } j_{D0} < j < J, \text{ and } j \notin \{z_2, z_3, \ldots, z_{Z-1}\} \tag{4.20}$$

$$\frac{dC^{D*}_j}{dt} = V_{D,mv,j-1} - V_{D,endo,j} \qquad \text{for } j \in \{z_2, z_3, \ldots, z_{Z-1}, z_Z\} \tag{4.21}$$

$$\frac{dC^{D*Frag}_j}{dt} = V_{D,endo,j} - V_{D,exo,j} \qquad \text{for } j \in \{z_2, z_3, \ldots, z_{Z-1}, z_Z\} \tag{4.22}$$

$$\frac{dC^D_j}{dt} = V_{D,exo,j} - V_{D,mv,j} \qquad \text{for } j \in \{z_2, z_3, \ldots, z_{Z-1}\} \tag{4.23}$$

$$\frac{dC^D_J}{dt} = V_{D,exo,J} - V_{D,T} \qquad \text{for } j = J \tag{4.24}$$

Due to equation (4.2), the time-dependent changes of all concentrations of mRNA base tripletts can be described by the following Z material balances:

$$\frac{dC^M_j}{dt} = -V_{D,endo,j} \quad \text{for } j \in \{z_1, z_2, \ldots, z_{Z-1}\} \tag{4.25}$$

$$\frac{dC^M_J}{dt} = -V_{D,T} \tag{4.26}$$

Reaction rates used in equations (4.19) to (4.26) have been presented in section 4.2.2.

4.2.4.2 mRNA degradation and translation

For a system comprising both mRNA degradation and ribosomal protein synthesis, additional balance equations need to be derived for the concentrations of mRNA-bound ribosomes. Under non-limiting growth conditions, metabolite pools are approximately buffered, and the concentrations of cellular catalysts involved in ribosomal translation may be viewed to be constant. The following simplified reaction kinetics can hence be formulated for the translational reaction steps:

$$V_{TLI,70\text{SIC}} = k_{TLI,70\text{SIC}}\, q^{R0}_{j_{R0}}\, C^M_{j_{R0}} \tag{4.27}$$

$$V_{TLI,\text{IF2D}} = k_{TLI,\text{IF2D}}\, C^{R*}_{j_{R0}} \tag{4.28}$$

$$V_{TLE,j} = k_{TLE,j}\, q^R_j\, C^R_j \qquad \text{for } j_{R0} \le j < K \tag{4.29}$$

$$V_{TLT} = k_{TLT}\, C^R_K \tag{4.30}$$

Equations (4.27) and (4.28) consider a two-step-mechanism for initiation of protein synthesis: The first step is characterized by 70S initiation complex formation at the translational start site. In a second step, the dissociation of initiation factor 2 (IF2) takes place. The kinetics for translation elongation and termination are given by equations (4.29) and (4.30), respectively. The material balance equations for the concentrations of ribosomes bound within the coding region of mRNA can thus be written as

$$\frac{dC^{R*}_{j_{R0}}}{dt} = V_{TLI,\mathrm{70SIC}} - V_{TLI,\mathrm{IF2D}} \qquad \text{for} \quad j = j_{R0} \tag{4.31}$$

$$\frac{dC^{R}_{j_{R0}}}{dt} = V_{TLI,\mathrm{IF2D}} - V_{TLE,j_{R0}} \qquad \text{for} \quad j = j_{R0} \tag{4.32}$$

$$\frac{dC^{R}_{j}}{dt} = V_{TLE,j-1} - V_{TLE,j} \qquad \text{for} \quad j_{R0} < j < K \tag{4.33}$$

$$\frac{dC^{R}_{K}}{dt} = V_{TLE,K-1} - V_{TLT} \qquad \text{for} \quad j = K \tag{4.34}$$

Symbol $C^{R*}_{j_{R0}}$ used in equation (4.31) refers to the concentration of 70S initiation complexes. After IF2 dissociation, the concentration of ribosomes bound to the translational start site is given by $C^{R}_{j_{R0}}$.

4.2.5 Model reduction

In order to improve the efficiency of computation, in particular for large gene sequences, it may be desirable to further reduce the complexity of a gene expression model. When a less detailed description of states is acceptable, a significant reduction in the number of state variables can be achieved by merging groups of base tripletts into one. This procedure is illustrated in Figure 4.10, where exemplary $n_{cod} = 4$ neighbouring base tripletts are at a time treated as one position in a new state representation. In the displayed example, the concentrations of four adjacent base tripletts (and analogously their loading states) are thus added up to give the concentration of a new state variable. In the following, systems representations with $n_{cod} = 1$ are referred to as 'resolution I'. Simplified models with $n_{cod} = 4$ are said to be of resolution IV.

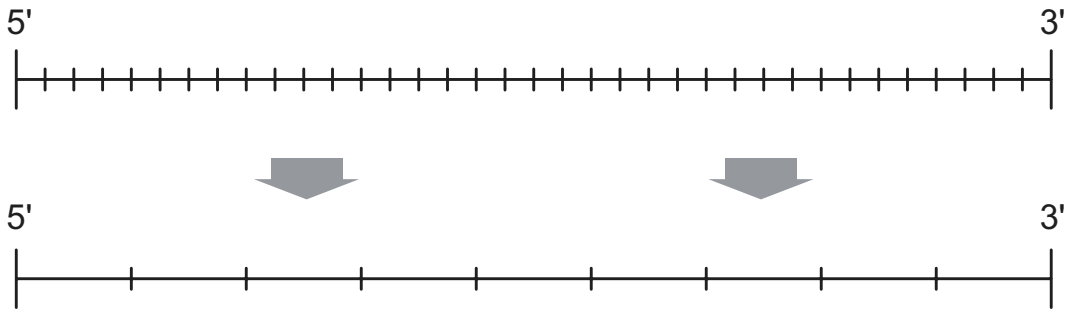

Figure 4.10: Degree of refinement of positional grid. In the shown example, base tripletts are merged into groups of four ($= n_{cod}$).

When applying this method of model reduction, several consistency checks need to be performed. It is important to ensure that the reading frame of the coding sequence remains

unaffected. Further, the consequences of the new systems representation need to be considered appropriately also for material balancing, as well as for the formulation of reaction kinetics and model parameters. In the case when translation elongation rates vary significantly in a codon-specific manner, material balancing of grouped base tripletts (and their states) becomes more cumbersome. This particular subject is treated in section 6.6.

4.3 Conclusions

The processes involved in mRNA degradation comprise an autonomous, separate modelling unit themselves. Nevertheless, care was taken to allow for the possibility to connect in a modular fashion the individual building blocks of a gene expression model, in order to describe the holistic performance of prokaryotic gene expression. The level of detail with which the connected units (say, e.g., translation or mRNA synthesis) are represented may vary with the modelling task. For the purpose of parameter estimation, greater emphasis was placed on modelling the mechanism of $5'$ to $3'$ mRNA degradation, while the kinetics of translation were treated in a simplistic manner in this chapter. Apart from transcript length, the number and position of endonucleolytic cleavage sites, the steps involved in exonucleolytic digestion of mRNA, and the mechanism of mRNA protection through ribosomal translation were included in the presented model of prokaryotic mRNA degradation.

At a later stage of model development, the described reaction sequence for mRNA degradation may be further augmented by additional reactions. For example, it is conceivable in future applications to consider the particular effects of secondary structures that may be encountered both within the $5'$ and the $3'$-region of the mRNA (see section 9.3), or that may form at intrinsic sites of mRNA when they are temporarily unoccupied by ribosomes.

A highly detailed, sequence-oriented description of mRNA degradation has very important implications for practical application. It would be extremely valuable, if, with the aid of such models, pseudo-first order rate constants for mRNA degradation could be inferred for each different type of mRNA.

5 mRNA Degradation: Parameter Identification for *lacZ* mRNA

The mathematical model of prokaryotic mRNA degradation presented in chapter 4 includes several model parameters, which need to be identified in order for this model to become applicable for predicton purposes. These parameters are subsequently estimated at the example of *lacZ* mRNA. This well-studied gene has been chosen here for investigation, since its mRNA is known to follow an exclusive $5'$ to $3'$ degradation pathway (Kennell and Riezman, 1977; Kennell, 1990; Liang et al., 1999).

5.1 *lac* operon

The *lacZ* gene is part of the lactose operon (see Figure 5.1) in *E. coli*. The operon encodes the enzymes required for lactose catabolism. Permease is responsible for lactose uptake into the cell. β-galactosidase carries out both of the reactions, the conversion of lactose into allolactose, and the cleavage of lactose into glucose and galactose. The latter monosaccharides are then further catabolized for cellular energy production. The physiological role of the transferase is unknown.

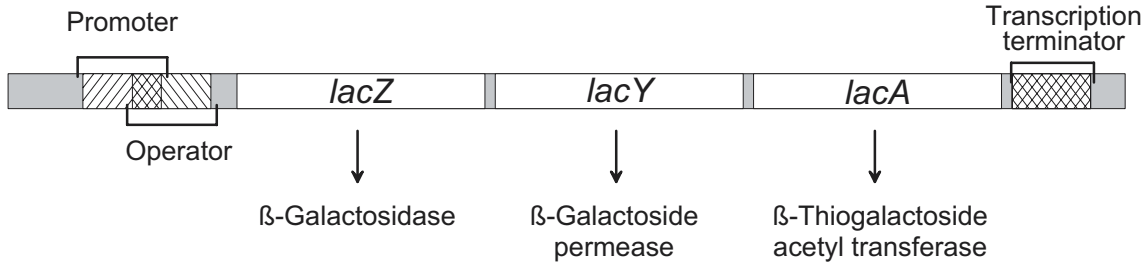

Figure 5.1: Scheme of the *lac* operon. *lac* promoter = binding site for RNA polymerase; operator = repressor binding site.

In an uninduced state, transcription initiation of the *lac* operon is physically blocked through binding of the repressor to the operator sequence. Allolactose functions as an inducer. Its binding to the repressor causes the repressor to dissociate from the operator region and thus enables transcription and subsequent translation of the encoded genes.

The sequence of the *lac*-operon was obtained for wild-type *Escherichia coli* K12 MG1655 from the European Molecular Biology Laboratory (EMBL) (accession number AE000141, release 66, version 6; http://www.embl-heidelberg.de/). *lacZ* mRNA contains 3,144 bases (= 1,048 base tripletts), considering the $5'$ and $3'$-ends reported earlier (Cannistraro and Kennell, 1985; Schulz and Reznikoff, 1990; McCormick et al., 1991). The coding region stretches from base tripletts 14 ($= j_{R0}$) to 1,037 ($= K$), and is thus 1,024 codons in length. Table 5.1 lists some of the main features of the *lac* operon.

Table 5.1: Characteristics of the *E. coli lac* operon. The spacer denotes the distance in number of bases between the Shine-Dalgarno sequence and the translation initiation codon.

	Gene [kb]	**mRNA** [kb]	**Shine-Dalgarno**	**Spacer**	**Translation start**
lacZ	3.072	3.144	AGGA	7	AUG
lacY	1.254	1.320	UAAGGA	8	AUG
lacA	0.612	0.671	GGAG	8	UUG

5.1.1 Half-lives of *lacZ* mRNA

Chemical half-lives of the 5′ and 3′-end of *lacZ* mRNA have been reported for various growth conditions of *E. coli* (see Table 5.2). The tabulated data can be distinguished according to the presence or absence of translation during the measurement period. For a system, in which translation initiation was inhibited, a half-life of 0.5 min was given for the 5′-terminal *lacZ* mRNA (Schneider et al., 1978). In the presence of an active translational machinery, the 5′-end is significantly stabilized and exhibits a chemical half-life of 1.9 min (Liang et al., 1999). In the same study, the 3′-end of *lacZ* mRNA was shown to be degraded also with a half-life of 1.9 min, albeit after a 1-min-delay as compared to the 5′-terminus. From these half-lives, the rate constants for exponential decay can be readily derived according to

$$k_{d,\mathrm{mRNA}} = \frac{\ln 2}{t_{1/2}} \tag{5.1}$$

Table 5.2: Chemical half-lives of *lacZ* mRNA for systems with/without translation. In either system, transcription initiation was turned off.

Translation	$t_{1/2}$ [min]	$k_{d,\mathrm{mRNA}}$ [min^{-1}]	**Remark**	**Source**
−	0.5	1.386	5′-end	Schneider et al. (1978)
+	1.9	0.365	5′-end	Liang et al. (1999)
+	1.9	0.365	3′-end, $\Delta t = 1$ min	dto.

5.1.2 Bounding regions for parameter range

The 1-min-time gap noted between 5′ and 3′-end degradation of *lacZ* mRNA in the presence of ribosomal translation (cf. Table 5.2) denotes the cumulative time needed for each degradosome to travel along a full-length transcript molecule and to perform endonuclease and exonuclease activities during this propagation. This Δt imposes severe constraints on the mean duration of each of the reaction steps during mRNA degradation. The average time required for each step is given by the reciprocal of the corresponding rate constant. The sum of all time steps taken

in the ordered process of mRNA degradation may thus be written as

$$\Delta t \;=\; \frac{J - j_{D0} - 1}{k_{D,mv}} + \frac{J-1}{k_{D,exo}} + \frac{Z}{k_{D,endo}} \tag{5.2}$$

Model constants used in equation (5.2) were explained in chapter 4. Since full-length *lacZ* mRNA consists of 3.144 kb, the overall degradosome propagation (including all of the above mentioned steps) thus proceeds in average at a rate of 17.5 base tripletts per second. Applying a limit case study, in which only one rate-limitation at a time is considered to occur, it is possible to estimate lower boundary values for each of the rate constants given above. That is,

$$\lim_{\substack{k_{D,exo}\to\infty \\ k_{D,endo}\to\infty}} \Delta t \;=\; \frac{J - j_{D0} - 1}{k_{D,mv}} \quad\Longleftrightarrow\quad k_{D,mv} \;\geq\; 17.5 \;\; \mathrm{s}^{-1} \tag{5.3}$$

$$\lim_{\substack{k_{D,mv}\to\infty \\ k_{D,endo}\to\infty}} \Delta t \;=\; \frac{J-1}{k_{D,exo}} \quad\Longleftrightarrow\quad k_{D,exo} \;\geq\; 17.5 \;\; \mathrm{s}^{-1} \tag{5.4}$$

$$\lim_{\substack{k_{D,mv}\to\infty \\ k_{D,exo}\to\infty}} \Delta t \;=\; \frac{Z}{k_{D,endo}} \quad\Longleftrightarrow\quad k_{D,endo} \;\geq\; \frac{Z}{60} \;\; \mathrm{s}^{-1} \tag{5.5}$$

j_{D0} was taken to be equal to 1 in this rough estimation. The total number of endonucleolytic cleavage sites is not exactly known for *lacZ* mRNA. From a close inspection of previously identified endonucleolytic cleavage sites, an empirical strategy has been devised (see next section) to estimate putative RNase E recognition sites. With this method, $Z = 20$ sites in total were predicted for *lacZ* mRNA to be susceptible to RNase E attack. Hence, and with equation (5.5), the rate constant for endonucleolytic cleavage ($k_{D,endo}$) calculates to be greater or equal to 0.3 s^{-1}.

5.1.3 Number of endonucleolytic cleavage sites

A number of five primary endonucleolytic cleavage sites has been verified experimentally for the 5′ and 3′-termini of *lacZ* mRNA (Cannistraro et al., 1986; Subbarao and Kennell, 1988; McCormick et al., 1991; Yarchuk et al., 1991). However, no such data exist for the major internal section of this mRNA. A detailed examination of the identified cleavage sites reveals that these sites share in common a region of minimally 8 nucleotides in length and a content of both G and C of at maximum 12.5 %. Under the premise that this concept of identifying endonucleolytic cleavage sites applies also for the remainder of the *lacZ* mRNA, the nucleotide sequence has been scanned for putative endonucleolytic cleavage sites according to this search pattern. The outcome of this analysis is shown in Table 5.3. In addition to the experimentally verified five endonucleolytic cleavage sites for *lacZ* mRNA, further 14 such regions have been uncovered, which proposedly function as RNase E recognition sites. Considering one additional cleavage site at the ultimate 3′-tail of *lacZ* mRNA, a total of 20 sites for endonucleolytic cleavage by RNase E were thus predicted. In average, one endonucleolytic cleavage site is suggested for about every 160 nucleotides.

Table 5.3: Estimated sites of endonucleolytic cleavage for wild-type *lacZ* mRNA. The position indicates the start of an A/U-rich stretch relative to native full-length mRNA. Sequences given are at minimum 8 bases long and interrupted at most once by one of the bases guanine or cytosine. When the exact sites of cleavage were reported, they are marked by a straight line in the respective sequence. References 1 = Subbarao and Kennell (1988), 2 = Yarchuk et al. (1991), 3 = Cannistraro et al. (1986), 4 = McCormick et al. (1991).

No.	Position [nt]	Length [nt]	G/C [%]	Sequence	Reference
1	13	10	10.0	AU|AACAAUUU	1, 2
2	70	8	12.5	UUUU|AC|AA	1, 2
3	109	8	12.5	AACUU|AAU	1
4	419	10	10.0	|AUUUAAUGUU	1
5	461	13	7.7	AAUUAUUUUUGAU	
6	732	9	11.1	UUUAAUGAU	
7	814	9	11.1	UUUCUUUAU	
8	869	9	11.1	UGAAAUUAU	
9	1050	9	11.1	AUUGAAAAU	
10	1188	8	12.5	AACUUUAA	
11	1281	10	10.0	AAUAUUGAAA	
12	1531	9	0.0	AUAUUAUUU	
13	1599	10	10.0	AUCAAAAAAU	
14	1691	8	12.5	UAAAUACU	
15	1765	11	9.1	UGAUUAAAUAU	
16	2356	11	9.1	AUAAAAAACAA	
17	2586	10	10.0	UUAUUUAUCA	
18	2869	11	9.1	AAUUGAAUUAU	
19	3106	14	0.0	AAAAAU|AAUAAUAA	3, 4

5.2 Dynamic simulation and nonlinear regression analysis

5.2.1 Initial conditions

The following conditions and assumptions were applied during the estimation procedure:

1. Throughout the experiment, mRNA synthesis is completely prevented through blocking of transcription initiation.

2. The degradosome diameter approximates the physical dimensions of the ribosome. I.e., $L_D = L_R = 12$ codons (Rauhut and Klug, 1999; Liou et al., 2001). The reference states for degradosome and ribosome, respectively, are $m_D = m_R = 7$.

3. The 5′-end of *lacZ* mRNA hosts both binding sites for degradosome and ribosome association. As can be seen from Figure 5.2, both sites are overlapping for the assumed ribosome and degradosome dimensions.

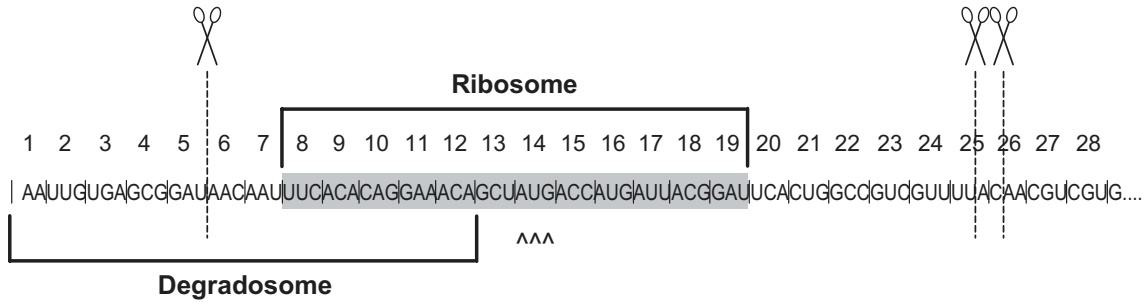

Figure 5.2: For wild-type *lacZ* mRNA, the sites of degradosome and ribosome association overlap. Base tripletts are sequentially numbered. The translational start codon is marked by arrows. Experimentally verified endonucleolytic cleavage sites (cf. Table 5.3) are indicated also.

4. Parameter $k_{TLI,\mathrm{IF2D}}$ was set to be equal to 48 min^{-1}, since this value was given for the effective frequency of translation initiation for wild-type *lacZ* mRNA under *in vivo* conditions (Liang et al., 1999).

5. In the case of *lacZ* mRNA, the average effective elongation rate of translating ribosomes, $\left(k_{TLE}\right)_{\mathrm{eff}}$, was reported to be 17.5 aa/s (Liang et al., 1999). In this value, sterical interactions among translating ribosomes are factored in, due to

$$\left(k_{TLE}\right)_{\mathrm{eff}} \;=\; q_j^R \; k_{TLE} \tag{5.6}$$

6. Termination of mRNA degradation was assumed to be a non-limiting reaction step. The rate constant $k_{D,T}$ was thus arbitrarily selected to be equal to 50 s^{-1}.

7. Simulation starts out with full-length mRNA. No degradation products of mRNA are present at this time ($t = t_0$). The initial concentration of each base triplett, $C_j^M(t_0)$, with $1 \leq j \leq J$ was chosen to be 0.05 μM. This value appears to be representative for the concentration of full-length *lacZ* mRNA in a growing cell (estimation given in Appendix F).

8. There are no degradosomes bound to full-length mRNA at simulation start. That is, $C_j^D(t_0) = 0$ μM for all j with $j_{D0} \leq j \leq J$.

9. For systems including ribosomal translation, the initial concentration of ribosomes bound to each codon j was taken to be equal to 2.3 nM.

10. Cell volume is regarded to be ideally mixed.

5.2.2 Performance index

With the measured chemical half-lives and the initial concentration of full-length mRNA, the time-dependent trajectory for 5′-terminal base tripletts of mRNA (i.e., base triplett $j = 1$) can be written as

$$C_1^M(t) \;=\; C_1^M(t_0) \; \exp\left[-\frac{\ln 2}{t_{1/2}} \cdot t\right] \tag{5.7}$$

The time-delayed first-order decay of the 3′-end of mRNA (i.e., base triplett $j = 1,048$) is described by

$$C^M_{1,048}(t) \;=\; \begin{cases} C^M_{1,048}(t_0) & \text{for } t \le \Delta t \\ C^M_{1,048}(t_0)\ \exp\left[-\dfrac{\ln 2}{t_{1/2}} \cdot (t - \Delta t)\right] & \text{for } t > \Delta t \end{cases} \tag{5.8}$$

Nonlinear regression analysis in combination with dynamic simulation was thus performed. The analysis was based on minimizing the sum of square relative errors. In these calculations, the setpoint concentrations of 5′ and 3′-terminal base tripletts were taken at discrete time points from equations (5.7) and (5.8), respectively. These setpoint concentrations served as artificially generated experimental data, which were derived from the reported chemical mRNA half-lives.

In addition to least square fit analysis, the following parameters were monitored during simulation as model outputs, in order to allow a further assessment of system performance. The average spacing between ribosomes can be calculated from

$$d_R \;=\; \frac{\sum_{j=j_{R0}}^{K} C^M_j}{\sum_{j=j_{R0}}^{K} C^R_j} \tag{5.9}$$

The average spacing between degradosomes is given by

$$d_D \;=\; \frac{\sum_{j=j_{D0}}^{J} C^M_j}{\sum_{j=j_{D0}}^{J} C^D_j} \tag{5.10}$$

For times at which all concentrations of mRNA-bound degradosomes differ from 0, the average effective rate constant of degradosome movement can be obtained from

$$\left(k_{D,mv}\right)_{\text{avg}} \;=\; \frac{n_{cod}}{J - j_{D0}} \sum_{j=j_{D0}}^{J-1} \frac{V_{D,mv,j}}{C^D_j} \tag{5.11}$$

5.2.3 Parameter estimation

In an attempt to identify model parameters with enhanced sensitivity, a sequential estimation procedure was applied. The identification of model parameters was initially carried out with a simplified state representation (see method described in section 4.2.5). At first, the concentrations of mRNA and positional loadings were derived for every four adjacent base tripletts ($n_{cod} = 4$). The results of this analysis were at a later stage compared to results obtained using the model with full state representation (with $n_{cod} = 1$).

Rate constant of degradosome association. From the degradation of 5′-terminal *lacZ* mRNA, when no translation was present, the rate constant of degradosome association, $k_{D,ass}$,

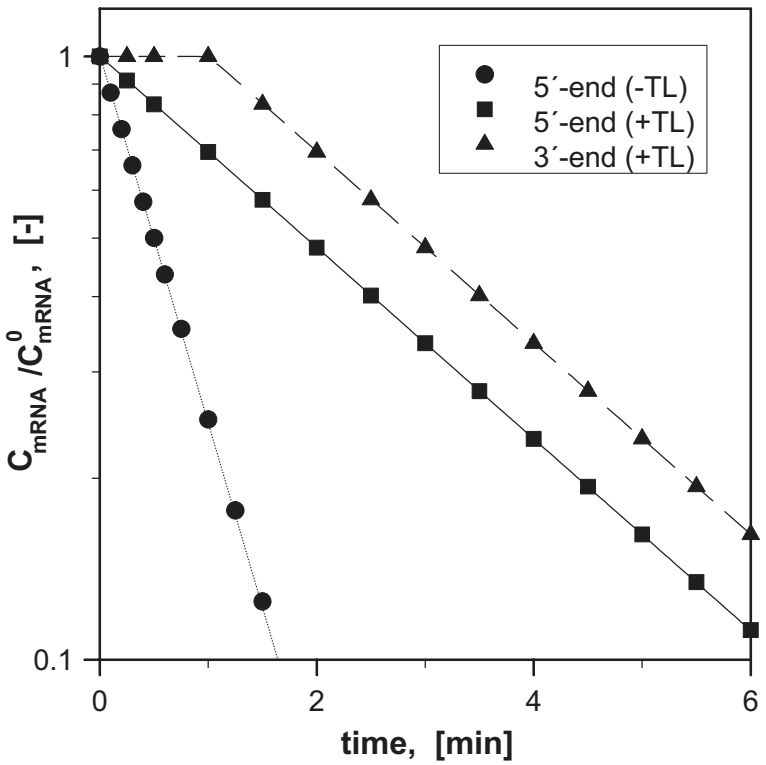

Figure 5.3: Comparison of simulated versus experimental time course of terminal regions of *lacZ* mRNA. Relative concentrations are given, normalized with respect to their initial concentration. Circles denote the 5′-end of mRNA in the absence of translation. Squares and triangles, and the corresponding lines refer to 5′-end and 3′-end of *lacZ* mRNA, respectively, in the presence of ribosomal translation. Experimental data were artificially generated from the mRNA half-life provided by Schneider et al. (1978) and Liang et al. (1999). Reduced model with $n_{cod} = 4$.

was estimated to be equal to 1.386 min^{-1}. The outcome of parameter estimation is given by the curve corresponding with the black circles in Figure 5.3. The same estimate could have been obtained from mere calculus and under the premise that the half-life of 5′-terminal *lacZ* mRNA is primarily governed by degradosome association.[6] The parameter value identified for $k_{D,ass}$ was kept fixed throughout the subsequent estimation procedure.

Rate constant of 70S initiation complex formation. Assuming that the increased mRNA stability due to translation is primarily caused by inhibited degradosome assocation, queueing factor q^{D0}_{jD0} can be estimated, as is outlined in the following. Using equation (4.3), the ratio of degradosome association rates of both systems with and without translation can

[6] The total concentration of RNase E is approximately 0.25 μM (see estimation given in Appendix F). Since the entirety of cellular mRNA concentration is roughly 1.5 μM (i.e., about 4.8 mM in nucleotide moles per cell volume (Schwarz, 1999)), the codon-specific concentrations of mRNA-bound degradosomes are expected to be very small. Consequently, degradosome movement will hardly be affected by interactions among mRNA-bound degradosomes. That is, q^{D0}_{jD0} and q^{D}_{j} will approximate 1 under conditions where ribosomal translation is discarded. The rate constant for degradosome association can then be calculated from equations (4.3) and (5.1) to give 1.386 min^{-1}.

be written as

$$\frac{\left(V_{D,ass}\right)_{(+TL)}}{\left(V_{D,ass}\right)_{(-TL)}} = \frac{\left(k_{D,ass}\; q_{jD0}^{D0}\; C_{jD0}^{M}\right)_{(+TL)}}{\left(k_{D,ass}\; q_{jD0}^{D0}\; C_{jD0}^{M}\right)_{(-TL)}} \tag{5.12}$$

If the concentration of *lacZ* mRNA (C_{jD0}^{M}) and the rate constant for degradosome association ($k_{D,ass}$) are the same, whether translation prevails or is excluded, a difference in the rate of 5′-mRNA degradation between both systems would be reflected solely by q_{jD0}^{D0}. From equation (5.12), it is then possible to derive the following relationship:

$$\frac{\left(V_{D,ass}\right)_{(+TL)}}{\left(V_{D,ass}\right)_{(-TL)}} = \frac{\left(q_{jD0}^{D0}\right)_{(+TL)}}{\left(q_{jD0}^{D0}\right)_{(-TL)}} = \frac{\left(t_{1/2}\right)_{(-TL)}}{\left(t_{1/2}\right)_{(+TL)}} \tag{5.13}$$

Using equation (5.13) and assuming $\left(q_{jD0}^{D0}\right)_{(-TL)} \approx 1$ (in the case where no ribosomes are attached to mRNA), $\left(q_{jD0}^{D0}\right)_{(+TL)}$ is calculated to be equal to 0.2632 (= 0.5/1.9). This is a rough estimate under the assumption of unimpaired degradosome association. Without the need for this simplification, parameter $\left(q_{jD0}^{D0}\right)_{(+TL)}$ was estimated from nonlinear regression analysis, as is detailed in the following.

The values which the queueing factor $\left(q_{jD0}^{D0}\right)_{(+TL)}$ takes are governed by the fractional occupancy of base tripletts in the direct vicinity of the ribosome binding site. These fractional loadings are a primary result of the relative rates of translation initiation versus translation elongation. In the investigated example, parameters $\left(k_{TLE}\right)_{\text{eff}}$ and $k_{TLI,\text{IF2D}}$ are fixed, as a result of experimental determination. The only model parameter left for influencing $\left(q_{jD0}^{D0}\right)_{(+TL)}$ is $k_{TLI,\text{70SIC}}$, which effectively determines the concentration of ribosomes attached to the ribosome binding site. Parameter $k_{TLI,\text{70SIC}}$ was estimated through fitting of simulation results to the setpoint trajectory of 5′-terminal mRNA in the presence of translation (square symbols and solid line in Figure 5.3). Through this procedure, the rate constant of 70S initiation complex formation ($k_{TLI,\text{70SIC}}$) was determined to be 14.2 s^{-1}. Given this parameter value, the queueing factor for degradosome association $\left(q_{jD0}^{D0}\right)_{(+TL)}$ was obtained to be equal to 0.2626, under pseudo-steady state conditions of mRNA degradation. The noted stability improvement of 5′-*lacZ* mRNA could thus be explained exclusively by mRNA-bound ribosomes physically preventing access to the degradosome binding site.

Rate constants of endonucleolytic cleavage, exonucleolytic cleavage, and degradosome movement. Through fitting the simulated time course of the 3′-terminal base triplett of *lacZ* mRNA to its setpoint trajectory, the rate constants for endonucleolytic cleavage ($k_{D,endo}$) was estimated to be 2.6 s^{-1}. Estimates for the rate constants of exonucleolytic cleavage ($k_{D,exo}$) and degradosome movement ($k_{D,mv}$) were determined to be equal to 680 nt s^{-1} and 95 nt s^{-1}, respectively. Figure 5.3 (triangles and dashed graph) illustrates the obtained time dependency for 3′-*lacZ* mRNA using the identified parameter set and in comparison to the experimentally measured 3′-terminal base triplett concentration. A consistency check demonstrates that these parameters estimated are located well above their earlier identified lower boundary values (cf. section 5.1.2).

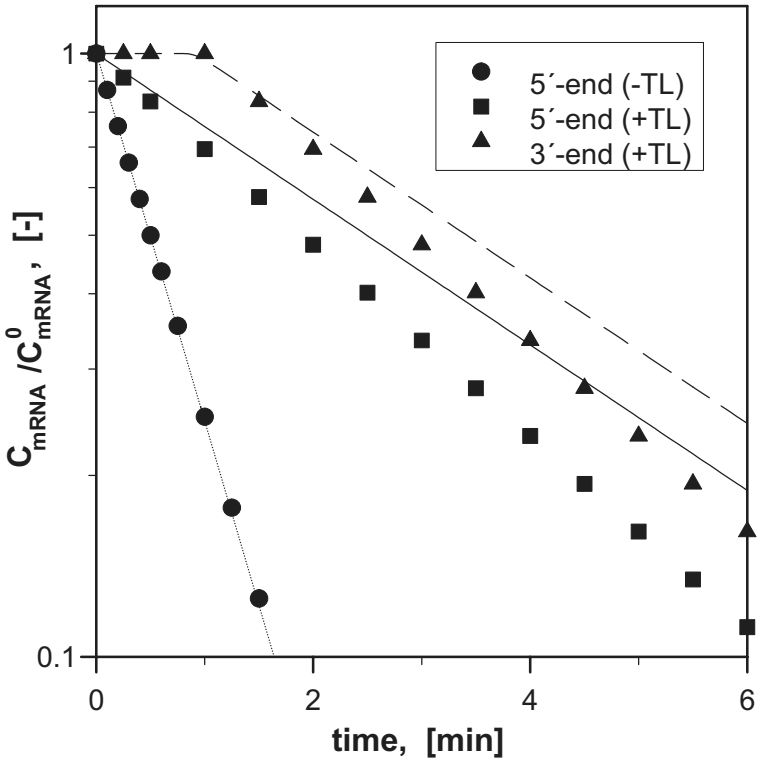

Figure 5.4: Comparison of simulated versus experimental time course of both 5′ and 3′-ends of *lacZ* mRNA in the presence of ribosomal translation. Relative concentrations are given, normalized with respect to their initial concentration. Experimental data were artificially generated from the mRNA half-life provided by Liang et al. (1999). Full model with $n_{cod} = 1$. All model constants were the same as identified for the system with $n_{cod} = 4$.

While the above parameter estimation has been conducted with a simplified model exhibiting a lower resolution of state variables ($n_{cod} = 4$), the applicability of these parameters was tested subsequently employing the model with full state representation ($n_{cod} = 1$). When the same parameter set as estimated for the simplified model is applied also for the full model (i.e., the model with resolution I), a mismatch between simulated time traces and experimental observation is noted for the system including ribosomal translation. The concentrations of mRNA base tripletts are in this case proposed to be higher than in comparison to the experiment (see Figure 5.4). Nevertheless, the 1-min-time delay between 5′ and 3′-end degradation appears to be predicted correctly by the model. This finding in combination whith the similarity noted between the set-off for both 5′ and 3′-terminal mRNA suggests mainly the degradosome association rate to be influenced by the effects of model reduction. When in a next step, the rate constant for 70S initiation complex formation was reevaluated keeping $n_{cod} = 1$, an improved fit between simulated and experimental time courses of both terminal mRNA base tripletts could be restored (see Figure 5.5). In this case, parameter $k_{TLI,70SIC}$ was estimated to be 4.3 s^{-1}. Thus, indeed degradosome association rate was shown to respond most sensitively among the parameters of the mRNA degradation model to changes in the state representation.

An explanation for the observed sensitivity becomes apparent from the implications of

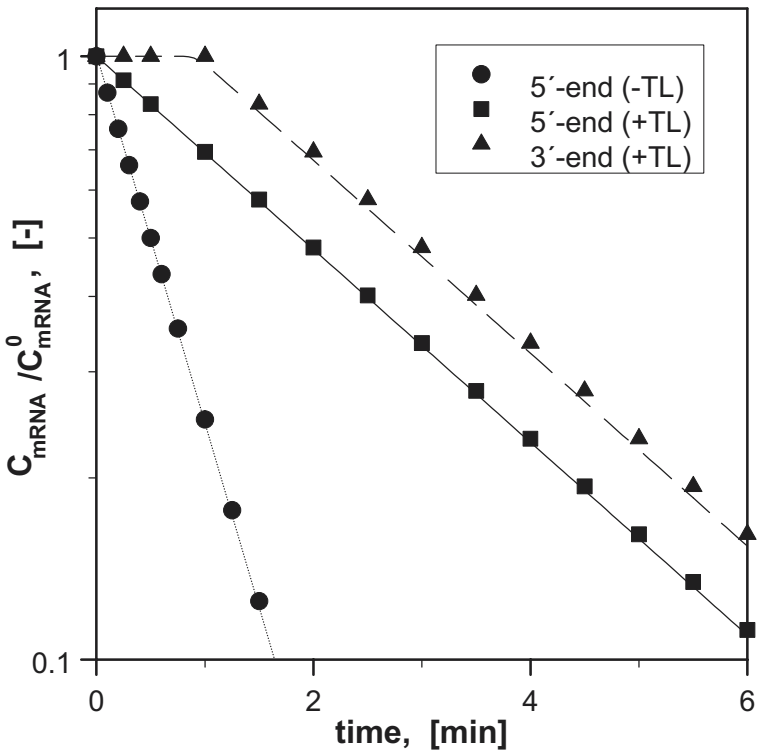

Figure 5.5: Comparison of simulated versus experimental time course for both $5'$ and $3'$-ends of *lacZ* mRNA in the presence of ribosomal translation.Relative concentrations are given, normalized with respect to their initial concentration. Experimental data were artificially generated from the mRNA half-life provided by Liang et al. (1999). Full model with $n_{cod} = 1$ and $k_{TLI,70\text{SIC}}$ equal to 4.3 s^{-1}.

reduced state representation. For $n_{cod} = 4$, ribosomes and degradosomes bound to mRNA cover a smaller number of positions at a time, namely 3 instead of 12 for the assumed case of resolution I (keeping in mind that the physical dimensions of ribosomes, degradosomes, and mRNA remain the same in either system representation). The queueing factor q^{R0}_{jR0} is then assembled of a smaller number of states of both ribosomes and degradosomes. These slight inaccuracies due to model simplification are shown to manifest themselves in an approximately 3-fold difference of factor q^{R0}_{jR0}. For $n_{cod} = 4$, q^{R0}_{jR0} equalled 0.0345, while it was equal to 0.1152 for $n_{cod} = 1$ (see Table 5.4). Both values were obtained under pseudo-steady state conditions of mRNA degradation. As a consequence of the above, parameter $k_{TLI,70\text{SIC}}$ was found to vary with the resolution of state representation.

Table 5.4 summarizes the effects of state resolution on characteristic quantities of the mRNA degradation model in combination with protein expression. In essence, it appears that lower resolutions lead to higher predicted concentrations of bound ribosomes, and consequently decreased values for queueing factors, average distances between ribosomes and degradosomes, respectively, and the average effective rate of degradosome propagation.

The fractional occupancy of codons with respect to ribosome loading is given by the ratio of ribosome concentration bound to a particular codon j and the concentration of this codon. That is, C^R_j/C^M_j. For $n_{cod} = 1$, this ratio is calculated to be 0.02 for all codons except for

Table 5.4: Model outputs from dynamic simulation and parameter identification. All quantities in this list refer to quasi-steady state conditions of mRNA degradation in the presence of translation. Parameter n_{cod} denotes the degree of codon refinement, as defined in section 4.2.5.

Parameter	**Unit**	$n_{cod} = 4$	$n_{cod} = 1$
$\left(q_{j_{R0}}^{R0}\right)_{\text{qss}}$	—	0.0345	0.1152
$\left(q_{j_{D0}}^{D0}\right)_{\text{qss}}$	—	0.2626	0.2632
$\left(q_{j}^{D}\right)_{\text{qss}}$	—	0.8563	0.9747
$\left(k_{D,mv}\right)_{\text{avg}}$	codons/s	26.8	30.6
$k_{D,mv}$	codons/s	31.5	31.4
d_R	nt	110	150
d_D	nt	8,600	9,300
$\left(C_j^R/C_j^M\right)_{\text{qss}}$	μM/μM	0.11	0.02
$\left((C_{j_{R0}}^{R*} + C_{j_{R0}}^{R})/C_{j_{R0}}^{M}\right)_{\text{qss}}$	μM/μM	0.73	0.65
$V_{D,ass}/V_{TLI,70\text{SIC}}$	(μM/min)/(μM/min)	0.01	0.01

the initiation codon (see Table 5.4). In contrast, the translational start site (at $j = j_{R0}$) is estimated to exhibit a by a factor of 32.5 higher ribosome loading, supporting the conception that ribosomal binding to the translation initiation site functions as an effective mechanism to block upstream propagating degradosomes from entering the coding region.

5.2.4 Comparison of estimated parameters with literature

Table 5.5 lists the outcome of parameter estimation for the mRNA degradation model. The only kinetic parameter found in the literature, to which the rate constant for degradosome movement may be compared, is the rate of DNA unwinding by PcrA helicase. The unwinding of DNA by this enzyme was reported to proceed at a rate of 50 nt/s (Soultanas and Wigley, 2001), which is about half the rate constant estimated in this study for degradosome movement (95 nt/s). The overall rate of endonucleolytic cleavage by endonuclease RNase P is equal to 0.3 s^{-1} (Hardt et al., 1995; Brännvall and Kirsebom, 1999). In this study, an about 10-fold higher rate constant for endonucleolytic cleavage was estimated. Parameter values estimated in this study are thus in reasonably good agreement with measured rates for similar enzymes.

Table 5.5: Estimated parameters for model of bacterial mRNA degradation employing *lacZ* mRNA in the presence of translation.

Parameter	**Unit**	**Value**
$k_{D,ass}$	s^{-1}	0.023
$k_{D,endo}$	s^{-1}	2.6
$k_{D,exo}$	nt s^{-1}	680
$k_{D,mv}$	nt s^{-1}	95
$k_{TLI,70\text{SIC}}$	s^{-1}	4.3

5.3 Conclusions

This chapter provided a framework for investigating the influence of ribosomal packing on mRNA protection against nucleolytic attack. At the example of *lacZ* mRNA, it was possible to estimate the model constants of the mRNA degradation model introduced in the previous chapter. The general applicability of the identified parameter values to span a variety of mRNAs that follow a $5'$ to $3'$-degradation pathway, however, needs to be further exploited.

An efficient translation initiation not only leads to high protein expression rates. The results obtained in this study demonstrate that the efficiency of translation initiation also functions to control the stability of a mRNA transcript, when it is conform with the investigated degradation mechanism involving the degradosome. In this case, high fractional loadings of the ribosome binding site effectively function as a road-block to keep upstream degradosomes from accessing endonucleolytic cleavage sites that are contained within the coding region.

Parameter estimation on the basis of a lower resolution of system representation can lead to an overestimation of queueing effects. A high sensitivity was observed for the association probabilities of both ribosomes and degradosomes dependent on the rate constant for 70S initiation complex formation.

Even if the concentration of bound ribosomes is in general expected to be orders of magnitude greater than the concentration of bound degrasosomes, it may become necessary - for technical reasons - to include the contribution of degradosomes in the queueing parameter for ribosome elongation. In particular, with progressing mRNA degradation, the imbalance between the concentrations of bound ribosomes versus bound degradosomes will shift towards an increased fraction of bound degradosomes, which may then add significantly to the occupational status of a mRNA.

In order to cover a broader range of mRNA degradation mechanisms and to improve the accuracy of predicted functional half-lives for arbitrary messages, RNA structural analysis needs to be included into the modelling scheme. An increased efficiency of ribosomal translation would manifest itself in terms of degradosome rejection from the $5'$-binding site. This experimental observation is correctly predicted by the model.

6 Prokaryotic Translation

6.1 Introduction

Translation is the polymerization process of protein biosynthesis, whereby the incorporation of the monomeric building blocks, the amino acids, is governed by the nucleotide sequence of the mRNA template. The rules according to which the genetic alphabet that is composed of the four bases Adenine (A), Cytosine (C), Guanine (G), and Thymine (T), is translated into amino acid sequence are provided by the genetic code. Thymine is replaced by Uracil (U) in mRNA. Each unit of three nucleic bases forms a codon. The genetic code (see Table 6.1) is **degenerate** in the sense that the majority of amino acids is characterized by more than one codon. That is, the 61 codons relate to the 20 naturally occurring amino acids. The three stop codons UAA, UAG, and UGA indicate translation termination. With a few exceptions, the same codon assignments apply in all organisms (universal genetic code). During protein biosynthesis, three different groups of ribonucleic acid (RNA) are involved.

Table 6.1: The bacterial genetic code. In general, AUG is the initiation codon. GUG, UUG, AUU, and CUG are alternative initiation codons. * denotes a termination codon.

UUU	F	Phe	UCU	S	Ser	UAU	Y	Tyr	UGU	C	Cys
UUC	F	Phe	UCC	S	Ser	UAC	Y	Tyr	UGC	C	Cys
UUA	L	Leu	UCA	S	Ser	UAA	*	Ter	UGA	*	Ter
UUG	L	Leu i	UCG	S	Ser	UAG	*	Ter	UGG	W	Trp
CUU	L	Leu	CCU	P	Pro	CAU	H	His	CGU	R	Arg
CUC	L	Leu	CCC	P	Pro	CAC	H	His	CGC	R	Arg
CUA	L	Leu	CCA	P	Pro	CAA	Q	Gln	CGA	R	Arg
CUG	L	Leu i	CCG	P	Pro	CAG	Q	Gln	CGG	R	Arg
AUU	I	Ile i	ACU	T	Thr	AAU	N	Asn	AGU	S	Ser
AUC	I	Ile i	ACC	T	Thr	AAC	N	Asn	AGC	S	Ser
AUA	I	Ile i	ACA	T	Thr	AAA	K	Lys	AGA	R	Arg
AUG	M	Met i	ACG	T	Thr	AAG	K	Lys	AGG	R	Arg
GUU	V	Val	GCU	A	Ala	GAU	D	Asp	GGU	G	Gly
GUC	V	Val	GCC	A	Ala	GAC	D	Asp	GGC	G	Gly
GUA	V	Val	GCA	A	Ala	GAA	E	Glu	GGA	G	Gly
GUG	V	Val i	GCG	A	Ala	GAG	E	Glu	GGG	G	Gly

Ribosomal RNA (rRNA) — Ribosomal RNA comprises about 80 % of the total cellular RNA content. Together with ribosomal proteins, rRNA assembles to the ribosome, the catalyst performing the translation reaction. Each ribosome is composed of two subunits, the smaller 30S subunit and the larger 50S subunit in prokaryotes, which are distinguished according to their sedimentation velocity (S = Svedberg is a unit measure for sedimentation rate under defined conditions). The association of both subunits leads to the formation of the 70S ribosome.

Transfer RNA (tRNA) — Each tRNA species has an exposed characteristic sequence of three bases, the anti-codon, which is complementary to a certain codon. In an enzyme-catalyzed

reaction, tRNA is charged at its 3′-end preferentially with its corresponding amino acid. tRNA functions as an adaptor molecule that transports amino acid to the ribosome. During the translation process, the anti-codon of the aminoacylated tRNA (aa-tRNA) interacts with the codon of the mRNA without the participation of the amino acid in codon recognition. At correct base-pairing, a correct amino acid can be incorporated into the nascent peptide chain.

Messenger RNA (mRNA) — The primary function of mRNA is to provide the sequence of amino acid addition during protein synthesis. Translation proceeds from the 5′-end to the 3′-end of mRNA. mRNA is the most unstable component among the three RNA species, as was discussed in chapter 4.

As illustrated in Figure 6.1, the polymerization reaction of ribosomal protein synthesis can be divided into three distinct reaction stages: **Initiation** denotes the binding of ribosome to the mRNA at a specific recognition site. Chain **elongation** is characterized by a successive progression of the bound ribosome along the mRNA template, where amino acids are linked by peptide bonds to generate protein. When a termination signal is reached on the mRNA, **termination** of protein synthesis is accomplished by the release of ribosomal subunits together with the fully synthesized gene product. Each of the described reaction stages itself involves several interrelated steps, at which additionally proteinogenic translation factors are catalytically engaged.

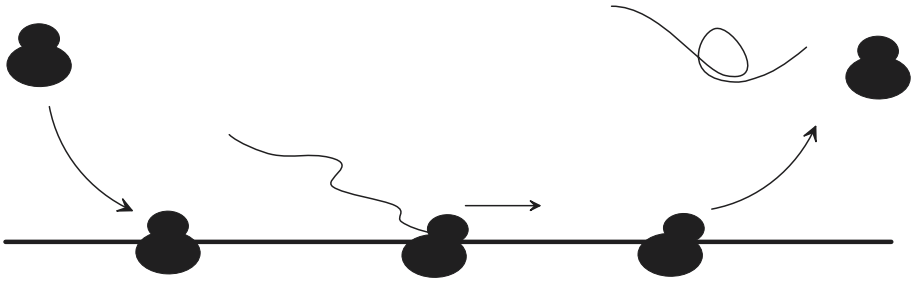

Figure 6.1: Schematic representation of prokaryotic protein synthesis, distinguished between initiation, elongation, and termination of translation.

Translation rates are known to vary with the protein product. It is generally accepted that codon composition, tRNA population and gene expressivity are strongly correlated (Gouy and Gautier, 1982). The concentration of cognate tRNA is known to be positively correlated with the frequency of codon usage (Ikemura, 1981a; 1981b). Abundant proteins were found to be translated at a higher rate than in comparison to rare proteins (Pedersen, 1984). Elongation rate for two neighbouring codons may be different by up to one order of magnitude (Liljenström and von Heijne, 1987). Synonymous codons sharing the same cognate tRNA showed noticably divergent elongation rates (Sørensen and Pedersen, 1991). Variations of elongation rate have been attributed to differences in tRNA availability (Varenne et al., 1984), and alternatively to the variability of binding constants for codon-anticodon interaction (Sørensen and Pedersen, 1991). Codon context was considered to be insignificant in determining elongation rates (Sørensen and Pedersen, 1991). An optimization of elongation rate along the mRNA can be accomplished through the preferential selection of synonymous codons matching those isoacceptor tRNAs that are largely available (Liljenström and von Heijne, 1987).

During ribosomal translation, typcially several ribosomes are bound in sequence along the mRNA, a phenomenon called polyribosomes (or polysomes). Queue formation among translating ribosomes has been demonstrated both *in vitro* (Wolin and Walter, 1988), and *in vivo*, the latter in *Escherichia coli* during amino acid starvation (Dahlberg et al., 1973). Stalled ribosomes can cause a situation similar to that observed during a traffic jam in car traffic. A temporal hold-up of ribosomes, as is visualized in Figure 6.2, may result from downstream ribosomes scanning for the correct aminoacylated tRNA. Another example is the clustering of rare codons, which leads to more densly spaced ribosomes upstream and causes more distant spacing among ribosomes downstream of the cluster (Zhang et al., 1994). Such effects can lead to significantly lower rates of ribosomal movement than may be inferred from substrate availability, and could ultimately cumulate in a breakdown of protein synthesis, when at least one amino acid is missing.

Figure 6.2: Interference between ribosomes travelling along a mRNA.

Due to the central role of gene expression on cell metabolism, protein biosynthesis has been a major target of mathematical modelling. While individual features of translation have been modelled in great detail, a mechanistic model combining the majority of involved key processes in one model is missing. This lack is of particular importance in the pursuit of a thorough understanding of the molecular basis of ribosomal interactions.

In this chapter, a kinetic model of the prokaryotic translation process is developed that substantiates on the profound biomolecular knowledge gathered over the past decades. The model distinguishes between initiation, elongation, and termination of protein polymerization, and features the key catalysts enrolled in these reactions. Moreover, mutual interactions among ribosomes organized within a polysome structure are taken into account.

6.2 Initiation

Translational initiation combines the series of reaction steps that lead up to the formation of the first peptide bond during ribosomal protein synthesis. In a complex multi-step process involving initiation factors IF1, IF2, and IF3, the binding of 30S ribosomal subunit to the initiator tRNA (fMet-tRNA$_f^{\mathrm{M}}$), and their association to the ribosome binding site (RBS) of the mRNA are accomplished (see also Figure 6.3).

The RBS of prokaryotic mRNA covers about 35 bases and is mostly characterized by the Shine-Dalgarno (SD) sequence, which base-pairs with the 16S rRNA during initiation site selection, and the number and base composition to the initiator codon. SD-sequences vary from one gene to another, but they all comprise a sequence of usually 4 to 6 bases (Stormo, 1990; Shultzaberger et al., 2001) that are complementary to the 3′-terminus of the 16S rRNA (5′-...-GAUCACCUCCUUA-3′). The initiation codon (typically AUG coding for methionine) is located approximately at 2/3 of this region (de Smit and van Duin, 1990a). Initiation

Table 6.2: Quantitative data of *Escherichia coli* protein synthesis versus specific growth rate μ. Data taken from Bremer and Dennis (1996) unless otherwise specified. [a]calculated from Bremer and Dennis (1996). [b]Gouy and Grantham (1980). [c]Howe and Hershey (1983), who detected an approximately constant molar ratio between translation factors and ribosomes with the growth rate. [d]Jakubowski and Goldman (1984). [e]calculated from data provided by Furano (1975). Values obtained through linear interpolation from data given by [f]Adamski et al. (1994). $70S_t$ = total ribosome concentration, ARS = aminoacyl-tRNA-synthetase. The ratio of IF3/$70S_t$ remains approximately constant with the growth rate and was reported to be 0.16, 0.14, 0.11, 0.14 for μ equal to 0.37, 0.78, 1.16, 1.46 h^{-1}, respectively (Liveris et al., 1991). The index t refers to total quantities.

μ (h^{-1})	0.4	0.7	1.0	1.4	1.7
peptide chain elongation rate, (aa/s)	12	16	18	20	21
spacing of 70S on mRNA, (nt)	79	85	65	52	41
$70S_t$/cell, (ribosomes/cell)	6,800	13,500	26,300	45,100	72,000
$70S_t$, (μM)	22.9[a]	26.1[a]	30.3[a]	35.0[a]	41.5[a]
$tRNA_t/70S_t$	9.3	9.3	9.3	9.3	9.3
$(ARS)_t/70S_t$		2.1[b]	1.7[b]	1.5[b]	1.3[b]
(average ARS)/$70S_t$		0.06[d]	0.08		
$IF1_t/70S_t$			0.25[c]		
$IF2_t/70S_t$			0.30[c]		
$IF3_t/70S_t$			0.20[c]		
$EFTu_t/70S_t$		5.8[e]	5.4[c]	5.2[e]	
$EFG_t/70S_t$			0.8[c]		
$EFTs_t/70S_t$			0.18[c]		
$RF1_t/70S_t$	0.45[f]	0.39[f]	0.32[f]	0.24[f]	0.18[f]
$RF2_t/70S_t$	2.05[f]	1.52[f]	0.84[f]	0.87[f]	0.90[f]

is generally believed to be the rate-controlling step of translation (Gualerzi and Pon, 1990; Neidhardt et al., 1996). An initiation frequency of 0.8 s^{-1} was determined for *lacZ* mRNA (Liang et al., 1999), while bacterial translation elongation proceeds at a rate of about 10 to 20 aa/s (cf. Table 6.2). The particular importance of translation initiation is further emphasized since it represents the focal point of many regulatory control mechanisms of gene expression on the translational level (Gold et al., 1981).

6.2.1 Previous modelling

A number of mathematical models exist that describe various facets of protein synthesis initiation. Binding studies were carried out to determine the association constants for *E. coli* ribosomal subunit association and initiation factor binding at various ionic conditions (Spirin and Lishnevskaya, 1971; Naaktgeboren et al, 1977; Chaires et al., 1981; Weiel and Hershey, 1982; Goss et al., 1981; Zucker and Hershey, 1986; Gualerzi and Pon, 1990). Initial rate kinetics

of translational initiation were derived from an *in vitro* system, by assuming a rapid equilibrium ordered mechanism for initiator tRNA binding to the 30S ribosomal subunit and the subsequent mRNA association (Ellis and Conway, 1984). However, modelling results obtained in the latter study may not be generalized to include cellular conditions, because these authors were setting the overall rate of protein appearance - including elongation and peptide release - equal to initiation rate. In general, a random binding mechanism is favoured for ternary complex formation as part of the entire initiation reaction (see review by Gualerzi and Pon (1990)). Translation initiation kinetics have been studied for *E. coli* derived systems using stopped-flow techniques to elucidate individual conformational changes and to measure the respective rates of elementary reactions (Wintermeyer and Gualerzi, 1983; Tomšic et al., 2000).

6.2.2 Reaction scheme and kinetics

From the above cited studies, the reaction scheme of bacterial translation initiation shown in Figure 6.3 was derived. The initiation process distinguishes the steps of dissociation of ribosomal subunits (step (1) in Figure 6.3), association of initiation factors to 30S (step (2)), binding of ribosomal subunits to mRNA (steps (3) to (6)), and dissociation of IF2 from the mRNA-bound ribosome (step (7)). The reaction mechanisms of the individual steps are subsequently further elucidated.

6.2.2.1 Dissociation of ribosomal subunits

The ribosomal subunits need to fall apart prior to ribosome binding to the mRNA. Under physiological conditions, the thermodynamic equilibrium of association of ribosomal subunits

$$30S + 50S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 70S \tag{6.1}$$

is shifted to 70S formation. The association constant for reaction equation (6.1) was given to be $K_{70S} = k_1/k_{-1} = 5.3{\cdot}10^7\ \mathrm{M}^{-1}$ (Zucker and Hershey, 1986). With increasing Mg^{2+} concentration, dropping concentrations of monovalent ions, reduced pH and temperature, further association of ribosomal subunits is promoted (Spirin and Lishnevskaya, 1971; Naaktgeboren et al., 1977). Importantly, the location of the equilibrium is greatly affected by the individual and combined effects of initiation factor presence.

The physiological roles of initiation factors IF1, IF2, and IF3 are summarized in Table 6.3. Factor IF1 is known to accelerate ribosomal dissociation without affecting the position of the equilibrium. Initiation factor IF3 displaces the association equilibrium in favour of the free subunits, however, without influencing the dissociation rate. IF2 participates in productive binding of fMet-$tRNA_f^M$ to the 30S subunit, as well as ribosome-dependent GTP hydrolysis. IF2 was suggested to exist mostly complexed with GTP under *in vivo* conditions (Tomšic et al., 2000). A preliminary estimation undertaken in this study essentially confirms this proposal (see Appendix F.3). It seems thus valid to regard the concentration of uncomplexed IF2 to be negligible at sufficiently high GTP concentrations. The activities of all initiation factors can be altered synergistically through cooperative binding to the native 30S subunit.

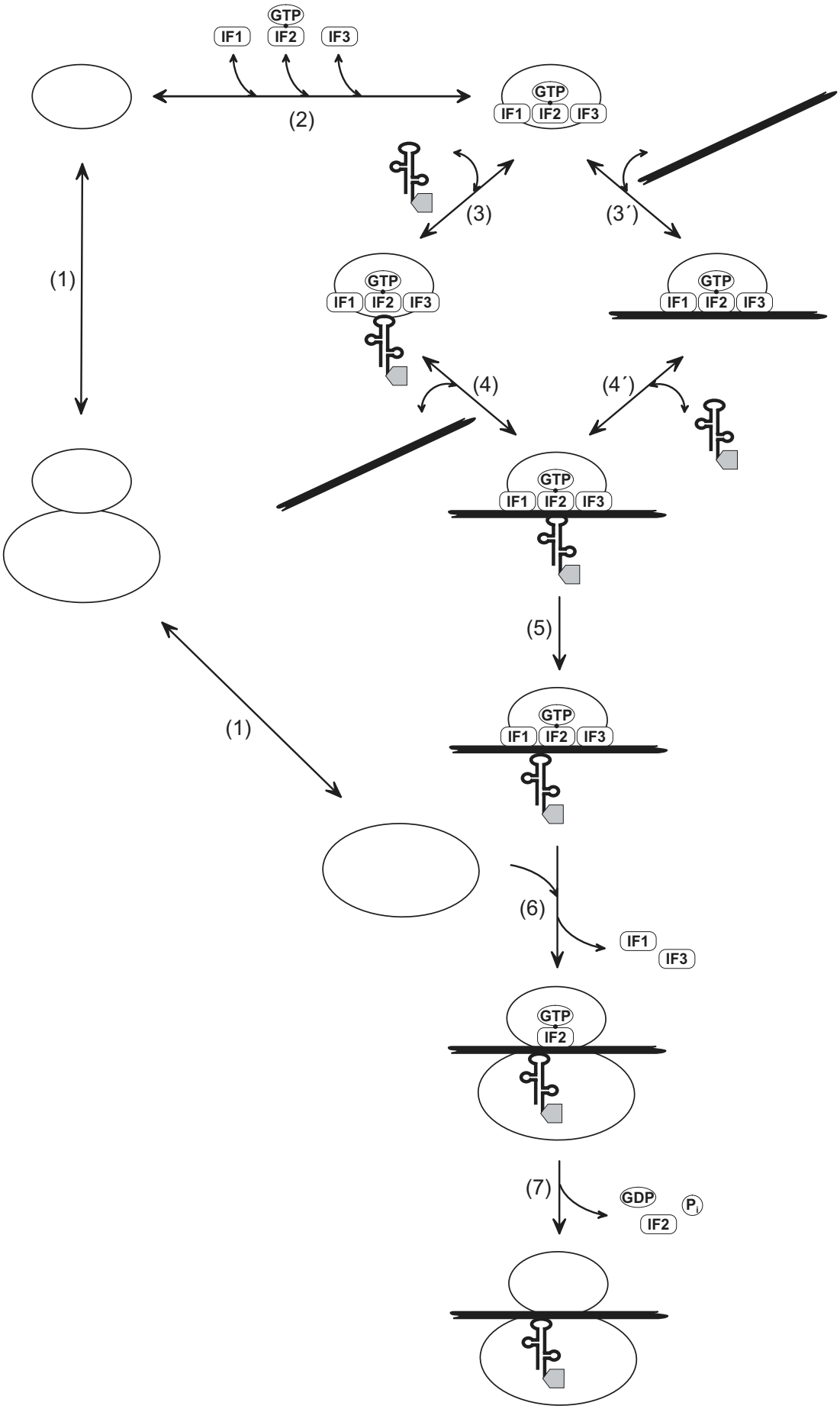

Figure 6.3: Principle reaction scheme of prokaryotic translation initiation.

Table 6.3: Role of prokaryotic initiation factors. Initiation factor IF2 is expressed as two proteins of different lengths in *E. coli*. However, the resulting mechanistic implications are not further considered in this study.

Factor	Properties and function
IF1	• increases the rate of exchange of the ribosomal subparticles through raising both k_1 and k_{-1} (see equation (6.1))
	• stimulates IF2 and IF3 activities
	• increases the affinity of 30S for IF2 and IF3
IF2	• promotes fMet-tRNA$_f^{\mathrm{M}}$ binding to 30S and its correct positioning in the P-site
	• protects fMet-tRNA$_f^{\mathrm{M}}$ against spontaneous deacylation
	• sole target for GTP binding, exerts ribosome-dependent GTPase activity
	• increases the affinity of 30S for IF1
IF3	• acts as an anti-association factor in the formation of 70S ribosomes by decreasing k_1 used in equation (6.1), while keeping k_{-1} unaltered
	• 30S is the only target for IF3 binding
	• increases the affinity of 30S for IF1 and IF2

6.2.2.2 Association of initiation factors to 30S

The binding of initiation factors IF1, IF2, and IF3 to ribosomal subunit 30S appears to occur rapidly and in a random fashion (as reviewed by Gualerzi and Pon (1990) and displayed in Figure 6.4). The net reaction for initiation factor binding to the 30S ribosomal subunit is given by:

$$30\mathrm{S} + \mathrm{IF1} + \mathrm{IF2}\cdot\mathrm{GTP} + \mathrm{IF3} \;\rightleftharpoons\; \underbrace{30\mathrm{S}\cdot\mathrm{IF1}\cdot\mathrm{IF2}\cdot\mathrm{GTP}\cdot\mathrm{IF3}}_{30\mathrm{S}\cdot\mathrm{IF}\cdot\mathrm{GTP}} \tag{6.2}$$

Complex $30\mathrm{S}\cdot\mathrm{IF1}\cdot\mathrm{IF2}\cdot\mathrm{GTP}\cdot\mathrm{IF3}$ used in equation (6.2) is furtheron denoted by $30\mathrm{S}\cdot\mathrm{IF}\cdot\mathrm{GTP}$ as well. The effective formation of $30\mathrm{S}\cdot\mathrm{IF}\cdot\mathrm{GTP}$ is crucial for the subsequent reaction steps of overall translation initiation. Although translation initiation may still proceed in the absence of several or all initiation factors, the rate of translation initiation is markedly enhanced only at sufficient levels of all three initiation factors (Wintermeyer and Gualerzi, 1983; Pon and Gualerzi, 1984; Canonaco et al., 1986; Gualerzi and Pon, 1990).

Conservation laws. Recalling the above implications, an estimation of the various ribosomal complexes occurring during initiation site selection can be obtained from mass balancing and using the corresponding association constants. The conservation relations for ribosomes and initiation factors are then obtained to the following:

$$\begin{aligned} C_{30\mathrm{S},t} \;=\; & C_{30\mathrm{S}} + C_{30\mathrm{S}\cdot\mathrm{IF1}} + C_{30\mathrm{S}\cdot\mathrm{IF2}\cdot\mathrm{GTP}} + C_{30\mathrm{S}\cdot\mathrm{IF3}} + C_{30\mathrm{S}\cdot\mathrm{IF1}\cdot\mathrm{IF2}\cdot\mathrm{GTP}} \\ & + C_{30\mathrm{S}\cdot\mathrm{IF1}\cdot\mathrm{IF3}} + C_{30\mathrm{S}\cdot\mathrm{IF2}\cdot\mathrm{GTP}\cdot\mathrm{IF3}} + C_{30\mathrm{S}\cdot\mathrm{IF}} + C_{70\mathrm{S}} + \sum_{j\,=\,j_{R0}}^{K} C_{70\mathrm{S},j} \end{aligned} \tag{6.3}$$

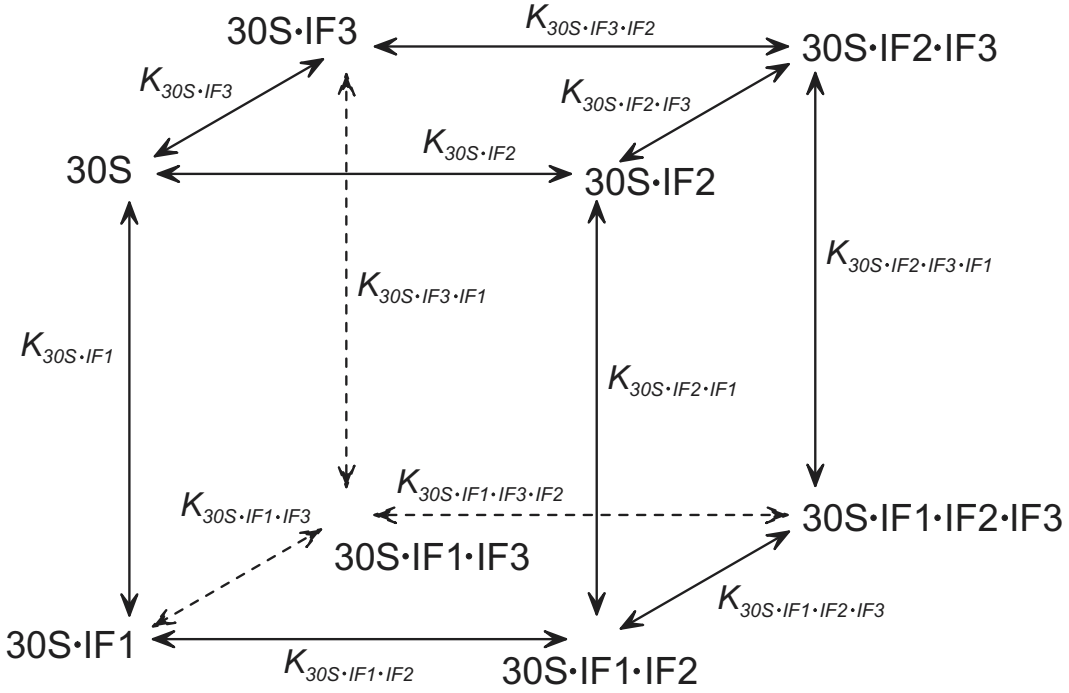

Figure 6.4: Random order of binding of IF1, IF2, and IF3 to 30S. For reasons of clarity, the preferred appearance of freely dissolved IF2 in a complexed form with GTP is omitted in this representation.

$$C_{50S,t} = C_{50S} + C_{70S} + \sum_{j=j_{R0}}^{K} C_{70S,j} \tag{6.4}$$

$$C_{IF1,t} = C_{IF1} + C_{30S\cdot IF1} + C_{30S\cdot IF1\cdot IF2\cdot GTP} + C_{30S\cdot IF1\cdot IF3} + C_{30S\cdot IF} \tag{6.5}$$

$$C_{IF2,t} = C_{IF2\cdot GTP} + C_{30S\cdot IF2\cdot GTP} + C_{30S\cdot IF1\cdot IF2\cdot GTP} + C_{30S\cdot IF2\cdot GTP\cdot IF3} + C_{30S\cdot IF} \tag{6.6}$$

$$C_{IF3,t} = C_{IF3} + C_{30S\cdot IF3} + C_{30S\cdot IF1\cdot IF3} + C_{30S\cdot IF2\cdot GTP\cdot IF3} + C_{30S\cdot IF} \tag{6.7}$$

The summation term used in equations (6.3) and (6.4) denotes the sum of ribosomes bound to mRNA (with K = number of base tripletts within the coding region). Total concentrations of 30S and 50S ribosomal subunits are believed to exist in equal stoichiometric amounts in the reaction system. Total concentrations of initiation factors can be obtained from their relative proportion to ribosome concentration (data given in Table 6.2). Initiation factor binding to 50S and 70S ribosomal subunits has been neglected owing to the reported low binding affinities (Pon et al., 1985; Gualerzi and Pon, 1990). Substituting the association constants from Table 6.4 into equations (6.3) to (6.7) leads to a set of nonlinear algebraic equations, which was then solved iteratively for the concentrations of uncomplexed species using OptdesX (Version 2.0.4, Design Synthesis, Inc.: Simulated annealing algorithm) and minimizing the sum of squared relative errors. This procedure was applied for computation of initial conditions to be used in dynamic simulation of protein production.

Table 6.4: Association constants used for the computation of levels of ribosomal complexes bound to initiation factors. Constants involving more than one initiation factor were derived using: $1.1{\cdot}10^8$ M^{-1} for the binding of IF1 to 30S in the presence of IF2 (Zucker and Hershey, 1986), $3.6{\cdot}10^7$ M^{-1} for the binding of IF1 to 30S incubated with IF3 (Zucker and Hershey, 1986), $1.2{\cdot}10^8$ M^{-1} for the binding of IF3 to 30S, when IF1 and IF2 were present (Chaires et al., 1981), $1.8{\cdot}10^8$ M^{-1} and $1.0{\cdot}10^8$ M^{-1} for the binding of IF2 and IF3, respectively, to 30S in the presence of both other initiation factors (Gualerzi and Pon, 1990).

K_{70S}	$=$	$\frac{C_{70S}}{C_{30S}\,C_{50S}}$	$= 5.3{\cdot}10^7$ M^{-1}	Zucker and Hershey (1986)
$K_{30S{\cdot}IF1}$	$=$	$\frac{C_{30S{\cdot}IF1}}{C_{30S}\,C_{IF1}}$	$= 5.0{\cdot}10^5$ M^{-1}	dto.
$K_{30S{\cdot}IF2}$	$=$	$\frac{C_{30S{\cdot}IF2{\cdot}GTP}}{C_{30S}\,C_{IF2{\cdot}GTP}}$	$= 2.7{\cdot}10^7$ M^{-1}	Weiel and Hershey (1982)
$K_{30S{\cdot}IF3}$	$=$	$\frac{C_{30S{\cdot}IF3}}{C_{30S}\,C_{IF3}}$	$= 3.1{\cdot}10^7$ M^{-1}	dto.
$K_{30S{\cdot}IF1{\cdot}IF2{\cdot}GTP}$	$=$	$\frac{C_{30S{\cdot}IF1{\cdot}IF2{\cdot}GTP}}{C_{30S}\,C_{IF1}\,C_{IF2{\cdot}GTP}}$	$= 4.3{\cdot}10^{14}$ M^{-2}	This study
$K_{30S{\cdot}IF1{\cdot}IF3}$	$=$	$\frac{C_{30S{\cdot}IF1{\cdot}IF3}}{C_{30S}\,C_{IF1}\,C_{IF3}}$	$= 5.6{\cdot}10^{14}$ M^{-2}	dto.
$K_{30S{\cdot}IF2{\cdot}GTP{\cdot}IF3}$	$=$	$\frac{C_{30S{\cdot}IF2{\cdot}GTP{\cdot}IF3}}{C_{30S}\,C_{IF2{\cdot}GTP}\,C_{IF3}}$	$= 8.4{\cdot}10^{14}$ M^{-2}	dto.
$K_{30S{\cdot}IF}$	$=$	$\frac{C_{30S{\cdot}IF1{\cdot}IF2{\cdot}GTP{\cdot}IF3}}{C_{30S}\,C_{IF1}\,C_{IF2{\cdot}GTP}\,C_{IF3}}$	$= 3.7{\cdot}10^{23}$ M^{-3}	dto.

6.2.2.3 70S initiation complex formation

The net reaction of 70S initiation complex formation is given by

$$\begin{aligned} 30S{\cdot}IF{\cdot}GTP + fMet\text{-}tRNA_f^M + RBS + 50S \longrightarrow \\ \underbrace{70S{\cdot}IF2{\cdot}GTP{\cdot}fMet\text{-}tRNA_f^M{\cdot}RBS}_{70S\text{-}IC} + IF1 + IF3 \end{aligned} \tag{6.8}$$

Reaction equation (6.8) comprises a multi-step mechanism, which was assumed to obey the following scheme:

$$30S{\cdot}IF{\cdot}GTP + fMet\text{-}tRNA_f^M + RBS \overset{K_A,K_B}{\rightleftharpoons} [30S{\cdot}IF{\cdot}GTP{\cdot}fMet\text{-}tRNA_f^M{\cdot}RBS] \tag{6.9}$$

$$[30S{\cdot}IF{\cdot}GTP{\cdot}fMet\text{-}tRNA_f^M{\cdot}RBS] \overset{k_{TLI,70SIC,1}}{\longrightarrow} \underbrace{30S{\cdot}IF{\cdot}GTP{\cdot}fMet\text{-}tRNA_f^M{\cdot}RBS}_{30S\text{-}IC} \tag{6.10}$$

$$30S\text{-}IC + 50S \overset{k_{TLI,70SIC,2}}{\longrightarrow} 70S\text{-}IC + IF1 + IF3 \tag{6.11}$$

A preinitiation complex is formed through the association of the ribosomal 30S subunit with initiator tRNA and the ribosome binding site (denoted by square brackets in equation (6.9)). Binding of fMet-$tRNA_f^M$ and the RBS, respectively, were assumed to be reversible and to take

place randomly. A simplification inherently made is to consider the binding of either ligand to be unaffected by the binding of the other substrate. A slow rearrangment of this complex leads to the 30S initiation complex, here termed 30S-IC (equation (6.10)). The rate constant $k_{TLI,70\mathrm{SIC},1}$ was reported to be 0.1 s^{-1} (Wintermeyer and Gualerzi, 1983).

Association of a 50S subparticle to the 30S initiation complex (equation (6.11)) leads to the formation of the 70S initiation complex (70S-IC). During this reaction step, the positioning of fMet-tRNA$_f^{\mathrm{M}}$ in the ribosomal P-site takes place together with a concomittant liberation of IF1 and IF3. Rate constant $k_{TLI,70\mathrm{SIC},2} = 8.4{\cdot}10^6$ $\mathrm{M}^{-1}\mathrm{s}^{-1}$ was taken from Blumberg et al. (1979). From equations (6.9) to (6.11), the following rate expression was derived (see Appendix A.2):

$$V_{TLI,70\mathrm{SIC}} = \frac{q_{jR0}^{R0}\; V_{TLI,70\mathrm{SIC}}^{max}}{1 + \dfrac{K_{M,\mathrm{fMet\text{-}tRNA}_f^{\mathrm{M}}}}{C_{\mathrm{fMet\text{-}tRNA}_f^{\mathrm{M}}}} + \dfrac{K_{M,\mathrm{RBS}}}{C_{\mathrm{RBS}}} + \dfrac{K_{M,50\mathrm{S}}}{C_{50\mathrm{S}}} + \dfrac{K_{M,\mathrm{fMet\text{-}tRNA}_f^{\mathrm{M}}} K_{\mathrm{RBS}}}{C_{\mathrm{fMet\text{-}tRNA}_f^{\mathrm{M}}}\, C_{\mathrm{RBS}}}} \tag{6.12}$$

Parameter q_{jR0}^{R0} denotes the probability of the RBS to be unoccupied. It was explained earlier in section 4.2.3.2. Other model parameters exhibit the following mathematical dependence on the rate constants and association constants of the elementary reactions:

$$V_{TLI,70\mathrm{SIC}}^{max} = k_{TLI,70\mathrm{SIC},1}\; C_{30\mathrm{S}\cdot\mathrm{IF}} \tag{6.13}$$

$$K_{M,\mathrm{fMet\text{-}tRNA}_f^{\mathrm{M}}} = K_{\mathrm{fMet\text{-}tRNA}_f^{\mathrm{M}}} \tag{6.14}$$

$$K_{M,\mathrm{RBS}} = K_{\mathrm{RBS}} \tag{6.15}$$

$$K_{M,50\mathrm{S}} = \frac{k_{TLI,70\mathrm{SIC},1}}{k_{TLI,70\mathrm{SIC},2}} \tag{6.16}$$

The affinity constants for initiator tRNA ($K_{M,\mathrm{fMet\text{-}tRNA}_f^{\mathrm{M}}}$) and mRNA ($K_{M,\mathrm{RBS}}$) were reported to be 0.05 μM and 0.009 μM, respectively (Gualerzi et al., 1977; Pon et al., 1985). $K_{M,50\mathrm{S}} = 12$ nM was calculated using the rate constants cited above. Throughout this study, the concentration of ribosome binding site (C_{RBS}) was taken to be equal to the concentration of the initiation codon (C_{jR0}^{M}).

6.2.2.4 IF2-dependent GTP hydrolysis

The ejection of IF2 from the 70S initiation complex is accompanied by GTP hydrolysis due to

$$\text{70S-IC} \xrightarrow{k_{TLI,\mathrm{IF2D}}} 70\mathrm{S}\cdot\mathrm{fMet\text{-}tRNA}_f^{\mathrm{M}}\cdot\mathrm{RBS} + \mathrm{IF2} + \mathrm{GDP} + \mathrm{P}_i \tag{6.17}$$

This reaction was considered to follow first-order kinetics according to

$$V_{TLI,\mathrm{IF2D}} = k_{TLI,\mathrm{IF2D}}\; C_{70\mathrm{SIC}} \tag{6.18}$$

The rate constants for IF2-dependent GTP hydrolysis and the release of inorganic phosphate were provided to be 30 s^{-1} and 1.5 s^{-1}, respectively (Tomšic et al., 2000). In the assumed mechanism, both reaction steps were combined into one step using a rate constant of 1.5 s^{-1}, in order to account for the slower of the both reaction steps.

6.3 Elongation

During translation elongation, amino acids are successively added to a growing peptide chain, where the sequence of addition is dictated by the mRNA. Under physiological conditions, chain elongation proceeds at a rate of 10 to 20 aa/s (cf. Table 6.2). However, the rate of elongation may be found to vary greatly along the mRNA (Talkad et al., 1976; Chaney and Morris, 1979; Pedersen, 1984; Varenne et al., 1984). Elongation rate is kinetically influenced by (a) substrate availability (e.g., abundance of amino acids and tRNA (Ikemura, 1981a; 1981b)), modulated by (b) codon usage (Jakubowski, 1988; Solomovici et al., 2000) and the strength of codon-anticodon interaction (Sørensen and Pedersen, 1991), affected by (c) sterical hindrance between ribosomes travelling further downstream (Dahlberg et al., 1973), and additionally regulated by (d) mRNA secondary structure (de Smit and van Duin, 1994a). Furthermore, elongation factors catalyzing various steps of translation elongation are critically needed for maintaining high elongation rates. In the absence of elongation factors, the rate of protein synthesis is reduced by up to a factor of 10^4 (Nierhaus, 1996). Elongation factors are also suggested to determine the effective direction of the elongation reaction, so that protein synthesis proceeds as an irreversible polymerization process (Nierhaus, 1996). The elongation cycle can be divided into the three phases of (1) binding of aminoacyl-tRNA, (2) transpeptidation, and (3) translocation. These steps are shown in Figure 6.5.

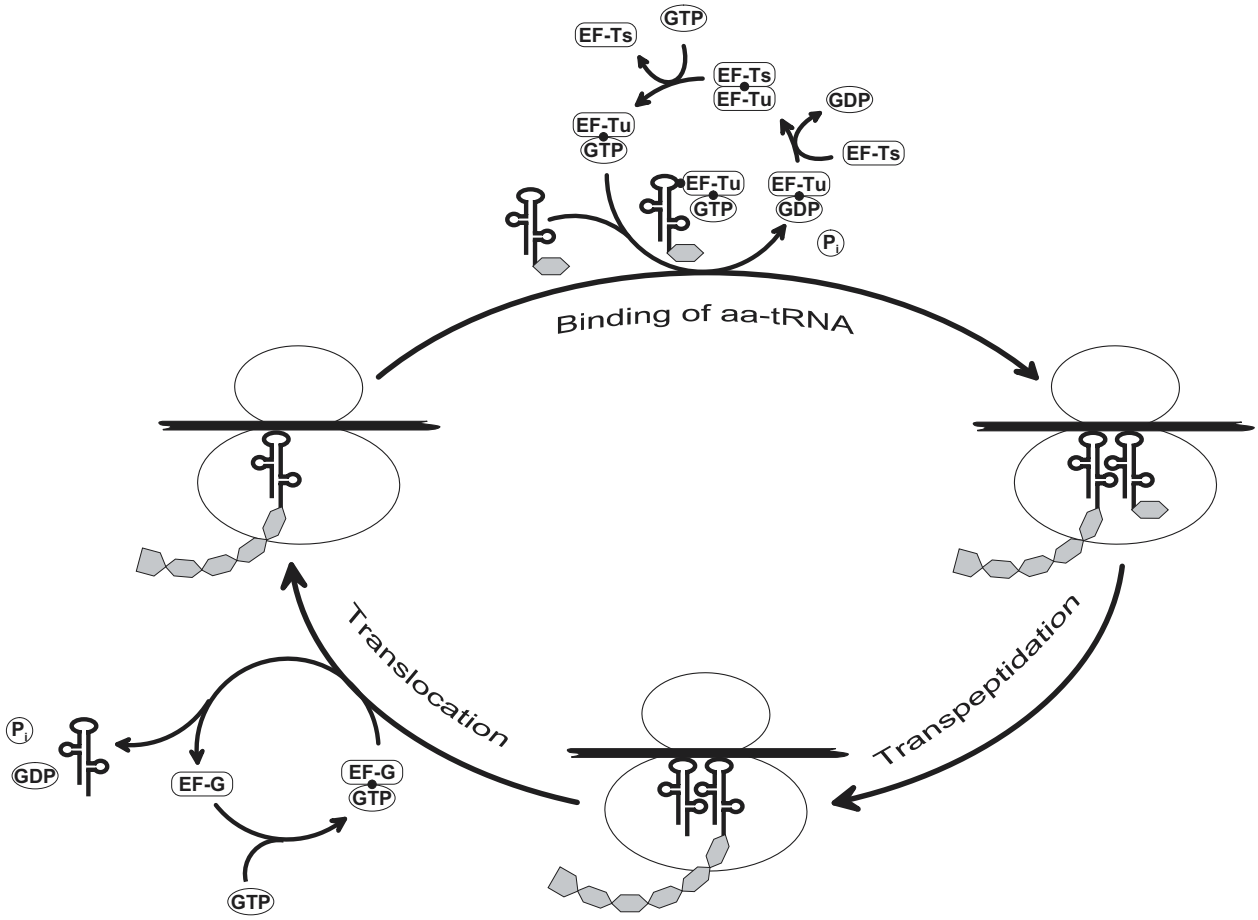

Figure 6.5: Principle reaction scheme of bacterial translation elongation.

Table 6.5: Adjacent binding sites of tRNA on ribosomes.

Site	Description
A-site	(aminoacyl-tRNA); recognition site for selective binding of ternary complex; here, decoding takes place
P-site	(peptidyl-tRNA, carrying the nascent polypeptide chain at its C-terminal end); this site always exhibits a high affinity for tRNA; initiator tRNA binds to the P-site
E-site	(exit); specific for deacylated tRNA prior to its ejection

The ribosome possesses three sites for tRNA binding (see Table 6.5). The P-site exhibits a high affinity for tRNA binding throughout elongation, in order to hinder the dissociation of unfinished product protein. Always two tRNAs are bound to the ribosome in an elongation cycle. At all stages of elongation are elongation factors catalytically enrolled. The specific functions of selected elongation factors are briefly outlined in Table 6.6 and are in the following explained in more detail.

Table 6.6: Role of prokaryotic elongation factors. The specific mechanistic role of elongation factors EFP and SELB is not further considered in this study.

Factor	Properties and function
EFTu	• propels selection and binding of cognate aa-tRNA to the ribosomal A-site • catalyzes the transition from the post-translocational state to the pre-translocational state • exhibits GTPase activity • involved in mechanisms determining translational accuracy (tRNA selection and proofreading)
EFTs	• accelerates GDP dissociation from EFTu • regeneration of EFTu through exchange of bound GDP with GTP
EFG	• positioning of peptidyl-tRNA in the P-site • catalyzes the ribosomal transition from the pre-translocational state to the post-translocational state • exhibits GTPase activity
EFP	• stimulates efficient peptide bond formation
SELB	• selenocysteinyl-tRNA-incorporating variant of EFTu

The important physiological role of **elongation factor Tu (EFTu)** is stressed by the fact that this protein is encoded by two highly expressed unlinked genes. This factor ranks among the most abundant gene products in *E. coli* (Bremer And Dennis, 1996). EFTu binds with GTP and aa-tRNA, generally to form a ternary complex, which functions as a substrate for translation elongation. EFTu concentration corresponds *in vivo* approximately with the concentration of total cellular aa-tRNA, taking into account that a percentage of 83 % of total tRNA were estimated to be acylated in *E. coli* (Gouy and Grantham (1980), and cf. Table 6.2).

EFTu appears to be highly unstable in an uncomplexed state. A half-life time of 5 min at

30°C was reported by Arai et al. (1974). Its stability is significantly increased in complexation with GDP, GTP, EFTs (see below), or within a ternary complex bound with both GTP and aa-tRNA. *In vitro*, EFTu can occur in a phosphorylated state, in which it does apparently not interact with aa-tRNA (Alexander et al., 1995).

Since EFTu·GDP is generated in each elongation cycle and owing to the slow dissociation rate of GDP from EFTu, an additional catalyst, **elongation factor Ts (EFTs)**, is required for an efficient regeneration of EFTu. EFTs binds to EFTu·GDP and causes the exchange of GDP against GTP.

The primary role of **elongation factor G** is to modulate the affinity of the ribosome binding sites A and E towards tRNA. EFG catalyzes the transition from the pre-translocational ribosomal state to the post-translocational state. Binding of EFG and EFTu are mutually exclusive.

6.3.1 Previous modelling

Individual features of the elongation reaction have been the subject of extensive mathematical modelling. Earlier studies were concerned with the determination of thermodynamic parameters for binding of EFTu and EFG with guanylates (Baca et al., 1976; Pingoud et al., 1977; Chau et al., 1981; Romero et al., 1985). The kinetics of GTP hydrolysis by EFG bound to ribosomes was studied (Rohrbach and Bodley, 1976). The formation rate of EFTu·GTP at EFTu regeneration was modelled kinetically and used for parameter estimation of substrate affinities (Hwang and Miller, 1985). The tRNA cycle was modelled in a probabilistic approach assigning mean duration times for various reaction steps (Gouy and Grantham, 1980).

Intricate kinetic models for tRNA charging have been developed to account for a functional dependency on Mg^{2+} ion concentration (Airas, 1990) and the inhibitory influence of byproduct inorganic pyrophosphate (Airas, 1992). Due to the large number of parameters associated with these rate expressions, such an approach of extreme detail is impractical from a standpoint of comprehensive gene expression modelling. Further, not all of the constants considered in such a detailed model may add sensitively to a description of protein synthesis behaviour.

In modelling ternary complex formation between EFTu, GTP and aa-tRNA, a few studies have distinguished among the affinities of EFTu for the different amino acids. A negative correlation of the abundance of aa-tRNA families and their affinities for EFTu·GTP was determined (Jakubowski, 1988). The same author indicated a tendency for those aa-tRNAs existing at lower concentrations to bind more firmly with EFTu·GTP, and thus to raise the effective concentration of the corresponding ternary complex. Dissociation constants for the reaction of ternary complex formation were provided for 18 different aa-tRNAs from *E. coli* (Ott et al., 1990). The ranking of affinities given by these authors is essentially compliant with the one tabulated by Jakubowski (1988). However, a full set of affinities covering all known bacterial tRNA species could not be found while scanning the literature.

Pavlov and Ehrenberg (1996) expressed the overall rate constant of elongation in terms of the total concentrations of EFTu and EFG. Their approach may be a valid approximation under homeostatic growth conditions, where roughly the same stoichiometries are maintained

for translation factors relative to ribosome concentration (cf. Table 6.2). These authors did, however, not give a dependency on the concentrations of substrate complexes actually required for translation elongation (e.g., ternary complex, and EFG complexed with GTP). Thus, their model is unable to describe dynamically changing growth conditions, where a functional connection between elongation rate and the corresponding substrate concentrations is essential.

A reaction scheme of the entire elongation cycle has been proposed containing the regeneration of EFTu and EFG (Gast, 1987; Pingoud et al., 1990). This model was then used to investigate the process of error correction during steady state protein synthesis. Various ordered and random steady-state kinetic mechanisms were analyzed theoretically for both factorless and factor-dependent translation elongation (Saifullin and Potapov, 1995a, 1995b). It was suggested by these authors to regard elongation rate as a reversible reaction, the favoured direction of which is dependent on substrate and product concentrations. These authors argue also that the reverse reaction is surpressed when energy levels are sufficiently high such that nucleotides exist mostly in their 3-fold phosphorylated state.

In a statistical model, a matrix of translational efficiencies was derived (Bagnoli and Liò, 1995)). The matrix elements denote efficiencies with which each aa-tRNA anticodon pairs with a codon. In the same context, Solomovici et al. (1997) computed elongation rates of synonymous codons given the hypothesis of an optimized (i.e., most economical) translation process.

Very detailed kinetic studies using stopped-flow techniques investigated elongation kinetics and identified rate constants for various steps of ligand association and catalytic isomerization (Rodnina et al., 1996; Pape et al., 1998). This model was later used for describing erroneous protein synthesis and translational accuracy (Pape et al., 1999).

6.3.2 Reaction scheme and kinetics

The subsequent model of translation elongation accounts for the processes of (1) ternary complex formation, (2) translation elongation, (3) EFTu regeneration, and (4) EFG regeneration.

6.3.2.1 Ternary complex formation

EFTu needs to combine with GTP, before it can synthesize to ternary complex EFTu·GTP·aa-tRNA$_j$ (furtheron denoted also by symbol T3$_j$). The index j denotes any of the tRNA species. Free EFTu can bind with either GTP or GDP, according to

$$\text{EFTu} + \text{GTP} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{EFTu·GTP} \tag{6.19}$$

$$\text{EFTu} + \text{GDP} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \text{EFTu·GDP} \tag{6.20}$$

The respective binding constant together with the rate constants for the elementary steps of association and dissiociation were given by Romero et al. (1985) for both GTP ($8.0{\cdot}10^6$ M^{-1}, $2.0{\cdot}10^5$ M^{-1}s^{-1} and $2.5{\cdot}10^{-2}$ s^{-1}) and for GDP ($5.3{\cdot}10^8$ M^{-1}, $9.0{\cdot}10^5$ M^{-1}s^{-1} and $1.7{\cdot}10^{-3}$ s^{-1}), respectively. GTP induces a strong affinity of EFTu towards aa-tRNA$_j$, leading to ternary

complex formation according to

$$\text{EFTu·GTP} + \text{aa-tRNA}_j \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \text{T3}_j \tag{6.21}$$

In contrast to the general perception of ternary complex formation, the existence of a pentameric complex, (EF-Tu·GTP)$_2$·aa-tRNA, was claimed (Bilgin and Ehrenberg, 1995). However, more recent studies suggest the formation of ternary complexes to be the common mechanism. A quinary complex seems to form only under certain conditions (Rodnina and Wintermeyer, 1995), like for example, when using poly(U)-mRNA or heteropolymeric mRNA containing contiguous phenylalanine codons (UUU).

The rate of ternary complex formation was derived for the forward and reverse reaction according to second-order kinetics on the basis of general collision theory (Romero et al., 1985)

$$V_{\text{T3,Form},j} \;=\; k_{\text{T3,Form},j}\, C_{\text{EFTu·GTP}}\, C_{\text{aa-tRNA},j} - k_{\text{-T3,Form},j}\, C_{\text{T3},j} \tag{6.22}$$

Rate constants for association and dissociation used in equation (6.22) may be discriminated against the type of aa-tRNA species. However, due to lack of information, they were taken in this study to be the same for each sort of aa-tRNA. The values applied were $k_{\text{T3,Form}} = 5.0{\cdot}10^7$ M^{-1}s^{-1} and $k_{\text{-T3,Form}} = 1$ s^{-1}, respectively, which had been determined earlier for Trp-tRNA (Pingoud et al., 1977; Pingoud and Urbanke, 1980; Gast, 1987)). The binding of complex EFTu·GDP to aa-tRNA was found to be weak in comparison to EFTu·GTP binding, with an association constant of at least two orders of magnitude below the association constant for EFTu·GTP binding (Romero et al., 1985). Hence, EFTu·GDP binding to aa-tRNA was omitted here.

6.3.2.2 Translation elongation

In an elongation cycle, the nascent peptide chain is prolonged by one amino acid requiring an energy expenditure of two phosphate groups being released. During this reaction, the ribosome propagates from codon j to codon $j+1$ along the mRNA, while the deacylated tRNA of the previous elongation cycle is being liberated. The net reaction for the elongation cycle is:

$$\begin{aligned} 70\text{S}_j + \text{EFTu·GTP·aa-tRNA}_{j+1} + \text{EFG·GTP} &\overset{k_{TLE,j}}{\longrightarrow} \\ 70\text{S}_{j+1} + \text{EFTu·GDP} + \text{EFG·GDP} + 2\ \text{P}_i + \text{tRNA}_j & \end{aligned} \tag{6.23}$$

The cycle is catalyzed by the bound ribosome. Complexes involving translation factors EFTu and EFG are treated as substrates and products of the overall reaction. For the derivation of a kinetic rate expression describing translation elongation, the entire cycle can be divided into

the following reaction steps (Gast, 1987; Pingoud et al., 1990):

$$\left(\mathrm{P}_j\text{-}\mathrm{T}^{\mathrm{P}}_j\right)\cdot 70\mathrm{S}_j + \text{aa-}\mathrm{T}_{j+1}\cdot\mathrm{EFTu}\cdot\mathrm{GTP} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \left(\mathrm{P}_j\text{-}\mathrm{T}^{\mathrm{P}}_j\right)\cdot 70\mathrm{S}_j\cdot\text{aa-}\mathrm{T}^{\mathrm{A}}_{j+1}\cdot\mathrm{EFTu}\cdot\mathrm{GTP} \tag{6.24}$$

$$\left(\mathrm{P}_j\text{-}\mathrm{T}^{\mathrm{P}}_j\right)\cdot 70\mathrm{S}_j\cdot\text{aa-}\mathrm{T}^{\mathrm{A}}_{j+1}\cdot\mathrm{EFTu}\cdot\mathrm{GTP} \xrightarrow{k_2} \left(\mathrm{P}_j\text{-}\mathrm{T}^{\mathrm{P}}_j\right)\cdot 70\mathrm{S}_j\cdot\text{aa-}\mathrm{T}^{\mathrm{A}}_{j+1}\cdot\mathrm{EFTu}\cdot\mathrm{GDP} + \mathrm{P}_i \tag{6.25}$$

$$\left(\mathrm{P}_j\text{-}\mathrm{T}^{\mathrm{P}}_j\right)\cdot 70\mathrm{S}_j\cdot\text{aa-}\mathrm{T}^{\mathrm{A}}_{j+1}\cdot\mathrm{EFTu}\cdot\mathrm{GDP} \xrightarrow{k_3} \left(\mathrm{P}_j\text{-}\mathrm{T}^{\mathrm{P}}_j\right)\cdot 70\mathrm{S}_j\cdot\text{aa-}\mathrm{T}^{\mathrm{A}}_{j+1} + \mathrm{EFTu}\cdot\mathrm{GDP} \tag{6.26}$$

$$\left(\mathrm{P}_j\text{-}\mathrm{T}^{\mathrm{P}}_j\right)\cdot 70\mathrm{S}_j\cdot\text{aa-}\mathrm{T}^{\mathrm{A}}_{j+1} \xrightarrow{k_4} \mathrm{T}^{\mathrm{P}}_j\cdot 70\mathrm{S}_j\cdot\left(\mathrm{P}_{j+1}\text{-}\mathrm{T}^{\mathrm{A}}_{j+1}\right) \tag{6.27}$$

$$\mathrm{T}^{\mathrm{P}}_j\cdot 70\mathrm{S}_j\cdot\left(\mathrm{P}_{j+1}\text{-}\mathrm{T}^{\mathrm{A}}_{j+1}\right) + \mathrm{EFG}\cdot\mathrm{GTP} \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} \mathrm{T}^{\mathrm{P}}_j\cdot 70\mathrm{S}_j\cdot\left(\mathrm{P}_{j+1}\text{-}\mathrm{T}^{\mathrm{A}}_{j+1}\right)\cdot\mathrm{EFG}\cdot\mathrm{GTP} \tag{6.28}$$

$$\mathrm{T}^{\mathrm{P}}_j\cdot 70\mathrm{S}_j\cdot\left(\mathrm{P}_{j+1}\text{-}\mathrm{T}^{\mathrm{A}}_{j+1}\right)\cdot\mathrm{EFG}\cdot\mathrm{GTP} \xrightarrow{k_6} \left(\mathrm{P}_{j+1}\text{-}\mathrm{T}^{\mathrm{P}}_{j+1}\right)\cdot 70\mathrm{S}_{j+1}\cdot\mathrm{EFG}\cdot\mathrm{GDP} + \mathrm{T}_j + \mathrm{P}_i \tag{6.29}$$

$$\left(\mathrm{P}_{j+1}\text{-}\mathrm{T}^{\mathrm{P}}_{j+1}\right)\cdot 70\mathrm{S}_{j+1}\cdot\mathrm{EFG}\cdot\mathrm{GDP} \xrightarrow{k_7} \left(\mathrm{P}_{j+1}\text{-}\mathrm{T}^{\mathrm{P}}_{j+1}\right)\cdot 70\mathrm{S}_{j+1} + \mathrm{EFG}\cdot\mathrm{GDP} \tag{6.30}$$

Symbol $70\mathrm{S}_j$ used in the above equations denotes a ribosome which carries a peptide of j amino acids (P_j) that is attached to the tRNA in the ribosomal P-site ($\mathrm{T}^{\mathrm{P}}_j$). The association of ternary complex (aa-T_{j+1}·EFTu·GTP) takes place to a vacant ribosomal A-site (equation (6.24)). The act of ternary complex binding is reversible, which is of vital importance for a correct tRNA selection and for proofreading. In a next step, the ribosome-bound ternary complex undergoes GTP hydrolysis (equation (6.25)). Several conformational changes take place prior to EFTu·GDP release (Pape et al., 1998). These isomerizations are summarized in reaction equation (6.26). Through peptide bond formation, the growing polypeptide is prolonged by one amino acid (equation (6.27)). During this step, the polypeptide chain attached to the tRNA in the P-site is handed over to the aa-tRNA located in the A-site. After this very rapid reaction step, a deacylated tRNA remains in the P-site. Binding of EFG·GTP (equation (6.28)) is required to provide the energy needed for subsequent translocation. During translocation (equation (6.29)), peptidyl-tRNA is transferred back into the P-site with the simultaneous release of the discharged tRNA (symbol T_j). This reaction is accompanied by GTP hydrolysis and by the propagation of the ribosome to the next codon on the mRNA. The dissociation of EFG·GDP (equation (6.30)) completes the elongation cycle.

From the above reaction scheme (equations (6.24) to (6.30)), and additionally considering the fact that codons can be recognized by more than one tRNA anticodon, steady state kinetics for the elongation cycle were derived using symbolic computation (Appendix A.3):

$$V_{TLE,j} = \frac{q^R_j\, V^{max}_{TLE,j}}{1 + \dfrac{K_{M,\mathrm{T3}_j}}{\sum\limits_i C_{\mathrm{T3}_j,i}} + \dfrac{K_{M,\mathrm{EFG\cdot GTP}}}{C_{\mathrm{EFG\cdot GTP}}}} \tag{6.31}$$

Probability q^R_j, for codon $j+1$ to be unoccupied, was introduced in section 4.2.3.2. Other model parameters of equation (6.31) are composed of the rate constants for the elementary reaction steps (equations (6.24) to (6.30)). Substituting the elementary rate constants provided by

Gast (1987), $K_{M,\text{EFG·GTP}}$ takes a value of 0.22 μM (as calculated by Schmid (1999)). Total cellular contents of 44 tRNA species (out of the 46 tRNAs known to exist in *E. coli*) were provided by Dong et al. (1996). Knowing from these authors that the total intracellular tRNA concentrations range from about 1 to 30 μM, $K_{M,\text{T3}_j}$ was selected to be equal to 0.4 μM.

The summation term depicted in equation (6.31) stands for the sum of ternary complexes with tRNA species that carry a correct amino acid corresponding to codon j and that are recognized by this codon. An example for which the summation term comprises more than one element is codon UUG. This base triplett is matched by both $\text{tRNA}^{\text{Ser1}}$ and $\text{tRNA}^{\text{Ser5}}$ (Dong et al., 1996). The rate of translation elongation at codon UUG is thus influenced by the concentrations of the respective ternary complexes corresponding to both of these tRNAs.

The maximum rate of translation elongation (symbol $V_{TLE,j}^{max}$ in equation (6.31)) is denoted by the concentration of ribosomes bound to codon j, and a codon-specific rate constant ($k_{TLE,j}$), according to

$$V_{TLE,j}^{max} \;=\; k_{TLE,j}\, C_j^R \tag{6.32}$$

Codon-specificity may arise, for example, due to different binding strengths of codon-anticodon interaction for different tRNAs. Constant $k_{TLE,j}$ was calculated from

$$k_{TLE,j} \;=\; f_j\, k_{TLE}^{max} \tag{6.33}$$

The efficiency factor, f_j, was adopted from Solomovici et al. (1997), who tabulated values of this parameter for all 61 sense codons. Unless otherwise stated, a maximum rate constant for translation elongation (k_{TLE}^{max}) equal to 24 codons/s was applied throughout this study.

In summary, the kinetic rate expression for translation elongation accounts for (a) individual tRNA abundance of natural types of bacterial tRNA, (b) a codon-specific efficiency of translation elongation, (c) sterical interference among translating ribosomes, and (d) a possibility to consider different affinities ($K_{M,\text{T3}_j}$) for ternary complex selection at codon j.

6.3.2.3 EFTu regeneration

The overall reaction scheme of EFTu regneration by EFTs is given according to:

$$\text{EFTu·GDP} + \text{GTP} \;\overset{\text{EFTs}}{\rightleftharpoons}\; \text{EFTu·GTP} + \text{GDP} \tag{6.34}$$

Considering reversible ping-pong bi-bi kinetics (as suggested by Romero et al. (1985)), the rate equation for the EFTu recycle can be derived to give

$$V_{\text{EFTu-Reg}} \;=\; \frac{V_f \left(C_A\, C_B - \dfrac{C_P\, C_Q}{K_{eq,\text{EFTu}}} \right)}{D} \tag{6.35}$$

with

$$\begin{aligned} D \;=\;& K_{M,B}\, C_A + K_{M,A}\, C_B + \frac{V_f\, K_{M,Q}}{V_r\, K_{eq,\text{EFTu}}}\, C_P + \frac{V_f\, K_{M,P}}{V_r\, K_{eq,\text{EFTu}}}\, C_Q + C_A\, C_B \\ & + \frac{V_f\, K_{M,Q}}{V_r\, K_{eq,\text{EFTu}}\, K_{iA}}\, C_A\, C_P + \frac{V_f}{V_r\, K_{eq,\text{EFTu}}}\, C_P\, C_Q + \frac{K_{M,A}}{K_{iQ}}\, C_B\, C_Q \end{aligned}$$

Kinetic constants of equation (6.35) are listed in Table 6.7. The maximum forward rate is

$$V_f \;=\; k_{\mathrm{EFTs},f}\, C_{\mathrm{EFTs},t} \tag{6.36}$$

Symbol $C_{\mathrm{EFTs},t}$ is the total concentration of EFTs. The maximum rate of the reverse reaction was calculated to be $V_r = k_{\mathrm{EFTs},r}\, C_{\mathrm{EFTs},t}$. Constants $k_{\mathrm{EFTs},f}$ and $k_{\mathrm{EFTs},r}$ were reported to be 30 s^{-1} and 10 s^{-1}, respectively (Ruusala et al., 1982a). The equilibrium constant $K_{eq,\mathrm{EFTu}}$ equalled 0.19 with the rate constants published by Romero et al. (1985).

Table 6.7: Kinetic constants of EFTu regeneration were calculated from the rate constants for the individual reaction steps given by Romero et al. (1985) unless otherwise noted. Other parameter values were taken from [a]Ruusala et al. (1982a) and [b]Hwang and Miller (1985).

	A EFTu·GDP	B GTP	P GDP	Q EFTu·GTP
K_M (μM)	2.5[a]	50	3[b]	1
K_i (μM)	5.6	6.5	15	1

6.3.2.4 EFG regeneration

The regeneration of elongation factor EFG takes place spontaneously according to

$$\mathrm{EFG{\cdot}GDP} \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} \mathrm{EFG} + \mathrm{GDP} \tag{6.37}$$

$$\mathrm{EFG} + \mathrm{GTP} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{EFG{\cdot}GTP} \tag{6.38}$$

Values used for the association and dissociation rate constants of GDP binding were $2.7{\cdot}10^7$ $\mathrm{M}^{-1}\mathrm{s}^{-1}$ and 100 s^{-1}, respectively (Gast, 1987). The rate constants for the forward and reverse reaction of equation (6.38) were reported to be $1.0{\cdot}10^7$ $\mathrm{M}^{-1}\mathrm{s}^{-1}$ and 400 s^{-1}, respectively (Gast, 1987).

6.3.2.5 Mass conservation

Neglecting any uncomplexed EFTu, the total mass balance for elongation factors and involved guanylates can be represented by

$$C_{\mathrm{EFTu},t} \;=\; C_{\mathrm{EFTu{\cdot}GTP}} + C_{\mathrm{EFTu{\cdot}GDP}} + \sum_{j=1}^{A} C_{\mathrm{T3},j} \tag{6.39}$$

$$C_{\mathrm{EFG},t} \;=\; C_{\mathrm{EFG}} + C_{\mathrm{EFG{\cdot}GTP}} + C_{\mathrm{EFG{\cdot}GDP}} \tag{6.40}$$

$$C_{\mathrm{GTP},t} \;=\; C_{\mathrm{GTP}} + C_{\mathrm{EFTu{\cdot}GTP}} + C_{\mathrm{EFG{\cdot}GTP}} \tag{6.41}$$

$$C_{\mathrm{GDP},t} \;=\; C_{\mathrm{GDP}} + C_{\mathrm{EFTu{\cdot}GDP}} + C_{\mathrm{EFG{\cdot}GDP}} \tag{6.42}$$

A is the number of different types of amino acids (usually 20). Elongation factor EFTs was regarded to function as a pure catalyst, whose concentration of uncomplexed EFTs is at any

instant in time taken to be approximately given by the total concentration of this factor. Considering the binding constants provided in Table 6.8, the set of equations (6.39) to (6.42) was solved to yield the respective equilibrium concentrations of uncomplexed components together with their complexed counterparts. The outcome of this analysis served as initial conditions for dynamic simulation.

Table 6.8: Binding constants for computation of concentrations of uncomplexed elongation factors, GTP, and GDP.

$K_{\mathrm{EFG\cdot GTP}}$	=	$\frac{C_{\mathrm{EFG}}\, C_{\mathrm{GTP}}}{C_{\mathrm{EFG\cdot GTP}}}$	= $2.5{\cdot}10^4$ M^{-1}	Gast (1987)
$K_{\mathrm{EFG\cdot GDP}}$	=	$\frac{C_{\mathrm{EFG}}\, C_{\mathrm{GDP}}}{C_{\mathrm{EFG\cdot GDP}}}$	= $2.7{\cdot}10^5$ M^{-1}	dto.
$K_{eq,\mathrm{EFTu}}$	=	$\frac{C_{\mathrm{EFTu\cdot GTP}}\, C_{\mathrm{GDP}}}{C_{\mathrm{EFTu\cdot GDP}}\, C_{\mathrm{GTP}}}$	= 0.19	Romero et al. (1985)

6.4 Termination

6.4.1 Reaction scheme

Translational termination is the reaction step which leads to the disruption of the translation complex and to the release of a fully synthesized protein molecule (see Figure 6.6). The overall reaction stoichiometry is given by

$$70\mathrm{S}_K + \mathrm{GTP} + \mathrm{H_2O} \xrightarrow{k_{TLT}} 70\mathrm{S} + \mathrm{mRNA} + \mathrm{Protein} + \mathrm{tRNA}_K + \mathrm{GDP} + \mathrm{P}_i \tag{6.43}$$

This reaction occurs at specific termination sites, which are signalled by the nonsense codons UAA, UAG, and UGA. Release factors 1 (RF1) and 2 (RF2) assist in recognizing these termination sites. RF1 is enrolled in termination at stop codons UAG and UAA, RF2 causes termination at codons UGA and UAA. Highly expressed genes were found to use mostly UAA as the stop codon (Brown et al., 1990), which is recognized by both RF1 and RF2. The only role of RF3 is to stimulate RF1 or RF2 dissociation from the ribosome subsequent to the hydrolysis of peptidyl-tRNA and thus to enhance ribosome recycling (Pavlov et al., 1997b). In addition to these three release factors, also factors RRF and RFH have been identified to be important for efficient translation termination. The latter two factors are, however, disregarded in this study, due to the limited information about their mechanistic involvement. Table 6.9 summarizes the functional role of known bacterial release factors.

Termination efficiency is further affected by sequence context of the stop codons (Pavlov et al., 1998). Translational read-through may occur in rare cases (Neidhardt et al. (1996)). The necessary conditions for this phenomenon are, however, unclear.

6.4.2 Reaction kinetics

Allowing for a random order of substrate binding according to Figure 6.6, and taking the reactions of substrate association to be rapid, the kinetic rate equation for translational termination

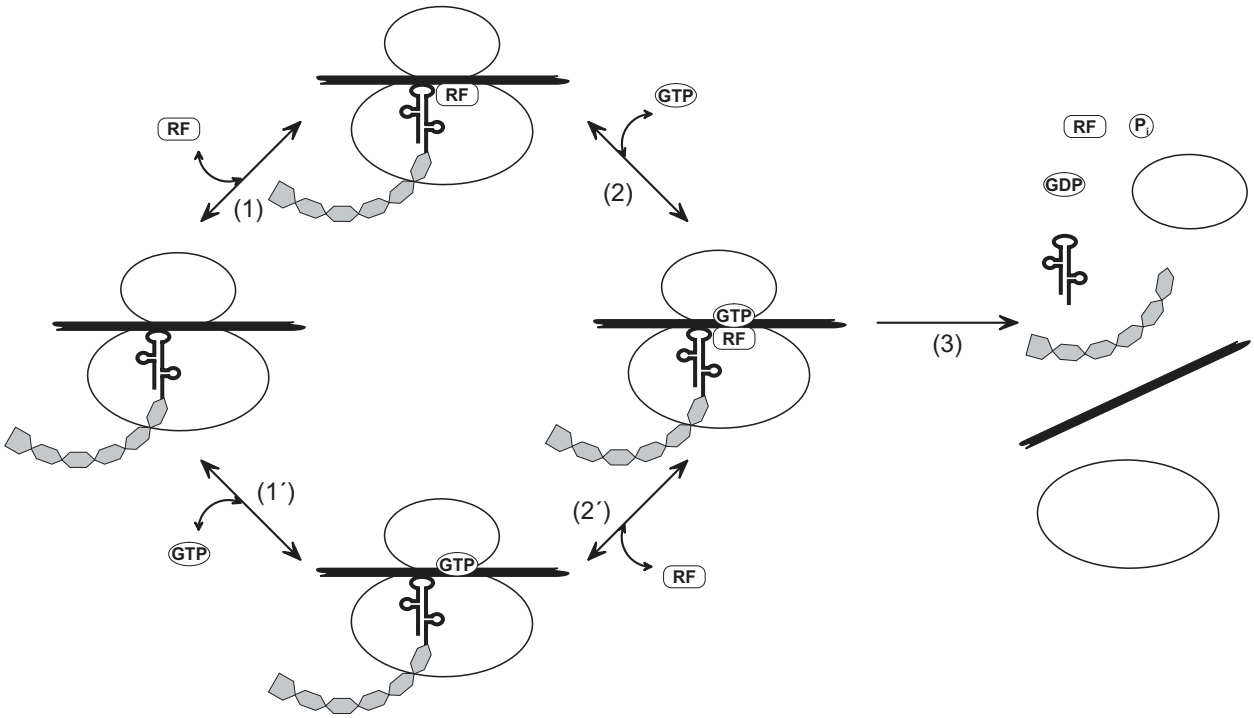

Figure 6.6: Principle reaction scheme of bacterial translation termination.

can be derived as follows

$$V_{TLT} = \frac{V_{TLT}^{max}}{1 + \frac{K_{M,R_K}}{C_K^R} + \frac{K_{M,\mathrm{GTP}}}{C_{\mathrm{GTP}}} + \frac{K_{M,R_K}\, K_{M,\mathrm{GTP}}}{C_K^R\, C_{\mathrm{GTP}}}} \tag{6.44}$$

The maximum termination rate $V_{TLT}^{max} = k_{TLT}\, C_{\mathrm{RF}}$. Symbol C_{RF} represents the concentration of the proper release factor corresponding to the particular stop codon of the termination site. C_K^R is the concentration of ribosomes bound to codon K. The rate constants for termination were reported to be $0.25\ \mathrm{s}^{-1}$ for RF1, and $0.5\ \mathrm{s}^{-1}$ for RF2 (Freistroffer et al., 1997). The affinity constant of ribosomes with respect to RF1 was given to be $K_{M,\mathrm{RF1}} = 8.3$ nM (Freistroffer et al., 1997). Under the assumption that this parameter equals the dissociation rate constant, the same value was taken for parameter K_{M,R_K} used in equation (6.44). An affinity constant

Table 6.9: Role of prokaryotic termination factors. The specific mechanistic role of termination factors RRF and RFH is not further considered in this study.

Factor	Properties and function
RF1	• directs translation termination at codons UAG and UAA
RF2	• invokes translation termination at codons UGA and UAA
RF3	• stimulates activities of RF1 and RF2 • enhances the formation of termination complexes • belongs to GTP-binding proteins
RRF	• ribosome release factor, release of ribosome from mRNA at termination • recycle of ribosomes from termination to initiation
RFH	• involved in translation termination, yet its role remains obscure

for GTP could not be found while scanning the literature. Constant $K_{M,\mathrm{GTP}} = 20\ \mu\mathrm{M}$ was thus chosen arbitrarily. Owing to the observed GTP concentration *in vivo* (in the mM range (Pingoud et al., 1983)), the termination rate may be saturated with GTP concentration.

6.5 tRNA charging

The charging of tRNA with amino acids includes the steps of aminoacylation and esterification. This reaction is catalyzed by the enzymes aminoacyl-tRNA-synthetases (ARS), thereby consuming ATP. The net stoichiometry reads

$$\mathrm{aa} + \mathrm{tRNA} + \mathrm{ATP} \overset{\mathrm{ARS}}{\rightleftharpoons} \mathrm{aa\text{-}tRNA} + \mathrm{AMP} + \mathrm{PP}_i \tag{6.45}$$

Through hydrolysis of byproduct inorganic pyrophosphate, PP_i, the thermodynamic equilibrium of equation (6.45) is shifted to the right. For each amino acid, there exists at least one corresponding ARS (Voet and Voet, 1994). These enzymes exhibit highly specific independent binding sites for the amino acid, tRNA, and ATP, respectively. The affinity of ARS to non-cognate tRNA can vary between 0 and the affinity of the cognate tRNA. All isoaccepting tRNAs (which are tRNAs with a specificity for the same amino acid) are recognized by the same ARS.

6.5.1 Reaction kinetics

Assuming a rapid equilibrium binding of substrates and neglecting product inhibition terms, the following rate equation was considered to apply for the reaction of tRNA charging:

$$V_{\mathrm{ARS},i,k} = \frac{V^{max}_{\mathrm{ARS},i,k}}{1 + \dfrac{K_{M,\mathrm{ARS},\mathrm{aa}_j}}{C_{\mathrm{aa},j}} + \dfrac{K_{M,\mathrm{ARS},\mathrm{ATP}}}{C_{\mathrm{ATP}}} + \dfrac{K_{M,\mathrm{ARS},\mathrm{tRNA}_j}}{C_{\mathrm{tRNA}_j}}} \tag{6.46}$$

In a simplifying assumption, the kinetic constants displayed in equation (6.46) were taken to be the same for all tRNA species, and for all aa-tRNA synthetases. These constants were derived from the data listed in Table 6.10.

6.5.2 Formyl-methionine

A formylation of the methionine bound to initiator tRNA is necessary to effectively dictate the synthesis direction of peptide chain polymerization. The formylation reaction is conducted by the enzyme methionyl-$\mathrm{tRNA}^{\mathrm{M}}_f$ *N*-formyltransferase (E.C. 2.1.2.3) according to the following equation:

$$\mathrm{fTHF} + \mathrm{Met\text{-}tRNA}^{\mathrm{M}}_f \rightleftharpoons \mathrm{fMet\text{-}tRNA}^{\mathrm{M}}_f + \mathrm{THF} \tag{6.47}$$

Symbol fTHF stands for Formyl-H_4-Folate, and THF for H_4-Folate. Kinetic constants $k_{cat} = 20\ \mathrm{s}^{-1}$, $K_{M,\mathrm{Met\text{-}tRNA}^{\mathrm{M}}_f} = 0.35\ \mu\mathrm{M}$, and $K_{M,\mathrm{fTHF}} = 13.5\ \mu\mathrm{M}$ were determined by Blanquet et al. (1984) for this reaction. Due to lack of information about intracellular levels of substrate fTHF, the formylation reaction was disregarded throughout this study. In simulation analyses, $\mathrm{Met\text{-}tRNA}^{\mathrm{M}}_f$ was supplied initially in sufficient amounts and then consumed over the course of the reaction.

Table 6.10: Apparent affinity constants of selected aminoacyl-tRNA synthetases of *E. coli*.

ARS	$K_{M,\mathrm{aa}}$ [μM]	$K_{M,\mathrm{tRNA}}$ [μM]	$K_{M,\mathrm{ATP}}$ [μM]	k_{cat} [s^{-1}]	**References**
Ala	–	2.2	–	1.0	Schulman (1991)
Arg	12	2.5	900	–	Lin et al. (1988)
Asp	–	0.33	–	1.0	Schulman (1991)
Glu	60	0.16	250	–	Kern and Lapointe (1981)
Gly	0.16	–	80	1.0	Francis and Nagel (1976), Schulman (1991)
His	–	0.5	–	1.7	Yan et al. (1996)
Ile	20	20	50	–	Freist et al. (1982)
Lys	20	0.7	110	–	Hirshfield and Yeh (1976)
Met_f	17	0.24	125	1.0	Ron et al. (1978), Schulman (1991)
Met_m	17	0.25	125	1.0	Ron et al. (1978), Schulman (1991)
Phe	–	2.1	–	–	Goodman and Schwartz (1988)
Thr	–	0.06	–	1.0	Schulman (1991)
Trp	0.13	–	–	1.0	Schulman (1991)
Tyr	0.32	–	–	1.0	Schulman (1991)
Val	1.0	0.5	–	1.0	Schulman and Pelka (1988), Schulman (1991)
aa_j	20	0.5	100	1.0	This study

6.6 Model reduction

Applying the model simplification of merging groups of codons, which was introduced in section 4.2.5, causes a profound effect on material balancing of variables enrolled in the translation process. In this case, the rate of translation elongation summarizes multiple (say n_{cod}) elongation cycles together. The reaction stoichiometry then reads

$$70\mathrm{S}'_j + \sum_{k=1}^{n_{cod}} \mathrm{EFTu{\cdot}GTP{\cdot}aa\text{-}tRNA}_{j+1,k} + n_{cod}\,\mathrm{EFG{\cdot}GTP} \xrightarrow{k_{TLE,\,j}}$$
$$70\mathrm{S}'_{j+1} + n_{cod}\,\mathrm{EFTu{\cdot}GDP} + n_{cod}\,\mathrm{EFG{\cdot}GDP} + 2\,n_{cod}\,\mathrm{P}_i + \sum_{k=1}^{n_{cod}} \mathrm{tRNA}_{j,k} \tag{6.48}$$

The prime refers to state variables of the new codon grid. Combining multiple rounds of the reaction scheme given in equations (6.24) to (6.30), it can be shown (see Appendix A.3) that the overall kinetics of n_{cod} elongation steps may be described mathematically by

$$V'_{TLE,j} \;=\; \frac{q_j^R\; k'_{TLE,j}\; C_j^{R'}}{1 + \sum_{k=1}^{n_{cod}} \dfrac{K_{M,\mathrm{T3}_j}}{\sum_i C_{\mathrm{T3}_j,i,k}} + \dfrac{K_{M,\mathrm{EFG{\cdot}GTP}}}{C_{\mathrm{EFG{\cdot}GTP}}}} \tag{6.49}$$

Parameter $k'_{TLE,j}$ was calculated from the smallest of the efficiency factors within each group of n_{cod} codons in the reduced state representation, according to

$$k'_{TLE,j} \;=\; \min(f_{j,k})\,\frac{k_{TLE}^{max}}{n_{cod}} \qquad \text{with} \;\; k = 1 \text{ to } n_{cod} \tag{6.50}$$

The sum of elongations consuming a particular ternary complex k is given by

$$V_{\mathrm{SumT3},k} \;=\; \sum_{j=j_{R0}}^{K-1} \alpha_{j,\,k}\; V_{TLE,j} \tag{6.51}$$

Parameter $\alpha_{j,k}$ denotes the fraction of translational elongation rates j at which the kth ternary complex is consumed. $\alpha_{j,k}$ typically equals 1, when only one cognate ternary complex exists. $\alpha_{j,k}$ takes values between 0 and 1 when codons are matched by more than one tRNA. $\alpha_{j,k}$ equals 0 for codons j that do not relate to the kth tRNA. This parameter was furtheron approximated by the ratio of total concentration of the kth ternary complex involved in elongation at a particular codon j to the sum of total concentrations of ternary complexes recognized by this codon, i.e.,

$$\alpha_{j,k} \approx \frac{C_{\mathrm{T3},j,\,k}}{\sum\limits_{i} C_{\mathrm{T3},j,i}} \qquad \text{for} \quad j_{R0} \le j \le K-1 \tag{6.52}$$

Analogously to equation (6.51), the sum of elongation rates releasing an uncharged tRNA species k may be written as

$$V_{\mathrm{SumT},k} \;=\; \sum_{j=j_{R0+1}}^{K} \alpha_{j,k}\; V_{TLE,j} \tag{6.53}$$

6.7 Material balances

The following material balances cover the time-dependent changes of protein product, concentrations of ribosomes freely dissolved and in diverse states of complexation with translation factors, as well as when they are bound to mRNA in different positions. Material balancing further includes the full sets of amino acids, tRNA species, aminoacylated tRNAs, ternary complexes EFTu·GTP·aa-tRNA$_j$, and balances of energy components consumed during translation.

$$\frac{dC_{\mathrm{Protein}}}{dt} = V_{TLT} \tag{6.54}$$

$$\frac{dC^{R*}_{j_{R0}}}{dt} = V_{TLI,\mathrm{70SIC}} - V_{TLI,\mathrm{IF2D}} \tag{6.55}$$

$$\frac{dC^{R}_{j_{R0}}}{dt} = V_{TLI,\mathrm{IF2D}} - V_{TLE,j_{R0}} \tag{6.56}$$

$$\frac{dC^{R}_{j}}{dt} = V_{TLE,j-1} - V_{TLE,j} \qquad \text{for} \quad j_{R0} < k < K \tag{6.57}$$

$$\frac{dC^{R}_{K}}{dt} = V_{TLE,K-1} - V_{TLT} \tag{6.58}$$

$$\frac{dC_{\mathrm{aa}_i}}{dt} \;=\; -\sum^{k} V_{\mathrm{ARS},i,k} \qquad \text{for} \quad 1 \le i \le A \tag{6.59}$$

$$\frac{dC_{\mathrm{T}_k}}{dt} \;=\; \sum_{j=j_{R0}+1}^{K} \alpha_{j,k} V_{TLE,j} - V_{\mathrm{ARS},i,k} \qquad \text{for} \quad 1 \le k \le T \tag{6.60}$$

$$\frac{dC_{\text{aa}_i\text{-tRNA}_k}}{dt} = V_{\text{ARS},i,k} - V_{\text{T3Form},k} \qquad \text{for} \quad 1 \leq k \leq T \tag{6.61}$$

$$\frac{dC_{\text{T3}_k}}{dt} = V_{\text{T3Form},k} - \sum_{j=j_{R0}}^{K-1} \alpha_{j,k} V_{TLE,j} \qquad \text{for} \quad 1 \leq k \leq T \tag{6.62}$$

6.8 Conclusions

In this chapter, a thorough mathematical description of the key processes involved in prokaryotic translation was presented, which - to a high degree - makes use of gene sequence information. Detailed kinetic expressions for the rates of translation initiation, elongation, and termination were developed, which include a functional dependence on substrate, product, and effector concentrations. At the same time, the catalytic role and recycle of major translation factors was taken into account in the overall reaction scheme. The sequence-oriented nature of this model is reflected in particular by the codon-specificity of elongation rates and the sterical interactions among ribosomes, as they propagate along the template mRNA.

The impact of mRNA secondary structure on translation efficiency, although significant as pointed out earlier at the example of initiation (de Smit and van Duin, 1990b), needed at present to be left out of the modelling frame. The translation model derived, however, is in principle compatible to distinguish between folded and unfolded regions of mRNA. For example, with respect to the rate expression developed for initiation, it is conceivable to replace the currently formulated substrate dependency on the ribosomal binding site in general (cf. equation (6.12)), by a more specific dependency solely on the fraction of RBS that is unfolded. The portion of unfolded RBS would then follow directly from the free energy computation for this site (using equation (2.2)). Nevertheless, the issue of incorporating information upon mRNA secondary structure into the model constitutes an entire study itself including systematic experimental investigation, which is beyond the scope of this study and subject of on-going research.

A main contribution of this work arises from the fact that the model allows to study the various nonlinear interactions among key antagonists of the bacterial translation machinery, both in a static and dynamically changing environment. Simulation analysis employing this model may reveal potentially rate-limiting steps, caused on both system-immanent and process-operational levels. With this model at hand, it is also feasible to determine optimum stoichiometric catalyst relations, that are difficult to estimate based on purely heuristic grounds. This issue is of particular importance for designing customized *in vitro* gene expression systems, where catalyst levels may be readily adjusted by their selective supplementation. Moreover, due to the chosen sequence-oriented modelling approach, it may ultimately be possible to predict steady state protein synthesis rates specifically in dependence of their gene sequence. Further fields of application include the use of this model as part of a comprehensive gene expression model, in order to study the mutual interactions and combined effects on expression performance. In this context, the translation model may serve as an analysis tool for custom-design of both *in vivo* and *in vitro* recombinant gene expression systems.

7 Results of Simulating Prokaryotic Translation

As demonstrated in the previous chapter, the processes enrolled in ribosomal translation are highly complex and mutually coupled by the effects of translation factor involvement and sterical interactions among translating ribosomes. Due to the large number of catalysts participating in translation and their interconnectedness, an optimization of translation factor levels is difficult to develop purely based on intuitive grounds. Instead, a model-based approach is taken in the sequel to estimate optimum initiation factor stoichiometries. The outcome of such an analysis may be readily valuable for optimization of reconstituted protein biosynthesis systems consisting of all purified catalyst and substrate concentrations (Shimizu et al., 2001). In a second stage of this chapter, the impact of codon-specificity on variability of elongation rate is examined theoretically. An estimate of the rate constant of translation elongation is provided for each codon.

7.1 Stoichiometry of translation initiation factors

Ribosomal complexes with initiation factors were calculated using the material balances given in section 6.2.2. The results of this computation are shown in Table 7.1 for typical growth conditions of *E. coli* (system I) and three selected bacterial cell-free gene expression systems (II, III, IV). Total concentrations displayed in this table were used as model entries (index *in*) in the equilibrium computation. In this analysis using equations (6.3) to (6.7), the concentrations of uncomplexed ribosomal subunits and initiation factors were estimated for the respective systems.

When subsequent to this estimation procedure, the concentrations of umcomplexed components were resubstituted to calculate all concentrations of complexed species and to check the consistency of material balancing, it was noted that the resultant total mass of initiation factors IF1, IF2, and IF3 (index *out*) did not exactly match the respective total concentrations initially introduced to the calculation. An explanation for this phenomenon may be given based on the sensitivity of computation results with regard to the association constants employed. Small errors introduced with these parameters, that were determined in a variety of studies and under isolated reaction conditions (see Table 6.4), obviously lead to inaccuracies due to non-additive effects when the combined system is examined. For those cases where total concentrations of IF1, IF2, and IF3 (index *out*) were calculated to be below their respective model entries, estimated concentrations of complexed ribosomal subunits provided in Table 7.1 represent lower boundary values to the situation in the real biological system. Importantly, this means that in the real system (i.e., when employing the concentrations denoted by index *in*), ribosomal subunits are driven further towards full complexation at the same time with all three initiation factors.

In vivo, roughly 80 % of the total ribosome concentration are found as polyribosomes (case I

in Table 7.1). The remaining 20 % ribosomes (i.e., $C_{70S_{un}} = 6\ \mu M$) are not bound to a mRNA and are during this stage not participating in translation. An approximately equal concentration of initiation factors IF1, IF2, and IF3 was measured in relation to the concentration of non-translating ribosomes (Table 7.1). From out of the non-translating ribosomes, 75 % (i.e., 4.52 μM) are estimated to be complexed at the same time with all three of these initiation factors (complex 30S·IF1·IF2·GTP·IF3). This finding indicates that *in vivo*, approximately 3/4 of the non-translating ribosomes are in a state that serves as a substrate for productive translation initiation. The estimate provided in this study is in accordance to previous belief, as reviewed by Gualerzi and Pon (1990).

Table 7.1: Estimated concentrations of ribosomal subunits complexed with initiation factors. Stoichiometric ratios with respect to the total concentration of non-translating ribosomes ($C_{70S_{un}}$) are given also. I - *E. coli* grown at $\mu = 1.0\ h^{-1}$. Total concentrations taken from Bremer and Dennis (1996). II - Initial conditions of S30 extract under standard conditions (Pratt, 1984), which are roughly 20-fold diluted in comparison to the cell (Siemann, 2001). Total concentrations of initiation factors are computed with the molar relations given by Howe and Hershey (1983) and under the condition of no loss or further concentrating during lysate preparation. III - Initial conditions of system used by Pavlov and Ehrenberg (1996) applying purified ribosomes and translation factors. IV - Optimized conditions for system II. Predicted relative concentrations of complexes 30S·IF1, 30S·IF2·GTP, 30S·IF3, 30S·IF1·IF2·GTP, 30S·IF1·IF3, and 30S·IF2·GTP·IF3 were each below 2 % for either of the cases I to IV, respectively (data not provided).

Concentrations	**I**		**II**		**III**		**IV**	
	C_j (μM)	$\frac{C_j}{C_{70S_{un}}}$ (–)	C_j (μM)	$\frac{C_j}{C_{70S_{un}}}$ (–)	C_j (μM)	$\frac{C_j}{C_{70S_{un}}}$ (–)	C_j (μM)	$\frac{C_j}{C_{70S_{un}}}$ (–)
$70S_t$	30.00	5.00	1.40	1.00	3.00	1.00	1.40	1.00
$70S_{un}$	6.00	1.00	1.40	1.00	3.00	1.00	1.40	1.00
$(IF1_t)_{in}$	7.50	1.25	0.35	0.25	4.00	1.33	1.75	1.25
$(IF2_t)_{in}$	9.00	1.50	0.42	0.30	3.00	1.00	2.10	1.50
$(IF3_t)_{in}$	6.00	1.00	0.28	0.20	4.00	1.33	1.40	1.00
$(IF1_t)_{out}$	4.87	0.81	0.35	0.25	2.78	0.93	1.67	1.19
$(IF2_t)_{out}$	5.09	0.85	0.42	0.30	2.46	0.82	1.28	0.91
$(IF3_t)_{out}$	4.72	0.79	0.28	0.20	2.94	0.98	1.65	1.18
30S·IF1·IF2·GTP·IF3	4.52	0.75	0.25	0.18	2.36	0.79	1.20	0.85
30S	≈ 0	≈ 0	0.01	≈ 0	≈ 0	≈ 0	≈ 0	≈ 0
50S	4.62	0.77	0.32	0.23	2.44	0.81	1.24	0.88
70S	1.38	0.23	1.08	0.77	0.57	0.19	0.16	0.12
IF1	0.30	0.05	0.07	0.05	0.36	0.12	0.45	0.32
IF2·GTP	0.49	0.08	0.11	0.08	0.08	0.03	0.07	0.05
IF3	0.15	0.02	0.01	0.01	0.51	0.17	0.42	0.30

As is shown in the same diagram, albeit for case II, there may be a completely different scenario under *in vitro* conditions of dilute systems. In the displayed scenario, the concentrations of cellular components correspond to a situation where an *E. coli* cell extract has been generated according to standard protocol (Pratt, 1984). Under these conditions, a roughly 20-fold dilution of all ribosomes and proteins is reached in comparison to the cell (Siemann (2001), personal communication). Considering the stoichiometric ratios determined for growth conditions and taking into account a 20-fold dilution, absolute initiation factor concentrations were calculated for *in vitro* conditions (as given in Table 7.1).

At reaction start of standard cell-free protein synthesis conditions, 18 % of freely dissolved ribosomes are predicted to be complexed with all three initiation factors simultaneously. That is, approximately only 1/5 of all non-translating ribosomes are in a stage to efficiently start translation initiation. In comparison, *in vivo* this ratio is about 3/4. Hence, the presented *in vitro* conditions clearly show an inferior synthesis capacity - *ab initio* - compared to *in vivo*, which is caused entirely by dilution effects.

At limiting initiation factor concentrations, the concentration of the kinetically favourable 30S·IF1·IF2·GTP·IF3 species is approximated roughly by the respective lowest concentration among the initiation factors. About 77 % of all ribosomes are predicted under these conditions to occur as 70S complexes that have no initiation factors bound to them, which all together constitutes a reversal to the conditions forecast in the growing cell.

In vitro, the stoichiometric relation between initiation factors and non-translating ribosomes ($70S_{un}$) is particularly disadvantageous at reaction start, when all ribosomes are freely dissolved and not bound to mRNA. Such a situation may thus potentially lead to a significant competition among the various complexed species of 30S subunits, all competing for association to the ribosome binding sites. Binding of 30S subunits other than the kinetically favourable 30S·IF1·IF2·GTP·IF3 complex leads to a reduced overall initiation rate (Spirin, 1986). In the likely situation, where cellular lysates are deprived of translation initiation factors, similarly as was demonstrated for translation elongation factors (Schindler, 1999), the equilibrium is shifted further towards uncomplexed 30S subunits. Translation initiation rates are diminshed in this case.

When applying purified components *in vitro* and under the conditions of case III described in Table 7.1, total concentrations of initiation factors can be suitably chosen such that unbound ribosomes are forced to exist almost exclusively in a state where both ribosomal subunits are dissociated, and moreover, the entirety of 30S subunits is fully complexed with all necessary initiation factors simultaneously. In this case, the concentration of complex 30S·IF1·IF2·GTP·IF3 is predicted to comprise 79 % of the non-translating ribosomes, which is very similar to the *in vivo* conditions predicted. Due to their apparently high binding capacity to ribosomal 30S subunits, initiation factors need not be added much in excess. Approximately equal molar amounts of each IF to total ribosome concentration appear to be sufficient, in order to reach full 30S complexation simultaneously with all of the three initiation factors considered. It is noteworthy to mention that the average specific elongation rate was 7 aa/s for case III (Pavlov and Ehrenberg, 1996), which is of the same order of magnitude as obtained *in vivo* (cf. Table 6.2).

Finally, the comparatively inferior conditions applied in system II may be improved by appropriately raising the concentrations of initiation factors. Scenario IV describes the predicted concentrations for a system in which the same stoichiometric ratios for initiation factors were selected (with respect to $C_{70S_{un}}$) as identified in the growing cell (scenario I). Under these conditions, 85 % of all 30S subunits that are not bound to mRNA are estimated to exist as complex 30S·IF1·IF2·GTP·IF3. This is a more than 4-fold predicted enhancement in comparison to standard *in vitro* reaction conditions (scenario II). Therefore, sufficient levels of initiation factors, and in general all translation factors, appear to be mandatory prerequisites for eliminating system-intrinsic limitations that adversely affect protein production rate.

7.2 Effect of codon-specificity

First, the effects of codon-specificity are investigated for individual codons and their specific elongation rate. In a second stage of this section, these effects are examined with respect to overall translation elongation rate.

7.2.1 Estimation of effective codon-specific rate constant of translation elongation

7.2.1.1 Definitions and assumptions

In order to investigate the individual contribution of codon-specificity on translation elongation rate, an effective rate constant, $(k_{TLE,j})_{\text{eff}}$, was estimated separately for each of the 61 sense codons and under the premise of sparse ribosome loading. Parameter $(k_{TLE,j})_{\text{eff}}$ indicates the effective rate with which a ribosome is predicted to elongate at a particular codon j. This parameter is defined according to

$$(k_{TLE,j})_{\text{eff}} = \frac{V_{TLE,j}}{C_j^R} \tag{7.1}$$

Substituting equations (6.31) to (6.33) into equation (7.1) and after rearrangement, it follows

$$(k_{TLE,j})_{\text{eff}} = \frac{q_j^R \; f_j \; k_{TLE}^{max}}{1 + \dfrac{K_{M,\text{T3}_j}}{\sum\limits_i C_{\text{T3}_j,i}} + \dfrac{K_{M,\text{EFG·GTP}}}{C_{\text{EFG·GTP}}}} \tag{7.2}$$

Since translation elongation rate ($V_{TLE,j}$) itself is a function of EFG·GTP concentration and the concentrations of ternary complexes matching this codon j, and depends on substrate affinities $K_{M,\text{T3}_j}$ and $K_{M,\text{EFG·GTP}}$ (see equation (6.31)), as well as on the codon-specific efficiency factor (f_j defined in equation (6.33)), parameter $(k_{TLE,j})_{\text{eff}}$ also shows a dependency on these quantities. The subsequent assumptions were made in the procedure for estimating $(k_{TLE,j})_{\text{eff}}$:

1. At extremely low ribosome densities, any adverse impact on elongation rate caused by ribosomal interference vanishes. Further, the effects of mRNA degradation on

elongation rate are neglected in this analysis. In this case, the queueing factor q_j^R (see equation (4.16)) approaches a value of 1 for all codons j.

2. The concentrations of ternary complexes entering translation elongation rate were taken to be the same as the respective total concentrations of tRNA, which were given by Dong et al. (1996) for a specific growth rate of $\mu = 1.0$ h^{-1}.

3. The concentration of complex EFG·GTP was taken to be 0.72 μM.

4. Model parameters $K_{M,\mathrm{T3}_j}$ and $K_{M,\mathrm{EFG\cdot GTP}}$ were 0.4 μM and 0.22 μM, respectively (see section 6.3.2.2 on page 79).

5. All values for $(k_{TLE,j})_{\mathrm{eff}}$ were expressed as relative quantities to the maximum computed $(k_{TLE,j})_{\mathrm{eff}}$. The latter was obtained for codon CUG (coding for leucine) and was set arbitrarily to equal 24 codons/s.

The fractional total codon usage (symbol η_j^t) denotes the usage (c_j) of a particular codon j with respect to total codon usage according to

$$\eta_j^t = \frac{c_j}{\sum_{k=1}^{64} c_k} \tag{7.3}$$

The sum of η_j^t across all 61 sense codons together with the fractional usage of the 3 stop codons adds up to unity. The summation term used in the denominator of equation (7.3) stands for the total (absolute) usage of base tripletts of the strain under investigation.

Moreover, symbol $\eta_j^{aa_i}$ was introduced, in order to express the fractional usage of codon j with respect to only those codons that represent the same amino acid i. Constant a_i denotes the number of base tripletts that code for the ith amino acid. $\eta_j^{aa_i}$ is further on also termed amino acid specific codon usage. It is defined by

$$\eta_j^{aa_i} = \frac{\eta_j^t}{\sum_{k=1}^{a_i} \eta_k^t} \tag{7.4}$$

The sum of all $\eta_j^{aa_i}$ for a particular amino acid i adds up to 1.

7.2.1.2 Results

Table 7.2 lists the specific rate constants of translation elongation, $(k_{TLE,j})_{\mathrm{eff}}$, predicted for all of the 61 sense codons. Tabulated data were calculated from the equations given in the previous section and are ordered alphabetically with respect to the encoded amino acid. Fractional codon usage in wild-type *E. coli* genome is provided also in this table, both in relation to total codon usage (symbol η_j^t) and the usage of this codon with respect to the encoded amino acid ($\eta_j^{aa_i}$). The former has been adopted from the Codon Usage Database at the Kazusa DNA Research Institute in Japan (http://www.kazusa.or.jp/codon/).

As can be viewed from Table 7.2, estimated values for parameter $(k_{TLE,j})_{\text{eff}}$ show a distinct response to codon-related effects. All values calculated for $(k_{TLE,j})_{\text{eff}}$ are located between 8.6 and 24 codons/s (for base tripletts GUA (Val) and CUG (Leu), respectively). The lowest value estimated for parameter $(k_{TLE,j})_{\text{eff}}$ comprises about 1/3 of the assumed maximum specific elongation rate.

An attempt was made to try and correlate the calculated codon-specific translation elongation rates with the fractional total codon usage (columns $(k_{TLE,j})_{\text{eff}}$ and η_j^t in Table 7.2). Linear regression analysis gave a regression coefficient of $r = 0.0148$ (graphical representation not provided), which is a clear indication that no linear relationship can be deduced between both quantitites. In addition to a codon-related divergence of tRNA concentration, also different affinities towards substrate ternary complexes may conceivably be significant in determining specific elongation rate. These influences are presently unaccounted for by the model, since parameter $K_{M,\text{T3}_j}$ was kept constant in the analysis. The rate expression of translation elongation could, however, be appropriately augmented, when these affinities were known for the entire set of sense codons. Moreover, nonlinear effects caused by codon context, i.e., ribosomal queueing, which are neglected in this analysis, may also be important in determining codon-specific elongation rates, and may serve to explain the fact that no linear relationship was found between parameters $(k_{TLE,j})_{\text{eff}}$ and η_j^t.

An investigation of amino acid related effects shows that also no direct correlation exists between the predicted values for $(k_{TLE,j})_{\text{eff}}$ and the amino acid specific fractional codon usage. For example, in the case of leucine (Leu), codon CUA shows a roughly 17-fold higher usage in *E. coli* in comparison to codon CUC, but they are both predicted to be elongated at the same rate of 22.1 aa/s (cf. Table 7.2). Similarly, codon GUG is used in average about 15-times more often than GUA to code for the amino acid valine (Val). Its specific elongation rate is, however, forecast to be only about 1.5-fold higher in comparison to GUA.

Groups of codons, with each group referring to a same amino acid, may be classified according to their predicted codon-specific elongation rates. Relatively low values for parameter $(k_{TLE,j})_{\text{eff}}$ were obtained, e.g., for amino acids Ala, Ser, and Val. In average, $(k_{TLE,j})_{\text{eff}}$ values of 10.6, 14.3, and 10.5 aa/s are predicted for these amino acids (values calculated as the geometric mean of the respective $(k_{TLE,j})_{\text{eff}}$). On the other hand, base tripletts coding, for example, for amino acids Gly, Ile, and Leu, are predicted to be elongated in average with a relatively higher specific rate (23.5, 23.6, and 22.9 aa/s, respectively), somewhat independent of the type of matching codon. According to model prediction, there are also sets of codons encoding for the same amino acid that span a range of codon-specific elongation rates. These are, for instance, represented by codons related to amino acids Arg, Pro, and Thr (with in average, 19.6, 19.6, and 19.9 aa/s, respectively). This latter group of codons is suggested to be particularly important from a stand-point of engineering recombinant protein expression. An expressivity improvement of poorly expressed genes may be pursued by replacing slowly elongated codons (corresponding to low values of $(k_{TLE,j})_{\text{eff}}$) by codons with a comparatively higher $(k_{TLE,j})_{\text{eff}}$, that still encode for the same amino acid.

Table 7.2: Variation of codon-specific rate constant for translation elongation. $(k_{TLE,j})_{\text{eff}}$ is expressed relative to the rate constant for codon CUG, which was arbitrarily set to 24 aa/s. η_j^t is the fractional codon usage with respect to total codon usage (equation (7.3)). $\eta_j^{aa_i}$ denotes the amino acid specific codon usage (equation (7.4)).

AA	**Codon**	$(k_{TLE,j})_{\text{eff}}$ [aa/s]	η_j^t [%]	$\eta_j^{aa_i}$ [%]	**AA**	**Codon**	$(k_{TLE,j})_{\text{eff}}$ [aa/s]	η_j^t [%]	$\eta_j^{aa_i}$ [%]
Ala	GCA	10.0	2.63	39.0	Leu	CUA	22.1	1.52	25.0
Ala	GCC	10.4	1.08	16.0	Leu	CUC	22.1	0.09	1.5
Ala	GCU	9.3	1.52	22.5	Leu	CUG	24.0	1.10	18.1
Ala	GCG	12.7	1.52	22.5	Leu	CUU	22.8	0.64	10.5
					Leu	UUA	22.5	1.38	22.7
Arg	AGA	22.1	1.03	11.0	Leu	UUG	23.7	1.36	22.3
Arg	AGG	17.5	0.87	9.3					
Arg	CGA	9.1	2.88	30.8	Lys	AAA	23.5	1.44	44.9
Arg	CGC	23.7	1.53	16.3	Lys	AAG	14.1	1.77	55.1
Arg	CGG	21.1	2.08	22.2					
Arg	CGU	24.0	0.97	10.4	Met	AUG	23.8	3.03	100.0
Asn	AAC	23.1	0.71	23.4	Phe	UUC	21.2	1.65	42.5
Asn	AAU	20.3	2.33	76.6	Phe	UUU	22.8	2.23	57.5
Asp	GAC	19.4	2.01	44.1	Pro	CCA	13.2	5.26	70.7
Asp	GAU	23.7	2.55	55.9	Pro	CCC	19.2	0.38	5.1
					Pro	CCG	23.0	0.70	9.4
Cys	UGC	23.2	0.02	9.5	Pro	CCU	23.1	1.10	14.8
Cys	UGU	20.0	0.19	90.5					
					Ser	AGC	23.0	3.36	38.1
Gln	CAA	22.3	2.31	64.2	Ser	AGU	17.2	2.16	24.5
Gln	CAG	22.8	1.29	35.8	Ser	UCA	10.7	0.71	8.0
					Ser	UCC	11.4	0.86	9.8
Glu	GAA	24.0	3.36	51.1	Ser	UCG	11.2	0.89	10.1
Glu	GAG	16.6	3.21	48.9	Ser	UCU	12.3	0.84	9.5
Gly	GGA	23.6	1.78	17.6	Thr	ACA	22.8	2.78	42.1
Gly	GGC	23.9	3.94	39.0	Thr	ACC	22.9	0.43	6.5
Gly	GGG	23.6	2.47	24.5	Thr	ACG	17.9	0.89	13.5
Gly	GGU	22.7	1.90	18.8	Thr	ACU	15.8	2.51	38.0
His	CAC	20.6	0.84	60.9	Trp	UGG	22.6	0.51	100.0
His	CAU	22.2	0.54	39.1					
					Tyr	UAC	21.4	1.22	43.1
Ile	AUA	23.9	0.54	17.5	Tyr	UAU	23.5	1.61	56.9
Ile	AUC	22.9	0.35	11.4					
Ile	AUU	23.9	2.19	71.1	Val	GUA	8.6	0.12	3.2
					Val	GUC	9.3	0.21	5.6
					Val	GUG	12.7	1.82	48.5
					Val	GUU	11.3	1.60	42.7

7.2.2 Estimation of average specific rate of translation elongation

Sørensen and Pedersen (1991) computed transit times for ribosomes translating various *lacZ* mRNAs, in which synthetic nucleotide sequences had been inserted additionally. These inserts differed in their relative accumulation of codons GAA (coding for Glu), GAG (Glu), CCG (Pro), and CGA (Arg), and were either 34 or 64 codons in length (see Table 7.3). An effective specific rate constant of translation, $(k_{TLE})_{\text{eff}}$, may be calculated from the results of the original study for each of the gene products, according to

$$(k_{TLE})_{\text{eff,exp}} = \frac{L}{\Delta T} \tag{7.5}$$

Parameter L represents the coding length, while ΔT is the measured time interval for a ribosome to propagate the distance between the translational start codon and termination site.

Table 7.3: Comparison of effective specific rates of translation elongation determined both from experimental data (Sørensen and Pedersen, 1991) and simulation analysis. Simulations were based on codon-specific elongation rates according to equation (6.33) and keeping $k_{TLI,\text{70SIC},1} = 0.02810\ \text{s}^{-1}$. ΔT is the experimentally determined transit time for ribosomes to travel between the translational start site and stop codon of the respective plasmid.

Plasmid	**Coding length** L [codons]	**Experiment** ΔT [s]	$(k_{TLE})_{\text{eff,exp}}$ [codons/s]	**Simulation** $(k_{TLE})_{\text{eff,sim}}$ [codons/s]
wt-*lacZ*	1,024	82.3	12.45	12.44
pMAS-48GAA	1,088	89.9	12.11	11.93
pMAS-24GAA	1,058	88.1	12.01	12.47
pMAS-12(GAACCG)	1,058	88.4	11.97	12.47
pMAS-24GAG	1,058	90.1	11.74	12.37
pMAS-12(GAGCCG)	1,058	90.3	11.72	12.42
pMAS-12(GAGCGA)	1,058	91.2	11.61	12.32
pMAS-48GAG	1,088	98.1	11.10	11.81

As can be seen from Table 7.3, the average rate of ribosome propagation was 12.45 codons/second for wild-type *lacZ* mRNA. All other plasmids, containing an additional insert within the coding region, were associated with a reduced average rate of ribosome elongation. The lowest measured rate of ribosome propagation was 11.10 codons/second (for plasmid pMAS-48GAG). This number constitutes a 10.8 % reduction in specific rate of β-galactosidase synthesis in comparison to the wild-type gene.

In addition to experimental determination, the effective specific rate of translation elongation $(k_{TLE})_{\text{eff,sim}}$ was estimated in this study by dynamically simulating prokaryotic translation using the same vector systems as examined by Sørensen and Pedersen (1991). In the simulation analysis, both mRNA synthesis and degradation were excluded, component inactivation was omitted, and numerical integration was performed until a pseudo-steady state in protein synthesis was reached (numerical integrator: ACSL (Mitchell & Gauthier Asociates Inc.)). Model constants kept fixed were $k_{TLI,\text{IF2D}} = 0.8\ \text{s}^{-1}$ (equation (6.17)) and $k_{TLE}^{max} = 19.3\ \text{s}^{-1}$

(equation (6.33)). These parameters had been determined earlier for protein synthesis on *lacZ* mRNA (Liang et al., 1999). In a preliminary analysis[7], parameter $k_{TLI,70SIC,1}$ (equation (6.13)) was varied until the following two conditions were met: (a) predicted translation elongation rates reached similar levels as in the experiment, and (b) in order to establish conditions of only negligible interference among elongating ribosomes, which had been suggested to prevail in the original study (Sørensen and Pedersen, 1991). These conditions were approached for $k_{TLI,70SIC,1}$ equal to 0.02810 s^{-1}. Measured and predicted values for the effective specific translation elongation rates, $(k_{TLE})_{eff,exp}$ and $(k_{TLE})_{eff,sim}$, respectively, are compared in Table 7.3.

With the chosen parameter vector, a similar value for $(k_{TLE})_{eff,sim} = 12.44$ codons/s is obtained for wild-type *lacZ* mRNA in comparison to the experiment (12.45 codons/s; see Table 7.3). For the other vector systems, the agreement between experimentally determined and predicted average protein synthesis rates is, however, much lower. Nonetheless, these rates are forecast to be in the same order of magnitude as for the experiment. Plasmid pMAS-48GAG shows the lowest predicted specific rate of ribosome propagation. That is, 11.81 codons/second were achieved for this plasmid, which corresponds to a 5.1 % reduction compared to wild-type *lacZ* translation.

An improvement between measured and predicted values for $(k_{TLE})_{eff}$ was sought by raising parameter $k_{TLI,70SIC,1}$. Elevated levels of $k_{TLI,70SIC,1}$ imply higher rates of translation initiation associated with higher ribosome densities and are thus expected to show a higher sensitivity towards codon-related effects. Simulation results were indeed noted to respond very sensitively, even to a slight increase of this parameter (by 0.06 % from 0.02810 s^{-1} to 0.02812 s^{-1}). While with the raised $k_{TLI,70SIC,1}$ value, the same average specific rates of ribosome elongation were predicted during pseudo-steady state translation as for the lower of the two $k_{TLI,70SIC,1}$ values (cf. Table 7.3) - using the plasmids with wt-*lacZ*, pMAS-24GAA, pMAS-12(GAACCG), pMAS-24GAG, pMAS-12(GAGCCG), and pMAS-12(GAGCGA), respectively -, simulations with plasmids pMAS-48GAA and pMAS-48GAG demonstrated a very divergent behaviour. Computed transients for these latter two plasmids were unable to reach a steady state within 2,000 and 7,000 min of simulated process time, respectively (data not shown). At the same time, numerical integrations were found to exhibit extremely small step-sizes associated with extensive computation times. For instance, the simulation using plasmid pMAS-48GAG and keeping $k_{TLI,70SIC,1} = 0.02812$ s^{-1} took more than one week on a Sun Ultra 60 workstation. On the other hand, computations with $k_{TLI,70SIC,1} = 0.0281$ s^{-1} run on the same machine had all been completed within less than one hour.

Observed excessive computational loads could be clearly attributed to effects caused by ribosomal crowding. With increasing ribosome densities, the sterical hinderance among elongating ribosomes is elevated also. These adverse interactions are reflected likewise by dropping queueing factors. Queueing factors close to a value of 0 imply significant crowding, while values approaching 1 indicate only negligible sterical hindrance. For the example of plasmid pMAS-

[7] Although an attempt was made, a thorough and rigorous optimization strategy on the basis of nonlinear regression analysis, e.g., by minimizing the sum of square relative errors, could not be pursued in this case (see discussion later in this section on page 95). Parameter $k_{TLI,70SIC,1}$ was varied manually instead.

48GAG and with $k_{TLI,70SIC,1} = 0.02812\ s^{-1}$, the smallest value among the queueing factors q_j^R was equal to 0.0121 (at $t = 7,000$ min). In contrast, the minimum q_j^R equalled 0.9948 for those plasmids where a pseudo-steady state synthesis was reached.

If a queueing factor, which enters translation elongation rate (and initiation rate) in a multiplicative fashion, tends to 0, then the corresponding rate is found to vanish. This rate is then - in the limiting case - associated with an infinitely large time constant. Hence, the difference between this time constant and the time constant of any other rate that is non-zero becomes huge. The coincidence of fast and slow transients is known to lead to stiff systems with inefficient step-size of numerical integration, which altogether explains the excessive computation times noted in the case of plasmids pMAS-48GAG and pMAS-48GAA.

In essence, although the model shows correctly a tendency to distinguish between codon-specific differences among the inserts, simulation results yielded, in general, a comparatively reduced variability among the examined vector systems than the experiment. A better match between prediction results and experiment may be obtained by further adjustment of model parameters. Parameter identification on the basis of nonlinear regression analysis is, however, a non-trivial task, due to the nature of the model to show a highly non-linear response behaviour to parameter perturbation.

7.3 Conclusions

The availability of ribosomal 30S subunits complexed with initiation factors may impose severe constraints on translational initiation rate due to the noted inefficiency of translation initiation in the absence of individual or all initiation factors. This is likely to be particularly important under *in vitro* conditions using cellular extracts that lack sufficient concentrations of translation factors. The appropriate choice of initial conditions was shown to be crucial for driving binding equilibria towards the active components needed for improved translation efficiency. At least approximately equal molar amounts of initiation factors IF1, IF2, and IF3 in relation to the concentration of freely-dissolved (i.e., non-translating) ribosomes appear to be a necessary prerequisite for establishing maximum volumetric rates of translation initiation.

It is noteworthy to mention that in addition to the considered major initiation factors and their binding to ribosomal subunits, further proteins are known to play a role in translation initiation. E.g., protein SmpB has been identified recently to bind in a complex with ribosomal protein S1, EFTu, mRNA, and transfer-messenger RNA (tmRNA) (Karzai et al., 1999; Wower et al., 2000). Obviously, a consideration of additional key compounds involved in translation initiation will shift the equilibrium concentrations calculated in this study. As knowledge about the mechanism of these interactions increases and the necessary kinetic parameters are identified, this information could, in principle, be added to the analysis scheme outlined above. Further mechanistic details could then allow for a refinement of prediction results.

This study provides for each of the 61 sense codons a rough estimate of its codon-specific elongation rate. From the results obtained, the translation model is evidently capable of reflecting codon-specific effects in a qualitative manner. However, quantitative predictions accounting

for codon-specificity are at present of limited value. This is primarily a consequence of the uncertainties associated with the numerous model parameters. Further experimental and theoretical analyses are, however, crucial in order to improve the predictability of the model. Nevertheless, this study provides a mechanistic basis for interpreting the individual contributions of several factors on codon-specificity.

In general, an adaptation of codon usage of a heterologous protein is favoured to match the preferred codon usage of the host strain. Such a design strategy appears to be over-simplified, due to the noted multi-factorial impact on codon-specific elongation rate. In particular, nonlinear effects related to sterical hindrance of translating ribosomes are usually neglected. Whether these adverse interactions may be significant for a given coding sequence, depends in a non-obvious way on the sequence context. The mathematical model of translation elongation developed in this study provides a basis for analysis of the combined effects of rate-influencing factors.

Codon-specific phenomena significantly affecting numerical computation times can manifest themselves already in a single simulation run. For analysis routines, however, that require the sequential calling of the numerical integrator (such as, e.g., nonlinear regression analysis), total computation time can reach untractable levels, that may hamper the general applicability of such methods for analysis of codon-specificity. Dependent on the chosen set of parameters, numerical problems may occur or not. It is, however, difficult to define - a priori to a simulation - parameter windows, within which these problems can be thoroughly excluded. The boundaries of these windows are not fixed, but are instead expected to vary in conjunction with the other parameters of the translation model. In particular, the rate constants of translation initiation, elongation, termination, and substrate concentrations are suggested to affect these rates. Additional complications arise from the fact that the system under investigation is usually far from an equilibrium at simulation start. This means that the problem of divergent time constants and small integrator step-size may not be readily apparent from the initial system state, as it may evolve over the course of a simulation. Future studies, employing system theoretic analyses, like e.g., stability analysis and bifurcation analysis, may provide further clues upon suitable, system-intrinsic boundaries on the parameter space of the gene expression model.

8 Model Validation with Experiments from Cell-free Protein Biosynthesis

8.1 Introduction

Cell-free protein synthesis systems are ideal, simplified exploration tools for gene expression analysis. Main advantages arise from their reduced complexity in comparison to a growing organism and their convenient accessibility. In these *in vitro* systems, protein production is typically achieved on the basis of cellular lysates, which contain the required biocatalysts extracted from the living cell. Through the choice of substrate composition, it is possible to selectively activate the endogenous gene expression pathway, whereas the majority of regulatory mechanisms, like e.g., induction and repression encountered *in vivo*, are switched off. Employing recombinant DNA technology, the synthesis capacity and energy expenditures usually spent on cell growth can thus in principle be redirected towards the production of single or a few gene products.

According to Figure 8.1, cell-free gene expression systems represent a subset of generic protein biosynthesis. By the selective addition of substrates and specific inhibitors to the reaction assay, it is possible to study the contributions of individual reactions on expression dynamics, as well as their nonlinear interactions. Moreover, cell-free protein synthesis systems are innovative production methods of particularly high potential for recombinant applications. Cytotoxic and novel peptides following from the incorporation of unnatural amino acids, that are not expressed *in vivo*, have been synthesized in mg-amounts in these cell extracts (Noren et al., 1989;). Recent practical examples of cell-free protein expression methods cover their use in functional genomics and evolutionary studies, such as for example ribosomal display (Hanes and Plückthun, 1997). Cell-free protein synthesis may be combined with tagging technology in conjunction with affinity chromatography to facilitate the recovery of product protein (Lamla et al., 2001).

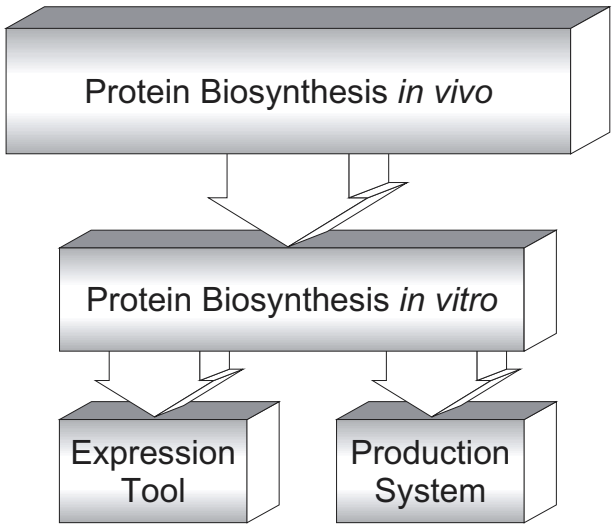

Figure 8.1: Cell-free protein biosynthesis is a subordinate system of *in vivo* gene expression.

Although *in vitro* protein production has been employed over several decades, many of the original constraints limiting both production rates and process duration remain unresolved. While various modifications were made to improve the commonly used systems (Zubay, 1973; Pratt, 1984), e.g., by applying condensed extracts (Kim et al., 1996) and continuous substrate supplementation via dialysis membrane technology, the problem of poor volumetric productivities still exists. Typical protein synthesis rates achieved in *E. coli* cell extracts are about 0.5 mg/ml/h (Patnaik and Swartz, 1998). This value is roughly 300-fold lower than in comparison to the *in vivo* synthesis rate of total protein at a specific growth rate of $\mu = 1.0\ \mathrm{h}^{-1}$ (Schmid (1999), calculated from Bremer and Dennis (1996)). The particular causes for this discrepancy between *in vitro* and *in vivo* synthesis rates are unclear.

Although cell-free protein synthesis systems serve as meaningful applications for probing gene expression models, they differ in some important aspects from the *in vivo* situation. For balanced growth, gene expression settles in a steady state, which is characterized by static pool concentrations and a constant renewal of the involved biocatalysts. On the other hand, cell-free gene expression systems suffer from a continuous catabolysis of supplied substrates and a gradual loss of biocatalytic activity. Counter measures commonly include the use of an energy regeneration system, as well as the addition of protease and RNase inhibitors. Nevertheless, degradation processes affecting the translation apparatus cannot be completely ruled out. At the same time, the initial lysate composition in terms of absolute and relative concentrations of translational key players is altered in comparison to *in vivo* conditions. This is caused mainly by the various processing steps and dilutions applied during lysate production, which typically add up to an approximately 20-fold dilution in comparison to the living cell (see section 7.1), as well as due to the supplementation of selected components such as translation factors and tRNA. A mathematical description of *in vitro* protein biosynthesis therefore needs to take into account all of the *in vitro* specific properties.

While in the previous chapters, kinetic models have been developed to describe the processes of transcription, mRNA degradation, and ribosomal translation individually as self-contained modules, modelling cell-free protein biosynthesis requires the formulation of a comprehensive gene expression model. It is a strategic question to design these former modelling tools and their interfaces appropriately, in order to yield a most representative picture of the real biological interactions for simulation of versatile system conditions. This is schematically demonstrated in Figure 8.2. The sequential scheme displayed on the left hand side of this figure constitutes a much too simplified picture of reality. When coupling the modelling units of gene expression, further, non-additive effects arise. An example for the nonlinearity of modular interactions is the feedback regulation of translational fidelity affecting mRNA degradation rate (see right hand side of Figure 8.2). Translating ribosomes are capable of providing a barrier to RNases trying to access endonucleolytic cleavage sites (cf. chapter 4). In order to account for these phenomena in a gene expression system, it is necessary to adequately modify the earlier defined stand-alone modelling units.

In the following, the necessary model adjustments to be made in order to arrive at a combined gene expression model are presented. Moreover, the effects of energy regeneration,

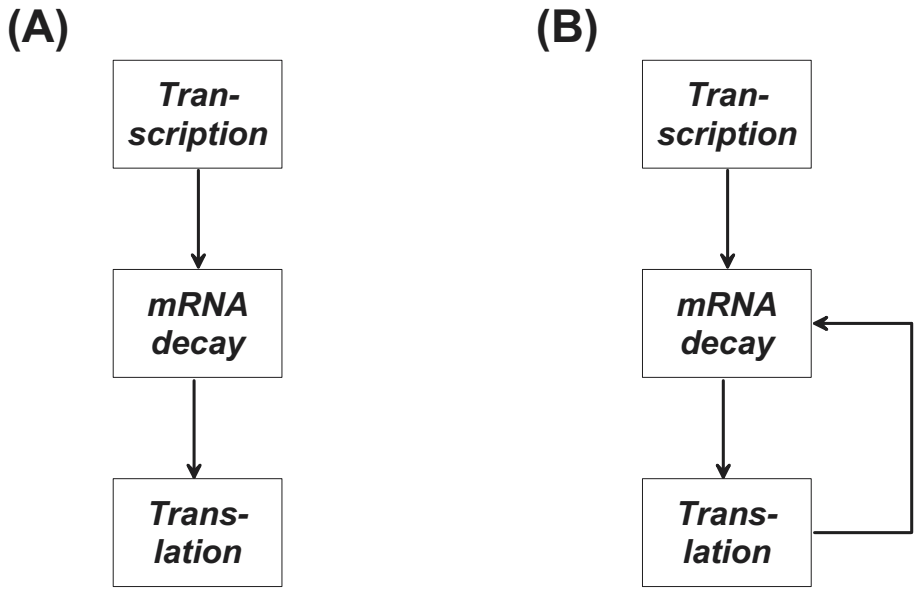

Figure 8.2: Coupling of modelling tools. (A) Unidirectional information flow. (B) Feedback interaction.

lysate composition, and inactivation kinetics on modelling of gene expression are outlined. For the purpose of model verification, the augmented model is subsequently applied to simulate the performance of cell-free protein expression. Such an approach aims to explore the predictability of the model by comparing simulation results with experimentally observed gene expression behaviour.

8.2 Combined gene expression model

A comparison of transcription rates both under simultaneous transcription/translation conditions and optimum *in vitro* transcription conditions revealed a reduction by 80 % of the apparent maximum transcription rate, for the case when the combined system was applied (Werner, 1997). Thus, the maximum rate of mRNA synthesis was reduced by a factor of 4 for simulations where transcription was embedded in cell-free gene expression conditions.

mRNA synthesis rate for each base triplett j can be acquired considering uniformly distributed RNA polymerases along the coding region. A time delay between initiation of transcript synthesis and the time point, when a particular base triplett j is synthesized, was neglected in this analysis. It was argued that the specific transcription rate of T7 RNA polymerase is much faster (100 nucleotides per second) in comparison to *E. coli* RNA polymerase (40 to 55 nucleotides per second, see chapter 3), and that both $5'$ and $3'$-ends of mRNA are thus synthesized approximately simultaneously and with the same rate.

Since the processes of transcription and translation are highly energy-dependent, all aspects of protein synthesis need to be viewed within the context of energy recycle systems (cf. Figure 8.3). Energy regeneration fulfills the task to continuously restore the pools of energy-carriers (such as, e.g., ATP and GTP), as they are constantly depleted over the course of protein synthesis. While these processes are maintained in the living cell as a result of catabolism, phosphor

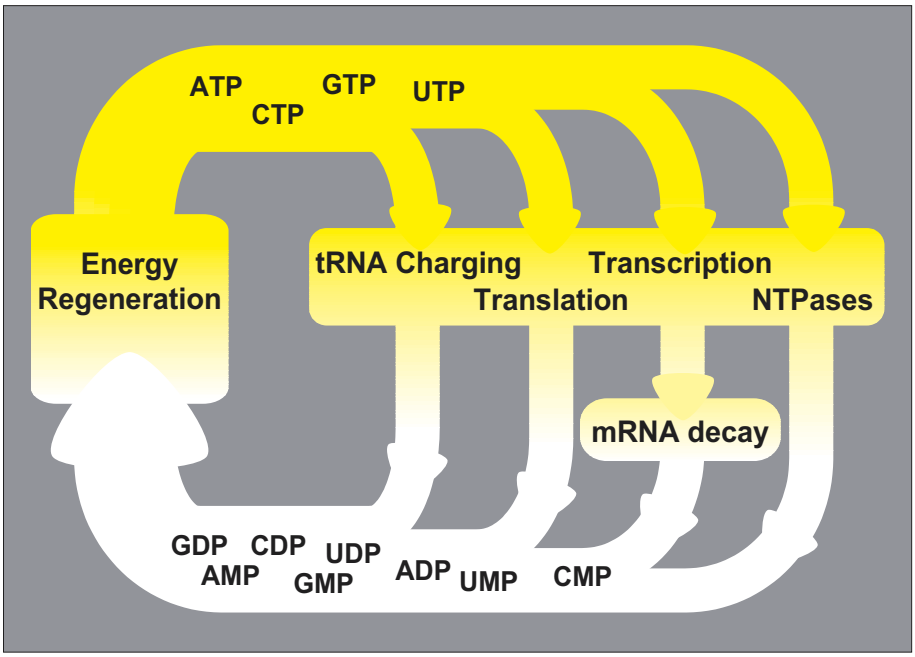

Figure 8.3: Turnover of energy components during gene expression.

donors need to be added specifically to cell-free systems. In addition, it is necessary to supply also the enzymes needed for regeneration, unless the regeneration machinery relies solely on endogenous enzymes that are already present in the native cellular extract. Since the issue of regenerating energy components constitutes an integral part of describing *in vitro* protein biosynthesis, material balance equations of, in particular, nucleotides need to be modified appropriately. The relevant modifications are outlined in the next section.

8.3 Energy regeneration

Gene expression is believed to comprise about 75 % of the energy expenditure in a growing bacterial cell (Neijssel et al., 1996). A preliminary estimation for *in vitro* protein synthesis shows that less than 1 % of the energy budget is directed towards gene expression, while the remainder is wasted by endogenous ATPase activity (Knapp, 1998). Hence, a sufficient supply and regeneration of nucleotide levels is a specific requirement for extended process longevity during cell-free protein production. Furthermore, raised energy levels have been shown to correspond with reduced error frequencies during translation (Jelenc and Kurland, 1979).

In an economic study comparing various enzymatic systems for ATP regeneration, Langer et al. (1976) identified acetyl phosphate (AcP)/acetate kinase (Ack) as one of the most cost-efficient solutions. Metabolic Flux Analysis (MFA) of the purine and pyrimidine pathways showed that, when employing AcP as phosphor donor, the endogenous regeneration system in an *E. coli* cell extract is capable of maintaining an energy charge of around 0.95 for at least 50 h of continuous-flow cell-free (CFCF) process operation (Knapp, 1998; Thiele, 1999).

8.3.1 Energy charge

The energy charge as defined by Atkinson (1968) provides a measure to assess the energy status of biological systems. It is calculated by the following empirically derived relationship:

$$\varphi \;=\; \frac{2\,C_{\mathrm{ATP}} + C_{\mathrm{ADP}}}{2\,(C_{\mathrm{ATP}} + C_{\mathrm{ADP}} + C_{\mathrm{AMP}})}. \tag{8.1}$$

φ is a dimensionles quantity ranging between 0 and 1. Typical values under growth conditions are distributed around an energy charge of 0.8 (Reich and Sel'kov, 1981).

8.3.2 Reaction model

The enzyme **acetate kinase** reversibly catalyzes the phosphorylation of ADP to form ATP, while AcP is converted to acetate (Ac). A kinetic rate expression for *E. coli* acetate kinase was derived in this study from the data given by Janson and Cleland (1974). The kinetics are assumed to obey a rapid equilibrium random bi bi mechanism with additional formation of dead-end inhibition complexes EBQ (= E·AcP·ATP) and EBP (= E·AcP·Ac) according to

$$V_{Ack} \;=\; \frac{V_f\,V_r\left(C_{\mathrm{ADP}}\,C_{\mathrm{AcP}} - \dfrac{C_{\mathrm{ATP}}\,C_{\mathrm{Ac}}}{K_{eq}}\right)}{D} \tag{8.2}$$

with

$$\begin{aligned} D \;=\;& V_r\,(K_{i,\mathrm{ADP}}\,K_{M,\mathrm{AcP}} + K_{M,\mathrm{AcP}}\,C_{\mathrm{ADP}} + K_{M,\mathrm{ADP}}\,C_{\mathrm{AcP}} + C_{\mathrm{ADP}}\,C_{\mathrm{AcP}}) \\ & + \frac{V_f}{K_{eq}}\left[C_{\mathrm{Ac}}\,C_{\mathrm{ATP}} + (K_{M,\mathrm{ATP}}\,C_{\mathrm{Ac}} + K_{M,\mathrm{Ac}}\,C_{\mathrm{ATP}})\left(1 + \frac{C_{\mathrm{AcP}}}{K_{i,\mathrm{AcP}}}\right)\right]. \end{aligned}$$

The enzyme **adenylate kinase** (Adk) performs the reaction converting AMP and ATP into two molecules of ADP. The following reversible rate equation was assumed to be representative for describing this reaction:

$$V_{Adk} \;=\; \frac{V_f\,C_{\mathrm{AMP}}\,C_{\mathrm{ATP}}}{(K_{M,\mathrm{AMP}} + C_{\mathrm{AMP}})\,(K_{M,\mathrm{ATP}} + C_{\mathrm{ATP}})} - \frac{V_r\,(C_{\mathrm{ADP}})^2}{(K_{M,\mathrm{ADP}} + C_{\mathrm{ADP}})^2} \tag{8.3}$$

Parameter values for model constants used in equations (8.2) and (8.3) are listed in Table D.2.1 in the Appendix D. Apart from this enzyme, there exist further nucleoside monophosphate kinases (Nmk) in *E. coli* to perform the reaction

$$\mathrm{N_1MP} + \mathrm{N_2TP} \rightleftharpoons \mathrm{N_1DP} + \mathrm{N_2DP}. \tag{8.4}$$

Nucleoside diphosphate kinases (Ndk) catalyze the reaction

$$\mathrm{N_1DP} + \mathrm{N_2TP} \rightleftharpoons \mathrm{N_1TP} + \mathrm{N_2DP} \tag{8.5}$$

Enzymes Ndk and Nmk form a network of near-equilibrium reactions, with both enzyme types exhibiting equilibrium constants close to unity (Reich and Sel'kov, 1981). Thus, and in order to mathematically implement a possibility to regenerate each of the four ribonucleoside mono- and

diphosphates, respectively, modelling assumed that there exist three further enzymes analogous to acetate kinase that are capable of regenerating nucleotides CDP, GDP, and UDP, respectively. For the same reasoning, rate expressions were derived additionally for three putative enzymes that were assumed to perform a reaction similar to the adenylate kinase reaction, with the difference of replacing AMP with one of the nucleoside monophosphates CMP, GMP, and UMP, respectively. Moreover, non-enzymatic chemical hydrolysis of acetyl phosphate (Oestreich and Jones, 1966), was taken into account by a first-order decay reaction according to

$$V_{d,\mathrm{AcP}} \;=\; k_{d,\mathrm{AcP}}\, C_{\mathrm{AcP}} \tag{8.6}$$

8.4 Catalyst inactivation

Catalyst inactivation takes place inherently in cell-free protein synthesis systems. In particular, a significant reduction of ribosomal protein S1 has been observed experimentally by proteome analysis (Schindler, 2000), as well as independently by measuring the time-dependency of total ribosome activity[(8)], and has thus been accounted for in the modelling scheme. The inactivation of ribosomal protein S1 (RP-S1) was included in the model in terms of a first-order inactivation of the maximum rate of 70S initiation complex formation:

$$V^{max}_{TLI,70\mathrm{SIC}} \;=\; k_{TLI,70\mathrm{SIC},1}\, e^{(-k_{d,\text{RP-S1}}t)}\, C_{30\mathrm{S{\cdot}IF}} \tag{8.7}$$

The time-dependent decrease of both EFTu and EFTs was modelled as first-order decay affecting their respective total concentrations, according to

$$C_{\mathrm{EFTu},t} \;=\; C_{\mathrm{EFTu},t}\, e^{(-k_{d,\mathrm{EFTu}}t)} \tag{8.8}$$

$$C_{\mathrm{EFTs},t} \;=\; C_{\mathrm{EFTs},t}\, e^{(-k_{d,\mathrm{EFTs}}t)} \tag{8.9}$$

First-order degradation constants used in the above equations were taken from Table 8.1. Equations (8.7) to (8.9) were then substituted into equations (6.13), (6.36), and (6.39), respectively.

Table 8.1: Half-life times of selected translational components calculated from experimental data. Measurements with 2D-gel electrophoresis within the initial 50 min of cell-free protein biosynthesis under batch conditions (Schindler, 2000). RP-S1 = ribosomal protein S1.

Component	**half-life** [min]	k_d [1/min]
RP-S1	13	0.05382
EFTu	51	0.01364
EFTs	59	0.01166

In addition, the same inactivation of protein T7 RNA polymerase as identified for the isolated enzyme (see section 3.4.4), was assumed to apply also for conditions of simultaneous

[(8)] Ribosome binding to poly-A columns are found to drop with increasing duration of cell-free protein synthesis (Siemann, personal communication).

transcription and translation. It remains unclear whether this assumption is valid, because experimental conditions of both systems may not be comparable, e.g., with respect to total ion concentration and total protein concentration.

8.5 Dynamic simulation

Figures 8.4 to 8.20 show the simulated time traces of selected quantities (mostly concentrations and reaction rates) characterizing cell-free synthesis of green fluorescent protein (GFP)[(9)] under batch conditions. The model applied combines reactions involved in

- mRNA synthesis
- mRNA degradation
- ribosomal translation
- energy regeneration
- and inactivation kinetics of proteins S1, EFTu, EFTs, and T7 RNA polymerase.

For those components where measurements were made (namely protein GFP, total mRNA, AcP, ATP, ADP, AMP, GTP, energy charge, EFTu, and EFTs), simulation results are compared to their experimentally determined counterparts (as described in Figures 8.4, 8.5, 8.6, 8.9, and 8.10, respectively).

The primary intention of this analysis was to investigate the predictability of the model in comparison to experimental data. Due to the number of states and parameters contained in the model, and the uncertainty afflicted with model constants taken from the literature, the ability to qualitatively predict measured results was of greater concern to the analysis, rather than a quantitative description of system behaviour. For this reason, no particular parameter estimation procedure was performed here. Nevertheless, some parameter values have been chosen manually such that they yielded an improved match between simulated versus experimentally determined results. All balance equations and parameters used together with the initial conditions of balanced concentrations are given in the Appendix D. Initial conditions were obtained by considering a 20-fold dilution of proteins and ribosomes in cell-free systems in comparison to a growing *E. coli* cell (Bremer and Dennis, 1996) (see also section 7.1 on page 89).

As can be seen from Figures 8.4 and 8.5, the predicted time-dependency of concentrations of protein GFP, full-length mRNA, and acetyl phosphate correspond quite favourably with the experiment. Both GFP and mRNA concentration increase, as they are synthesized with the time. Protein concentration is shown to level off after about one hour into the experiment. This is a consequence primarily due to the measured inactivation of ribosomal protein S1, with a half-life of 13 min (cf. Table 8.1). The concentration of acetyl phosphate is demonstrated to continuously diminish with the time, mainly caused by acetyl phosphate consumption through the acetate kinase reaction and its equivalents.

[(9)] Proteolysis is known to be negligible for protein GFP (Spirin, personal communication).

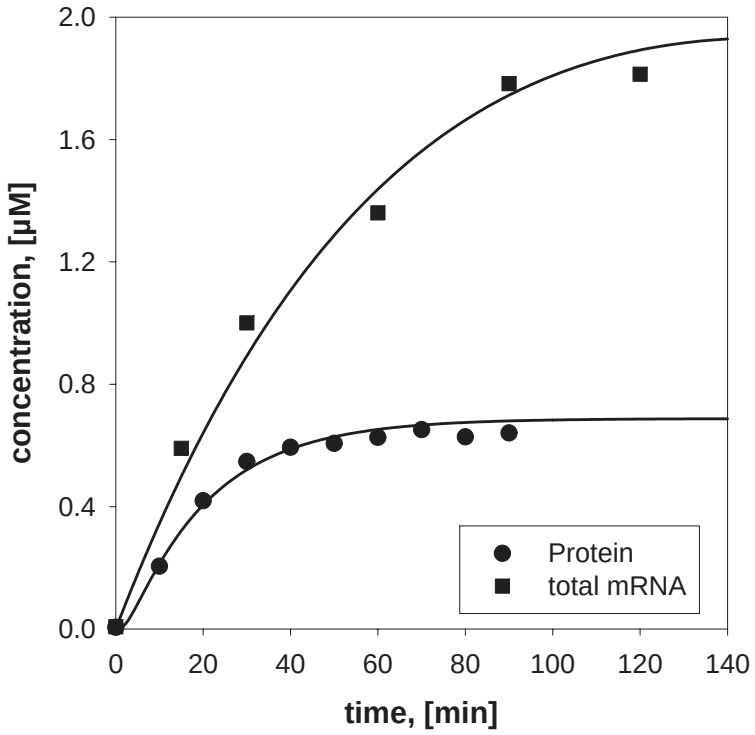

Figure 8.4: Time course of measured and predicted levels of protein GFP and full-length mRNA.

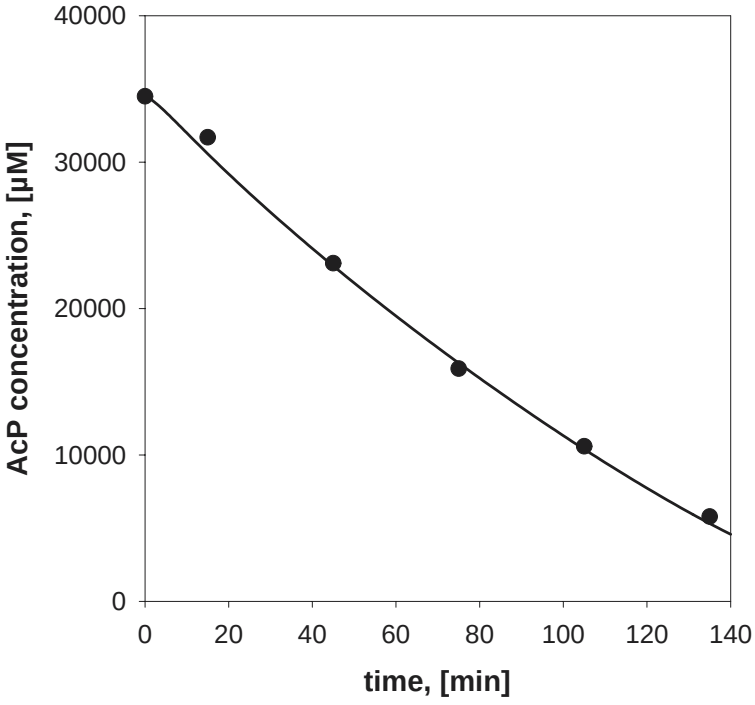

Figure 8.5: Time course of measured and predicted acetyl phosphate concentration.

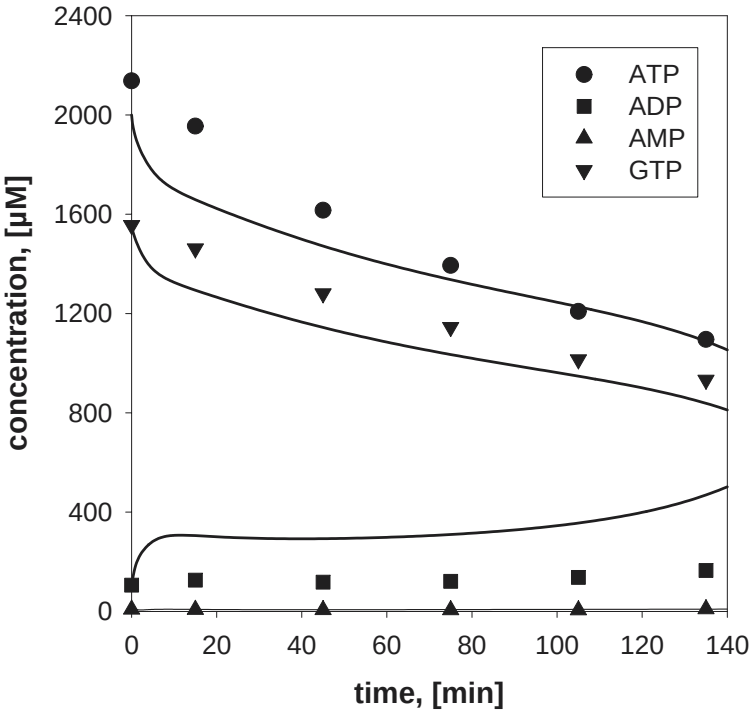

Figure 8.6: Time course of concentrations of ATP, ADP, AMP, and GTP. Comparison between simulation and experiment.

As a consequence of energy regeneration, it is possible to maintain sufficiently high levels of nucleotide concentrations. This is demonstrated in Figure 8.6, where the time course of concentrations of adenylates and GTP are displayed. In contrast to the results shown in this figure, for systems lacking energy regeneration, nucleotide concentrations are depleted within a few minutes only (Otto, 1996). Although prediction results exhibit a noticeable off-set from experimental data, tendencies and the order of magnitude of the displayed concentration courses are in agreement with the experiment. Furthermore, the model suggests an accelerated drop of ATP and GTP concentration, roughly within the initial 10 min of process time. Such a decrease is not mimicked by the corresponding experimental concentration curves. The observed discrepancy may be explained by a displacement of binding equilibria considered for the system at simulation start, and are thus a result of the chosen initial conditions. In particular, the sum of aminoacylation reactions (cf. Figure 8.7) appears to be responsible for the observed sharp decrease of NTP concentration. This finding may give some indication that the initial conditions of tRNA charging are probably over-estimated by the model.

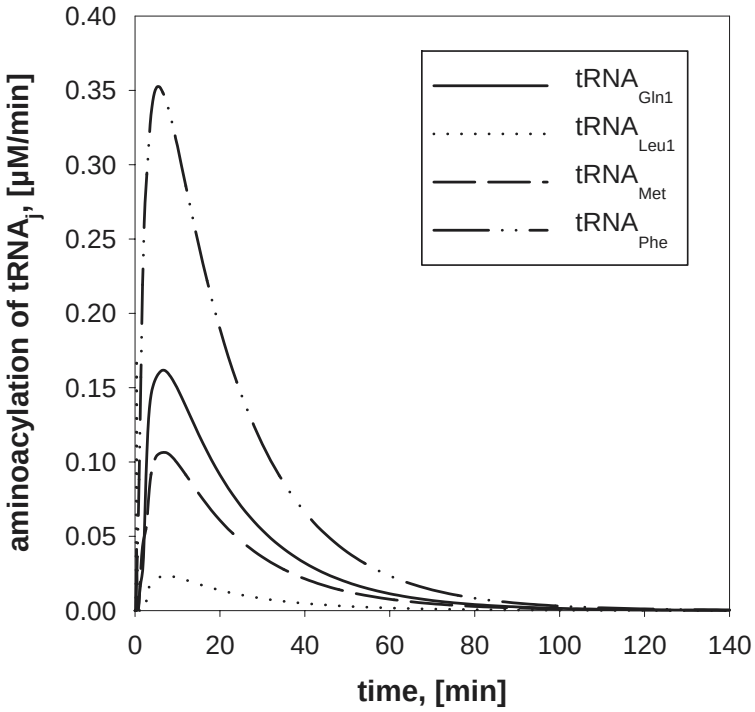

Figure 8.7: Time course of predicted rates of aminoacylation for selected tRNAs.

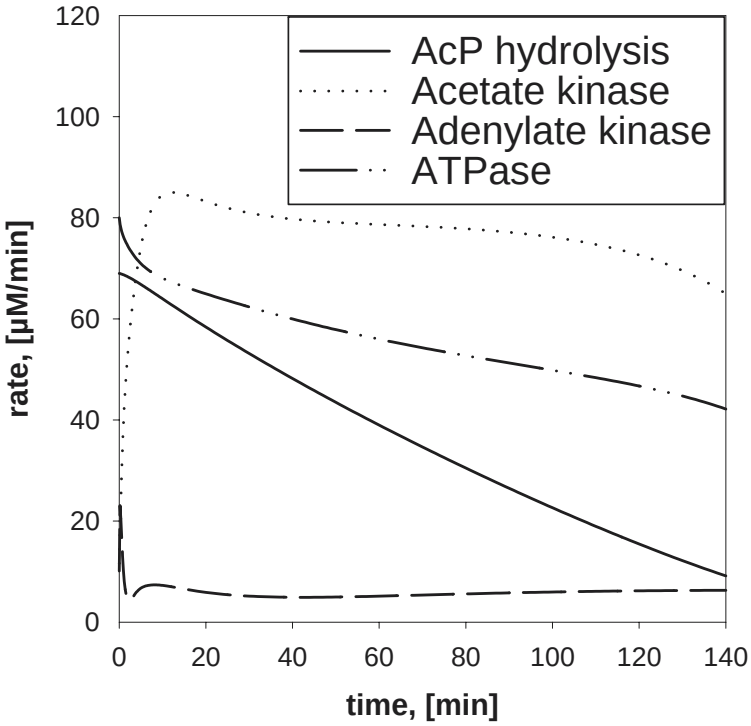

Figure 8.8: Time course of predicted rates involved in energy consumption and regeneration.

Figure 8.8 plots the predicted rates for selected reactions of the energy regeneration network. Both the rates of AcP hydrolysis and ATPase reaction are found to decrease with the time. On the other hand, the rates of acetate kinase and adenylate kinase are shown to remain approximately constant over two hours of process duration. Hence, the endogenous energy regeneration system is demonstrated to be capable of providing sufficient energy levels for at least two hours of process duration. This view is supported by the fact that the energy charge obtained from experimental data remained above 0.92 throughout the process (Figure 8.9). As a consequence of the mismatch described above between simulated and experimental nucleotide concentrations, the energy charge is slightly under-estimated by the model.

In Figure 8.10, the time-dependent trajectories of measured versus predicted total concentrations of elongations factors EFTu and EFTs are illustrated (cf. Schindler (2000) and Table 8.1). Both quantities show an exponential decay with the time due to inactivation. The low absolute levels of these elongation factors are striking, when compared to *in vivo* conditions. Under balanced growth, the concentrations of EFTu, EFTs, and EFG are by a factor of about 150, 20, and 20, respectively, higher than compared to the initial conditions of the investigated *in vitro* system (cf. Tables 6.2 and D.4.3). While the discrepancies for initial EFTs and EFG levels can be explained primarily by the dilution steps employed during lysate preparation, the preparation procedure apparently leads to a selective deprivation by EFTu concentration (Schindler, 2000). With progressing production time, the mismatch to ribosome concentration becomes increasingly severe, due to the noted inactivation of EFTu and EFTs, respectively.

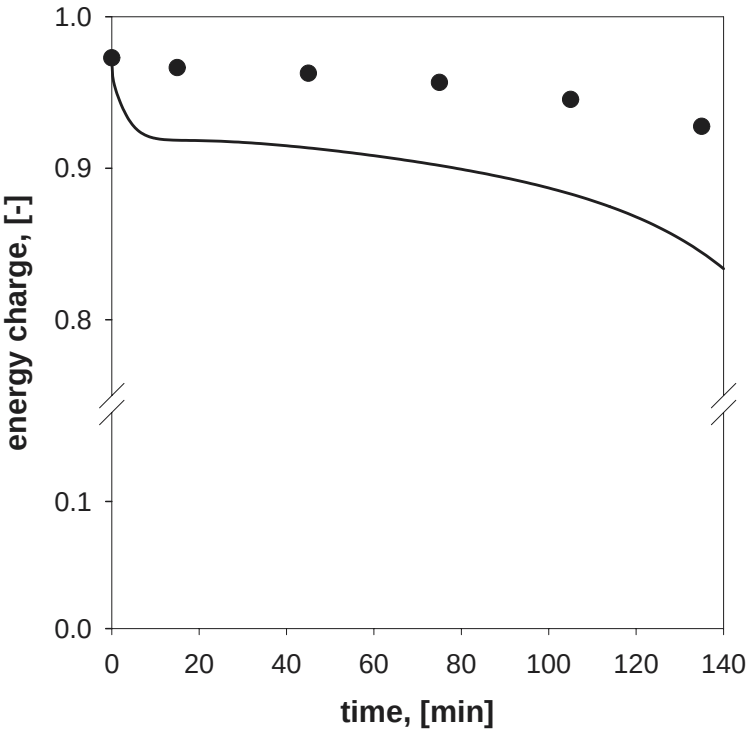

Figure 8.9: Time course of energy charge as defined by Atkinson (1968). Comparison between simulation and experiment.

The consequences of reduced EFTu levels are further reflected in Figure 8.11, where the simulated concentration courses of the various forms of $tRNA_{Leu5}$ are given versus the time. The sum of the displayed concentrations together with the corresponding tRNA-species bound to elongating ribosomes add up to roughly 0.26 μM at any instant during the process time (for there is no tRNA degradation considered here). As is obvious from this figure, the split ratio between Leu-$tRNA_{Leu5}$ and its corresponding ternary complex is very large. It increases from 16 to 115 over the course of the experiment. The predominant conformation in which this tRNA is predicted to exist is the aminoacylated form. This holds true also for the other 34 tRNA species considered (data not provided). In other words this means a highly unfavourable situation for elongation kinetics, since tRNA is required in terms of ternary complexes to serve as a substrate at each step of translation elongation. The average specific rate of ribosomal elongation, as sketched in Figure 8.12, is thus predicted to decline from about 2 aa/s to roughly 0.3 aa/s within almost 2.5 hours of experiment duration. *In vivo*, the average specific rate of peptide bond formation ranges between 10 to 20 aa/s. Hence, there exists an approximate 5 to 60-fold difference between specific protein synthesis rates obtained *in vivo* and the investigated *in vitro* system. These findings together strongly suggest the need for an appropriate supplementation of purified translation factors, here most importantly of EFTu, in order to maintain their catalytically active forms at levels necessary for efficient translation elongation.

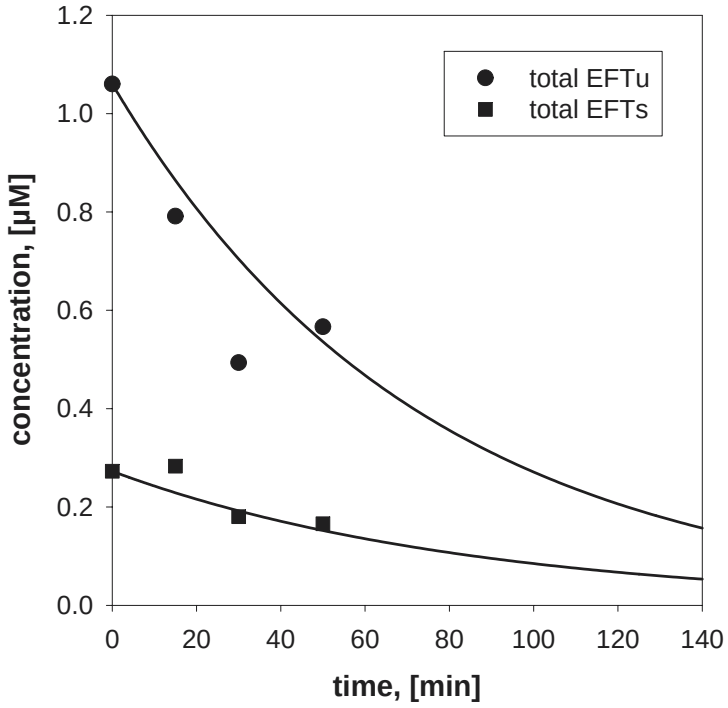

Figure 8.10: Time course of measured and simulated total EFTu and EFTs levels (see Table 8.1).

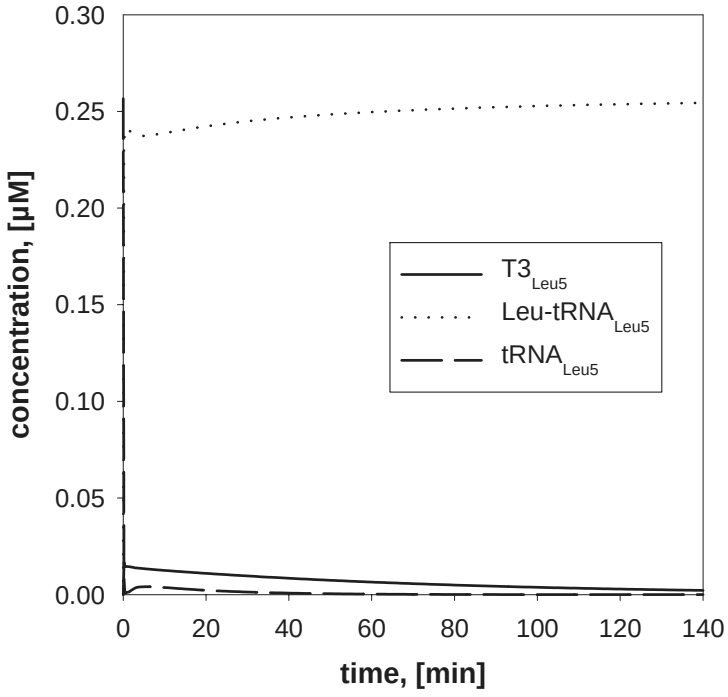

Figure 8.11: Time course of predicted concentrations of $tRNA_{Leu5}$ in its uncomplexed form, aminoacylated state (Leu-$tRNA_{Leu5}$), and as ternary complex ($T3_{Leu5}$). Initial concentrations (at $t = 0$) were 0, 0, and 0.2566 μM for $T3_{Leu5}$, Leu-$tRNA_{Leu5}$, and $tRNA_{Leu5}$, respectively.

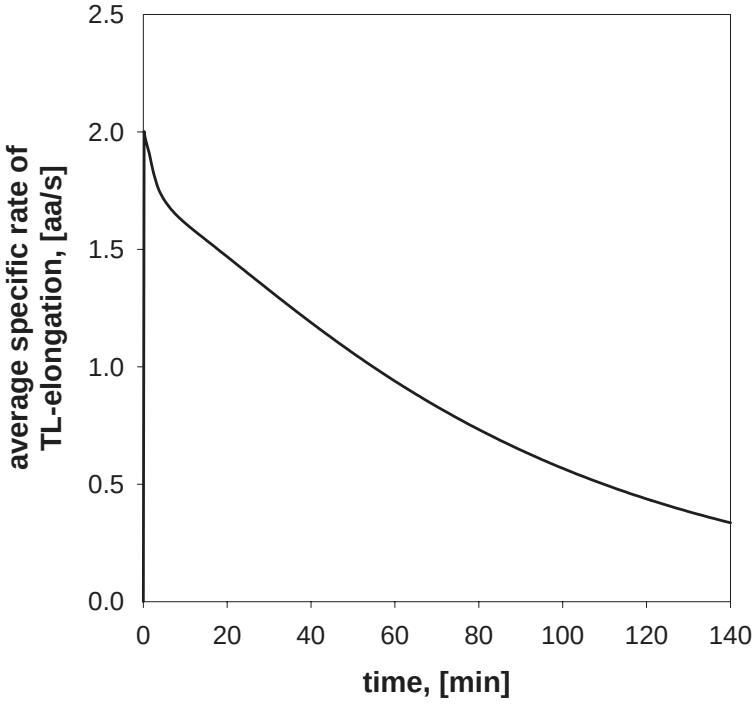

Figure 8.12: Predicted time course of average specific rate of translation elongation (per mRNA-bound ribosome). At $t = 0$, this rate is not defined (since there are initially no ribosomes bound to mRNA). It was taken to be equal to 0 then.

The rates of mRNA synthesis and degradosome association are both depicted in Figure 8.13. With declining nucleotide concentrations and due to inactivation of the enzyme T7 RNA polymerase (compare with section 3.4.4), the rate of transcription is found to diminish with the time. However, it is shown to remain above the rate of degradosome association throughout the displayed time period. Contrarily, the rate of degradosome association increases with the time. As can be viewed from the similarity to the time curve of mRNA concentration (see Figure 8.4), this rate is dictated by mRNA availability. The average specific rate of degradosome movement was predicted to be 31.7 codons/s in the investigated system and remained essentially constant across the entire process (data not shown).

After an initial period of about 10 minutes of the experiment, the predicted average gap between degradosomes settled at 690 codons (Figure 8.14). This means that in average one degradosome was bound roughly per two molecules of full-length mRNA (consisting of 357 base tripletts each). On the other hand, average ribosome densities indicated at maximum one ribosome bound per three native mRNA transcripts. This situation corresponds with the local minimum of ribosome spacing at $t = 3$ min displayed in Figure 8.15. During subsequent process times, ribosome spacing was found to increase exponentially, in agreement with the exponential slow-down of translation initiation introduced into the model (equation (8.7)). The average distance of translating ribosomes was at all times during the process predicted to be greater

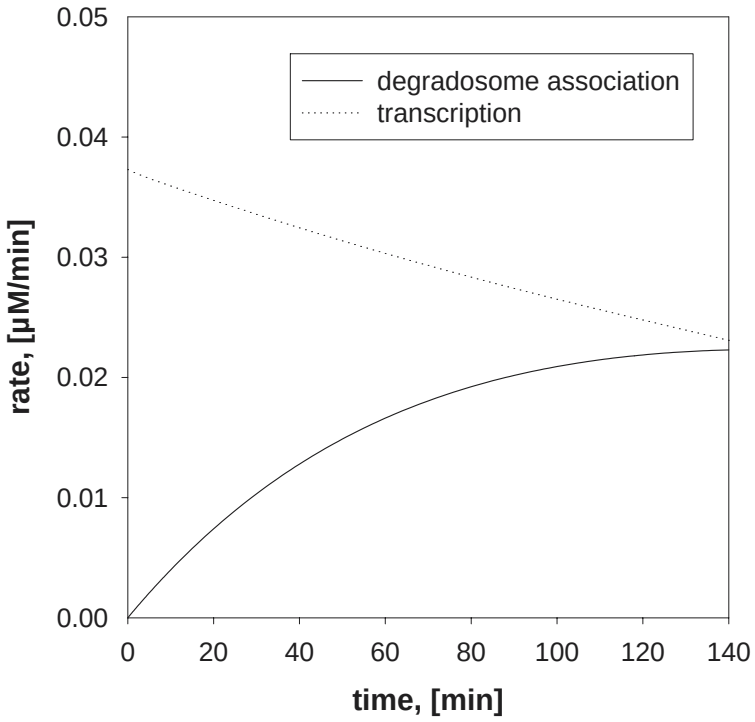

Figure 8.13: Time course of predicted rates of transcription and degradosome asociation.

than the average spacing between mRNA-bound degradosomes. At process termination after 140 min, there was only one ribosome bound per about 7,000 mRNA molecules according to the model (data not shown). These values compare to average ribosome distances of about 40 to 80 codons in a growing *E. coli* cell (see Table 6.2), a factor of about 100 lower than predicted for the *in vitro* system.

In the above, transcription rate was demonstrated to be able to compensate for the endogenous mRNA degradation processes. The choice of T7 RNA polymerase concentration added to the system even appears to be over-dimensioned, since lower mRNA levels in conjunction with higher ribosome densities could have well been tolerated. Higher ribosome loadings can function as an effective protection mechanism against ribonucleolysis (see chapter 5). In fact, excessive mRNA levels may not be desirable, since mRNA synthesis is highly energy consuming. Further, the pool of transcripts constitutes a significant sink for nucleotides. Material balancing revealed that the reduction in total nucleotide levels matched the nucleotide requirements for generating the measured mRNA concentration (data not provided). Therefore, even in the presence of a functioning co-factor regeneration system, that pushes nucleotide concentrations to their most phosphorylated state, the total sum of nucleotides is noted to decrease also with the time (see Figure 8.16). Hence, the noted drop of both concentrations of ATP and GTP (see Figure 8.6), and also of CTP and UTP (data not shown), can be explained with their incorporation into mRNA, instead of them being degraded.

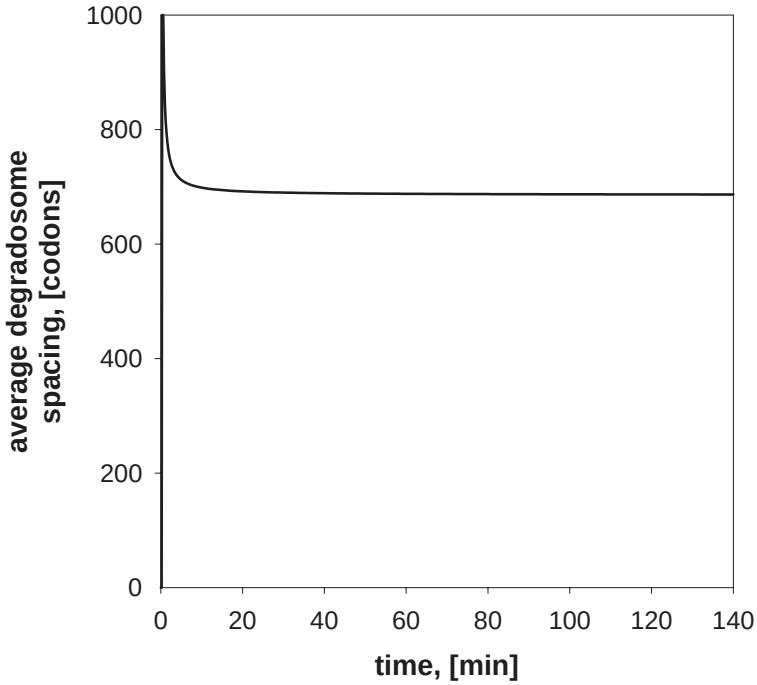

Figure 8.14: Predicted time-dependent average spacing between mRNA-bound degradosomes.

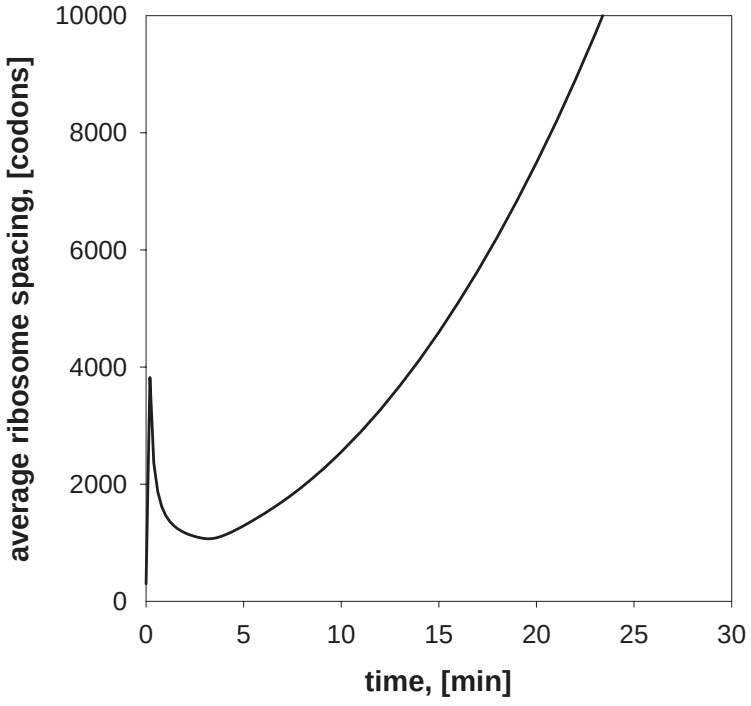

Figure 8.15: Predicted time-dependent spacing among mRNA-bound ribosomes.

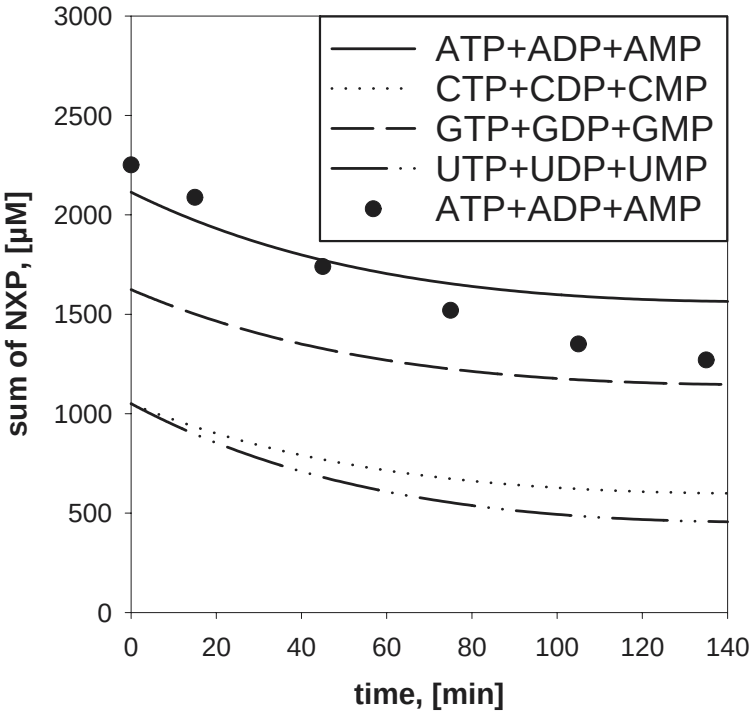

Figure 8.16: Time course of the predicted sum of concentrations of adenylates, cytidylates, guanylates, and uridylates, respectively. Measured total adenylate concentration is given also.

Low ribosome densities imply negligible sterical effects among translating ribosomes. This is in agreement with ribosomal queueing factors being predicted to be close to unity. As a representative constituent of all queueing factors for translation elongation, the time course of factor q_{14}^{R} is displayed in Figure 8.17. This factor demonstrates to remain virtually equal to 1 throughout the process. The only exception among all queueing factors where a significant difference from 1 was observed, at least temporarily, is the queueing factor for translation initiation (q_{22}^{R0}, depicted in Figure 8.17). This factor, denoting the probability for the ribosome binding site to be unoccupied, is shown to increase from about 0.80 at simulation start to a value of about 1 within the initial 10 minutes of process time. During this time interval, the concentration of mRNA is low, so that the fraction of occupied ribosome binding sites is greater than at subsequent process times, which then correspond with higher mRNA levels.

When investigating the dynamics involved in the loading process of an initially naked mRNA, interesting phenomena can be noted. As is visualized in Figure 8.18, the rates of translation initiation, elongation, and termination are shown to increase initially, as ribosomes are getting loaded onto the (earlier naked) mRNA. Elongation rates at codons 107 and 207 (as well as at the termination site (codon 273)) show a time-delayed response, which corresponds to the time gap needed for ribosomes to travel the distance between the initiation site and the respective codon (i.e., codons 107, 207, and 273). The trajectories of the rates of 70S initiation complex formation and IF2-dissociation are indistinguishable in this graph. Both of these rates

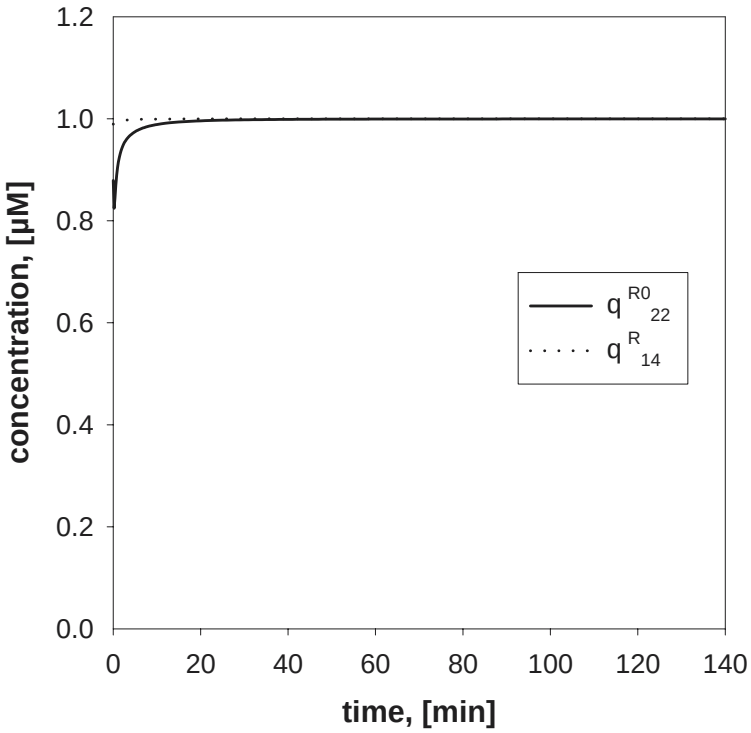

Figure 8.17: Predicted time courses of two selected queueing factors. q_{22}^{R0} denotes the probability for the ribosome binding site to be unoccupied. q_{14}^{R} represents the probability for forward movement onto codon 15.

reach a maximum, when the contribution of inactivation of ribosomal protein S1 just equals the effects of substrate availability on 70S initiation complex formation rate (cf. equation (6.12)), and are found to drop afterwards.

The step-wise propagation of ribosomes along the mRNA causes temporally spaced processes to take place, which are, e.g., reflected in the codon-specific elongation rates. Viewing the trajectory of each elongation rate as a frequency distribution, the mean of the distribution moves to higher values with increasing codon number, while the profile is smoothening. This is a behaviour generally observed for Poisson distributions, as was pointed out earlier (MacDonald et al., 1969).

Figure 8.19 shows the concentrations of ribosomes bound to the initiation codon (number 22), and to codon positions immediately after the start codon. As can be seen from this figure, the concentration of ribosomes representing 70S-initiation complexes (symbol R*22) is shown to be higher than the concentration of ribosomes that are bound to this position after IF2-dependent GTP hydrolysis (state R22). Ribosomes occupying the initiation site, thus effectively function as a road-block, in a sense that they prevent upstream propagating degradosomes from getting access to endonucleolytic cleavage sites contained within the coding region (cf. chapter 5).

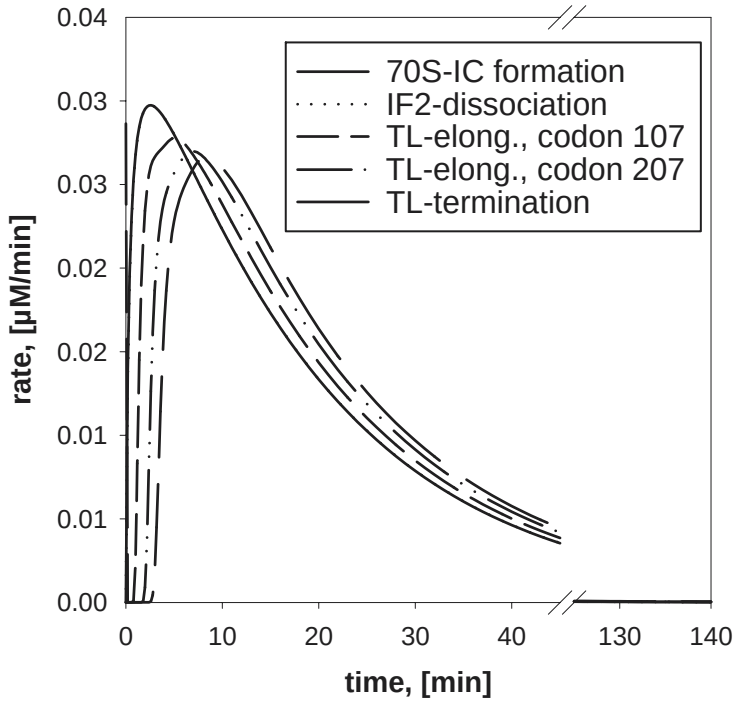

Figure 8.18: Time course of rates of translation initiation, elongation, and termination.

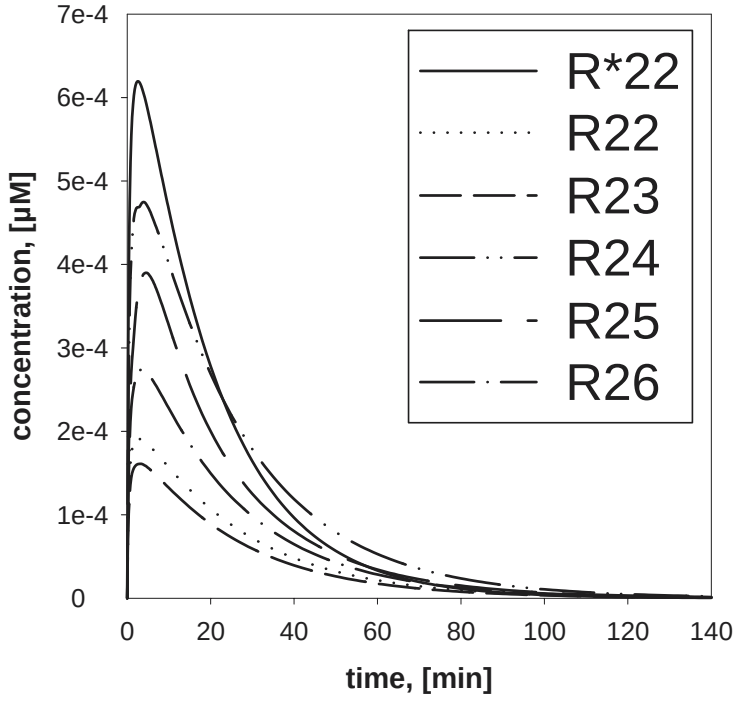

Figure 8.19: Time course of concentrations of mRNA-bound ribosomes at selected codons in the vicinity of the start codon (number 22). Symbols R*22 and R22 distinguish ribosomes bound to the initiation codon prior and subsequent to IF2-dissociation, respectively.

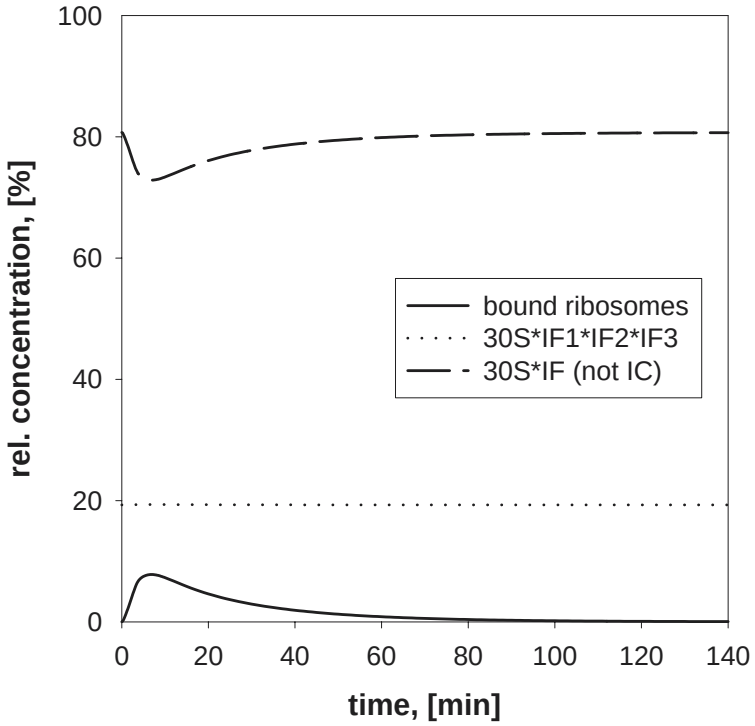

Figure 8.20: Time course of relative ribosome concentrations.

Furthermore, it should be mentioned that time profiles displayed in Figure 8.19 are not exactly Poisson-distributed, as could have been intuitively thought. This follows as a direct consequence of variable codon-specific elongation rates. Ribosomal loading patterns will thus evolve that compensate for these codon-specific differences. Explicitly this means that codons corresponding to relatively lower specific elongation rates will show a higher ribosome loading, in order to maintain volumetric elongation rates that are equal for all codons j during pseudo-steady state synthesis conditions.

In Figure 8.20, the predicted relative concentrations of ribosomes bound to mRNA, ribosomal subunits bound to all three initiation factors simultaneously (symbol $30S \cdot IF1 \cdot IF2 \cdot IF3$), and the remainder of 30S subunits (i.e., freely dissolved and complexed with any one or multiple, but not all initiation factors simultaneously) are plotted. The sum of these three quantities adds up to 1 at any point of time during the process. Over the entire time course, about 80 % of all ribosomes are predicted to be in a state that is neither bound to mRNA, nor complexed at the same time with all three initiation factors. The concentration profile of this pool shows a slight drop roughly within the initial 20 minutes, as ribosomes get loaded onto mRNA. Most noticeably, the concentration of complex $30S \cdot IF1 \cdot IF2 \cdot IF3$ stays virtually unaffected by the dynamics of translation. It takes a value of about 20 % of total ribosome concentration. The concentration of $30S \cdot IF1 \cdot IF2 \cdot IF3$, however, enters the rate of 70S initiation complex formation in a linear fashion (see equation (6.12)). As the optimization analysis carried out in section 7.1 suggested, the equilibrium between $30S \cdot IF1 \cdot IF2 \cdot IF3$ and the non-active forms

of 30S (i.e., complexed with less than all three initiation factors) could be favourably shifted at higher levels of initiation factors, so that ideally all ribosomes unbound from mRNA would exist as complex 30S·IF1·IF2·IF3. In this case, the initial volumetric rate of protein synthesis could theoretically be raised by a factor of maximally 5, unless further rate-limitations exist.

8.6 Optimization of translation factor levels

As one of the results obtained from simulating cell-free GFP production in the previous section, dilute translation factor levels were proposed to be the primary cause for the observed low protein production rates. In order to further investigate this hypothesis and to check whether higher total translation factor levels would lead to a performance improvement, the previously described model was subjected to a sequence of raised initial concentrations of total translation factors, and the resulting system dynamics were simulated. The reference, to which elevated initial translation factor concentrations are compared, is the same as for the cell-free protein synthesis system described in section 8.5.

In the following analysis, the impact of selectively increasing (A) the concentration of elongation factor EFTu, (B) the concentrations of all elongation factors simultaneously, (C) the concentration of all initiation factors, and (D) at the same time the concentrations of both all initiation factors and elongation factors was investigated. The initial conditions of the respective simulations are compared in Table 8.2. Importantly, all other reaction conditions and initial concentrations were kept the same as in the reference system. The time-dependent inactivation of selected compounds identified earlier (see Table 8.1) was considered here also.

8.6.1 Effect of elongation factor concentration

Figure 8.21 shows the predicted time traces of average specific rate of translation elongation obtained for varied total EFTu concentrations. As can be seen from this graph, increasing levels of EFTu are proposed to lead to a significant enhancement of average specific ribosome propagation rate. Already a doubling of EFTu concentration at simulation start, is predicted to give a - by a factor of 1.8 - higher average specific elongation rate at $t = 0$ (dotted line) than in comparison to the reference condition (solid line). This finding indicates an almost 1:1 improvement and suggests that in the earlier scenario, indeed EFTu concentration was most limiting this rate. At EFTu levels equal to and higher than by a factor of 20 in comparison to the reference system (section 8.5), the average rate of ribosome elongation is predicted to reach a maximum of 11.5 aa/s. This rate lies in the order of *in vivo* specific rates of peptide bond formation (10 to 20 aa/s). Thus, by increasing EFTu concentration, the stringent limitations of specific elongation rate noted earlier could in theory be successfully overcome, until further rate-limitations apply (that set the upper-boundary threshold shown in Figure 8.21). The causes for why the average specific rate of translation elongation reaches a maximum in this case, have not been further investigated in this study. Equipped with the aid of the gene expression model, however, it would now be interesting to analyze step-by-step subsequent rate-limitations as they occur during cell-free protein synthesis systems, and to suggest further counter-measures

Table 8.2: Initial conditions applied for simulating cell-free protein synthesis during optimization of translation factor concentration. The reference condition refers to the simulation study of section 8.5. A - 30-fold EFTu concentration in comparison to the reference state. B - All EF concentrations are raised by a factor of 30, respectively. C - Optimum IF levels taken from scenario IV of Table 7.1. D - Simultaneous increase of both concentrations of initiation factors and elongation factors.

Concentration (μM)	**Reference**	**A**	**B**	**C**	**D**
$C_{30S_{tot}}$	1.40	1.40	1.40	1.40	1.40
$C_{EFG_{tot}}$	1.21	1.21	36.4	1.21	36.4
$C_{EFTu_{tot}}$	1.06	31.8	31.8	1.06	31.8
$C_{EFTs_{tot}}$	0.27	0.27	8.18	0.27	8.18
$C_{IF1_{tot}}$	0.38	0.38	0.38	1.67	1.67
$C_{IF2_{tot}}$	0.45	0.45	0.45	1.28	1.28
$C_{IF3_{tot}}$	0.30	0.30	0.30	1.65	1.65
C_{30S}	0.0065	0.0065	0.0065	0.003	0.003
C_{50S}	0.32	0.32	0.32	1.24	1.24
C_{IF1}	0.07	0.07	0.07	0.45	0.45
C_{IF2}	0.11	0.11	0.11	0.07	0.07
C_{IF3}	0.01	0.01	0.01	0.42	0.42
C_{EFG}	0.02	0.02	0.66	0.02	0.66
$C_{EFG\cdot GTP}$	0.78	0.78	24.8	0.78	24.8
$C_{EFG\cdot GDP}$	0.41	0.41	10.9	0.41	10.9
$C_{EFTu\cdot GTP}$	0.71	25.6	26.2	0.71	26.2
$C_{EFTu\cdot GDP}$	0.35	6.26	5.62	0.35	5.62
C_{GTP}	1548.5	1530.0	1505.0	1548.5	1505.0
C_{GDP}	75.2	72.2	61.3	75.2	61.3

on the basis of simulation analysis.

When in addition to EFTu concentration, also the initial levels of elongation factors EFG and EFTs were raised by a factor of 30 (scenario B in Table 8.2), no further performance improvement was noted. The final concentration of protein product, as well as translation initiation rate, the specific rate of translation elongation, and the fractional splitting among ribosomes were predicted to be the same as for the system of increasing EFTu concentration only (cf. Table 8.3).

Notably, time profiles of concentration of protein product GFP, are the same for the investigated systems of raised EFTu concentrations and for the system of increasing all EF concentrations simultaneously (data not provided). They are all virtually identical to the time profile of synthesized GFP that is displayed in Figure 8.4. Also, the final concentration of protein product achieved after 140 minutes of process time is predicted to be approximately identical (i.e., equal to 0.70 μM) across all of the different systems with elevated EF concentrations. The effects of raising total EF concentration showed exclusively at an increased specific translation elongation rate. This finding simply means that elongating ribosomes travel relatively faster along the mRNA under conditions of raised EF concentration. The number of mRNA-bound ribosomes remains, however, unchanged to the system of non-elevated EF concentration, and

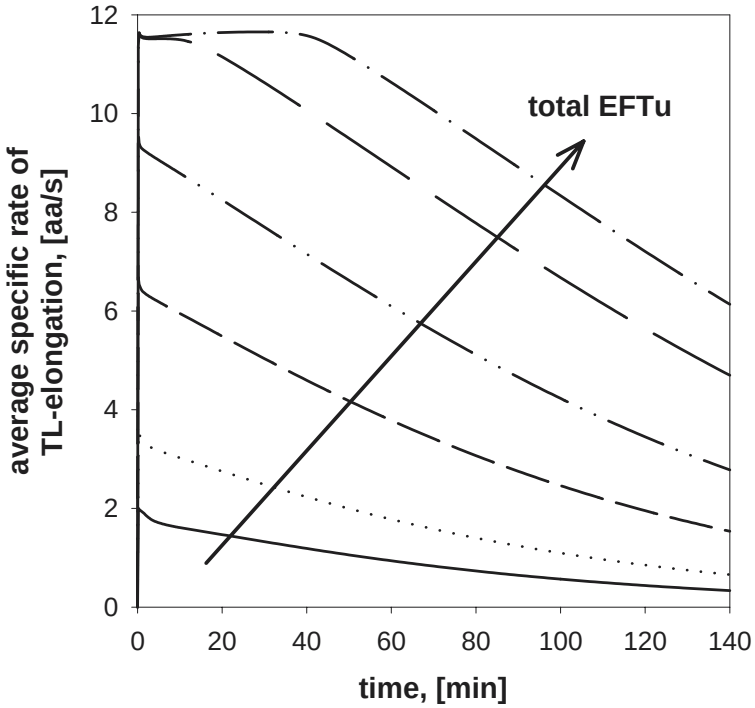

Figure 8.21: Impact of EFTu concentration on the average specific rate of translation elongation (i.e., per mRNA-bound ribosome). The solid line is replotted from Figure 8.12. The other trajectories correspond with initial total EFTu concentration increased by a factor of 2, 5, 10, 20, and 30, respectively, in comparison to the reference conditions described in section 8.5.

per unit of time, the same number of GFP molecules is completed.

As demonstrated, an enhancement of *specific* protein synthesis rate is not necessarily sufficient to ensure at the same time improved *volumetric* protein production rates. Raising volumetric productivities is generally achieved by further concentrated catalyst levels. In the case of protein synthesis, this is equivalent to driving ribosomes to a mRNA-bound state. Higher ribosome densities are expected to occur at higher rates of translation initiation. Due to the earlier noted excess of freely dissolved ribosomes in contrast to their active form as a complex with initiation factors, raised IF concentrations are expected to yield higher rates of translation initiation (as predicted in section 7.1). Thus, the impact of increasing initiation factor concentration on protein synthesis rate is examined in the sequel.

8.6.2 Effect of initiation factor concentration

In section 7.1, an improvement of volumetric protein production rate was suggested to be obtained through raising initiation factor levels in an appropriate stoichiometric relation to total ribosome concentration. This working hypothesis was subsequently tested by simulating cell-free protein synthesis dynamics with the set of optimum initial concentrations suggested for scenario IV in that earlier section (which corresponds to condition C in Table 8.2).

Table 8.3: Results of simulating cell-free protein synthesis during optimization of translation factor concentrations. C_{Prot} is the protein concentration at $t = 140$ min. Other quantities displayed were taken at time $t = 2$ min, respectively. All of these quantities remained essentially constant throughout the process, except for the average specific rate of elongation $(k_{TLE})_{avg}$, which decreased with the process time. A - 30-fold EFTu concentration in comparison to the reference state. B - All EF concentrations are raised by a factor of 30, respectively. C - Optimum IF levels taken from scenario IV of Table 7.1. D - Simultaneous increase of both concentrations of initiation factors and elongation factors.

Condition	C_{Prot} (μM)	$V_{TLI,70\mathrm{SIC}}$ (μM/min)	$(k_{TLE})_{avg}$ (aa/s)	$\frac{C_{30\mathrm{S}\cdot\mathrm{IF1}\cdot\mathrm{IF2}\cdot\mathrm{IF3}}}{C^R_{tot}}$ (%)	$\frac{C^R_{bound}}{C_{70\mathrm{S}tot}}$ (%)
Reference	0.69	0.03	1.9	19.3	3.4
A: EFTu	0.70	0.03	11.5	19.3	0.8
B: EF	0.70	0.03	11.5	19.3	0.8
C: IF	n.d.	0.10	1.7	85.2	11.2
D: IF + EF	3.17	0.13	10.7	88.5	4.1

Under these conditions, indeed an improved rate of 70S initiation complex formation is noted. This translation initiation rate of 0.10 μM/min is predicted to be 3.5-fold higher than the corresponding rate of the reference simulation (0.03 μM/min) (see Table 8.3). As can be viewed from further data provided in this table, the enhancement can be explained by a favourable shift of non-translating ribosomes towards full complexation simultaneously with all three initiation factors considered (an increase from 19.3 % to 85.2 %). This compound enters the rate of 70S initiation complex formation in a linear fashion (equation (6.13)). Interestingly, however, numerical integration was found to cease only after 4 min of simulated process time. In the applied situation, ribosomes showed a tendency to get stalled when bound to mRNA, due to a lack of sufficient amounts of elongation factors that would promote the rate of translation elongation. Apparently, sterical interactions among translating ribosomes were noted to propagate backwards to the ribosome binding site (data not provided), which altogether ultimately led to a premature simulation termination. This finding indicates that at higher rates of translation initiation, sufficiently high specific rates of translation elongation become increasingly important, for they can ensure a sufficiently high rate of clearance of the ribosome binding site.

Consequently, in a next step of the opimization strategy, both the concentrations of initiation factors and elongation factors were raised simultaneously (scenario D in Table 8.2). Under these conditions, a tremendous improvement of cell-free protein synthesis was predicted. Figure 8.22 shows a comparison of predicted time profiles of product protein concentration for the reference simulation in comparison to the situation where both levels of translation initiation and elongation factors were optimized. As can be seen from this figure, the final concentration of protein product was simulated to reach a level of 3.17 μM (in contrast to 0.69 μM obtained in the reference system). The initial rate of translation initiation was 0.13 μM/min (compared to

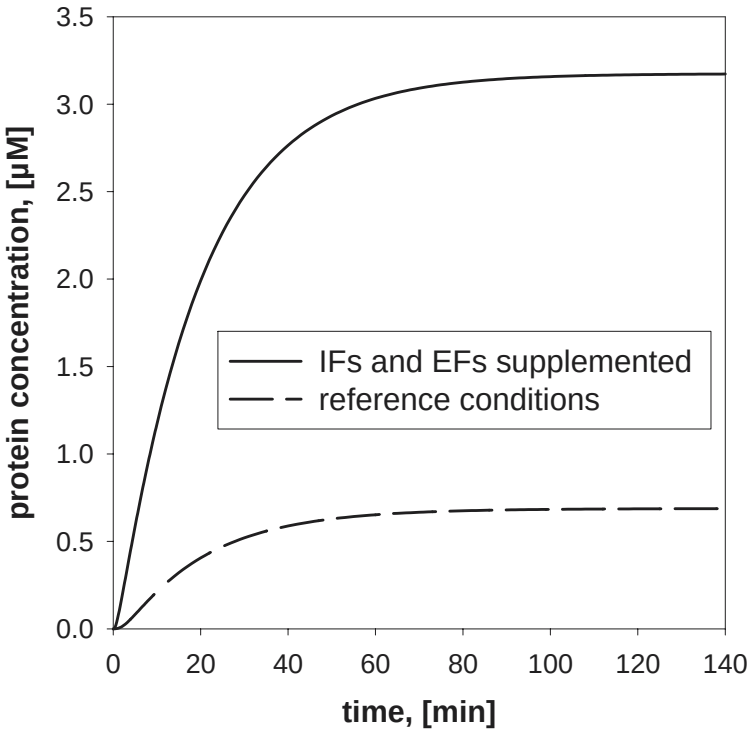

Figure 8.22: Time profile of protein concentration under reference conditions and for a system with combined supplementation of initiation factors (IF1, IF2, and IF3) and elongation factors (EFTu, EFG, and EFTs).

the reference rate of 0.03 μM/min). The concentration of 30S ribosomal subunits that exists in a complex with all initiation factors taken into account simultaneously calculates to be 88.5 % in this case (formerly 19.3 % in the reference system). All three quantities, C_{Prot}, $V_{TLI,70\mathrm{SIC}}$, and the fractional amount of complex 30S·IF1·IF2·IF3, showed a 4.6-fold increase in comparison to the reference condition (Table 8.3). In this case, the average specific rate of translation elongation was suggested to be 10.7 aa/s, which reaches up to *in vivo* levels (10 to 20 aa/s, cf. Table 6.2).

In summary, only a combination of simultaneously increasing the levels of both translation initiation and elongation factors (in comparison to the reference) is suggested by the model to significantly improve both specific and volumetric protein production rates.

8.7 Conclusions

A sufficiently detailed dynamic model is indispensable to allow for both a distinction among and quantification of individual contributions of reactions affecting the overall measured protein synthesis rate. The combined gene expression model provides time courses for a number of key state variables, reaction rates, and characteristic process parameters for performance assessment, that are extremely tedious to access on a large scale experimentally.

Mathematical modelling was shown to be a valuable tool for the design of protein production processes, e.g., involving an optimization of stoichiometric catalyst concentrations. As an immediate result of dynamic simulation, some of the factors usually speculated to limit protein production rate, could be excluded in the investigated cell-free system, as is outlined in the following.

Endogenous acetate kinase, a key enzyme in the energy regeneration network, was demonstrated to be sufficiently active to compensate for the observed high rates of nucleotide consumption. Under these conditions, a virtually complete regeneration of all necessary ribonucleotides could be achieved, maintaining an energy charge of ≈ 0.92. Furthermore, mRNA concentration was kept at high levels almost throughout the process and thus did not directly impose any restriction on protein synthesis rate. The issue of sterical interactions among translating ribosoms was suggested to be of lesser importance in the described system, which was deduced from the sparse ribosome densities predicted. The total fractional portion of mRNA-bound ribosomes was suggested to be elevated under conditions, where translational initiation and elongation factors are raised.

The problem of improving protein production rates is not only a matter of establishing higher total catalyst concentrations. Their appropriate relative ratios were shown to be essential for efficient translation. The question whether it is in theory possible to achieve higher protein synthesis rates *in vitro* than are measured under *in vivo* conditions, could not be fully resolved in this study. Nevertheless, when all necessary substrates and translation factors are supplied at saturating conditions, specific elongation rates are likely to be bounded by a fixed limit that is given by thermodynamic constraints. Whether such a limit is reached under *in vivo* conditions, remains unclear.

On the other hand, when considering volumetric protein production rates *in vitro*, it may be possible to reach up to the level of synthesis rates measured in the growing cell. The volumetric productivity is crucially dependent on the concentration of main catalysts. In gene expression systems, that is in particular the concentration of mRNA-bound ribosomes, given substrates and translation factor levels are provided in excess. The combined effects of appropriate relative levels of all translation factors are suggested in this study to be necessary for obtaining both maximum specific and maximum volumetric protein synthesis rates. It is intriguing to speculate that under conditions of ideally matching translation factor ratios and employing at the same time further concentrated *in vitro* systems, productivities may reach even beyond cellular volumetric protein synthesis rates.

9 mRNA Secondary Structure Predictions

mRNA secondary structure plays a critical part in determining both mRNA stabilities (Coburn and Mackie, 1999), and translation efficiency (de Smit and van Duin, 1990a). Experimental detection methods for elucidating RNA secondary structure are extremely tedious, typically involving X-ray crystallography (Baeyens et al., 1996) or NMR spectroscopy (Cheong et al., 1996). Alternatively, RNA secondary structures may be predicted on the basis of calculating the minimum free energy of structure formation (see section 2.5).

Due to the significance of mRNA secondary structure for gene expression performance, are the effects of structure prediction on protein synthesis rate investigated in the following. In an initial part of this chapter, the impact of sequence length on the results of structure prediction is analyzed. Subsequently, an attempt is made to draw a relationship between predicted mRNA secondary structures and their influence on measured functional mRNA half-lives. A substantial part of this chapter is devoted to an assessment of secondary structural properties encountered within the ribosome binding site and their effects on protein expression rate.

9.1 The *mfold* program

All RNA secondary structures predicted in this study were computed using the *mfold* package developed by Zuker et al. (1999). This program uses the thermodynamic rules of the nearest neighbour model and is based on updated energy parameters (Mathews et al., 1999). In addition to the optimum (i.e., minimum) free energy structure, a multitude of suboptimal foldings is typically output with this method. Up to 50 suboptimal foldings are calculated with the default program settings. Among these foldings, the degree of conservation of certain double-stranded regions may serve as a good estimate for mRNA secondary structure predictions.

9.2 Effect of length of unshielded mRNA

Prior to structure analysis, a choice needs to be made whether the full length of a particular mRNA should be considered or only an excerpt of the entire sequence. Answering this question may not be a trivial task. In special cases only, like for instance for *in vitro* RNA synthesis systems employing an isolated RNA polymerase, is the full-length transcript product expected to take a fully-folded conformation. For most of the cases, i.e., under conditions of simultaneous ribosomal protein synthesis, some regions of mRNA sequence are shielded by bound ribosomes. Evidently, ribosome binding prevents locally from structure formation, and may also affect longer-distance base-pair interactions. Altered conformations of overall mRNA secondary structure are obtained in this case, depending on whether they can fold in an unhindered fashion, or else are influenced by the dynamics of ribosomal translation.

In order to investigate the impact of a chosen sequence length that enters RNA secondary structure prediction, lengths between 30 to 90 nucleotides were examined for the 5′-terminal regions of both *lacZ* and *lacY* mRNA, respectively. The results of this analysis are shown in Fig-

ures 9.1 and 9.2. For *lacZ* mRNA, the great majority of structures suggest the 5′-motif displayed in Figure 9.1(a) to remain conserved. This holds true for sequence lengths 30 to 70 nucleotides, and also for 90 nucleotides, respectively. However, for a sequence length of 80 nucleotides, a total of two structural conformations was obtained as a result of structure prediction. The minimum free energy structure of these two conformations is given in Figure 9.1(h). Both the optimum and suboptimal folding structure showed pronounced base-pairing interaction with further distant regions of the mRNA molecule, which led to structure predictions that did not contain the ultimate 5′-stem-loop. In contrast to *lacZ* mRNA, the ultimate 5′-terminal region of *lacY* mRNA is predicted to be the same, independent of sequence lengths ranging from 30 to 90 nucleotides (see Figure 9.2).

In summary, clearly sequence length showed a significant influence on the outcome of RNA secondary structure predictions. The degree of conservation of particular stem-loop structures is expected to depend on sequence context and thus to vary with the nucleotide sequence under investigation.

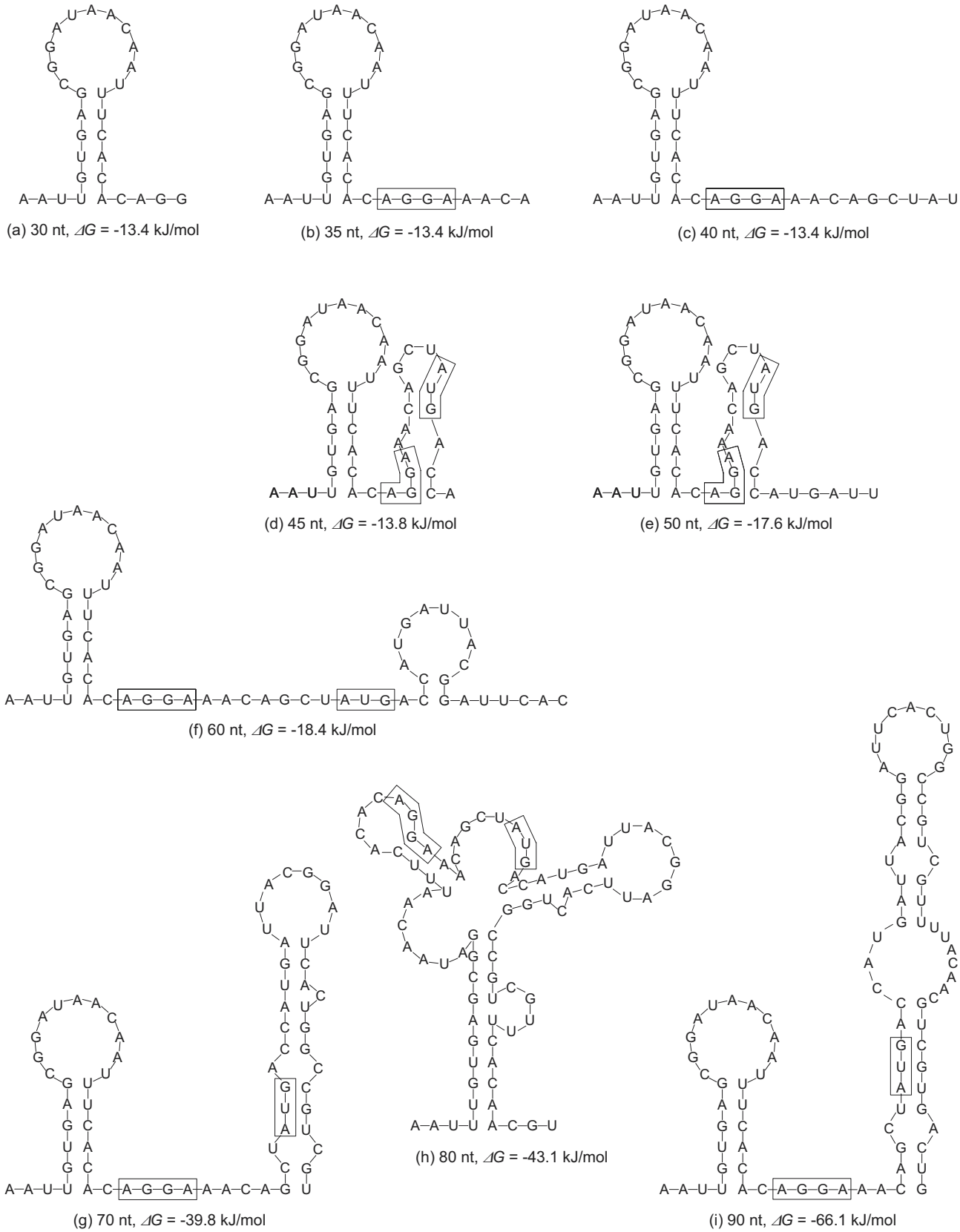

Figure 9.1: Predicted secondary structures for 5′-end of *lacZ* mRNA (varied sequence length). Shine-Dalgarno sequence and translational start codon are indicated by boxes.

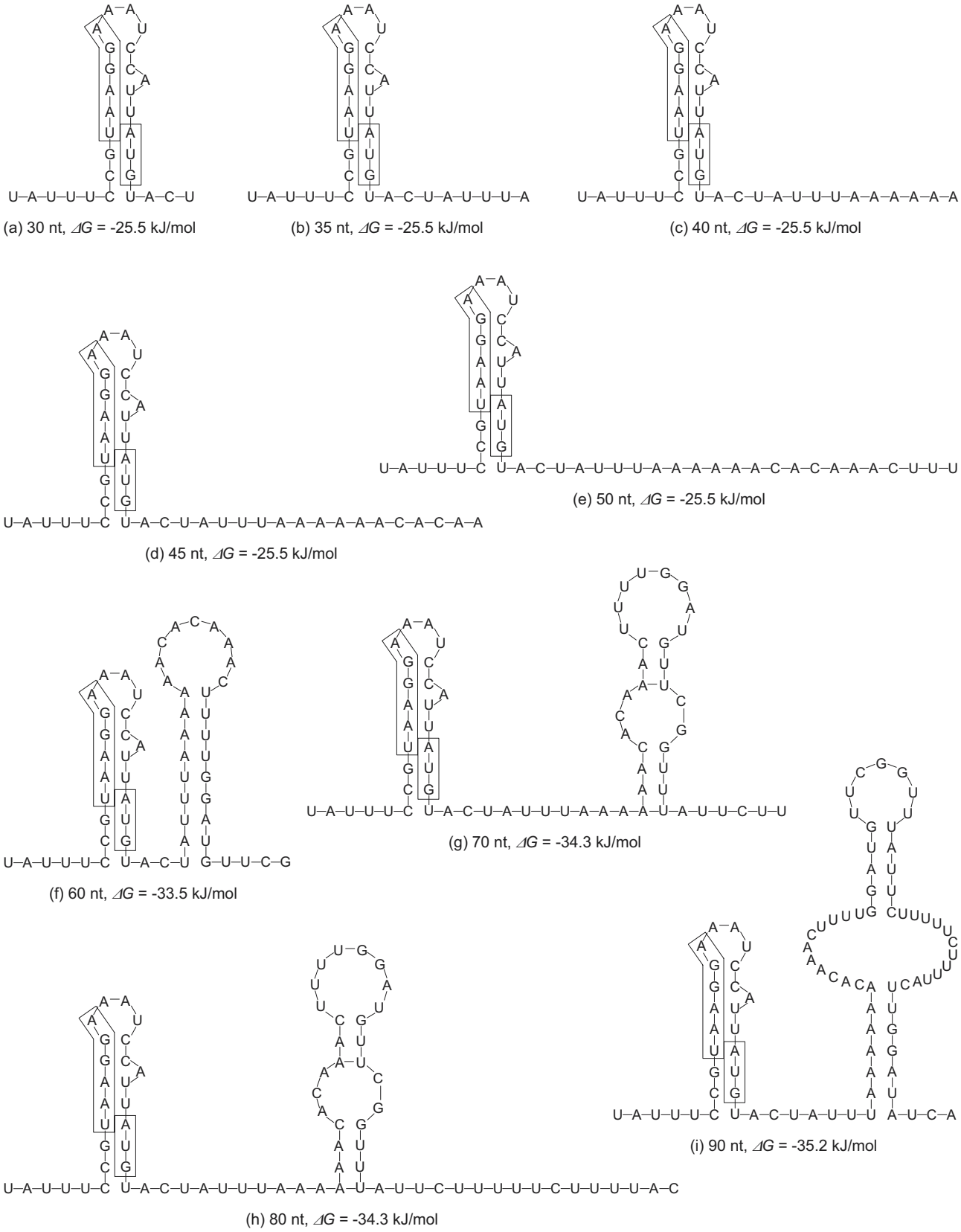

Figure 9.2: Predicted secondary structures for 5′-end of *lacY* mRNA (varied sequence length). Shine-Dalgarno sequence and translational start codon are indicated by boxes.

9.3 Effect of secondary structure at 5′-end

The impact of 5′-mRNA secondary structure on mRNA stability was investigated previously by Carrier and Keasling (1999). These authors created a library of 5′-untranslated regions and tested their influence on functional message stability, when inserted into a *lacZ*-containing vector system. The experimentally observed functional mRNA half-lives together with the free energies of secondary structure formation provided in that study are revisited in Table 9.1. As can be seen from that table, mRNA half-lives observed within the investigated hairpin library spanned almost one order of magnitude (from 2.1 min to 19.8 min). Although a direct correlation between measured functional mRNA half-lives and predicted free energies of secondary structure formation could not be made, some trends within different classes of mRNA structures were detected in the original study.

Table 9.1: Juxtaposition of free energy computation with functional mRNA half-lives. Predictions using MulFold 2.0 (Jaeger et al., 1989) together with the presented half-lives were reported by Carrier and Keasling (1999). ΔG computations with *mfold*3.1 (Zuker et al., 1999) are a result of this study.

Plasmid	$t_{1/2}$ [min]	MulFold 2.0 ΔG [kJ/mol]	*mfold*3.1 ΔG [kJ/mol]
pHP15	2.1	-39.8	-39.8
pHP14	4.9	-57.3	-53.6
pHP9	5.5	-87.1	-74.1
pHP5	5.6	-110.0	-117.6
pHP8	6.1	-75.3	-87.9
pHP18	6.3	-65.7	-64.0
pHP10	6.8	-91.3	-99.0
pHP4	8.3	-69.1	-69.1
pHP16	12.5	-104.7	-113.0
pHP17	19.8	-66.6	-67.4

When applying *mfold*3.1, a more recent algorithm for RNA secondary structure predictions than the program MulFold 2.0 (Jaeger et al., 1989) used in the earlier study, similar (but not identical) folding energies are obtained. As can be viewed from Table 9.1, the largest deviations are seen for plasmids pHP9 and pHP8 with a 13.0 and -12.6 kJ/mol absolute difference, respectively, in their computed free energies. Although these differences may be important, when considering that a stability increase by -5.9 kJ/mol changes the equilibrium between folded and unfolded RNA conformation by approximately one order (de Smit and van Duin, 1994a), the standard free energies calculated with the new prediction method are likewise not able to explain the observed range of functional mRNA stabilities. Hence, one has to conclude that functional mRNA stabilities can at present not be derived - in a straightforward manner - from computed free energies of folding.

Discrepancies between measured half-lives and predicted folding energies were attributed earlier to the putative prevalance of additional degradative reactions that were not considered in

the assumed mRNA degradation pathway (Carrier and Keasling, 1999). It was then speculated that RNA processing by RNase III could be a significant factor in modifying mRNA stabilities. This enzyme is known to cleave double-stranded RNA that may often contain an internal unpaired bubble, which characterizes the preferred site of cleavage (Nicholson, 1999). Due to the high substrate specificity of RNase III (Krinke and Wulff, 1990), however, this enzyme may not in all of the above investigated cases apply to explain the deviations between observed mRNA half-lives and predicted free energies. It is likely that further mechanisms play a role in rendering mRNA stabilities. Recent studies suggest transient conformations attained over the course of RNA folding to significantly influence RNA functionality (Treiber and Williamson, 2001; van Meerten et al., 2001).

Secondary structure predictions typically calculate the free energies of a fully equilibrated RNA conformation, that is the most stable form in which this molecule may exist. The dynamics involved during the folding process, however, are neglected. Reported folding times range from micro- to milliseconds for tRNA (Crothers et al., 1974) up to several minutes, as for phage MS2 RNA of maturation protein (van Meerten et al., 2001). In the latter study, a metastable intermediate conformation has been verified experimentally, that functions as a kinetic trap over the course of RNA folding. Kinetic traps have also been proposed to be a major cause for misfolding (Treiber and Williamson, 1999).

At present, conformational changes of RNA need to be detected experimentally employing the method of fluorescence resonance energy transfer (FRET). With this analytical method, conformational fluctuations in the order of nanoseconds can be resolved (Klostermeier and Millar, 2001). On the other hand, computational algorithms that take into account transient RNA conformations, e.g., induced by ligand binding, are still in their infancies (as reviewed by Major and Griffey (2001)). Further advances in the prediction of RNA folding dynamics are expected to yield an improved correlation between computed free energies and measured RNA half-lives.

9.4 Effect of secondary structure within RBS

Predicted structural features of the ribosome binding site and their effects on measured protein expression rates are investigated in this section. In contrast to the previous sections of this chapter, where secondary structure predictions were performed for an excerpt of the mRNA sequence, the following analysis is based on structure predictions spanning the entire length of mRNA. The rationale behind this procedure is the fact that the vector systems under investigation are all controlled by a T7 promoter. mRNA synthesis employing T7 RNA polymerase is known to proceed approximately 5-fold faster in comparison to *E. coli* RNA polymerase (Golomb and Chamberlin, 1974). It is assumed in the following, that mRNA secondary structure formation is essentially equilibrated, before ribosome binding during translation initiation takes place. Partial unfolding of mRNA (e.g., due to ribosome binding) is thus neglected in this study.

As a consequence of the above, the computed free energy of structure formation is then a

global parameter to assess the structural stability of the entire mRNA molecule. But it does not provide details about the stability of particular regions of mRNA, such as e.g., the ribosome binding site or the 5′-end. An interpretation of the local stability of such sites is all the more complicated, when these regions are involved in long-distance base pairings, thus adding to the complexity and nonlinearity of mechanisms that determine RNA secondary structure. For this reasoning, an alternative method was applied to analyze the outcome of mRNA secondary structure predictions, as is outlined in the following.

File ss-count, which is generated by the RNA structure prediction method *mfold*3.1, contains for each base of the RNA sequence, the number of predicted foldings in which this particular base is predicted to be single-stranded. When this number was below one half of the total number of foldings, then the corresponding base was considered to be double-stranded. At values above 50 % of single-strandedness, the respective base was taken to be unpaired. According to this method, the output of mRNA secondary structure predictions was categorized and formed the basis for the diagrams shown in Figures 9.5 and 9.8.

9.4.1 Vector systems

The ribosome binding site stretches over about 36 nt (Stormo, 1990). All plasmids examined (see Table 9.2) contain the same SD-sequence (AAGGAG), and their translational start codon is AUG. Unless otherwise noted, all of these vectors exhibit the same spacing, d, of 8 nucleotides with the same sequence (AUAUACAU) between the SD-sequence and the translational start codon. pJOE-derivatives are regulated by transcription terminators rrnB T1 and rrnB T2. Contrarily, mRNA synthesis using the pET-derived vectors is ceased at transcription terminator Tϕ. The precise location of transcription termination at either of the applied termination signals is known from Lyakhov et al. (1997).

9.4.2 Relative protein expression rate

Measurements of *in vitro* protein synthesis in *E. coli*-derived cellular extracts were received from the "Molecular Biotechnology Lab" of Dr. Martin Siemann (Institute of Biochemcial Engineering, University of Stuttgart). All plasmids investigated (see Table 9.2) were applied in various projects studying particular aspects of both *in vitro* and *in vivo* protein expression. The results of these studies are reused here, in order to compare mRNA structure predictions with protein synthesis performance. Protein synthesis was monitored by radioactive amino acid labelling ([^{14}C]-Leucine) (Sambrook et al., 1989). Under standard reaction conditions, protein synthesis rate remained constant within the initial 40 minutes of synthesis duration and for all plasmids used (data not shown). Relative protein expression rates, V_P, were normalized with respect to the protein expression rate obtained from vector pJOE-malE-eGFP.

Table 9.2: List of plasmids considered for mRNA secondary structure analysis. Plasmid sequences and expression results were adopted from Siemann (2001). malE = maltose binding protein E, GFP = green fluorescent protein, scFv = single chain antibody Fv against atrazine, pelB = pelB leader sequence (length: 63 bp; signifies excrection into periplasm), *gene10* = *gene10* leader, *lacZ* = *lacZ* leader, hyd = hydantoinase, car = carbamoylase, RP-S1 = ribosomal protein S1.

Plasmid	**Size** [bp]	**Features**
pJOE-derivatives		
pJOE-malE-eGFP (pTST101)	6,464	malE-eGFP fusion protein (69.3 kDa)
pJOEK(+)	5,512	pelB-scFv-HisTag (31.2 kDa)
pJOEK(+)(-AT)	5,510	pelB-scFv-HisTag (31.2 kDa)
pJOEK(-)(-AT)	5,447	scFv-HisTag (31.2 kDa)
pJOE-4027.2	5,339	*gene10*+2 Bam-GFP (28.1 kDa)
pJOE-4056.1	5,336	*lacZ*-GFP (28.3 kDa)
pJOE-D-car (MW1)	5,570	*D*-car-HisTag (44 kDa)
pJOE-D-hyd (MW10)	5,993	*D*-hyd-StrepTag (51 kDa)
pJOE-nD-hyd (pJavier2)	5,993	*D*-hyd variant *N*-terminal StrepTag (51 kDa)
pJOE-D-car-D-hyd (pJavier3)	6,955	*D*-car-*D*-hyd fusion protein (95 kDa)
pJOE-D-car-D-hyd (pJavier4)	6,958	polycistron: *D*-car, *D*-hyd
pJOE-D-car-nD-hyd (pJavier5)	6,958	polycistron: *D*-car, *D*-hyd variant
pJOE-RPS1	6,296	RP-S1-HisTag (67.2 kDa)
pET-derivatives		
pETK(-)	4,369	scFv-HisTag (29.0 kDa)
pETK(-)(-AC)	4,367	scFv-HisTag (29.0 kDa)
pETKE(-)	5,084	scFv-GFP (54.4 kDa)
pETKE(-)(-AC)	5,082	scFv-GFP-HisTag (54.4 kDa)
pETK(+)	4,433	pelB-svFv-HisTag (31.2 kDa)
pETK(+)(+AC)	4,435	pelB-scFv-HisTag (31.2 kDa)
pETKE(+)	5,148	pelB-scFv-GFP (56.5 kDa)
pETKE(+)(+AC)	5,150	pelB-scFv-GFP (56.5 kDa)
pET-L-car (pBW8)	6,880	*L*-car (44 kDa)
pET-L-hyd (pMW10*)	7,018	*L*-hyd (49.7 kDa)

9.4.3 Results

For all plasmids listed in Table 9.3, secondary structures were calculated considering full-length mRNA. Two examples for predicted secondary structures, one each for a pET-derived vector and pJOE-vector, are given in Figures 9.3 and 9.4. Figure 9.3 shows the optimum predicted folding for full-length mRNA of plasmid pETK(+). In Figure 9.4, the minimum free energy structure of full-length mRNA of plasmid pJOE-4027.2 is displayed. In both figures, a magnification of the ribosome binding site and 5′-terminal region is provided also. Structure predictions were subsequently analyzed according to the method described earlier in this study, yielding Figures 9.5 and 9.8. From the analysis, the following conclusions can be drawn.

Structure conservation. Although several folding conformations were predicted for each plasmid, respectively, most of the total structures showed a high degree of conservation of several local stem-loop structures, respectively (data not shown). Among the investigated plasmids, this holds true in general within the regions of ribosome binding, as well as concerning the 5′ and 3′-end of mRNA. The secondary structure of transcription terminator Tϕ was predicted to be approximately the same for all investigated mRNAs carrying Tϕ. Similarly, the structures of transcription terminators rrnB T1 and rrnB T2 were computed to remain essentially fixed across all comparable plasmids. Predicted structures were consistent with the secondary structures postulated earlier for terminators Tϕ, rrnB T1 and rrnB T2 (Lyakhov et al., 1997), respectively, thus lending further credibility to the prediction method applied.

5′ and 3′-end of mRNA. Secondary structural elements formed within the 5′ and 3′-region of mRNA remain in general conserved across the predicted optimal and suboptimal foldings (plurality of predicted results are explained in section 9.1). When the ribosome binding site is located in close proximity to such fixed moieties, as is the case for the pET-derived vectors (SD-sequence about 50 bases from the 5′-end), structure predictions within the RBS are found to be delivered with only little variation. In contrast, mRNA derived from pJOE-vectors, where the RBS is positioned roughly 150 bases from the 5′-end, showed more structural variability concerning the ribosome binding site.

Shine-Dalgarno sequence. For all predicted structures, the SD-sequence exhibits a high degree of double-strandedness, independent of the respective protein expression rate observed. Further, in almost all of the investigated vectors, there exists a region of at least 4 unpaired bases right upstream of the SD-sequence. It is thus intriguing to believe that double-strandedness may play a role in SD-sequence recognition during the process of translation initiation.

Translational start codon (here AUG). According to the predicted structures displayed in Figure 9.5(A), 6 out of the 7 plasmids giving maximum relative protein expression rates, namely plasmids pJOE-malE-eGFP, pET-L-car, pETKE(+), pJOE-RPS1, pETK(+), and pMW10*, show the translational start codon to be unpaired. Consistent with this observation, plasmids with relative protein expression rates below 30 % exhibit a preference (5 out of 6 plasmids) for having double-stranded predicted translation initiation codons (Figure 9.5(B)). Hence, single-stranded translational start codons seem to be important for obtaining improved protein expression rates.

Table 9.3: Correlation between the initiation codon being single-stranded and protein expression rate. d = spacing between SD-sequence and translational start codon. f = fraction of single-stranded bases within the 23 bases subsequent to the SD-sequence (cf. Figure 9.6). V_P = relative protein expression rate within the initial 40 min of *in vitro* protein synthesis under standard conditions (experimental results kindly provided by Siemann (2001)). Protein expression rate obtained from plasmid pJOE-malE-eGFP served as the reference value for normalization. For plasmids pJavier3, pJavier4, and pJavier5, two protein synthesis rates were given each. The first of these rates refers to protein carbamoylase, the second to hydantoinase. n.d. = not determined.

Plasmid	**mRNA** [kb]	d [# nt]	**Total number of foldings**	f	V_P *in vitro* [%]
pJOE-malE-eGFP (pTST101)	2.473	8	44	0.96	100
pJOEK(+)(-AT)	1.519	8	29	0.87	n.d.
pET-L-car (pBW8)	1.422	8	31	0.83	100
pETKE(+)	1.759	8	37	0.74	100
pJOE-RPS1	2.313	8	21	0.74	95
pETK(+)	1.044	8	27	0.70	100
pJOE-D-car (pMW1)	1.579	8	28	0.57	100
pET-L-hyd (MW10*)	1.560	8	29	0.57	100
pJOE-D-car/D-hyd (pJavier3)	2.964	8	49	n.d.	100/0
pJOE-D-car-D-hyd (pJavier4)	2.967	8	46	n.d.	95/5
pJOE-D-car-nD-hyd (pJavier5)	2.967	8	50	n.d.	95/5
pETKE(-)(-AC)	1.693	8	35	0.57	30
pETK(-)(-AC)	0.978	8	23	0.57	n.d.
pJOE-nD-hyd (pJavier2)	2.002	8	31	0.52	20
pJOE-D-hyd (pMW10)	2.002	8	24	0.61	5
pJOE-4027.2	1.348	8	25	0.43	5
pJOEK(-)(-AT)	1.453	8	28	0.30	5
pJOE-4056.1	1.345	8	24	0.22	5
pETK(+)(+AC)	1.046	10	24	0.78	75
pETKE(+)(+AC)	1.761	10	36	0.78	75
pJOEK(+)	1.521	10	33	0.87	70
pETK(-)	0.980	10	26	0.52	30
pETKE(-)	1.695	10	35	0.52	30

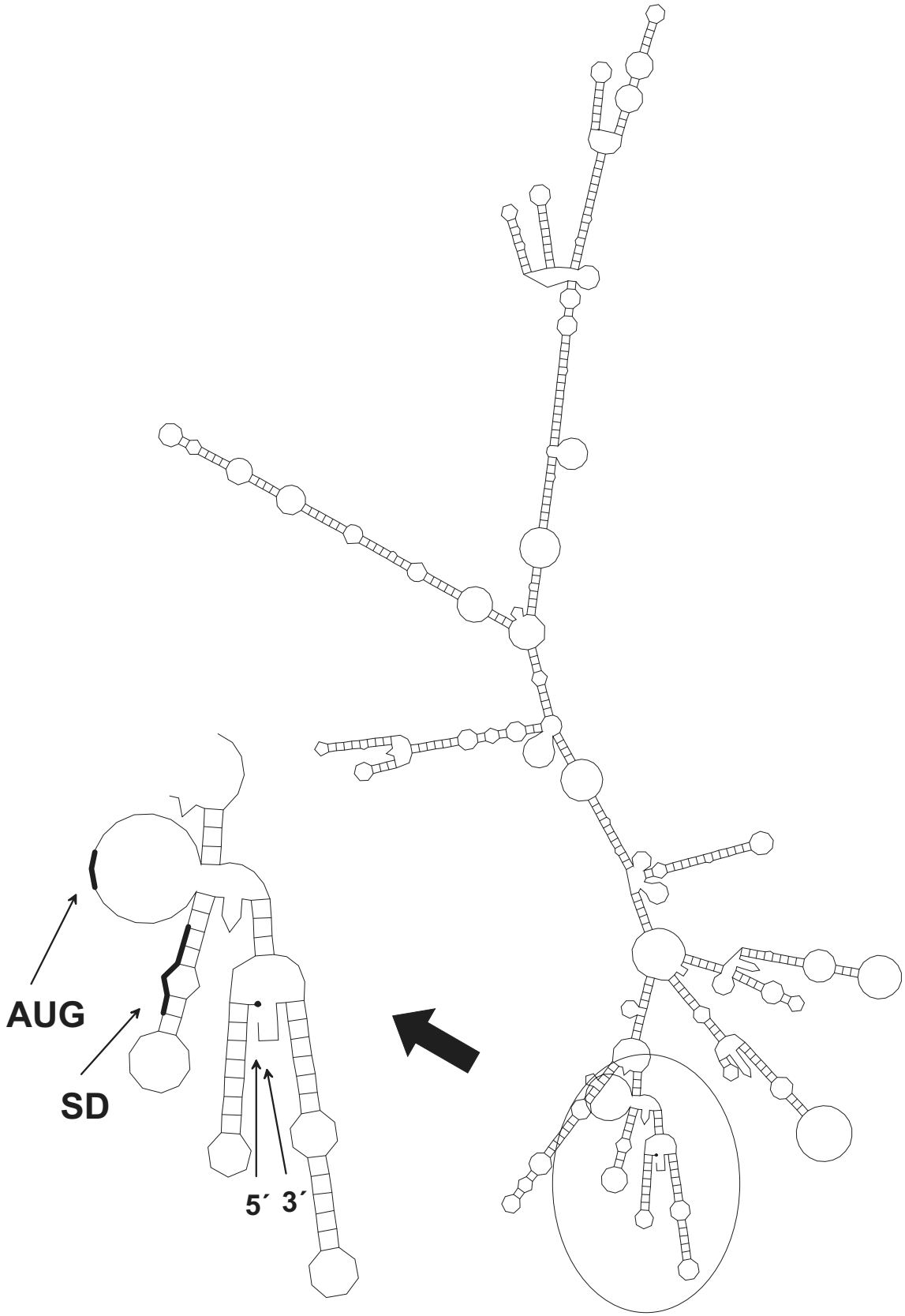

Figure 9.3: Predicted minimum free energy structure of full-length mRNA obtained from plasmid pETK(+), with ΔG = -1,486.5 kJ/mol.

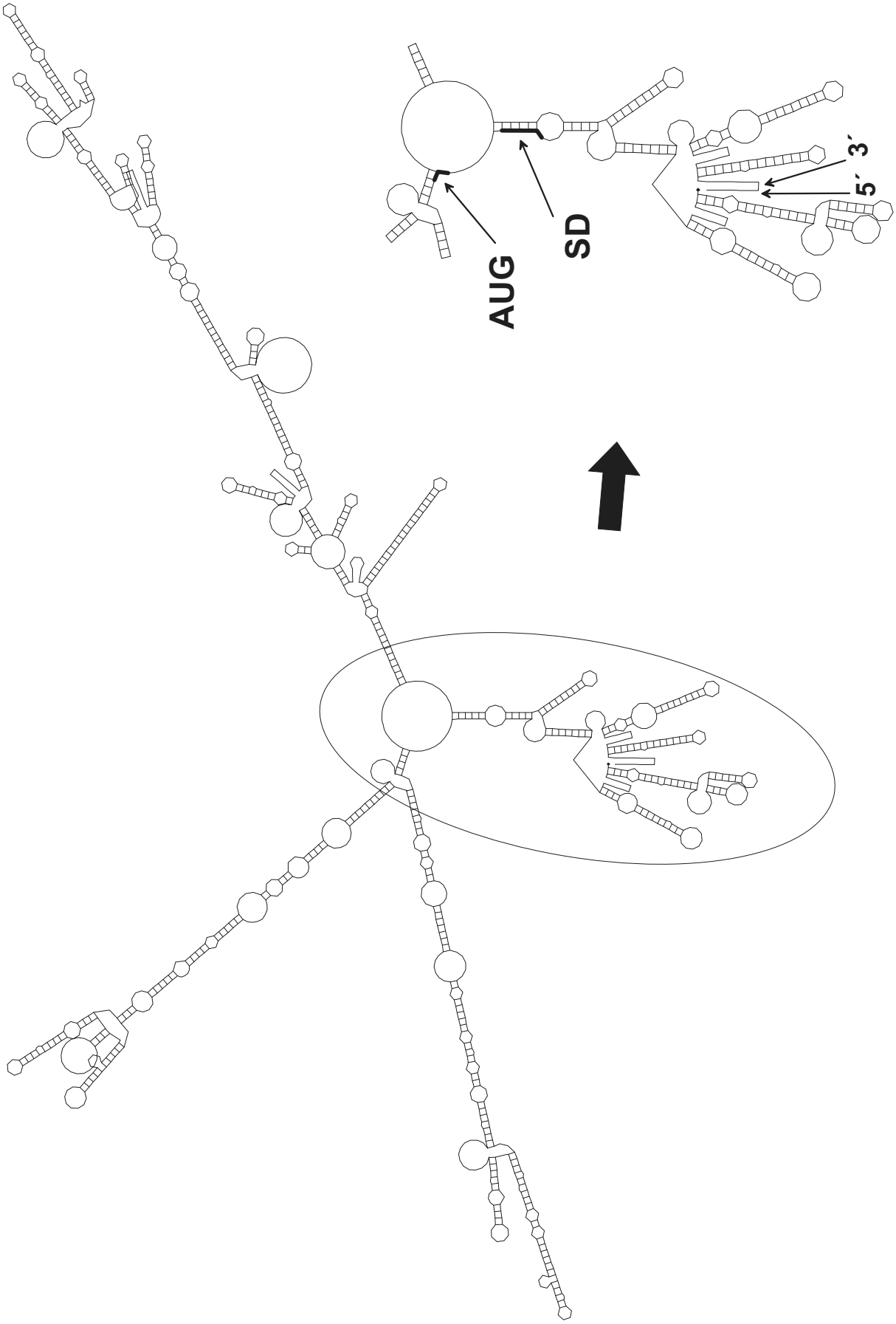

Figure 9.4: Predicted minimum free energy structure of full-length mRNA obtained from plasmid pJOE-4027.2, ΔG = -2,119.4 kJ/mol.

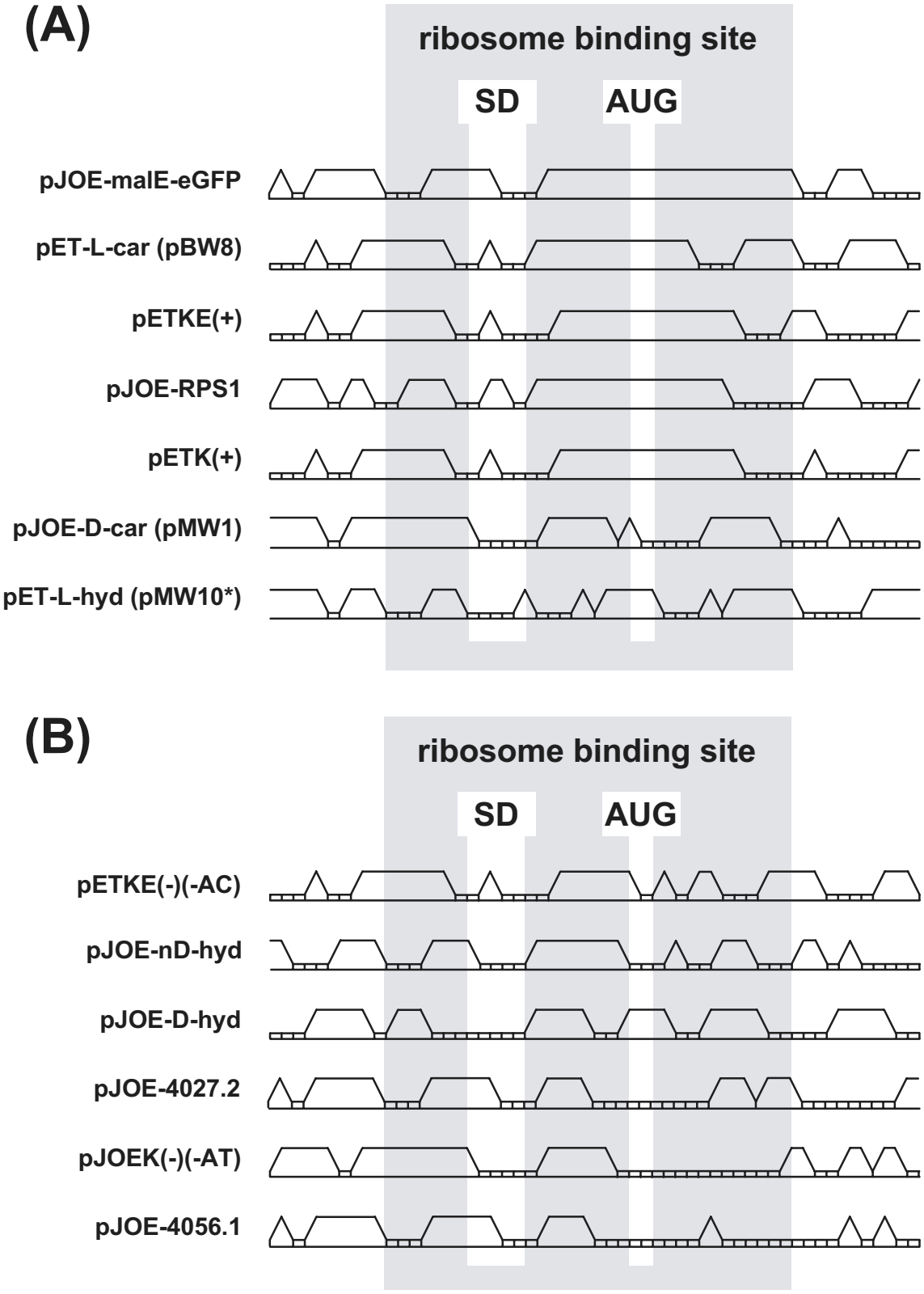

Figure 9.5: Predictions of single-stranded versus double-stranded bases within the ribosome binding site for the given plasmids. (A) 100 % relative protein expression rate, and (B) less than or equal to 30 % of relative protein expression rate.

Region of RBS downstream of the SD-sequence. As can be seen from Figure 9.5, a high degree of single-strandedness downstream of the SD-sequence and extending throughout the remainder of the ribosome binding site appears in general to coincide with raised protein expression rates (Table 9.3). In order to illustrate this point, parameter f was introduced in this study. f is defined as the portion of single-stranded bases within the 23 bases immediately following downstream of the SD-sequence (see Figure 9.6). A graphical representation of the measured relative protein synthesis rate as a function of f is provided in Figure 9.7. As can be seen from this graph, relative protein expression rate shows a sigmoidal behaviour with f. Despite the displayed variation, plasmids with values of f greater than approximately 0.6 are indicated to yield maximum protein synthesis rates.

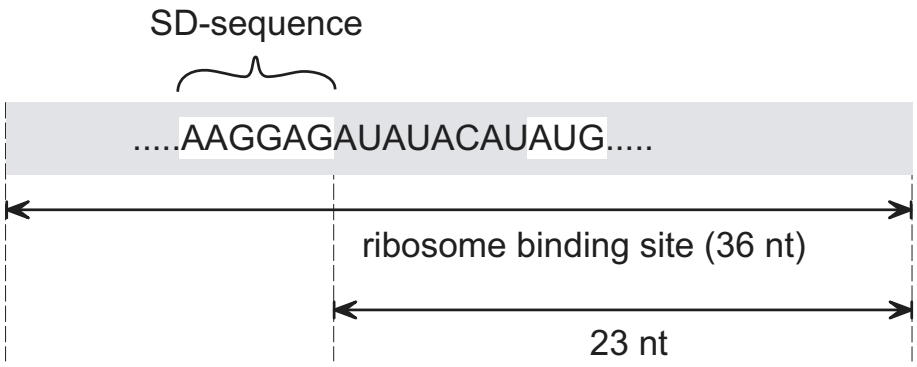

Figure 9.6: Definition of f. This parameter denotes the fraction of single-stranded bases within the 23 nucleotides downstream of the Shine-Dalgarno sequence. The SD-sequence and the translational start codon are highlighted in white.

Distance between Shine-Dalgarno sequence and AUG. Raising the spacing from 8 to 10 nucleotides usually resulted in an extended loop of single-stranded bases, at the same time keeping the predicted structural appearance of the surrounding region unaffected. While an insertion of 2 additional bases between the SD-sequence and the translational start codon AUG (Figure 9.8) thus caused no significant effect on the structural features of the ribosome binding site, it led in several cases to a reduction of relative protein expression rate, namely by a factor of about 1/4 (see Table 9.3). Examples hereof are vector pairs pETK(+)/pETK(+)(+AC), pETKE(+)/pETKE(+)(+AC), and pJOEK(+)(-AT)/pJOEK(+). The latter pair is based on predictions made in this study. For vectors pETK(-) and pETK(-)(-AC) and their counterparts containing eGFP fused to the encoded single-chain antibody, pETKE(-)/pETKE(-)(-AC), no significant change in relative protein expression rate could be detected. These latter vectors showed, in general, a reduced protein expression rate (30 %), and a putative reduction of expression rate might have been undistinguishable from measurement error in these experiments.

pelB leader sequence. Experimentally investigated vectors pETK(+), pETKE(+), pETK(+)(+AC), and pETKE(+)(+AC) containing an N-terminal pelB leader sequence gave maximum protein expression rates. That is 100 % and 75 % of relative protein expression rate were obtained for vectors with a spacing of 8 nucleotides and 10 nucleotides, respectively, between the SD-sequence and the translational start codon. These vector systems showed very similar properties of the ribosome binding site with respect to predicted base pairing

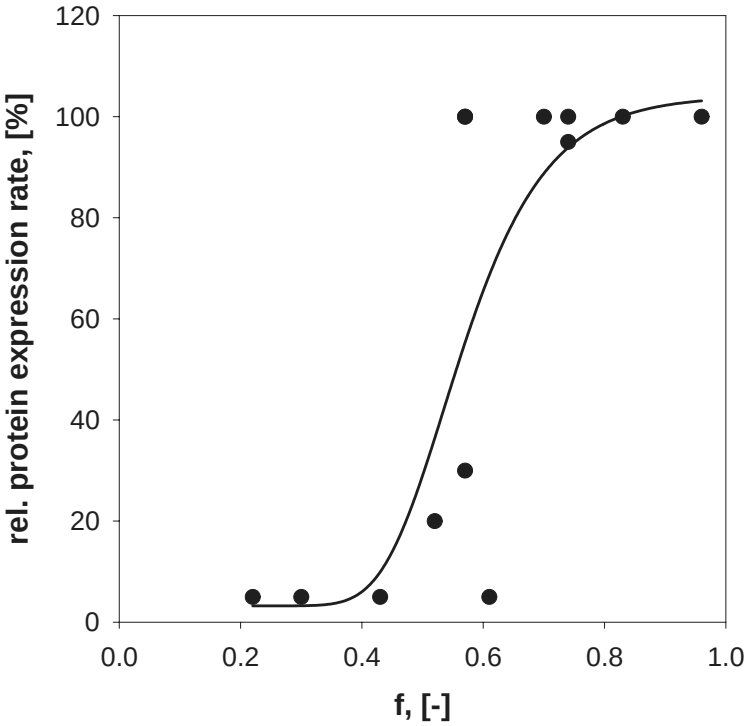

Figure 9.7: Relative protein expression rate measured *in vitro* versus parameter f.

(Figure 9.8). The insertion of an N-terminal pelB leader sequence appears to yield improved protein expression rates. The results obtained in this study suggest that due to the insertion, a favourable secondary structural conformation of the RBS can be maintained that allows for efficient translation initiation.

Structural requirements surrounding the RBS. From the selected plasmids, it is impossible to conclude whether structural features of mRNA surrounding the ribosome binding site are necessary for efficient translation initiation. Further investigation including, e.g., site-directed mutagenesis is necessary to explore this issue, which is beyond the scope of this study.

Sequence length. With increasing sequence length, mRNA structure predictions showed a tendency to produce ambiguous results. This was the case, in particular, for plasmids pJOE-malE-eGFP, pJavier3, pJavier4, and pJavier5. All of these plasmids code for fusion proteins with C-terminal eGFP and thus show larger-sized mRNAs than in comparison to the non-fusion protein encoding plasmids examined. The greater the mRNA length, the higher the number of combinations of base-pairings is expected to be. This is also reflected in Figure 9.9, where the number of computed foldings is displayed versus mRNA sequence length. Linear regression analysis showed an increase of 6.5 predicted suboptimal foldings roughly per every 500 bases. The results obtained in this study suggest an upper boundary threshold of about 2.5 to 3.0 kb in mRNA size for application of the chosen method of RNA structure prediction. Beyond this level, an interpretation of prediction results becomes increaslingly ambiguous, unless further constraints like known rigid stem-loop structures are kept fixed in the analysis.

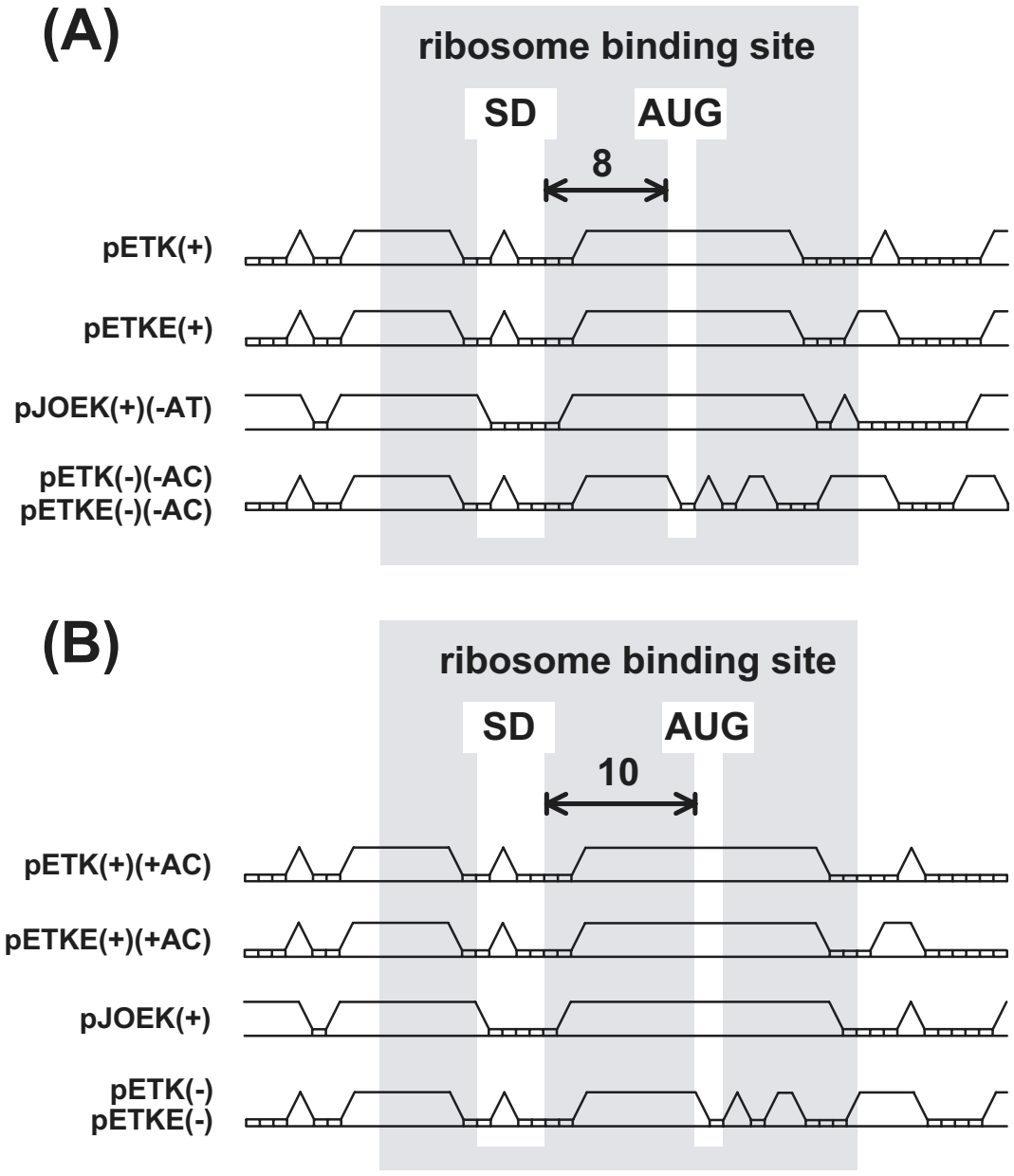

Figure 9.8: Impact of variation of spacing between Shine-Dalgarno sequence and translational start codon on the predictions of mRNA secondary structure.

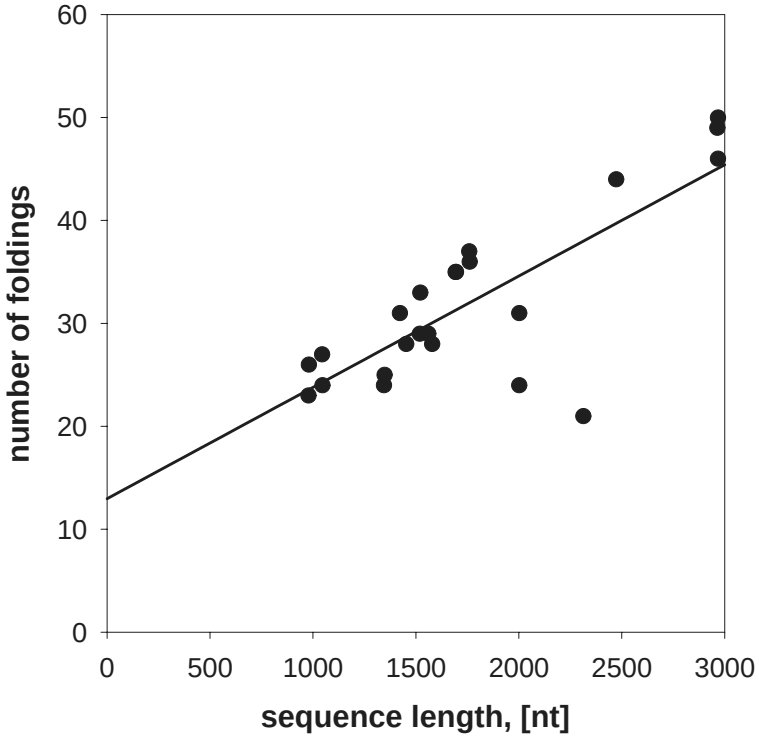

Figure 9.9: Number of predicted RNA foldings versus sequence length. Linear regression analysis yielded $y = 12.9672 + 0.0108 \cdot x$, with regression coefficient $r = 0.79$.

9.5 Conclusions

The results of this chapter suggest a correlation between predicted secondary structural properties regarding the ribosome binding site and experimentally detected translation efficiencies. A high degree of single-strandedness surrounding the translational start codon was shown to correspond with enhanced, experimentally determined protein expression rates. Albeit, a universal relationship between predicted free energies of structure formation and measured functional half-lives of mRNA could not be found.

Furthermore, this study demonstrated that sequence length entering structure predictions strongly affects the analysis results. With increasing RNA lengths, prediction results were found to be dramatically deteriorated, leading to a raised plurality of solutions. This finding is in accordance with earlier studies (Schulz and Reznikoff, 1990; Carrier and Keasling, 1999). On the other hand, for transcript lengths below about 2.5 kb in size, mRNA secondary structure predictions delivered in general convergent results that showed a tendency to predict at least locally conserved stem-loop structures.

The spacing between the Shine-Dalgarno sequence and the start codon is known to influence the codon-anticodon interaction with fMet-tRNA (Ringquist et al., 1992). Structure predictions conducted in this study are in agreement with the earlier results. For the investigated vector systems, a distance between the SD-sequence and translational start codon of 8 nucleotides

yielded higher protein expression rates, by a factor of approximately 1/4, in comparison to a spacing of 10 nucleotides.

In general, a static model to describe mRNA conformation is assumed. This premise may constitute a significant limitation on current methods for mRNA structure prediction. Intermediary states adopted during the temporary folding and unfolding of mRNA are known to have a significant impact also on the structural functionality of mRNA. It is possible that structures corresponding to other than global minimum free energies exist. Such a situation may arise for large RNAs, particularly when local structural elements (i.e., kinetic traps) are created. These kinetic traps may be stable enough to prohibit their conformational alteration during the folding process. Furthermore, since folding is known to occur co-transcriptionally, mechanisms like, e.g., pausing of RNA polymerase may be an important factor in determining the folding conformation and thus the translatability of a mRNA (Pan et al., 1999). This issue needs further consideration, when heterologous RNA polymerases are employed in recombinant expression systems. For example, RNA polymerase from bacteriophage T7 is known to transcribe approximately 5-fold faster than the corresponding host enzyme in *E. coli* (see chapter 3, and Golomb and Chamberlin (1974)). The rate of mRNA synthesis clearly influences the mRNA folding pathway, which in turn is likely to affect also transcript stability.

Although the prediction of RNA secondary structures comprises various sources of inherent variation (e.g., due to the uncertainty associated with the thermodynamic constants for nucleotide stacking used in the estimation procedure, due to unknown putative long-range sequence context, as well as due to the neglect of transient folding conformations in the prediction procedure), such methods are nevertheless useful as a further means to assess the biochemical implications of nucleotide sequence modification.

The example of the pelB leader sequence demonstrated that an N-terminal insertion of a certain peptide (or protein) may assist in attaining a favourable secondary structure of the ribosome binding site, such that an efficient initiation of translation is promoted. Additionally, a selective manipulation of the coding region is conceivable, which allows a modification of nucleotide sequence, while the amino acid sequence remains unaltered. Further, the effects of point mutations (e.g., applied within the ribosome binding site) may be estimated. The method suggested in this study for analyzing the secondary structural appearance may help to predict *in silico* the success of such optimization strategies. Such predictive tools are extremely useful for the targeted engineering of genetic sequences within the application field of designer genomics.

10 Outlook

While the overall gene expression model developed in this study covers some of the main aspects determining expression rates, it raises at the same time several questions for model improvement that remain unresolved. At present, modelling efforts have focussed primarily on developing a functional dependence of gene expression rates on the nucleotide sequence of the encoded gene. In this initial approach, further regulatory aspects, in particular related to transcription regulation, were currently left out of the modelling scheme. A main reason for this neglect was due to the lack of information about global hierarchical regulation mechanisms. However, as knowledge increases about the co-regulation of genes obtained through cluster analysis of DNA chip experiments carried out for certain physiological conditions, the issue of adding gene regulation information to the current model formulation becomes an increasingly realistic task. Furthermore, the effects of post-translational modifications on the protein level can be suitably investigated also in a high-throughput manner, e.g., making use of recent advances in proteome analysis. The strength of combining both transcript and protein expression profiles with high-throughput metabolome analysis, is expected to provide with a highly detailed snapshot of whole-cellular metabolism, in order to further elucidate the complex structures of hierarchical gene expression regulation. These types of biomolecular data becoming increasingly available on a global, whole-cellular level, constitute an invaluable source of information that could be added for further model amendments.

Another process affecting the protein functionality is the proper folding conformation of the expressed protein. Protein folding is known to occur co-translationally. Similar as for mRNA synthesis, where transcriptional pause sites commonly function as a regulatory site to allow the proper folding of mRNA molecules, it is conceivable that analogous such sites may exist also on the translational level. In this context, the effects of codon-specific elongation rate on the protein folding pathway may be an interesting topic to investigate in future studies.

For more accurate model predictions, the expression model presented should be expanded appropriately to account for further reactions particularly enrolled in differential mRNA decay. While in this study, a mathematical model for the commonly believed primary degradation mechanism of bacterial mRNA was developed, this model is insufficient to describe the diversity of observed degradation pathways. Especially, the degradation of stem-loops and hairpin structures depending on their stability, as well as sequence-specific features for recognition by corresponding RNases have been neglected in this study. These mechanisms are clearly important in determining half-lives of differently structured mRNAs. Future modelling efforts in this field should account for these phenomena.

Furthermore, the efficiency of ribosmal binding may be additionally controlled by mRNA structural elements governing the accessibility of the initiation region. The impact of mRNA secondary structure formation on elongation rate could additionally be taken into account in the modelling scheme. The model developed in this study is compatible to the approach taken by von Heijne et al. (1977), which treated the inhibitory effects of hairpin formation on translation elongation rate. In this context, also the dynamics involved in forming mRNA

secondary structures may be significant.

While the gene expression model presented in this study was developed on the basis of prokaryotic systems, it would be extremely interesting to test the transferability of this kind of comprehensive modelling approach to also treat eukaryotic systems. There are many analogies between prokaryotic and eukaryotic gene expression, both involving ribosomal subunits and the principle of carrier-mediated transport of monomeric building blocks, as well as translation factor participation. On the other hand, the mechanisms and regulatory schemes of eukaryotic protein synthesis are much more complex than in bacterial systems. For instance, on the transcriptional level, a number of transcription factors and their concerted interplay are significant in determining transcript expression rates. Furthermore, eukaryote-specific properties like intron/exon structures of expressed sequence tags, as well as the phenomenon of splice variants lead to a much more complicated expression scenario than described for prokaryotes. In order to achieve a meaningful description of eukaryotic protein expression, the modelling scheme presented in this study would need to be augmented by the major of these additional reactions. Evidently, parameters identified for prokaryotic gene expression, cannot readily be applied to cover eukaryotic models. Instead, a thorough model parametrization is required when modelling gene expression in eukaryotes.

With a comprehensive dynamic model of gene expression at hand, it is valuable to use this model for design studies in process development. Model-based simulation analysis provides detailed time-dependent information of key compounds and catalysts involved in gene expression. Model predictions can help to determine most suitable reactor concepts and operating conditions, in order to establish improved synthesis rates and allow for an enhanced process duration.

An optimum procedure for improving gene expression rates should involve an optimum selection of coding sequence to improve both mRNA secondary structure (including stabilizing structural elements) and enable an optimum adaptation to tRNA concentration, while keeping the amino acid sequence unaltered. Due to the genetic code being degenerate, there is some variability possible in coding sequence. In principle, all possible permutations (combinations of coding sequence) can be calculated. This type of computation can easily be automated. Then, for the entire collection of nucleotide sequences, mRNA secondary structures could be calculated. From the set of structures, mRNA degradation rates should be predicted together with the expected protein synthesis rates. Resulting mRNA and protein levels would be obtained as model outputs in this case. The nucleotide sequences could finally be ranked according to the predicted results. Ranking criteria could include (a) highest levels of functional protein, (b) the avoidance of inclusion body formation, (c) stability against proteolysis, (d) economic energy consumption, and (e) improved mRNA stability, that should altogether be reflected in maximum production rates of functionally active target protein. And finally, the possibility to assess sequence-specific effects on protein synthesis rates may provide valuable information for optimum pathway design, where a modulation and fine-tuned relative adjustment of catalytic activities is desirable.

A Derivation of Enzymatic Rate Equations

A.1 Method description

Mathematical derivation of enzymic rate equations is very tedious for complex reaction mechanisms, in particular when involving a large number of reactants. For this reason, several computational tools have been developed to assist in this derivation procedure through an automation of kinetic modelling (Cornish-Bowden, 1977; Ishikawa et al., 1988; Straathof and Heijnen, 1997).

The scope of application of the here developed algorithm spans reactions in pseudo-steady state or rapid equilibrium, as well as partial equilibrium mechanisms. Part of this method has been described elsewhere (Mauch et al., 1997). Further editions are necessary to include metabolic channelling kinetics.

Rate equations are calculated on the basis of the sequence of elementary reaction steps. The functional dependency of a reaction rate is given in terms of total enzyme concentration, the concentrations of involved substrates and products, as well as the respective rate constants and equilibrium constants for each individual reaction step.

Kinetic rate expressions based on the quasi-stationary state assumption are derived according to the determinant method, e.g., described in Purich (1983). According to this method, the balance equations for all enzyme-containing species form a set of linear algebraic equations that can be solved simultaneously, e.g., through Gaussian elimination. Combined rapid equilibrium and steady state mechanisms can be reduced to the steady state method by treating the equilibrium segments as though they were single enzyme species at steady state with the other enzyme intermediates (Cha, 1968).

The kinetic rate equations may be arranged in such a way that kinetic parameters are represented according to the Cleland nomenclature (Cleland, 1963a, 1963b, 1963d; Segel, 1993). The computation of these kinetic constants is carried out for typical steady state rate equations. However, a generalized formulation of kinetic constants including more complex reaction mechanisms requires further research efforts and is currently impossible with the algorithm presented in this study.

It is noteworthy to point out that the number of kinetic constants like Michaelis-Menten constants obtained from a rate equation may be larger than the actual number of rate constants of this equation, so that constraints among kinetic parameters may be imposed (Straathof and Heijnen, 1996). Further efforts are necessary, however, to develop an overall generalization for the nomenclature of kinetic constants.

The program developed serves as a useful tool for model-building and automated derivation of complex enzymatic rate equations based on the quasi-stationary state assumption and partial rapid equilibrium treatment. It may be applied in both educational and research environment.

A.2 Rate expression of 70S initiation complex formation

For 70S initiation complex formation, a rapid equilibrium and random substrate association was considered, as described by equations (6.9) to (6.11). Using symbols $[E]_t$ = total concentration of complex $30S \cdot IF1 \cdot IF2{\cdot}GTP \cdot IF3$, $[A]$ = concentration of fMet-tRNA$_f^M$, $[B]$ = concentration of ribosome binding sites, $[C]$ = concentration of ribosomal subunit 50S, $[P]$ = concentration of IF1, $[Q]$ = concentration of IF3, the elementary reaction steps read:

$$E + A \overset{K_A}{\rightleftharpoons} EA \tag{A.1}$$

$$EA + B \overset{K_B}{\rightleftharpoons} EAB \tag{A.2}$$

$$E + B \overset{K_B}{\rightleftharpoons} EB \tag{A.3}$$

$$EB + A \overset{K_A}{\rightleftharpoons} EAB \tag{A.4}$$

$$EAB \overset{k_{TLI,70SIC,1}}{\longrightarrow} EPQ \tag{A.5}$$

$$EPQ + C \overset{k_{TLI,70SIC,2}}{\longrightarrow} E + P + Q \tag{A.6}$$

With the method presented in the previous section, the following rate equation was derived:

$$V_{TLI,70SIC} = \frac{k_{TLI,70SIC,1}\ k_{TLI,70SIC,2}\ [A][B][C][E]_t}{k_{TLI,70SIC,1}[A][B] + k_{TLI,70SIC,2}[C]\Big(K_A K_B + K_B[A] + K_A[B] + [A][B]\Big)} \tag{A.7}$$

Using equations (6.13) to (6.16), equation (A.7) can then be simplified to yield equation (6.12).

A.3 Rate expression of translation elongation

The following derivation is carried out for $n_{cod} = 3$ (cf. section 6.6). Symbols are $[E]_t$ = total concentration of ribosomes bound to mRNA at codon j; $[A]$, $[C]$, $[D]$ = concentrations of ternary complexes (T3$_j$); $[B]$ = concentration of EFG·GTP; $[P]$ = concentration of P$_i$; $[Q]$ = concentration of EFTu·GDP; $[R]$, $[M]$, $[O]$ = concentrations of tRNA species, and $[T]$ = concentration of EFG·GDP. The elementary reaction steps spanning 3 consecutive elongation cycles are represented by:

$$E + A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E1 \tag{A.8}$$

$$E1 \overset{k_2}{\longrightarrow} E2 + P \tag{A.9}$$

$$E2 \overset{k_3}{\longrightarrow} E3 + Q \tag{A.10}$$

$$E3 \overset{k_4}{\longrightarrow} E4 \tag{A.11}$$

$$E4 + B \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} E5 \tag{A.12}$$

$$E5 \overset{k_6}{\longrightarrow} E6 + R + P \tag{A.13}$$

$$E6 \overset{k_7}{\longrightarrow} E7 + T \tag{A.14}$$

$$E7 + C \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E8 \tag{A.15}$$

$$E8 \xrightarrow{k_2} E9 + P \tag{A.16}$$

$$E9 \xrightarrow{k_3} E10 + Q \tag{A.17}$$

$$E10 \xrightarrow{k_4} E11 \tag{A.18}$$

$$E11 + B \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} E12 \tag{A.19}$$

$$E12 \xrightarrow{k_6} E13 + M + P \tag{A.20}$$

$$E13 \xrightarrow{k_7} E14 + T \tag{A.21}$$

$$E14 + D \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E15 \tag{A.22}$$

$$E15 \xrightarrow{k_2} E16 + P \tag{A.23}$$

$$E16 \xrightarrow{k_3} E17 + Q \tag{A.24}$$

$$E27 \xrightarrow{k_4} E18 \tag{A.25}$$

$$E18 + B \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} E19 \tag{A.26}$$

$$E19 \xrightarrow{k_6} E20 + O + P \tag{A.27}$$

$$E20 \xrightarrow{k_7} E + T \tag{A.28}$$

Enzyme conformations are denoted by symbols E, $E1$, ..., $E20$ in equations (A.8) to (A.28). Under pseudo-steady state conditions, the reaction rate covering n_{cod} elongation cycles is expressed by:

$$V_{TLE,j} = \frac{[E]_t}{\frac{1}{k_2} + \frac{1}{k_3} + \frac{1}{k_4} + \frac{1}{k_6} + \frac{1}{k_7} + \frac{k_{-1} + k_2}{n_{cod} k_1 k_2}\left(\frac{1}{[A]} + \frac{1}{[C]} + \frac{1}{[D]}\right) + \frac{k_{-5} + k_6}{k_5 k_6 [B]}} \tag{A.29}$$

Substituting

$$k_{TLE,j} = \left(\frac{1}{k_2} + \frac{1}{k_3} + \frac{1}{k_4} + \frac{1}{k_6} + \frac{1}{k_7}\right)^{-1} \tag{A.30}$$

$$K_{M,\mathrm{T3}_j} = k_{TLE,j}\left(\frac{k_6 + k_{-5}}{k_5}\right) \tag{A.31}$$

$$K_{M,\mathrm{EFG{\cdot}GTP}} = \frac{k_{TLE,j}}{n_{cod}}\left(\frac{k_2 + k_{-1}}{k_1}\right) \tag{A.32}$$

into equation (A.29) yields equation (6.49) (for $n_{cod} \geq 1$), and equation (6.31) for the particular case, where $n_{cod} = 1$.

B Ionic Species Computation

The following iterative method based on Storer and Cornish-Bowden (1976) was applied for the computation of free ion concentrations in aqueous solutions assuming total dissociation of salts.

Total material balance equations for protons (H), magnesium (Mg), inorganic orthophosphate (Pi), inorganic pyrophosphate (PPi), and the sum of concentrations of nucleoside triphosphates (NTP) read:

$$\begin{aligned}[H]_t &= [HNTP]^{3-} + [H]^{+} + [HPi]^{-} + 2\,[H_2NTP]^{2-} + [MgHPi]^{+} \\ &\quad +[MgHNTP]^{-} + 2\,[H_2Pi]^{0} \qquad \text{(B.1)}\end{aligned}$$

$$\begin{aligned}[Mg]_t &= [MgNTP]^{2-} + [Mg]^{2+} + 2\,[Mg_2NTP]^{0} + [MgNTP_2]^{6-} \\ &\quad +[MgHPi]^{+} + [MgHNTP]^{-} \qquad \text{(B.2)}\end{aligned}$$

$$[Pi]_t = [Pi]^{2-} + [HPi]^{-} + [MgHPi]^{+} + [H_2Pi]^{0} \qquad \text{(B.3)}$$

$$[PPi]_t = [PPi]^{4-} + [MgPPi]^{2-} + [Mg_2PPi]^{0} \qquad \text{(B.4)}$$

$$\begin{aligned}[NTP]_t &= [NTP]^{4-} + [HNTP]^{3-} + [MgNTP]^{2-} + [H_2NTP]^{2-} + [Mg_2NTP]^{0} \\ &\quad +2\,[MgNTP_2]^{6-} + [MgHNTP]^{-} \qquad \text{(B.5)}\end{aligned}$$

The amount of magnesium bound to nucleic acids (both DNA and mRNA) was considered assuming in average 1 molecule of Mg^{2+} to be bound per each 2 phosphate groups of nucleic acid (Record et al., 1976).

The proton concentration is known for a given pH in a buffered system. When the total concentrations of magnesium, Pi, PPi, and nucleic acid are specified, free ion concentrations can be calculated from equations (B.1) to (B.5). In this procedure, an initial guess for each of the concentrations of uncomplexed Mg^{2+}, Pi^{2-}, PPi^{4-}, and NTP^{4-} was assumed, which is refined with every next iteration step. When finally the convergence criterion was fulfilled (absolute tolerance of 10^{-9} for all Mg^{2+}, Pi^{2-}, PPi^{4-}, and NTP^{4-}), or when the maximum number of iterations was reached (default value of 300), the concentrations of all ionic species complexes were output.

C Algorithm for Automated Code Generation

C.1 Software architecture

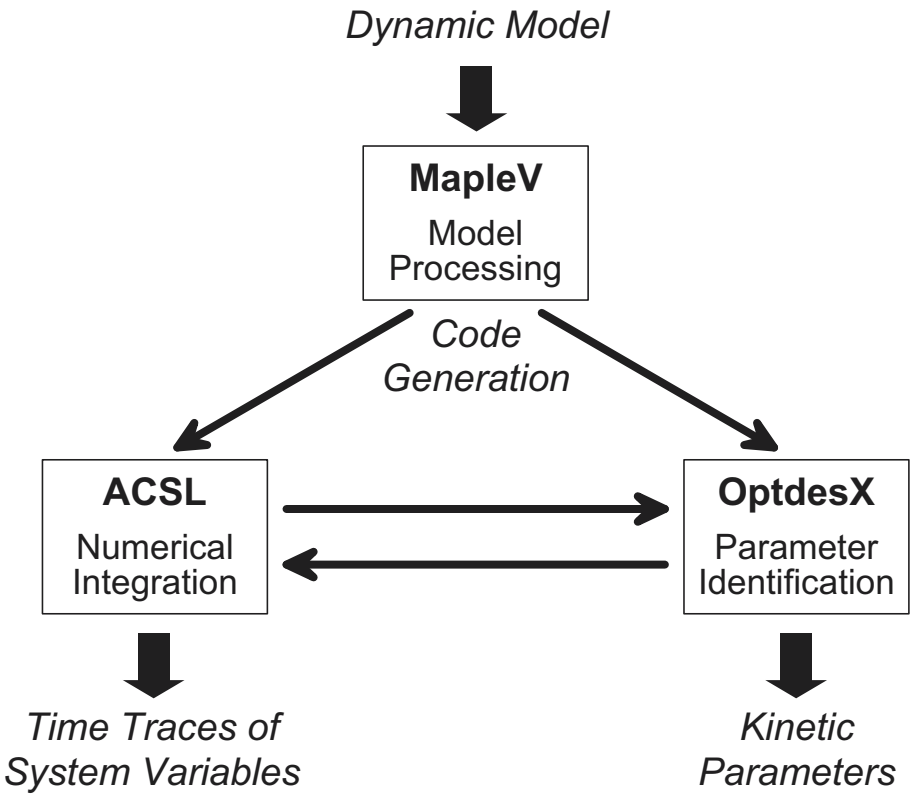

Figure C.1: Model building includes an automated generation of differential equations. This tool is implemented in a symbolic programming language (MapleV). The mathematical model is subsequently transferred in Fortran code, both to the numerical integrator (Advanced Continuous Simulation Language, ACSL) and the nonlinear regression package (OptdesX).

C.2 Initializing flags

The following logical variables can be set either true or false, prior to simulation analysis, in order to switch ON/OFF particular units of the comprehensive gene expression model.

`code`	when set equal to 1, code generation is performed; if set equal to 2, then no code generation occurs (which is the usual case during model building)
`ncod`	parameter denotes the degree of codon refinement as defined in section 4.2.5; parameter takes integer values typically between 1 to 4
`transcription`	mRNA synthesis
`mRNAdegradation`	mRNA degradation reactions
`translation`	codon-specific elongation rates
`translationFix`	uniform, codon-independent rate constant for translation elongation

`Eregeneration`	energy regeneration employing the reactions of acetate kinase and nucleoside monophosphate kinases
`queueing`	controls the general consideration of ribosomal and degradosomal queueing for both association and propagation of the respective catalyst
`qjqueueing`	queueing effects during template-bound propagation
`Dqueueing`	queueing effects of degradosomes
`InVitro`	affects the choice of tRNA concentration; when set to true, typical values for standard conditions of bacterial *in vitro* protein synthesis are applied; otherwise, tRNA concentrations are interpolated for a user-selected specific growth-rate from data tabulated by Dong et al. (1996)

C.3 Output during model building

Several flags control the standard output provided by the maple worksheet `sim.mws`.

`printall`	print out all variables, rates, material balances, initial conditions, and model parameters
`printout`	print selected quantitites, as specified below
`printR`	print reaction kinetics in general
`printARS`	print reactions involved in tRNA charging
`printRIB`	print kinetics of ribosomal translation
`printDEG`	print kinetics of mRNA degradation
`printTC`	print kinetics of mRNA synthesis
`printW`	print monitoring parameters
`printQ`	print queueing factors
`printDC`	print material balance equations
`printIC`	print initial conditions
`printPM`	print non-kinetic model parameters
`printRMPK`	print kinetic model parameters
`psOut`	write ps-PROCEDURES to file `acsl.cmd`

C.3.1 Main program files

`sim.mws`	- maple worksheet for Maple V Release 4 - execution of `sim.mws` calls files `sim.mpl` and `indat.mpl`
`sim.mpl`	- main program, reads `indat.mpl` - model processing: vectors of states, initial conditions, maximum rates, other kinetic constants, non-kinetic constants, experimental data - translates nucleotide sequence into amino acid sequence

	- assigns parameter values to kinetic constants, values taken from file `../bib/param.mpl`
	- automated generation of Fortran source code for user-specified files
`indat.mpl`	- reads gene sequence (`FileNT`), experimental data (multiple data sets possible), reads files `trna.mpl`, `dae.mpl`, `kin.mpl`
	- contains analytical functions
	- contains non-kinetic model parameters
	- contains ACSL-specific parameter settings
	- automated generation of codon-specific queueing factors for both ribosomes and degradosomes
	- calculates endonucleolytic cleavage sites of mRNA
`dae.mpl`	- contains ordinary differential equations and algebraic equations, both including initial conditions (multiple data sets possible)
	- automated generation of codon-specific balance equations
`kin.mpl`	- contains kinetic rate expressions
	- automated generation of codon-specific reaction rates
`trna.mpl`	- physiological concentrations of tRNA (taken from Dong et al. (1996))
`../bib/param.mpl`	- contains kinetic model constants (Vmax[], K[], k[])

C.3.2 Files being generated by source code generation

`./sim.csl`	- main program for numerical integration (ACSL)
	- reads subroutines `./src/model.inc`, `./src/pmod.inc`, `./src/pkin.inc`, `./src/rmax.inc`, `./src/icond.inc`, and `./src/mess.inc`
`./src/model.inc`	- kinetics and differential equations (vectors r, dc)
`./src/pmod.inc`	- non-kinetic constants (pm)
`./src/pkin.inc`	- kinetic constants (pk)
`./src/rmax.inc`	- maximum rates (rm)
`./src/icond.inc`	- initial conditions (ic)
`./src/mess.inc`	- experimental data (mdat)
`./acsl.cmd`	- called by ACSL for regular numerical integration
`./anasubF.f`	- used by OptdesX for parameter estimation
`./sim.cmd`	- called by ACSL in the case of parameter estimation
`./start.ana`	- used by OptdesX (initial guesses)

C.3.3 Files being generated by ACSL

`./optdes.dat`	file exchanges (optimized) parameters between ACSL (`sim.csl`) and OptdesX (`anasubF.f`)
`./acsl.dat`	file exchanges computed error criteria between ACSL (`sim.csl`) and OptdesX (`anasubF.f`)
`./ns.dat`	file exchanges the number of experimental data sets between ACSL (`sim.csl`) and OptdesX (`anasubF.f`)

C.3.4 General guidelines

1. State variables are defined as indexed variables c[].
   (a) This is a general rule:
   do NOT use nested indices like for example c[glc[ex]]
   use instead: c[glc_ex] or c[‘glc[ex]‘] (here, the index is defined as a string)
   (b) according to 1., reaction volume V can be expressed by c[V]!!
2. Reaction rates are defined as indexed quantities r[] (see 1.)).
3. Maximum rates are expressed with Vmax[].
4. Kinetic parameters are defined as K[], k[].
5. All other model parameters should be expressed as non-indexed quantities (e.g., fPulse, cFeed).
6. Multiple experimental data sets can be used (ns = <data_set_number>).
   (a) For simulation of a particular data set: ns can be set at runtime level of ACSL.
   (b) For parameter identification using a particular data set: Change ‘DO’-loop in `anasubF.f` accordingly.
7. In the current setup, all Vmax[], K[], k[] are subject to parameter estimation.
8. Initial conditions can be estimated also (if in Maple, `OptimizeIC := true:`).

D Dynamic Model of Prokaryotic Cell-free Protein Biosynthesis

The following conditions were applied for simulating cell-free synthesis of GFP, as presented in section 8.5.

D.1 Output of model processing

```
The gene expression model contains:
-----------------------------------
    Transcription
    mRNA degradation
    Ribosomal translation
    Energy regeneration
    Queueing in general
    Queueing caused by degradosomes
    Simplification: Clustering of codons with ncod=1

The entered gene sequence contains 1071 bases.
----------------------------------------------
    307  A,  ( 0.2866 )
    247  C,  ( 0.2306 )
    233  G,  ( 0.2175 )
    284  T,  ( 0.2651 )

Model reduction:  1 codon per cluster.
--------------------------------------
            original   clustered
    iAUG      22.000     22.000
    iSTOP    273.000    273.000
    KK       251.000    251.000
    LL        12.000     12.000
    DD        12.000     12.000
    MM       357.000    357.000
    mR         7.000      7.000
    mD         7.000      7.000
    i0R       22.000     22.000
    i0D        7.000      7.000
```

```
mRNA characteristics:
---------------------
    Length of synthesized mRNA:   MM  = 357
    Length of coding region:      KK  = 251
    Coding region:                AUG = 22  to  STOP = 273

Coding sequence with 251 Aa.
----------------------------
    11  Ala,  ( 0.0438 )       19  Leu,  ( 0.0756 )
     7  Arg,  ( 0.0278 )       21  Lys,  ( 0.0836 )
    13  Asn,  ( 0.0517 )        5  Met,  ( 0.0199 )
    18  Asp,  ( 0.0717 )       13  Phe,  ( 0.0517 )
     2  Cys,  ( 0.0079 )       12  Pro,  ( 0.0478 )
     8  Gln,  ( 0.0318 )       14  Ser,  ( 0.0557 )
    17  Glu,  ( 0.0677 )       16  Thr,  ( 0.0637 )
    23  Gly,  ( 0.0916 )        2  Trp,  ( 0.0079 )
    11  His,  ( 0.0438 )       11  Tyr,  ( 0.0438 )
    12  Ile,  ( 0.0478 )       16  Val,  ( 0.0637 )

Number of ODEs:                         nc = 1145
Number of algebraic equations:          na = 7
Total number of reaction kinetics:      nr = 1224

Number of maximum rates:                nrm = 9
Number of ki, Km, Ki, Keq:              npk = 64
Number of non-kinetic parameters:       npm = 14
```

D.2 Kinetic model constants

Parameter	Unit	Value	Source
Transcription			
$V^{max}_{\mathrm{T7RNAP}}$	μM/min	0.09	This study
$K_{M,\mathrm{ATP}}$	μM	76	dto.
$K_{M,\mathrm{CTP}}$	μM	34	dto.
$K_{M,\mathrm{GTP}}$	μM	76	dto.
$K_{M,\mathrm{UTP}}$	μM	33	dto.
$K_{M,\mathrm{DNA}}$	μM	$6.3{\cdot}10^{-3}$	dto.
$K_{i,\mathrm{GTP}}$	μM	0.025	Sen and Dasgupta (1993)
NTPase activity			
$k_{d,\mathrm{NTP}}$	s^{-1}	$6.7{\cdot}10^{-4}$	This study

Parameter	Unit	Value	Source
mRNA degradation			
$k_{D,ass}$	s^{-1}	$2{\cdot}10^{-4}$	This study
$k_{D,Term}$	s^{-1}	50	dto.
$k_{D,endo}$	s^{-1}	2.6	dto.
$k_{D,exo}$	nt s^{-1}	680	dto.
$k_{D,mv}$	nt s^{-1}	95	dto.
70S initiation complex formation			
$k_{\mathrm{TLI,70SIC}}$	s^{-1}	$2.5{\cdot}10^{-3}$	This study
$K_{M,\mathrm{50S}}$	μM	0.011	dto.
$K_{M,\mathrm{fMet\text{-}tRNA}_f^{\mathrm{M}}}$	μM	0.053	Gualerzi et al. (1977)
$K_{M,\mathrm{mRNA}}$	μM	0.01	dto.
IF2-dependent GTP hydrolysis			
$k_{\mathrm{TLI,IF2D}}$	s^{-1}	0.8	Liang et al. (1999)
Translation elongation			
$k_{\mathrm{TLE},j}$	s^{-1}	24	This study
$K_{M,\mathrm{T3}_j}$	μM	0.4	dto.
$K_{M,\mathrm{EFG{\cdot}GTP}}$	μM	0.22	Schmid (1999)
EFG regeneration			
$k_{\mathrm{EFG{\cdot}GTP}}$	$M^{-1}s^{-1}$	$1.0{\cdot}10^{7}$	Gast (1987)
$k_{\text{-EFG{\cdot}GTP}}$	s^{-1}	400	dto.
$k_{\mathrm{EFG{\cdot}GDP}}$	$M^{-1}s^{-1}$	$2.7{\cdot}10^{7}$	dto.
$k_{\text{-EFG{\cdot}GDP}}$	s^{-1}	100	dto.
Translation termination			
k_{TLT}	s^{-1}	24	This study
$K_{M,\mathrm{GTP}}$	μM	100	dto.
K_{M,R_K}	μM	$8.3{\cdot}10^{-3}$	Freistroffer et al. (1997)
Ternary complex formation			
k_{T3_j}	$M^{-1}s^{-1}$	$5{\cdot}10^{7}$	Gast (1987)
$k_{\text{-T3}_j}$	s^{-1}	1.0	dto.
tRNA charging			
V_{ARS}^{max}	μM/min	10	This study
$K_{M,\mathrm{ATP}}$	μM	100	dto.
K_{M,aa_j}	μM	20	dto.
K_{M,tRNA_j}	μM	0.5	dto.

Parameter	Unit	Value	Source
EFTu regeneration			
k_f	s^{-1}	30	Ruusala et al. (1982a)
k_r	s^{-1}	10	dto.
K_{eq}	-	0.4	This study
$K_{M,\mathrm{EFTu\cdot GTP}}$	μM	1.0	dto.
$K_{M,\mathrm{EFTu\cdot GDP}}$	μM	2.5	Ruusala et al. (1982a)
$K_{M,\mathrm{GTP}}$	μM	50	This study
$K_{M,\mathrm{GDP}}$	μM	3.0	Hwang and Miller (1985)
$K_{i,\mathrm{EFTu\cdot GTP}}$	μM	1.0	This study
$K_{i,\mathrm{EFTu\cdot GDP}}$	μM	5.6	dto.
Chemical hydrolysis of AcP			
$k_{d,\mathrm{AcP}}$	s^{-1}	$3.3\cdot10^{-5}$	This study
Acetate kinase			
$V^{max}_{\mathrm{Ack},f}$	μM/min	4,000	This study
$V^{max}_{\mathrm{Ack},r}$	μM/min	900	dto.
K_{eq}	-	114	Janson and Cleland (1974)
$K_{M,\mathrm{AcP}}$	μM	340	dto.
$K_{M,\mathrm{Ac}}$	μM	5,800	dto.
$K_{M,\mathrm{ATP}}$	μM	20	dto.
$K_{M,\mathrm{ADP}}$	μM	360	dto.
$K_{i,\mathrm{AcP}}$	μM	47	dto.
$K_{i,\mathrm{Ac}}$	μM	100,000	dto.
$K_{i,\mathrm{ATP}}$	μM	350	dto.
$K_{i,\mathrm{ADP}}$	μM	50	dto.
Adenylate kinase			
$V^{max}_{\mathrm{Adk},f}$	μM/min	80	This study
$V^{max}_{\mathrm{Adk},r}$	μM/min	12	dto.
$K_{M,\mathrm{ATP}}$	μM	51	Rose et al. (1991)
$K_{M,\mathrm{ADP}}$	μM	92	dto.
$K_{M,\mathrm{AMP}}$	μM	38	dto.
Inactivation kinetics			
$k_{d,\mathrm{TLI}}$	s^{-1}	$8.9\cdot10^{-4}$	This study
$k_{d,\mathrm{T7RNAP}}$	s^{-1}	$5\cdot10^{-5}$	dto.
$k_{d,\mathrm{EFTu}}$	s^{-1}	$2.3\cdot10^{-4}$	dto.
$k_{d,\mathrm{EFTs}}$	s^{-1}	$1.9\cdot10^{-4}$	dto.

D.3 Non-kinetic model constants

Parameter	Unit	Value	Source
f_{A}	-	0.2652	This study
f_{C}	-	0.2176	dto.
f_{G}	-	0.2306	dto.
f_{U}	-	0.2866	dto.

D.4 Initial conditions

Concentration	(μM)
C_{Protein}	0
C_j^D	0
C_j^M	0
C_j^R	0
C_{AcP}	34,500
C_{ATP}	2,000
C_{ADP}	106
C_{AMP}	8
C_{GTP}	1,550
C_{GDP}	75
C_{GMP}	0
C_{CTP}	1,000
C_{CDP}	50
C_{CMP}	0
C_{UTP}	1,000
C_{UDP}	50
C_{UMP}	0
$C_{30\mathrm{S},t}$	1.4
$C_{50\mathrm{S},t}$	1.4
C_{DNA}	$5{\cdot}10^{-3}$
$C_{\mathrm{EFG},t}$	1.2120
$C_{\mathrm{EFTu},t}$	1.0605
$C_{\mathrm{EFTs},t}$	0.2727
$C_{\mathrm{IF1},t}$	0.3788
$C_{\mathrm{IF2},t}$	0.4545
$C_{\mathrm{IF3},t}$	0.3030
$C_{\mathrm{RF},t}$	1.7574
$C_{30\mathrm{S}}$	0.0065
$C_{50\mathrm{S}}$	0.3159
C_{IF1}	0.0704

Concentration	(μM)
C_{aa_j}	250
$C_{\mathrm{T3},j}$	0
$C_{\mathrm{aa}_j\text{-tRNA}^k}$	0
$C_{\mathrm{fMet\text{-}tRNA}_f^{\mathrm{M}}}$	20
$C_{\mathrm{tRNA}^{\mathrm{Met}}}$	0.8678
$C_{\mathrm{tRNA}^{\mathrm{Ala1B}}}$	1.0957
$C_{\mathrm{tRNA}^{\mathrm{Ala2}}}$	0.1941
$C_{\mathrm{tRNA}^{\mathrm{Arg2}}}$	1.4002
$C_{\mathrm{tRNA}^{\mathrm{Arg3}}}$	0.1320
$C_{\mathrm{tRNA}^{\mathrm{Asn}}}$	0.3681
$C_{\mathrm{tRNA}^{\mathrm{Asp1}}}$	0.7232
$C_{\mathrm{tRNA}^{\mathrm{Cys}}}$	0.4303
$C_{\mathrm{tRNA}^{\mathrm{Gln1}}}$	0.2242
$C_{\mathrm{tRNA}^{\mathrm{Gln2}}}$	0.3025
$C_{\mathrm{tRNA}^{\mathrm{Glu2}}}$	1.4449
$C_{\mathrm{tRNA}^{\mathrm{Gly12}}}$	0.6594
$C_{\mathrm{tRNA}^{\mathrm{Gly3}}}$	1.2607
$C_{\mathrm{tRNA}^{\mathrm{His}}}$	0.2083
$C_{\mathrm{tRNA}^{\mathrm{Ile12}}}$	1.1365
$C_{\mathrm{tRNA}^{\mathrm{Leu1}}}$	1.3246
$C_{\mathrm{tRNA}^{\mathrm{Leu2}}}$	0.3013
$C_{\mathrm{tRNA}^{\mathrm{Leu3}}}$	0.2010
$C_{\mathrm{tRNA}^{\mathrm{Leu5}}}$	0.2566
$C_{\mathrm{tRNA}^{\mathrm{Lys}}}$	0.5545
$C_{\mathrm{tRNA}^{\mathrm{Phe}}}$	0.3063
$C_{\mathrm{tRNA}^{\mathrm{Pro1}}}$	0.2038
$C_{\mathrm{tRNA}^{\mathrm{Pro2}}}$	0.2275
$C_{\mathrm{tRNA}^{\mathrm{Pro3}}}$	0.1629
$C_{\mathrm{tRNA}^{\mathrm{Ser1}}}$	0.4333
$C_{\mathrm{tRNA}^{\mathrm{Ser2}}}$	0.0879

C_{IF2}	0.1137	$C_{\mathrm{tRNA^{Ser3}}}$	0.3430
C_{IF3}	0.0132	$C_{\mathrm{tRNA^{Ser5}}}$	0.2288
C_{EFG}	0.0202	$C_{\mathrm{tRNA^{Thr13}}}$	0.3402
$C_{\mathrm{EFG{\cdot}GTP}}$	0.7816	$C_{\mathrm{tRNA^{Thr2}}}$	0.1655
$C_{\mathrm{EFG{\cdot}GDP}}$	0.4102	$C_{\mathrm{tRNA^{Thr4}}}$	0.2933
$C_{\mathrm{EFTu{\cdot}GTP}}$	0.7135	$C_{\mathrm{tRNA^{Trp}}}$	0.2605
$C_{\mathrm{EFTu{\cdot}GDP}}$	0.3467	$C_{\mathrm{tRNA^{Tyr12}}}$	0.5800
C_{Ac}	136,000	$C_{\mathrm{tRNA^{Val1}}}$	1.0867
C_{Pi}	0	$C_{\mathrm{tRNA^{Val2A2B}}}$	0.3941

D.5 Material balance equations

$$d[Protein]/dt = r_{TLT}$$

$$d[AcP]/dt = -r_{d,AcP} - r_{Ack,ATP} - r_{Ack,CTP} - r_{Ack,GTP} - r_{Ack,UTP}$$

$$\begin{aligned} d[ATP]/dt = {} & -r_{d,ATP} - 1071.\,fA\,r_{TC357} - r_{SumARS,ATP} + r_{Ack,ATP} \\ & - r_{Adk,AMP} - r_{Adk,CMP} - r_{Adk,GMP} - r_{Adk,UMP} \end{aligned}$$

$$d[ADP]/dt = r_{d,ATP} - r_{Ack,ATP} + 2.\,r_{Adk,AMP} + r_{Adk,CMP} + r_{Adk,GMP} + r_{Adk,UMP}$$

$$\begin{aligned} d[AMP]/dt = {} & r_{SumARS,ATP} - r_{Adk,AMP} + fA(35\,r_{D,exo13} + 13\,r_{D,exo17} + 69\,r_{D,exo40} \\ & + 28\,r_{D,exo49} + 58\,r_{D,exo69} + 114\,r_{D,exo107} + 109\,r_{D,exo143} + 15\,r_{D,exo148} \\ & + 13\,r_{D,exo152} + 46\,r_{D,exo168} + 36\,r_{D,exo180} + 76\,r_{D,exo205} + 9\,r_{D,exo208} \\ & + 197\,r_{D,exo274} + 71\,r_{D,exo297} + 21\,r_{D,exo304} + 115\,r_{D,exo343} + 46\,r_{D,exo357}) \end{aligned}$$

$$\begin{aligned} d[GTP]/dt = {} & -r_{TLI,IF2D} - r_{EFTu,Reg} - r_{TLT} - r_{EFG*GTP,Ass} + r_{Ack,GTP} \\ & - r_{d,GTP} - 1071.\,fG\,r_{TC357} \end{aligned}$$

$$d[EFTutot]/dt = -r_{d,EFTu}$$

$$d[EFTstot]/dt = -r_{d,EFTs}$$

$$d[EFG*GTP]/dt = r_{EFG*GTP,Ass} - r_{TL,SumElong}$$

$$d[EFG*GDP]/dt = r_{TL,SumElong} - r_{EFG*GDP,Diss}$$

$$d[EFTu*GTP]/dt = r_{EFTu,Reg} - r_{SumT3form}$$

$$d[Ac]/dt = r_{d,AcP} + r_{Ack,ATP} + r_{Ack,CTP} + r_{Ack,GTP} + r_{Ack,UTP}$$

$$\begin{aligned} d[Pi]/dt = {} & r_{TLI,IF2D} + r_{TLT} + 2.\,r_{TL,SumElong} + r_{d,ATP} + r_{d,CTP} \\ & + r_{d,GTP} + r_{d,UTP} + 2136.\,r_{TC357} + r_{d,AcP} \end{aligned}$$

$$d[GDP]/dt = r_{TLI,IF2D} + r_{EFTu,Reg} + r_{EFG*GDP,Diss} + r_{d,EFTu} + r_{TLT} + r_{d,GTP} + r_{Adk,GMP} - r_{Ack,GTP}$$

$$\begin{aligned} d[GMP]/dt = -r_{Adk,GMP} + fG(&35\,r_{D,exo13} + 13\,r_{D,exo17} + 69\,r_{D,exo40} + 28\,r_{D,exo49} \\ &+ 58\,r_{D,exo69} + 114\,r_{D,exo107} + 109\,r_{D,exo143} + 15\,r_{D,exo148} + 13\,r_{D,exo152} \\ &+ 46\,r_{D,exo168} + 36\,r_{D,exo180} + 76\,r_{D,exo205} + 9\,r_{D,exo208} + 197\,r_{D,exo274} \\ &+ 71\,r_{D,exo297} + 21\,r_{D,exo304} + 115\,r_{D,exo343} + 46\,r_{D,exo357}) \end{aligned}$$

$$d[CTP]/dt = r_{Ack,CTP} - r_{d,CTP} - 1071.\,fC\,r_{TC357}$$

$$d[CDP]/dt = r_{Adk,CMP} - r_{Ack,CTP} + r_{d,CTP}$$

$$\begin{aligned} d[CMP]/dt = -r_{Adk,CMP} + fC(&35\,r_{D,exo13} + 13\,r_{D,exo17} + 69\,r_{D,exo40} + 28\,r_{D,exo49} \\ &+ 58\,r_{D,exo69} + 114\,r_{D,exo107} + 109\,r_{D,exo143} + 15\,r_{D,exo148} + 13\,r_{D,exo152} \\ &+ 46\,r_{D,exo168} + 36\,r_{D,exo180} + 76\,r_{D,exo205} + 9\,r_{D,exo208} + 197\,r_{D,exo274} \\ &+ 71\,r_{D,exo297} + 21\,r_{D,exo304} + 115\,r_{D,exo343} + 46\,r_{D,exo357}) \end{aligned}$$

$$d[UTP]/dt = r_{Ack,UTP} - r_{d,UTP} - 1071.\,fU\,r_{TC357}$$

$$d[UDP]/dt = r_{Adk,UMP} - r_{Ack,UTP} + r_{d,UTP}$$

$$\begin{aligned} d[UMP]/dt = -r_{Adk,UMP} + fU(&35\,r_{D,exo13} + 13\,r_{D,exo17} + 69\,r_{D,exo40} + 28\,r_{D,exo49} \\ &+ 58\,r_{D,exo69} + 114\,r_{D,exo107} + 109\,r_{D,exo143} + 15\,r_{D,exo148} + 13\,r_{D,exo152} \\ &+ 46\,r_{D,exo168} + 36\,r_{D,exo180} + 76\,r_{D,exo205} + 9\,r_{D,exo208} + 197\,r_{D,exo274} \\ &+ 71\,r_{D,exo297} + 21\,r_{D,exo304} + 115\,r_{D,exo343} + 46\,r_{D,exo357}) \end{aligned}$$

$$d[Met]/dt = -r_{sMet}$$

$$d[fMettRNA]/dt = r_{Mformyl} - r_{TLI70SIC}$$

$$d[fTHF]/dt = -r_{Mformyl}$$

$$d[T3Met]/dt = r_{T3Met} - r_{sT3Met}$$

$$d[Met - tRNAMet]/dt = r_{MetRS} - r_{T3Met} - r_{Mformyl}$$

$$d[tRNAMet]/dt = r_{stRNAMet} - r_{MetRS}$$

$$d[Ala]/dt = -r_{sAla}$$

$$d[Arg]/dt = -r_{sArg}$$

$$d[Asn]/dt = -r_{sAsn}$$

$$d[Asp]/dt = -r_{sAsp}$$

$$d[Cys]/dt = -r_{sCys}$$

$$d[Gln]/dt = -r_{sGln}$$

$$d[Glu]/dt = -r_{sGlu}$$

$$d[Gly]/dt = -r_{sGly}$$

$$d[His]/dt = -r_{sHis}$$

$$d[Ile]/dt = -r_{sIle}$$

$$d[Leu]/dt = -r_{sLeu}$$

$$d[Lys]/dt = -r_{sLys}$$

$$d[Phe]/dt = -r_{sPhe}$$

$$d[Pro]/dt = -r_{sPro}$$

$$d[Ser]/dt = -r_{sSer}$$

$$d[Thr]/dt = -r_{sThr}$$

$$d[Trp]/dt = -r_{sTrp}$$

$$d[Tyr]/dt = -r_{sTyr}$$

$$d[Val]/dt = -r_{sVal}$$

$$d[T3Ala1B]/dt = r_{T3Ala1B} - r_{sT3Ala1B}$$

$$d[Ala - tRNAAla1B]/dt = -r_{T3Ala1B} + r_{Ala1BRS}$$

$$d[tRNAAla1B]/dt = r_{stRNAAla1B} - r_{Ala1BRS}$$

$$d[T3Ala2]/dt = r_{T3Ala2} - r_{sT3Ala2}$$

$$d[Ala - tRNAAla2]/dt = -r_{T3Ala2} + r_{Ala2RS}$$

$$d[tRNAAla2]/dt = r_{stRNAAla2} - r_{Ala2RS}$$

$$d[T3Arg2]/dt = r_{T3Arg2} - r_{sT3Arg2}$$

$$d[Arg - tRNAArg2]/dt = -r_{T3Arg2} + r_{Arg2RS}$$

$$d[tRNAArg2]/dt = r_{stRNAArg2} - r_{Arg2RS}$$

$$d[T3Arg3]/dt = r_{T3Arg3} - r_{sT3Arg3}$$

$$d[Arg - tRNAArg3]/dt = -r_{T3Arg3} + r_{Arg3RS}$$

$$d[tRNAArg3]/dt = r_{stRNAArg3} - r_{Arg3RS}$$

$$d[T3Asn]/dt = r_{T3Asn} - r_{sT3Asn}$$

$$d[Asn - tRNAAsn]/dt = -r_{T3Asn} + r_{AsnRS}$$

$$d[tRNAAsn]/dt = r_{stRNAAsn} - r_{AsnRS}$$

$$d[T3Asp1]/dt = r_{T3Asp1} - r_{sT3Asp1}$$

$$d[Asp - tRNAAsp1]/dt = -r_{T3Asp1} + r_{Asp1RS}$$

$$d[tRNAAsp1]/dt = r_{stRNAAsp1} - r_{Asp1RS}$$

$$d[T3Cys]/dt = r_{T3Cys} - r_{sT3Cys}$$

$$d[Cys - tRNACys]/dt = -r_{T3Cys} + r_{CysRS}$$

$$d[tRNACys]/dt = r_{stRNACys} - r_{CysRS}$$

$$d[T3Gln1]/dt = r_{T3Gln1} - r_{sT3Gln1}$$

$$d[Gln - tRNAGln1]/dt = -r_{T3Gln1} + r_{Gln1RS}$$

$$d[tRNAGln1]/dt = r_{stRNAGln1} - r_{Gln1RS}$$

$$d[T3Gln2]/dt = r_{T3Gln2} - r_{sT3Gln2}$$

$$d[Gln - tRNAGln2]/dt = -r_{T3Gln2} + r_{Gln2RS}$$

$$d[tRNAGln2]/dt = r_{stRNAGln2} - r_{Gln2RS}$$

$$d[T3Glu2]/dt = r_{T3Glu2} - r_{sT3Glu2}$$

$$d[Glu - tRNAGlu2]/dt = -r_{T3Glu2} + r_{Glu2RS}$$

$$d[tRNAGlu2]/dt = r_{stRNAGlu2} - r_{Glu2RS}$$

$$d[T3Gly12]/dt = r_{T3Gly12} - r_{sT3Gly12}$$

$$d[Gly - tRNAGly12]/dt = -r_{T3Gly12} + r_{Gly12RS}$$

$$d[tRNAGly12]/dt = r_{stRNAGly12} - r_{Gly12RS}$$

$$d[T3Gly3]/dt = r_{T3Gly3} - r_{sT3Gly3}$$

$$d[Gly - tRNAGly3]/dt = -r_{T3Gly3} + r_{Gly3RS}$$

$$d[tRNAGly3]/dt = r_{stRNAGly3} - r_{Gly3RS}$$

$$d[T3His]/dt = r_{T3His} - r_{sT3His}$$

$$d[His - tRNAHis]/dt = -r_{T3His} + r_{HisRS}$$

$$d[tRNAHis]/dt = r_{stRNAHis} - r_{HisRS}$$

$$d[T3Ile12]/dt = r_{T3Ile12} - r_{sT3Ile12}$$

$$d[Ile - tRNAIle12]/dt = -r_{T3Ile12} + r_{Ile12RS}$$

$$d[tRNAIle12]/dt = r_{stRNAIle12} - r_{Ile12RS}$$

$$d[T3Leu1]/dt = r_{T3Leu1} - r_{sT3Leu1}$$

$$d[Leu - tRNALeu1]/dt = -r_{T3Leu1} + r_{Leu1RS}$$

$$d[tRNALeu1]/dt = r_{stRNALeu1} - r_{Leu1RS}$$

$$d[T3Leu2]/dt = r_{T3Leu2} - r_{sT3Leu2}$$

$$d[Leu - tRNALeu2]/dt = -r_{T3Leu2} + r_{Leu2RS}$$

$$d[tRNALeu2]/dt = r_{stRNALeu2} - r_{Leu2RS}$$

$$d[T3Leu3]/dt = r_{T3Leu3} - r_{sT3Leu3}$$

$$d[Leu - tRNALeu3]/dt = -r_{T3Leu3} + r_{Leu3RS}$$

$$d[tRNALeu3]/dt = r_{stRNALeu3} - r_{Leu3RS}$$

$$d[T3Leu5]/dt = r_{T3Leu5} - r_{sT3Leu5}$$

$$d[Leu - tRNALeu5]/dt = -r_{T3Leu5} + r_{Leu5RS}$$

$$d[tRNALeu5]/dt = r_{stRNALeu5} - r_{Leu5RS}$$

$$d[T3Lys]/dt = r_{T3Lys} - r_{sT3Lys}$$

$$d[Lys - tRNALys]/dt = -r_{T3Lys} + r_{LysRS}$$

$$d[tRNALys]/dt = r_{stRNALys} - r_{LysRS}$$

$$d[T3Phe]/dt = r_{T3Phe} - r_{sT3Phe}$$

$$d[Phe - tRNAPhe]/dt = -r_{T3Phe} + r_{PheRS}$$

$$d[tRNAPhe]/dt = r_{stRNAPhe} - r_{PheRS}$$

$$d[T3Pro1]/dt = r_{T3Pro1} - r_{sT3Pro1}$$

$$d[Pro - tRNAPro1]/dt = -r_{T3Pro1} + r_{Pro1RS}$$

$$d[tRNAPro1]/dt = r_{stRNAPro1} - r_{Pro1RS}$$

$$d[T3Pro2]/dt = r_{T3Pro2} - r_{sT3Pro2}$$

$$d[Pro - tRNAPro2]/dt = -r_{T3Pro2} + r_{Pro2RS}$$

$$d[tRNAPro2]/dt = r_{stRNAPro2} - r_{Pro2RS}$$

$$d[T3Pro3]/dt = r_{T3Pro3} - r_{sT3Pro3}$$

$$d[Pro - tRNAPro3]/dt = -r_{T3Pro3} + r_{Pro3RS}$$

$$d[tRNAPro3]/dt = r_{stRNAPro3} - r_{Pro3RS}$$

$$d[T3Ser1]/dt = r_{T3Ser1} - r_{sT3Ser1}$$

$$d[Ser - tRNASer1]/dt = -r_{T3Ser1} + r_{Ser1RS}$$

$$d[tRNASer1]/dt = r_{stRNASer1} - r_{Ser1RS}$$

$$d[T3Ser2]/dt = r_{T3Ser2} - r_{sT3Ser2}$$

$$d[Ser - tRNASer2]/dt = -r_{T3Ser2} + r_{Ser2RS}$$

$$d[tRNASer2]/dt = r_{stRNASer2} - r_{Ser2RS}$$

$$d[T3Ser3]/dt = r_{T3Ser3} - r_{sT3Ser3}$$

$$d[Ser - tRNASer3]/dt = -r_{T3Ser3} + r_{Ser3RS}$$

$$d[tRNASer3]/dt = r_{stRNASer3} - r_{Ser3RS}$$

$$d[T3Ser5]/dt = r_{T3Ser5} - r_{sT3Ser5}$$

$$d[Ser - tRNASer5]/dt = -r_{T3Ser5} + r_{Ser5RS}$$

$$d[tRNASer5]/dt = r_{stRNASer5} - r_{Ser5RS}$$

$$d[T3Thr13]/dt = r_{T3Thr13} - r_{sT3Thr13}$$

$$d[Thr - tRNAThr13]/dt = -r_{T3Thr13} + r_{Thr13RS}$$

$$d[tRNAThr13]/dt = r_{stRNAThr13} - r_{Thr13RS}$$

$$d[T3Thr2]/dt = r_{T3Thr2} - r_{sT3Thr2}$$

$$d[Thr - tRNAThr2]/dt = -r_{T3Thr2} + r_{Thr2RS}$$

$$d[tRNAThr2]/dt = r_{stRNAThr2} - r_{Thr2RS}$$

$$d[T3Thr4]/dt = r_{T3Thr4} - r_{sT3Thr4}$$

$$d[Thr - tRNAThr4]/dt = -r_{T3Thr4} + r_{Thr4RS}$$

$$d[tRNAThr4]/dt = r_{stRNAThr4} - r_{Thr4RS}$$

$$d[T3Trp]/dt = r_{T3Trp} - r_{sT3Trp}$$

$$d[Trp - tRNATrp]/dt = -r_{T3Trp} + r_{TrpRS}$$

$$d[tRNATrp]/dt = r_{stRNATrp} - r_{TrpRS}$$

$$d[T3Tyr12]/dt = r_{T3Tyr12} - r_{sT3Tyr12}$$

$$d[Tyr - tRNATyr12]/dt = -r_{T3Tyr12} + r_{Tyr12RS}$$

$$d[tRNATyr12]/dt = r_{stRNATyr12} - r_{Tyr12RS}$$

$$d[T3Val1]/dt = r_{T3Val1} - r_{sT3Val1}$$

$$d[Val - tRNAVal1]/dt = -r_{T3Val1} + r_{Val1RS}$$

$$d[tRNAVal1]/dt = r_{stRNAVal1} - r_{Val1RS}$$

$$d[T3Val2A2B]/dt = r_{T3Val2A2B} - r_{sT3Val2A2B}$$

$$d[Val - tRNAVal2A2B]/dt = -r_{T3Val2A2B} + r_{Val2A2BRS}$$

$$d[tRNAVal2A2B]/dt = r_{stRNAVal2A2B} - r_{Val2A2BRS}$$

$$d[M1]/dt = r_{TC1} - r_{D,\,endo13}$$

$$d[M2]/dt = r_{TC2} - r_{D,\,endo13}$$

$$\vdots$$

$$d[M356]/dt = r_{TC356} - r_{D,\,endo357}$$

$$d[M357]/dt = r_{TC357} - r_{D,T}$$

$$d[R*22]/dt = r_{TLI70SIC} - r_{TLI,IF2D}$$

$$d[R22]/dt = r_{TLI,IF2D} - r_{TLE22}$$

$$d[R23]/dt = r_{TLE22} - r_{TLE23}$$

$$\vdots$$

$$d[R272]/dt = r_{TLE271} - r_{TLE272}$$

$$d[R273]/dt = r_{TLE272} - r_{TLT}$$

$$d[D7]/dt = r_{D,Ass} - r_{D,mv7}$$

$$d[D8]/dt = r_{D,mv7} - r_{D,mv8}$$

$$d[D9]/dt = r_{D,mv8} - r_{D,mv9}$$

$$d[D10]/dt = r_{D,mv9} - r_{D,mv10}$$

$$d[D11]/dt = r_{D,mv10} - r_{D,mv11}$$

$$d[D12]/dt = r_{D,mv11} - r_{D,mv12}$$

$$d[D13*]/dt = r_{D,mv12} - r_{D,\,endo13}$$

$$d[D13*Frag13]/dt = r_{D,\,endo13} - r_{D,exo13}$$

$$d[D13]/dt = r_{D,exo13} - r_{D,mv13}$$

$$d[D14]/dt = r_{D,mv13} - r_{D,mv14}$$

$$\vdots$$

$$d[D356]/dt = r_{D,mv355} - r_{D,mv356}$$

$$d[D357*]/dt = r_{D,mv356} - r_{D,\,endo357}$$

$$d[D357*Frag357]/dt = r_{D,\,endo357} - r_{D,exo357}$$

$$d[D357]/dt = r_{D,exo357} - r_{D,T}$$

$$C_{30S} - C_{30S,t} + C_{30SIF1} + C_{30SIF2} + C_{30SIF3} + C_{30SIF12} + C_{30SIF13} + C_{30SIF23} + C_{30SIF123} + C_{70S} + C_{Rbound} = 0$$

$$C_{50S} - C_{50S,t} + C_{70S} + C_{Rbound} = 0$$

$$C_{IF1} - C_{IF1,t} + C_{30SIF1} + C_{30SIF12} + C_{30SIF13} + C_{30SIF123} = 0$$

$$C_{IF2} - C_{IF2,t} + C_{30SIF2} + C_{30SIF12} + C_{30SIF23} + C_{30SIF123} = 0$$

$$C_{IF3} - C_{IF3,t} + C_{30SIF3} + C_{30SIF13} + C_{30SIF23} + C_{30SIF123} = 0$$

$$C_{EFG} - C_{EFG,t} + C_{EFG*GTP} + C_{EFG*GDP} = 0$$

$$C_{EFTu*GDP} - C_{EFTu,t} + C_{EFTu*GTP} + C_{T3} = 0$$

E Derivation of Queueing Factors for System with Two Catalysts

The following paragraphs provide an extension of a model approach previously suggested by the working group of Gibbs for template-directed and enzyme-catalyzed polymerization (Pipkin and Gibbs, 1966; MacDonald et al., 1968; MacDonald and Gibbs, 1969). In the original study, sterical interactions among template-bound catalysts of the same type were considered. In particular, probabilities for movement from one site on the template to the next site were developed. In the following, an analogous derivation of these probabilities is given for the case of two types of catalysts (in multiple copies) bound to a same template. Further new aspects of this model arise owing to the transition from a fractional system description to employing molarities, and the resulting consequences on material balancing. Some part of the model derivation presented here has been published elsewhere (Schmid, 1999).

Nomenclature. Let L_D and L_R denote the lengths of two different catalysts, respectively. The index D may refer to a degradosome, the index R may signify a ribosome. Both catalysts are believed to propagate into the same direction and one site at a time along a linear template. The template is divided into J different sites, as is shown in Figure E.1.

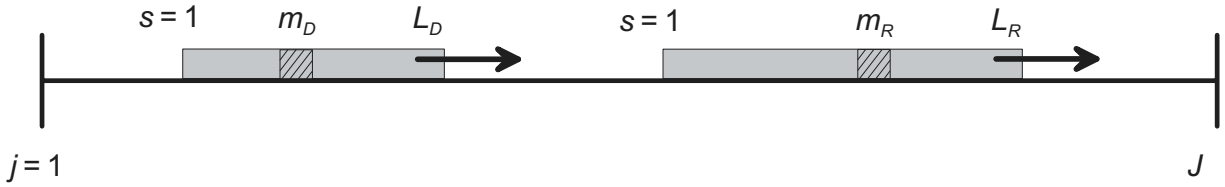

Figure E.1: Definition of states, when two different types of catalysts are bound to a template.

Parameters m_D (with $1 \leq m_D \leq L_D$) and m_R (with $1 \leq m_R \leq L_R$) characterize the positions of the catalytic center for catalysts D and R, respectively. If a site j is known to be covered by catalyst D, its surrounding $j - m_D + 1, \ \ldots, \ j - m_D + L_D$ sites are simultaneously blocked by this catalyst. Similarly, catalyst R covers at a time L_R sites within the vicinity of its binding site. Overlapping of catalysts is excluded (MacDonald and Gibbs, 1969). Although no restrictions are imposed with respect to maximum relative velocity of each catalyst, their propagation may be hindered due to physical obstruction.

The relative positions of a catalyst, while site j is in different states, are explained in Figure E.2. A site j on the template can be either empty (state $s = 0$), or in L_D different states of catalyst D, or L_R different states of catalyst R. In total, that makes $L_D + L_R + 1$ different states s for each site. The fractional occupancy of site j being occupied by catalyst D that is in state s is given by $n_j^{(s)}$. The fractional occupancy of this site with respect to catalyst R being in state s is denoted by $\tilde{n}_j^{(s)}$. The tilde is used here as an indicator for catalyst R. The summation over all the states for site j leads to unity, according to

$$n_j^{(0)} + \sum_{s=1}^{L_D} n_j^{(s)} + \sum_{s=1}^{L_R} \tilde{n}_j^{(s)} \ = \ 1 \tag{E.1}$$

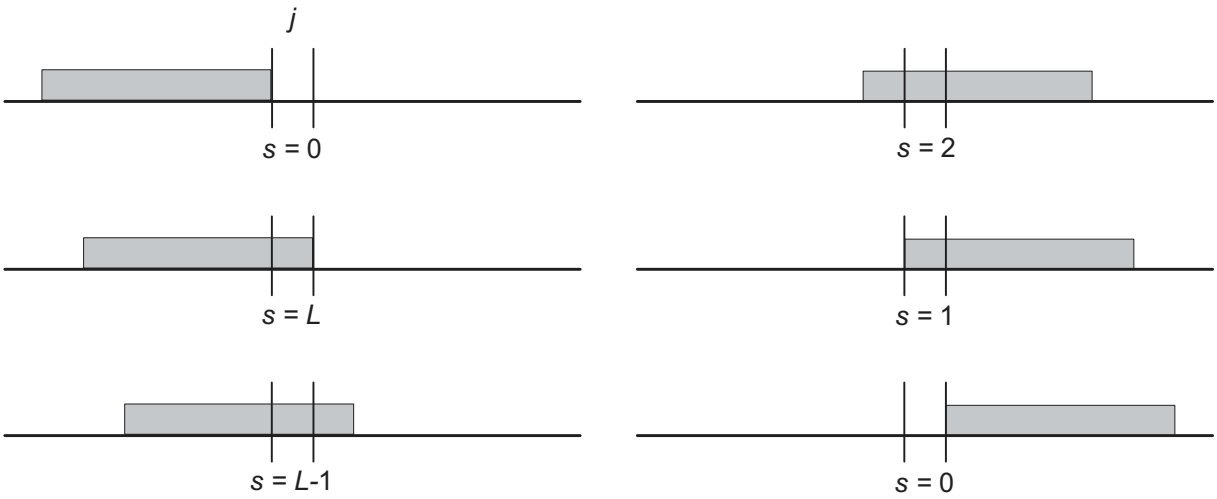

Figure E.2: Definition of the different states a template-bound catalyst can take.

Probabilities for unoccupied sites. Site $j+1$ can be empty only if site j is either in state 0, L_D, or state L_R, but not otherwise. Any other state s would cause a blocking of position $j+1$ and thus preclude the catalyst movement onto this site. Then, if site j is in either of the states 0, L_D, or L_R, site $j+1$ must take one of exactly three states: site $j+1$ is in this case either unoccupied ($s=0$), or in state 1 of either of the two catalysts.

Individual states of site j are distinguished together with the restrictions consequently imposed on site $j+1$. If site j is in state 0, then there are at the same time only three states possible for site j+1, namely in this case to be either empty (s=0), or to be in state 1 of catalyst D, or else to be in state 1 of catalyst R. It follows that if site j is in state L_D, then site $j+1$ can only take any one of the three states, either to be 0, or 1 (for either of the two catalysts). Thus, if site j is in any one of the states, 0, L_D, or L_R, respectively, then at the same time, site $j+1$ needs to be in any one of the three states 0 or 1 (for catalysts D and R, respectively). The converse is true, too. This leads to the following relation:

$$n_j^{(0)} + n_j^{(L_D)} + \tilde{n}_j^{(L_R)} \;=\; n_{j+1}^{(0)} + n_{j+1}^{(1)} + \tilde{n}_{j+1}^{(1)} \tag{E.2}$$

The sum of fractional loadings of site j in states 0, L_D, and L_R just equals the sum of fractions in states 0 and 1 of site $j+1$. Under the assumption that there exists no causal relationship for site $j+1$ to be empty whether site j is in state L_D, or L_R, or empty itself (MacDonald and Gibbs, 1969), the conditional probability, q_j, that site $j+1$ is empty may be expressed as

$$q_j \;=\; P\left(n_{j+1}^{(0)} \,\middle|\, n_j^{(0)} \cup n_j^{(L_D)} \cup \tilde{n}_j^{(L_R)}\right) \tag{E.3}$$

By using equation (E.2), equation (E.3) becomes

$$q_j \;=\; P\left(n_{j+1}^{(0)} \,\middle|\, n_{j+1}^{(0)} \cup n_{j+1}^{(1)} \cup \tilde{n}_{j+1}^{(1)}\right) \tag{E.4}$$

This is likewise represented by

$$q_j \;=\; \frac{n_{j+1}^{(0)}}{n_{j+1}^{(0)} + n_{j+1}^{(1)} + \tilde{n}_{j+1}^{(1)}} \tag{E.5}$$

Considering equation (E.1), equation (E.5) yields

$$q_j = \frac{1 - \sum_{s=1}^{L_D} n_{j+1}^{(s)} - \sum_{s=1}^{L_R} \tilde{n}_{j+1}^{(s)}}{1 - \sum_{s=1}^{L_D} n_{j+1}^{(s)} - \sum_{s=1}^{L_R} \tilde{n}_{j+1}^{(s)} + n_{j+1}^{(1)} + \tilde{n}_{j+1}^{(1)}} \tag{E.6}$$

A transformation of variables, as suggested earlier (MacDonald and Gibbs, 1969), leads to an expression for the state s relative to the states $s = L_D$ and $s = L_R$, respectively. I.e.,

$$n_j^{(s)} = n_{j-s+L_D}^{(L_D)} \qquad \text{for } 1 \le s \le L_D \tag{E.7}$$

and

$$\tilde{n}_j^{(s)} = \tilde{n}_{j-s+L_R}^{(L_R)} \qquad \text{for } 1 \le s \le L_R \tag{E.8}$$

With equation (E.7), it can be shown that the following relation holds for $1 \le s \le L_D$:

$$\sum_{s=1}^{L_D} n_j^{(s)} = \sum_{s=1}^{L_D} n_{j-s+L_D}^{(L_D)} = \sum_{s=1}^{L_D} n_{j+s-1}^{(L_D)} \tag{E.9}$$

Analogously, for $1 \le s \le L_R$:

$$\sum_{s=1}^{L_R} \tilde{n}_j^{(s)} = \sum_{s=1}^{L_R} \tilde{n}_{j-s+L_R}^{(L_R)} = \sum_{s=1}^{L_R} \tilde{n}_{j+s-1}^{(L_R)} \tag{E.10}$$

With equations (E.7) to (E.10), the fractional loadings used in equation (E.6) can be represented in terms of the states L_D and L_R, respectively. It then follows

$$q_j = \frac{1 - \sum_{s=1}^{L_D} n_{j+s}^{(L_D)} - \sum_{s=1}^{L_R} \tilde{n}_{j+s}^{(L_R)}}{1 - \sum_{s=1}^{L_D} n_{j+s}^{(L_D)} + n_{j+L_D}^{(L_D)} - \sum_{s=1}^{L_R} \tilde{n}_{j+s}^{(L_R)} + \tilde{n}_{j+L_R}^{(L_R)}} \tag{E.11}$$

or

$$q_j = \frac{1 - \sum_{s=1}^{L_D} n_{j+s}^{(L_D)} - \sum_{s=1}^{L_R} \tilde{n}_{j+s}^{(L_R)}}{1 - \sum_{s=1}^{L_D-1} n_{j+s}^{(L_D)} - \sum_{s=1}^{L_R-1} \tilde{n}_{j+s}^{(L_R)}} \tag{E.12}$$

For arbitrary reference states, m_D (with $1 \le m_D \le L_D$) and m_R (with $1 \le m_R \le L_R$), equation (E.12) reads

$$q_j = \frac{1 - \sum_{s=1}^{L_D} n_{j+s}^{(m_D)} - \sum_{s=1}^{L_R} \tilde{n}_{j+s}^{(m_R)}}{1 - \sum_{s=1}^{L_D-1} n_{j+s}^{(m_D)} - \sum_{s=1}^{L_R-1} \tilde{n}_{j+s}^{(m_R)}} \tag{E.13}$$

Strictly speaking, equations (E.11) and (E.12) are valid only for the particular situation that $L_D = L_R$ and $m_D = m_R$. In this case, q_j is the same for either of the two catalysts. On the other hand, if both catalysts show a divergence in lengths (i.e., when $L_D \neq L_R$), and when they have different reference states (i.e., $m_D \neq m_R$), q_j will differ with respect to the type of catalyst. This is demonstrated subsequently. First, q_j^D, is derived for catalyst D, before this term is elaborated analogously for catalyst R.

For convenience, L_D and L_R are now assumed to fulfill the condition that $L_D < L_R$. It may be further imposed that $m_D = m_R = 1$. These assumptions can be abandoned later in the section. As is displayed in Figure E.3, a movement of catalyst D located in site j to position $j+1$ is impeded by all the catalysts that are bound (with respect to their reference state) throughout the sites $j+1$ to $j+L_D$. All other catalysts, whose reference states are located beyond this interval (i.e., at sites greater than $j+L_D$, or at sites smaller than j), do not affect the movement of D from site j into site $j+1$. In particular, this means that the catalysts R bound to sites L_D+1 to L_R, obviously cause no impact onto the queueing of catalyst D. This may be taken into account for mathematically describing q_j for catalyst D. If additionally the assumption of equal reference states is dropped, so that $m_D \neq m_R$ is permitted, equation (E.13) may thus be modified to yield

$$q_j^D = \frac{1 - \sum_{s=1}^{L_D} n_{j+s}^{(m_D)} - \sum_{s=1}^{L_D} \tilde{n}_{j+s}^{(m_R)}}{1 - \sum_{s=1}^{L_D-1} n_{j+s}^{(m_D)} - \sum_{s=1}^{L_D-1} \tilde{n}_{j+s}^{(m_R)}} \tag{E.14}$$

The superscript D stands for catalyst D. Further on, the superscript indicating the reference state is neglected. Queueing parameters for catalysts D and R located in position j, respectively, can be rewritten in the following form:

$$q_j^D = \frac{1 - \sum_{s=1}^{L_D} n_{j+s} - \sum_{s=1}^{L_D} \tilde{n}_{j+s-m_D+m_R}}{1 - \sum_{s=1}^{L_D-1} n_{j+s} - \sum_{s=1}^{L_D-1} \tilde{n}_{j+s-m_D+m_R}} \tag{E.15}$$

$$q_j^R = \frac{1 - \sum_{s=1}^{L_R} n_{j+s-m_R+m_D} - \sum_{s=1}^{L_R} \tilde{n}_{j+s}}{1 - \sum_{s=1}^{L_R-1} n_{j+s-m_R+m_D} - \sum_{s=1}^{L_R-1} \tilde{n}_{j+s}} \tag{E.16}$$

Equations (E.15) and (E.16) denote the probabilities that site $j+1$ is accessible, when the respective catalyst (D or R) bound to site j moves on to the next site.

Catalyst association. Similarly, the previously derived probability for catalyst association (MacDonald and Gibbs, 1969), now needs to be modified, in order to accomodate for the situation where two different types of catalysts are considered. In this case, the binding site

(j_{D0}) for catalyst D may not coincide with the binding location for R (j_{R0}). For example, it may be assumed that $j_{D0} < j_{R0}$. That is, catalyst D is taken to bind further upstream than R. In this case, the binding of catalyst R would be hampered not only by the catalysts bound to sites j with $j_{R0} \leq j \leq j_{R0} + L_R$, but also by catalysts D bound within $L_D - 1$ sites upstream from j_{R0}. This situation is visualized in Figure E.4. If this additional interaction is taken into consideration, and without fixing *a priori* the positional order of binding, the probabilities for unoccupied binding sites can thus be derived for catalysts D and R, respectively. That is,

$$q_{j_{D0}}^{D0} \;=\; 1 - \sum_{s=1}^{L_D} n_{j_{D0}+s-1} - \sum_{s=1}^{L_D+L_R-1} \tilde{n}_{j_{D0}+s-m_D-L_R+m_R} \tag{E.17}$$

$$q_{j_{R0}}^{R0} \;=\; 1 - \sum_{s=1}^{L_D+L_R-1} n_{j_{R0}+s-m_R-L_D+m_D} - \sum_{s=1}^{L_R} \tilde{n}_{j_{R0}+s-1} \tag{E.18}$$

The probabilities of forward movement given the set of equations (E.15) to (E.18), generally hold for a situation with two different types of catalysts propagating simultaneously along a same template.

Transition to concentrations. A transition of fractions to concentrations can be made, as is explained at the example of a system comprising both degradosomes and ribosomes bound to a mRNA. The fractional loading of base tripletts j (with $1 \leq j \leq J$) is then given by

$$n_j \;=\; \frac{\sum^{i} C_{i,j}^{D}}{C_j^M} \qquad \text{and} \qquad \tilde{n}_j \;=\; \frac{C_j^R}{C_j^M} \tag{E.19}$$

The summation over index i indicates the sum of all concentrations of degradosome species bound to that particular j. At endonucleolytic cleavage sites, this sum is given by the following expression:

$$\sum^{i} C_{i,j}^{D} \;=\; C_j^{D*} + C_j^{D*Frag} + C_j^{D} \tag{E.20}$$

The transition to concentrations now permits to consider time-dependent changes not only for the concentrations of bound degradosomes and ribosomes, but also for the concentration, C_j^M, of each base triplett j.

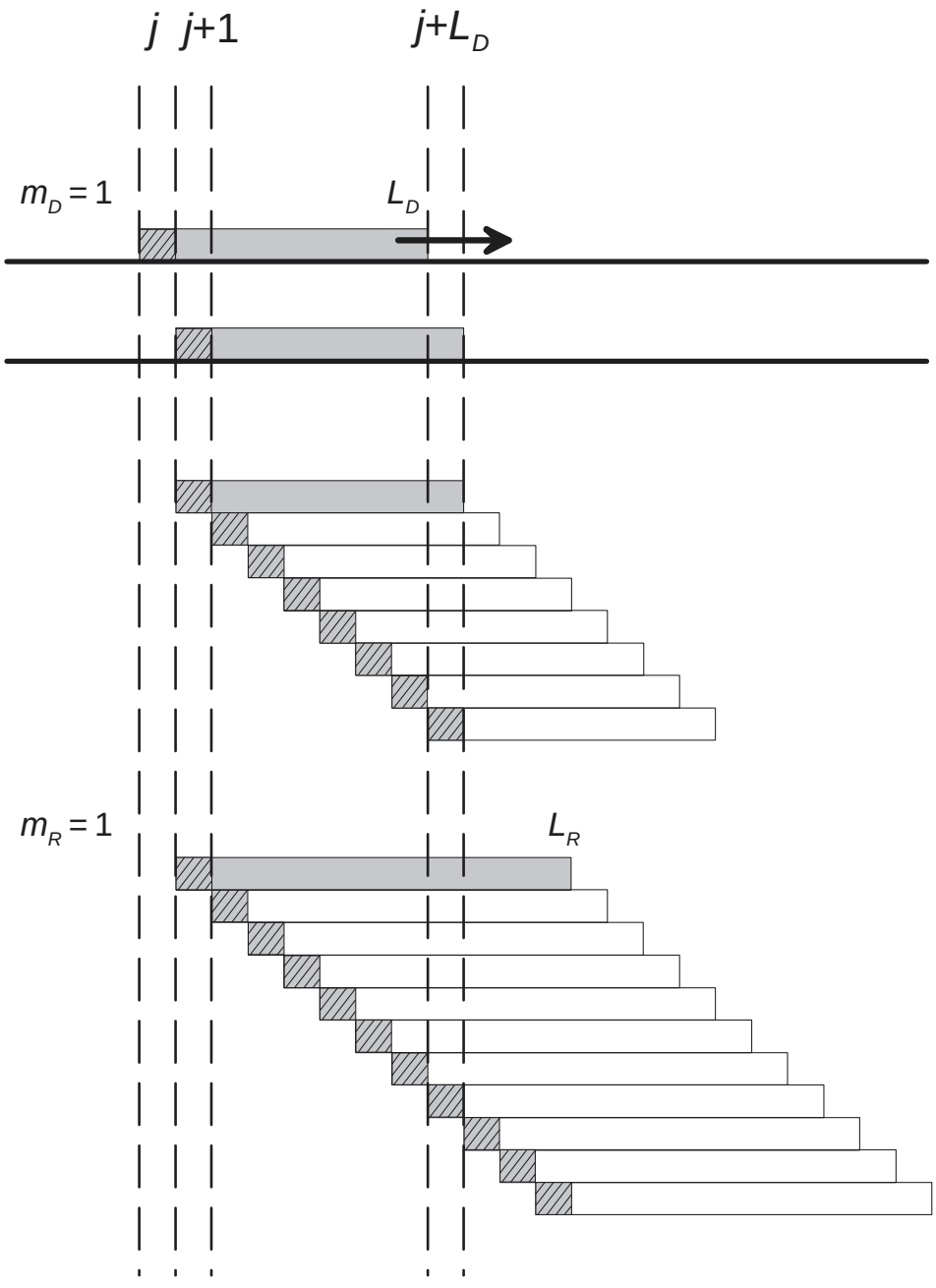

Figure E.3: Derivation of parameter q_j^D. Catalyst D moves from site j to site $j+1$.

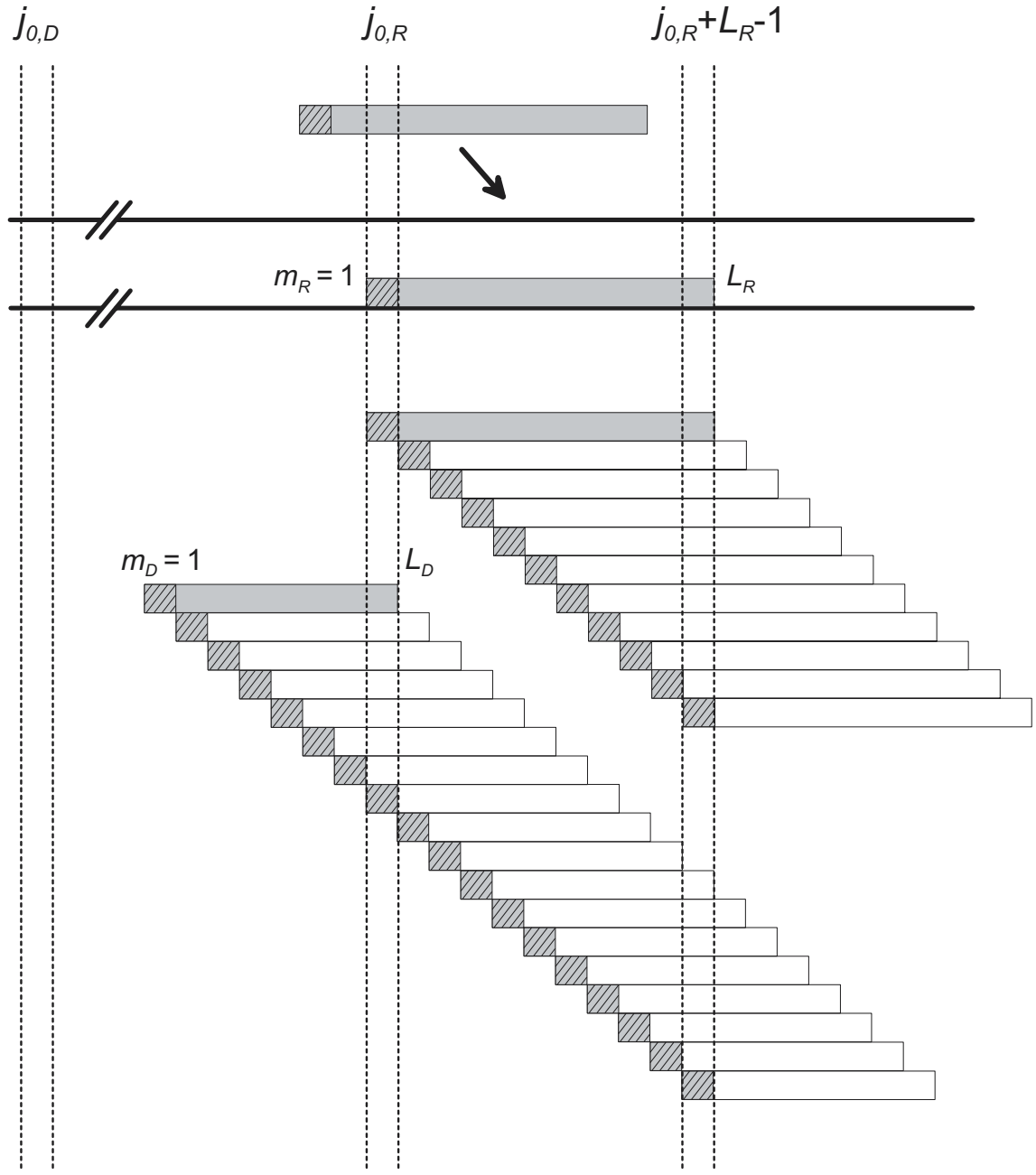

Figure E.4: Derivation of parameter q_{jR0}^{R0}. The binding site of catalyst R is located downstream of the binding site for catalyst D.

The queueing factors for a system combining both $5'$ to $3'$-directed mRNA degradation and translation, and at the same time allowing for arbitrary lengths (L_D and L_R) and reference states (m_D and m_R) of degradosome and ribosome, respectively, can be summarized as to the following:

Degradosomes

For $j_{D0} = m_D$:

$$q_{j_{D0}}^{D0} = 1 - \sum_{s=1}^{L_D} \frac{\sum^{i} C_{i,j_{D0}+s}^{D}}{C_{j_{D0}+s}^{M}} - \sum_{s=1}^{L_D+L_R-1} \frac{C_{j_{D0}+s-m_D-L_R+m_R}^{R}}{C_{j_{D0}+s-m_D-L_R+m_R}^{M}} \tag{E.21}$$

For $1 \leq j \leq J$:

$$q_j^D = \frac{1 - \sum_{s=1}^{L_D} \frac{\sum^{i} C_{i,j+s}^{D}}{C_{j+s}^{M}} - \sum_{s=1}^{L_D} \frac{C_{j+s-m_D+m_R}^{R}}{C_{j+s-m_D+m_R}^{M}}}{1 - \sum_{s=1}^{L_D-1} \frac{\sum^{i} C_{i,j+s}^{D}}{C_{j+s}^{M}} - \sum_{s=1}^{L_D-1} \frac{C_{j+s-m_D+m_R}^{R}}{C_{j+s-m_D+m_R}^{M}}} \tag{E.22}$$

Ribosomes

For $j_{R0} = j_{AUG}$:

$$q_{j_{R0}}^{R0} = 1 - \sum_{s=1}^{L_D+L_R-1} \frac{\sum^{i} C_{i,j_{R0}+s-m_R-L_D+m_D}^{D}}{C_{j_{R0}+s-m_R-L_D+m_D}^{M}} - \sum_{s=1}^{L_R} \frac{C_{j_{R0}+s-1}^{R}}{C_{j_{R0}+s-1}^{M}} \tag{E.23}$$

For $j_{AUG} \leq j \leq K$:

$$q_j^R = \frac{1 - \sum_{s=1}^{L_R} \frac{\sum^{i} C_{i,j+s-m_R+m_D}^{D}}{C_{j+s-m_R+m_D}^{M}} - \sum_{s=1}^{L_R} \frac{C_{j+s}^{R}}{C_{j+s}^{M}}}{1 - \sum_{s=1}^{L_R-1} \frac{\sum^{i} C_{i,j+s-m_R+m_D}^{D}}{C_{j+s-m_R+m_D}^{M}} - \sum_{s=1}^{L_R-1} \frac{C_{j+s}^{R}}{C_{j+s}^{M}}} \tag{E.24}$$

with

$$C_j^D = 0 \quad \text{for} \quad j < j_{D0} \quad \text{and} \quad j > J \tag{E.25}$$

$$C_j^R = 0 \quad \text{for} \quad j < j_{R0} \quad \text{and} \quad j > K \tag{E.26}$$

The index j (with $1 \leq j \leq J$) refers to all base tripletts of a mRNA molecule. J is the total number of (coding and non-coding) base tripletts of a mRNA. Base triplett $j = K$ signifies the last codon upstream of a translational stop codon.

F Auxiliary Calculations

F.1 RNase E concentration

A number of 2,248 RNase E molecules were detected within 124 slices (each 0.1 μm thick) of frozen *E. coli* cells, whose metabolism had been stopped during the exponential growth phase (Liou et al., 2001). That makes 18 such molecules per slice.

With a cellular width of 0.8 μm and a length of 1.6 μm for an average *E. coli* cell, the slice has a volume of $V_{slice} = 0.128$ μm^3 (and 1 μm^3 = 10^{-15} L).

Considering the Avogadro number ($N_A = 6.022 \cdot 10^{23}$ molecules/mol), the total cellular concentration of RNase E can then be calculated according to

$$\left(C_{\text{RNase E}}\right)_t = \frac{2248}{124}\,\frac{1}{V_{slice}}\,\frac{1}{N_A} = 235 \text{ nM} \tag{F.1}$$

Since this concentration is only a rough estimate, a total concentration of RNase E of 250 nM was assumed throughout this study. This value corresponds with an approximate number of 150 molecules RNase E per cell.

F.2 Concentration of *lacZ* mRNA

There are about 31 molecules of *lacZ* mRNA per *E. coli* cell under induced growth conditions (Kennell and Riezman, 1977). With a cellular volume of $V_{cell} = 1.015$ μm^3, the total concentration of *lacZ* mRNA can be estimated from

$$C_{lacZ}^{mRNA} = \frac{31}{V_{cell}}\,\frac{1}{N_A} = 49.9 \text{ nM} \tag{F.2}$$

Thus, the intracellular concentration of *lacZ* mRNA was taken to be 50 nM.

F.3 IF2 concentration

The association constant $K_{30S\cdot IF1}$ for IF2 binding to GTP equals $8.0{\cdot}10^4$ M^{-1} (taken from Pon et al. (1985)). Thus,

$$C_{\text{IF2}\cdot\text{GTP}} = K_{30S\cdot IF1}\, C_{\text{IF2}}\, C_{\text{GTP}} \tag{F.3}$$

Neglecting the presence of other reactants, the total mass balances for IF2 and GTP read

$$C_{\text{IF2},t} = C_{\text{IF2}\cdot\text{GTP}} + C_{\text{IF2}} \tag{F.4}$$

$$C_{\text{GTP},t} = C_{\text{IF2}\cdot\text{GTP}} + C_{\text{GTP}} \tag{F.5}$$

The total concentrations of cellular IF2 and GTP are approximately 10 μM and 1000 μM, respectively (Neidhardt et al., 1996).

Solving the above 3 algebraic equations for the 3 unknowns leads to $C_{\text{IF2}\cdot\text{GTP}} = 9.9$ μM, $C_{\text{IF2}} = 0.1$ μM. The concentration of uncomplexed GTP, C_{GTP}, then equals 990.1 μM. Thus, approximately 99 % of IF2 are complexed with GTP at total concentrations comparable to *in vivo* conditions.

F.4 Testing the continuum condition

The cellular concentration of codon-bound degradosomes, C_j^D, ranks among the smallest concentrations of state variables encountered in the gene expression model. With $C_j^D = 10^{-11}$ M (chapter 5), $V_{cell} = 1.015\ \mu\text{m}^3$, and $N_A = 6.022 \cdot 10^{23}$ molecules/mol, the total cellular number, $n_{j,cell}^D$, of degradosomes bound to a codon j is estimated to be $6.1{\cdot}10^{-3}$, according to

$$n_{j,cell}^D \;=\; C_j^D \; V_{cell} \; N_A \tag{F.6}$$

Thus, under the given premises, the continuum condition is shown to be violated with respect to a single cell. On the other hand, when considering the entirety of biomass under typical submerged cultivation conditions, the law of large numbers applies. This is outlined in the following. Furtheron, a concentration of wet biomass of 30 g per liter reaction volume is assumed. The biomass concentration is given by

$$C_X \;=\; \frac{m_X}{V} \tag{F.7}$$

Symbol V denotes the volume of aqueous phase in the fermentor. m_X is the mass of wet cells. Expressing biomass in terms of units of dry weight, equation (F.7) reads

$$C_X \;=\; \frac{n_{cells}\; m_{cell}}{(1 - f_{aq})\; V} \tag{F.8}$$

with m_{cell} denoting the dry weight per cell. f_{aq}, the fractional water content of a cell, was taken to be 0.7 (Neidhardt and Umbarger, 1996). For $\mu = 1.0\ \text{h}^{-1}$, cell dry weight can be estimated to be $4.33{\cdot}10^{-13}$ g (calculation based on data provided by Bremer and Dennis (1996)). After rearranging equation (F.8), the total number of cells, n_{cells}, is obtained to be equal to $2.1 \cdot 10^{13}$ per liter reaction volume. Using

$$n_j^D \;=\; n_{j,cell}^D \; n_{cells} \tag{F.9}$$

n_j^D is calculated to be $1.3{\cdot}10^{11}$, which means that there are a total of $1.3{\cdot}10^{11}$ degradosomes bound to the assumed codon j within a cell culture.

References

Adamski, F. M., McCaugham, K. K., Jørgensen, F., Kurland, C. G., and Tate, W. P. (1994). The Concentration of Polypeptide Chain Release Factors 1 and 2 at Different Growth Rates of *Escherichia coli*. *J. Mol. Biol.*, 238:302–308.

Aiba, S., Humphrey, A. E., and Millis, N. F. (1973). *Biochemical Engineering*. Academic Press, New York.

Airas, R. K. (1990). On the Roles of Magnesium and Spermidine in the Isoleucyl-tRNA Synthetase Reaction. An Analysis of the Reaction Mechanism by Total Rate Equation. *Eur. J. Biochem.*, 192:401–450.

Airas, R. K. (1992). Analysis of the Isoleucyl-tRNA Synthetase Reaction by Total Rate Equations. Magnesium and Spermidine in the tRNA Kinetics. *Eur. J. Biochem.*, 210:443–450.

Arai, K.-I., Kawakita, M., and Kaziro, Y. (1974). Studies on the Polypeptide Elongation Factors from *E. coli*. V. Properties of Various Complexes Containing EF-Tu and EF-Ts. *J. Biochem. (Tokyo)*, 76:293–306.

Arnold, S. G., Siemann, M., Scharnweber, K., Werner, M., Baumann, S., and Reuss, M. (2001). Kinetic Modelling and Simulation of *in vitro* Transcription by Phage T7 RNA Polymerase. *Biotechnol. Bioeng.*, 72:548–561.

Atkinson, D. E. (1968). The Energy Charge of the Adenylate Pool as a Regulatory Parameter. Interaction with Feedback Modifiers. *Biochemistry*, 7:4030–4034.

Baca, O. G., Rohrbach, M. S., and Bodley, J. W. (1976). Equilibrium Measurements of the Interactions of Guanine Nucleotides with *Escherichia coli* Elongation Factor G and the Ribosome. *Biochemistry*, 15:4570–4575.

Baeyens, K. J., de Bondt, H. L., Pardi, A., and Holbrook, S. R. (1996). A Curved RNA Helix Incorporating an Internal Loop with G·A and A·A Non-Watson-Crick Base Pairing. *Proc. Natl. Acad. Sci. USA*, 93:12851–12855.

Bagnoli, F. and Liò, P. (1995). Selection, Mutations and Codon Usage in a Bacterial Model. *J. theor. Biol.*, 173:271–281.

Bailey, J. E. (1998). Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities. *Biotechnol. Prog.*, 14:8–20.

Basu, S. and Maitra, U. (1986). Specific Binding of Monomeric Bacteriophage T3 and T7 RNA Polymerases to Their Respective Cognate Promoters Requires the Initiating Ribonucleoside Triphosphate (GTP). *J. Mol. Biol.*, 190:425–437.

Belasco, J. and Higgins, C. (1988). Mechanisms of mRNA Decay in Bacteria: A Perspective. *Gene*, 72:15–23.

Bergmann, J. E. and Lodish, H. F. (1979). A Kinetic Model of Protein Synthesis. *J. Biol. Chem.*, 254:11927–11937.

Biblia, T. A. and Flickinger, M. C. (1992). Use of a Structured Kinetic Model of Antibody Synthesis and Secretion for Optimization of Antibody Productin Systems: I. Steady-state Analysis. *Biotechnol. Bioeng.*, 39:251–261.

Bilgin, N. and Ehrenberg, M. (1995). Stoichiometry for the Elongation Factor Tu·Aminoacyl-tRNA Complex Switches with Temperature. *Biochemistry*, 34:715–719.

Blank, A., Gallant, J. A., Burgess, R. R., and Loeb, L. A. (1986). An RNA Polymerase Mutant with Reduced Accuracy of Chain Elongation. *Biochemistry*, 25:5920–5928.

Blanquet, S., Dessen, P., and Kahn, D. (1984). Properties and Specificity of Methionyl-tRNA$_f^{Met}$ Formyltransferase from *Escherichia coli*. *Methods Enzymol.*, 106:141–153.

Blumberg, B. M., Nakamoto, T., and Kézdy, F. J. (1979). Kinetics of Initiation of Bacterial Protein Synthesis. *Proc. Natl. Acad. Sci. USA*, 76:251–255.

Blundell, M., Craig, E., and Kennell, D. (1972). Decay Rates of Different mRNA in *E. coli* and Models of Decay. *Nature New Biol.*, 238:46–49.

Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M., and Yuan, Y. (1998). Predicting Function: From Genes to Genomes and Back. *J. Mol. Biol.*, 283:707–725.

Bouvet, P. and Belasco, J. G. (1992). Control of RNase E-mediated RNA Degradation by 5′-terminal Base Pairing in *E. coli*. *Nature*, 360:488–491.

Brännvall, M. and Kirsebom, L. A. (1999). Manganese Ions Induce Miscleavage in the *Escherichia coli* RNase P RNA-catalyzed Reaction. *J. Mol. Biol.*, 292:53–63.

Bremer, H. and Dennis, P. P. (1996). Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. In Neidhardt, F. C., Curtiss III., R., Ingraham, J. L., Lin, E. C. C., Brooks Low, K., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., Editors, *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Microbiology*, pages 1553–1569. American Society for Microbiology, Washington DC.

Brown, C. M., Stockwell, P. A., Trotman, C. N. A., and Tate, W. P. (1990). The Signal for the Termination of Protein Synthesis in Prokaryotes. *Nucl. Acids Res.*, 18:2079–2086.

Brutlag, D. L., Galper, A. R., and Millis, D. H. (1991). Knowledge-based Simulation of DNA Metabolism: Prediction of Enzyme Action. *Comput. Appl. Biosci.*, 7:9–19.

Cannistraro, V. J. and Kennell, D. (1985). The 5′ Ends of *Escherichia coli lac* mRNA. *J. Mol. Biol.*, 182:241–248.

Cannistraro, V. J., Subbarao, M. N., and Kennell, D. (1986). Specific Endonucleolytic Cleavage Sites for Decay of *Escherichia coli* mRNA. *J. Mol. Biol.*, 192:257–274.

Canonaco, M. A., Calogero, R. A., and Gualerzi, C. O. (1986). Mechanism of Translational Initiation in Prokaryotes. *FEBS Lett.*, 207:198–204.

Carpousis, A. J., Van Houwe, G., Ehretsmann, C., and Krisch, H. M. (1994). Copurification of *E. coli* RNase E and PNPase: Evidence for a Specific Association between Two Enzymes Important in RNA Processing and Degradation. *Cell*, 76:889–900.

Carrier, T. A. and Keasling, J. D. (1997). Mechanistic Modeling of Prokaryotic mRNA Decay. *J. theor. Biol.*, 189:195–209.

Carrier, T. A. and Keasling, J. D. (1999). Library of Synthetic $5'$ Secondary Structures to Manipulate mRNA Stability in *Escherichia coli*. *Biotechnol. Prog.*, 15:58–64.

Cha, S. (1968). A Simple Method for Derivation of Rate Equations for Enzyme-catalyzed Reactions under the Rapid Equilibrium Assumption or Combined Assumptions of Equilibrium and Steady State. *J. Biol. Chem.*, 243:820–825.

Chaires, J. B., Pande, C., and Wishnia, A. (1981). The Effect of Initiation Factor IF-3 on *Escherichia coli* Ribosomal Subunit Association Kinetics. *J. Biol. Chem.*, 256:6600–6607.

Chamberlin, M. and Ring, J. (1973). Characterization of T7-specific Ribonucleic Acid Polymerase. I. General Properties of the Enzymatic Reaction and the Template Specificity of the Enzyme. *J. Biol. Chem.*, 248:2235–2244.

Chaney, W. G. and Morris, A. J. (1979). Nonuniform Size Distribution of Nascent Peptides. The Effect of Messenger RNA Structure upon the Rate of Translation. *Arch. Biochem. Biophys.*, 194:283–291.

Chau, V., Romero, G., and Biltonen, R. L. (1981). Kinetic Studies on the Interactions of *Escherichia coli* K12 Elongation Factor Tu with GDP and Elongation Factor Ts. *J. Biol. Chem.*, 256:5591–5596.

Chela-Flores, J., Liquori, A. M., and Florio, A. (1988). A Kinetic Thermodynamic Phenomenological Approach to Genetic Expression of Heat-shock Proteins. *J. theor. Biol.*, 134:319–325.

Chen, W., Bailey, J. E., and Lee, S. B. (1991). Molecular Design of Expression Systems: Comparison of Different Repressor Control Configurations Using Molecular Mechanism Models. *Biotechnol. Bioeng.*, 38:679–687.

Cheong, H. K., Cheong, C., and Choi, B. S. (1996). Secondary Structure of the Panhandle RNA of Influenza Virus. A Study by NMR Spectroscopy. *Nucl. Acids Res.*, 24:4197–4201.

Chetouani, F., Monestié, P., Thébault, P., Gaspin, C., and Michot, B. (1997). ESSA: An Integrated and Interactive Computer Tool for Analysing RNA Secondary Structure. *Nucl. Acids Res.*, 25:3514–3522.

Christou, I., Felberg, R. A., Demchuk, A. M., Burgin, W. S., Malkoff, M., Grotta, J. C., and Alexandrov, A. V. (2002). Intravenous Tissue Plasminogen Activator and Flow Improvement in Acute Ischemic Stroke Patients with Internal Carotid Artery Occlusion. *J. Neuroimaging*, 12:119–123.

Cleland, W. W. (1963a). The Kinetics of Enzyme-catalyzed Reactions with Two or More Substrates or Products. I. Nomenclature and Rate Equations. *Biochim. Biophys. Acta*, 67:104–137.

Cleland, W. W. (1963b). The Kinetics of Enzyme-catalyzed Reactions with Two or More Substrates or Products. II. Inhibition: Nomenclature and Theory. *Biochim. Biophys. Acta*, 67:173–187.

Cleland, W. W. (1963c). The Kinetics of Enzyme-catalyzed Reactions with Two or more Substrates or Products. III. Prediction of Initial Velocity and Inhibition Patterns by Inspection. *Biochim. Biophys. Acta*, 67:188–196.

Coburn, G. A. and Mackie, G. A. (1999). Degradation of mRNA in *Escherichia coli*: An Old Problem with Some New Twists. *Proc. Nucleic Acid Res. Mol. Biol.*, 62:55–108.

Cornish-Bowden, A. (1977). An Automated Method for Deriving Steady-state Rate Equations. *Biochem. J.*, 165:55–59.

Court, D. (1993). RNA Processing and Degradation by RNase III. In Belasco, J. G. and Brawerman, G., Editors, *Control of Messenger RNA Stability*, pages 71–116. Academic Press Inc., San Diego.

Crothers, D. M., Cole, P. E., Hilbers, C. W., and Shulman, R. G. (1974). The Molecular Mechanism of Thermal Unfolding of *Escherichia coli* Formylmethionine Transfer RNA. *J. Mol. Biol.*, 87:63–88.

Cunningham, P. R. and Ofengand, J. (1990). Use of Inorganic Pyrophosphatase to Improve The Yield of *in vitro* Transcription Reactions Catalyzed by T7 RNA-Polymerase. *Biotechniques*, 9:713–714.

Dahlberg, A. E., Lund, E., and Kjeldgaard, N. O. (1973). Some Effects of Antibiotics on Bacterial Polyribosomes as Studied by Gel Electrophoresis. *J. Mol. Biol.*, 78:627–636.

de Smit, M. H. and van Duin, J. (1990a). Control of Prokaryotic Translational Initiation by mRNA Secondary Structure. *Proc. Nucleic Acid Res. Mol. Biol.*, 38:1–35.

de Smit, M. H. and van Duin, J. (1990b). Secondary Structure of the Ribosome Binding Site Determines Translational Efficiency: A Quantitative Analysis. *Proc. Natl. Acad. Sci. USA*, 87:7668–7672.

de Smit, M. H. and van Duin, J. (1994a). Control of Translation by mRNA Secondary Structure in *Escherichia coli*. A Quantitative Analysis of Literature Data. *J. Mol. Biol.*, 244:144–150.

de Smit, M. H. and van Duin, J. (1994b). Translational Initiation on Structured Messengers. *J. Mol. Biol.*, 235:173–184.

Design Synthesis, Inc. (1994). OptdesX, A Software System for Optimal Engineering Design. Users Manual.

D'haeseleer, R., Liang, S., and Somogyi, R. (2000). Genetic Network Interference: From Co-expression Clustering to Reverse Engineering. *Bioinformatics*, 16:707–725.

Dong, H., Nilsson, I., and Kurland, C. G. (1996). Co-variation of tRNA Abundance and Codon Usage in *Escherichia coli* at Different Growth Rates. *J. Mol. Biol.*, 260:649–663.

Draper, D. E. (1993). Mechanisms of Translational Initiation and Repression in Prokaryotes. In Nierhaus, K. H., Editor, *The Translational Apparatus*, pages 197–207. Plenum Press, NY.

Drew, D. A. (2001). A Mathematical Model for Prokaryotic Protein Synthesis. *Bull. Math. Biol.*, 63:329–351.

Ellis, S. and Conway, T. W. (1984). Initial Velocity Kinetic Analysis of 30 S Initiation Complex Formation in an *in vitro* Translation System Derived from *Escherichia coli*. *J. Biol. Chem.*, 259:7607–7614.

Evans, G. A. (2000). Designer Science and the "Omic" Revolution. *Nature Biotechnology*, 18:127.

Evers, D. and Giegerich, R. (1999). RNA Movies: Visualizing RNA Secondary Structure Spaces. *Comput. Appl. Biosci.*, 15:32–37.

Francis, T. A. and Nagel, G. M. (1976). Glycyl-tRNA Synthetase: Evidence for Two Enzyme Forms and Sigmoidal Saturation Kinetics. *Biochem. Biophys. Res. Commun.*, 70:862–868.

Freist, W., Sternbach, H., and Cramer, F. (1982). Isoleucyl-tRNA Synthetase from *Escherichia coli* MRE 600. *Eur. J. Biochem.*, 128:315–329.

Freistroffer, D. V., Pavlov, M. Y., MacDougall, J., Buckingham, R. H., and Ehrenberg, M. (1997). Release Factor RF3 in *E. coli* Accelerates the Dissociation of Release Factors RF1 and RF2 from the Ribosome in a GTP-dependent Manner. *EMBO J.*, 16:4126–4133.

Furano, A. V. (1975). Content of Elongation Factor Tu in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, 72:4780–4784.

Gast, F.-U. (1987). Mechanistische Untersuchungen zur Fehlerkorrektur bei der ribosomalen Proteinsynthese. Dissertation im Fachbereich Chemie der Universität Hannover.

Gerst, I. and Levine, S. N. (1965). Kinetics of Protein Synthesis by Polyribosomes. *J. theor. Biol.*, 9:16–36.

Godefroy-Colburn, T. and Thach, R. E. (1981). The Role of mRNA Competition in Regulating Translation. IV. Kinetic Model. *J. Biol. Chem.*, 256:11762–11773.

Gold, L., Pribnow, D., Schneider, D., Shinelding, S., Singer, B. S., and Stormo, G. (1981). Translational Initiation in Prokaryotes. *Ann. Rev. Microbiol.*, 35:365–403.

Golomb, M. and Chamberlin, M. (1974). Characterization of T7-specific Ribonucleic Acid Polymerase. *J. Biol. Chem.*, 249:2858–2863.

Good, N. E., Winget, G. D., Winter, W., Connelly, T. N., Izawa, S., and Singh, R. M. M. (1966). Hydrogen Ion Buffers for Biological Research. *Biochemistry*, 5:467–477.

Goodman, R. and Schwartz, I. (1988). Kinetic Analysis of an *E. coli* Phenylalanine-tRNA Synthetase Mutant. *Nucl. Acids Res.*, 16:7477–7486.

Gordon, R. (1969). Polyribosome Dynamics at Steady State. *J. theor. Biol.*, 22:515–532.

Goss, D. J., Parkhurst, L. J., and Wahba, A. J. (1982). Kinetic Studies on the Interaction of Chain Initiation Factor 3 with 70 S *Escherichia coli* Ribosomes and Subunits. *J. Biol. Chem.*, 257:10119–10127.

Götz, P. and Reuss, M. (1997). Dynamics of Microbial Growth: Modelling Time Delays by Introducing a Polymerization Reaction. *J. Biotechnol.*, 58:101–114.

Gouy, M. and Gautier, C. (1982). Codon Usage in Bacteria: Correlation with Gene Expressivity. *Nucl. Acids Res.*, 10:7055–7074.

Gouy, M. and Grantham, R. (1980). Polypeptide Elongation and tRNA Cycling in *Escherichia coli*: A Dynamic Approach. *FEBS Lett.*, 115:151–155.

Guajardo, R., Lopez, P., Dreyfus, M., and Sousa, R. (1998). NTP Concentration Effects on Initial Transcription by T7 RNAP Indicate that Translocation Occurs through Passive Sliding and Reveal that Divergent Promoters Have Distinct NTP Concentration Requirements for Productive Initiation. *J. Mol. Biol.*, 281:777–792.

Guajardo, R. and Sousa, R. (1997). A Model for the Mechanism of Polymerase Translocation. *J. Mol. Biol.*, 265:8–19.

Gualerzi, C. and Pon, C. L. (1990). Initiation of mRNA Translation in Prokaryotes. *Biochemistry*, 29:5881–5889.

Gualerzi, C., Risuleo, G., and Pon, C. L. (1977). Initial Rate Kinetic Analysis of the Mechanism of Initiation Complex Formation and the Role of Initiation Factor IF-3. *Biochemistry*, 16:1684–1689.

Guerra, C. F., Bickelhaupt, F. M., Snijders, J. G., and Baerends, E. J. (2000). Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment. *J. Am. Chem. Soc.*, 122:4117–4128.

Gunderson, S. I., Chapman, K. A., and Burgess, R. R. (1987). Interactions of T7 RNA Polymerase with T7 Late Promoters Measured by Footprinting with Methidiumpropyl-EDTA-Iron(II). *Biochemistry*, 26:1539–1546.

Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999). Correlation between Protein and mRNA Abundance in Yeast. *Mol. Cell. Biol.*, 19:1720–1730.

Hanes, J. and Plükthun, A. (1997). *In vitro* Seclection and Evolution of Functional Proteins by Using Ribosome Display. *Proc. Natl. Acad. Sci. USA*, 94:4937–4942.

Hardt, W.-D., Schlegl, J., Erdmann, V. A., and Hartmann, R. K. (1995). Kinetics and Thermodynamics of the RNase P RNA Cleavage Reaction: Analysis of tRNA 5′-end Variants. *J. Mol. Biol.*, 247:161–172.

Hargrove, J. L. and Schmidt, F. H. (1989). The Role of mRNA and Protein Stability in Gene Expression. *Faseb J.*, 3:2360–2370.

Harley, C. B., Pollard, J. W., Stanners, C. P., and Goldstein, S. (1981). Model for Messenger RNA Translation during Amino Acid Starvation Applied to the Calculation of Protein Synthetic Error Rates. *J. Biol. Chem.*, 256:10786–10794.

Hatzimanikatis, V. and Lee, K. H. (1999). Dynamical Analysis of Gene Networks Requires Both mRNA and Protein Expression Information. *Metabolic Engineering*, 1:275–281.

Heinrich, R. and Rapaport, T. A. (1980). Mathematical Modelling of Translation of mRNA in Eucaryotes: Steady State, Time-dependent Processes and Application to Reticulocytes. *J. theor. Biol.*, 86:279–313.

Hirshfield, I. N. and Yeh, F.-M. (1976). An *in vivo* Effect of the Metabolite L-Leucine on the Properties of the Lysyl-tRNA Synthetase from *Escherichia coli*. *Biochim. Biophys. Acta*, 435:306–314.

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. (1994). Fast Folding and Comparison of RNA Secondary Structures. *Monatshefte für Chemie (Chemical Monthly)*, 125:167–188.

Hofestädt, R. and Thelen, S. (1998). Quantitative Modeling of Biochemical Networks. *In Silico Biol.*, 1:0005.

Howe, J. G. and Hershey, J. W. B. (1983). Initiation Factor and Ribosome Levels Are Coordinately Controlled in *Escherichia coli* Growing at Different Rates. *J. Biol. Chem.*, 258:1954–1959.

Hwang, Y. W. and Miller, D. L. (1985). A Study of the Kinetic Mechanism of Elongation Factor Ts. *J. Biol. Chem.*, 21:11498–11502.

Ikeda, R. A. and Richardson, C. C. (1987). Enzymatic Properties of a Proteolytically Nicked RNA Polymerase of Bacteriophage T7. *J. Biol. Chem.*, 262:3790–3799.

Ikemura, T. (1981a). Correlations between the Abundance of *Escherichia coli* Transfer RNAs and the Occurrence of the Respective Codons in Its Protein Genes: A Proposal for a Synonymous Codon Choice that Is Optimal for the *E. coli* Translation System. *J. Mol. Biol.*, 151:389–409.

Ikemura, T. (1981b). Correlations between the Abundance of *Escherichia coli* Transfer RNAs and the Occurrence of the Respective Codons in the mRNA Genes. *J. Mol. Biol.*, 146:1–21.

Iost, I. and Dreyfus, M. (1995). The Stability of *Escherichia coli lacZ* mRNA Depends upon the Simultaneity of Its Synthesis and Translation. *EMBO J.*, 14:3252–3261.

Ishikawa, H., Maeda, T., and Miyatake, K. (1988). The Computerized Derivation of Rate Equations for Enzyme Reactions on the Basis of the Pseudo-steady State Assumption and the Rapid-equilibrium Assumption. *Biochem. J.*, 251:175–181.

Jaeger, J., Turner, D. H., and Zuker, M. (1989). Improved Predicitons of Secondary Structures for RNA. *Proc. Natl. Acad. Sci. USA*, 86:7706–7710.

Jakubowski, H. (1988). Negative Correlation between the Abundance of *Escherichia coli* Aminoacyl-tRNA Families and Their Affinities for Elongation Factor Tu-GTP. *J. theor. Biol.*, 133:363–370.

Jakubowski, H. and Goldman, E. (1984). Quantities of Individual Aminoacyl-tRNA Families and Their Turnover in *Escherichia coli*. *J. Bacteriol.*, 158:769–776.

Janson, C. A. and Cleland, W. W. (1974). The Inhibition of Acetate, Pyruvate, and 3-Phosphoglycerate Kinases by Chromium Adenosine Triphosphate. *J. Biol. Chem.*, 249:2567–2571.

Jelenc, P. C. and Kurland, C. G. (1979). Nucleoside Triphosphate Regeneration Decreases the Frequency of Translation Errors. *Proc. Natl. Acad. Sci. USA*, 76:3174–3178.

Jermutus, L., Ryabova, L. A., and Plückthun, A. (1998). Recent Advances in Producing and Selecting Functional Proteins by Using Cell-free Translation. *Curr. Opin. Biotechnol.*, 9:534–548.

Jia, Y. and Patel, S. S. (1997). Kinetic Mechanism of GTP Binding and RNA Synthesis during Transcription Initiation by Bacteriophage T7 RNA Polymerase. *J. Biol. Chem.*, 272:30147–30153.

Job, D., Soulié, J. M., and Job, C. (1988). Potential Memory and Hysteretic Effects in Transcription. *J. theor. Biol.*, 134:273–289.

Juan, V. and Wilson, C. (1999). RNA Secondary Structure Prediction Based on Free Energy and Phylogenetic Analysis. *J. Mol. Biol.*, 289:935–947.

Kamimura, R., Konstantinov, K., and Stephanopoulos, G. (1996). Knowledge-based Systems, Artificial Neural Networks and Pattern Recognition: Applications to Biotechnological Processes. *Curr. Opin. Biotechnol.*, 7:231–234.

Käpylä, J., Hyytiä, T., Lahti, R., Goldman, A., Baykov, A. A., and Cooperman, B. S. (1995). Effect of D97E Substitution on the Kinetic and Thermodynamic Properties of *Escherichia coli* Inorganic Pyrophosphatase. *Biochemistry*, 34:792–800.

Karzai, A. W., Susskind, M. M., and Sauer, R. T. (1999). SmpB, a Unique RNA-binding Protein Essential for the Peptide-tagging Activity of SsrA (tmRNA). *EMBO J.*, 18:3793–3799.

Kennell, D. and Riezman, H. (1977). Transcription and Translation Initiation Frequencies of the *Escherichia coli lac* Operon. *J. Mol. Biol.*, 114:1–21.

Kennell, D. and Simmons, C. (1972). Synthesis and Decay of Messenger Ribonucleic Acid from the Lactose Operon of *Escherichia coli* during Amino-acid Starvation. *J. Mol. Biol.*, 70:451–464.

Kennell, D. E. (1990). The Instability of Messenger RNA in Bacteria. In Reznikoff, W. and Gold, L., Editors, *Maximizing Gene Expression*, pages 101–142. Butterworths, Boston.

Kern, D. and Lapointe, J. (1981). The Catalytic Mechanism of Glutamyl-tRNA Synthetase of *Escherichia coli*. *Eur. J. Biochem.*, 115:29–38.

Kern, J. A. and Davis, R. H. (1999). Application of a Fed-Batch System To Produce RNA by *in vitro* Transcription. *Biotechnol. Prog.*, 15:174–184.

Kessling, G., Rawlings, P. K., and Eigen, M. (1976). Determining the Base Sequence of RNA Molecules Using Continuous Degradation Kinetics. *Biophys. Chem.*, 5:369–375.

Kim, D. M., ad C. Y. Choi, T. K., and Yokoyama, S. (1996). A Highly Efficient Cell-free Protein Synthesis System from *Escherichia coli*. *Eur. J. Biochem.*, 239:881–886.

Klostermeier, D. and Millar, D. P. (2001). RNA Conformation and Folding Studied with Fluorescence Resonance Energy Transfer. *Methods*, 23:240–254.

Klug, G. (1993). The Role of mRNA Degradation in the Regulated Expression of Bacterial Photosynthesis Genes. *Mol. Microbiol.*, 9:1–7.

Knapp, S. (1998). Stationäre Netzwerkanalyse eines Energieregenerierungssystems der zellfreien Proteinbiosynthese. Studienarbeit am Institut für Bioverfahrenstechnik, Universität Stuttgart.

Kolchanov, N. A., Ananko, E. A., Podkolodnaya, O. A., Ignatieva, E. V., Stepanenko, I. L., Kel-Margoulis, O. V., Kel, A. E., Merkulova, T. I., Goryachkovskaya, T. N., Busygina, T. V., Kolpakov, F. A., Podkolodny, N. L., Naumochkin, A. N., and Romashchenko, A. G. (1999). Transcription Regulatory Region Database (TRRD): Its Status in 1999. *Nucl. Acids Res.*, 27:303–306.

Krinke, L. and Wulff, D. L. (1990). The Cleavage Specificity of RNase III. *Nucl. Acids Res.*, 18:4809–4815.

Kukko-Kalske, E., Lintunen, M., Inen, M. K., Lahti, R., and Heinonen, J. (1989). Intracellular PPi Concentration Is Not Directly Dependent on Amount of Inorganic Pyrophosphatase in *Escherichia coli* K-12 Cells. *J. Bacteriol.*, 171:4498–4500.

Kuzmine, I. and Martin, C. T. (2001). Pre-steady-state Kinetics of Initiation of Transcription by T7 RNA Polymerase: A New Kinetic Model. *J. Mol. Biol.*, 305:559–566.

Lamla, T., Mammeri, K., and Erdmann, V. A. (2001). The Cell-free Protein Biosynthesis - Applications and Analysis of the System. *Acta Biochim. Pol.*, 48:453–465.

Langer, R. S., Hamilton, B. K., and Colton, C. K. (1977). Enzymatic Regeneration of ATP. II. Equilibrium Studies with Acetate Kinase and Adenylate Kinase. *AIChE J.*, 23:1–10.

Langer, R. S., Hamilton, B. K., Gardner, C. R., and Colton, C. K. (1976). Enzymatic Regeneration of ATP. I. Alternative Routes. *AIChE J.*, 22:1079–1090.

Ledley, T. S. and Ledley, F. D. (1994). Multicompartment, Numerical Model of Cellular Events in the Pharmacokinetics of Gene Therapies. *Hum. Gene Ther.*, 5:579–691.

Lee, P. S. and Lee, K. H. (2000). Genomic analysis. *Curr. Opin. Biotechnol.*, 11:171–175.

Lee, S. B. and Bailey, J. E. (1984a). Analysis of Growth Rate Effects on Productivity of Recombinant *Escherichia coli* Populations Using Molecular Mechanism Models. *Biotechnol. Bioeng.*, 26:66–73.

Lee, S. B. and Bailey, J. E. (1984b). Genetically Structured Models for *lac* Promoter-Operator Function in the *Escherichia coli* Chromosome and in Multicopy Plasmids: *lac* Operator Function. *Biotechnol. Bioeng.*, 26:1372–1382.

Li, K., Kisilevsky, R., Wasan, M. T., and Hammond, G. (1972). A Computer Simulation of *in vivo* Protein Synthesis. *Biochim. Biophys. Acta*, 272:451–462.

Liang, S.-T., Ehrenberg, M., Dennis, P., and Bremer, H. (1999). Decay of *rplN* and *lacZ* mRNA in *Escherichia coli*. *J. Mol. Biol.*, 288:521–538.

Liljenström, H. and Blomberg, C. (1987). Site Dependent Time Optimization of Protein Synthesis with Special Regard to Accuracy. *J. theor. Biol.*, 129:41–56.

Liljenström, H. and von Heijne, G. (1987). Translation Rate Modification by Preferential Codon Usage: Intragenic Position Effects. *J. theor. Biol.*, 124:43–55.

Lim, L. W. and Kennell, D. (1979). Models for Decay of *Escherichia coli lac* Messenger RNA and Evidence for Inactivating Cleavages between Its Messages. *J. Mol. Biol.*, 135:369–390.

Lin, S. X., Shi, J. P., Cheng, X. D., and Wang, Y. L. (1988). Arginyl-tRNA Synthetase from *Escherichia coli*, Purification by Affinity Chromatography, Properties and Steady State Kinetics. *Biochemistry*, 27:6343–6353.

Liou, G.-G., Jane, W.-N., Cohen, S. N., Lin, N.-S., and Lin-Chao, S. (2001). RNA Degradosomes Exist *in vivo* in *Escherichia coli* as Multicomponent Complexes Associated with the Cytoplasmic Membrane Via the N-terminal Region of Ribonuclease E. *Proc. Natl. Acad. Sci. USA*, 98:63–68.

Livadas, C., Lygeros, J., and Lynch, N. A. (1999). High-level Modeling and Analysis of TCAS. In *IEEE Real-Time Systems Symposium*, pages 115–125.

Liveris, D., Klotsky, R. A., and Schwartz, I. (1991). Growth Rate Regulation of Translation Initiation Factor IF3 Biosynthesis in *Escherichia coli*. *J. Bacteriol.*, 173:3888–3893.

Lopez, P. J., Marchand, I., Yarchuk, O., and Dreyfus, M. (1998). Translation Inhibitors Stabilize *Escherichia coli* mRNAs Independently of Ribosome Protection. *Proc. Natl. Acad. Sci. USA*, 95:6067–6072.

Lyakhov, D. L., He, B., Zhang, X., Studier, F. W., Dunn, J. J., and McAllister, W. T. (1997). Mutant Bacteriophage T7 RNA Polymerase with Altered Termination Properties. *J. Mol. Biol.*, 269:28–40.

Lyakhov, D. L., He, B., Zhang, X., Studier, F. W., Dunn, J. J., and McAllister, W. T. (1998). Pausing and Termination by Bacteriophage T7 RNA Polymerase. *J. Mol. Biol.*, 280:201–213.

MacDonald, C. T. and Gibbs, J. H. (1969). Concerning the Kinetics of Polypeptide Synthesis on Polyribosomes. *Biopolymers*, 7:707–725.

MacDonald, C. T., Gibbs, J. H., and Pipkin, A. C. (1968). Kinetics of Biopolymerisation on Nucleic Acid Templates. *Biopolymers*, 6:1–25.

Mahaffy, J. M. (1993). Variation in Concentrations of RNAs and Proteins Involved in Gene Expression of *Escherichia coli*. *J. theor. Biol.*, 162:153–186.

Major, F. and Griffey, R. (2001). Computational Methods for RNA Structure Determination. *Curr. Opin. Struct. Biol.*, 11:282–286.

Makarova, O. V., Makarov, E. M., Sousa, R., and Dreyfus, M. (1995). Transcribing of *Escherichia coli* Genes with Mutant T7 RNA Polymerases: Stability of *lacZ* mRNA Inversely Correlates with Polymerase Speed. *Proc. Natl. Acad. Sci. USA*, 92:12250–12254.

Martin, C. T. and Coleman, J. E. (1987). Kinetic Analysis of T7 RNA Polymerase-promoter Interactions with Small Synthetic Promoters. *Biochemistry*, 26:2690–2696.

Martin, C. T. and Coleman, J. E. (1989). T7 RNA Polymerase Does Not Interact with the 5′-Phosphate of the Initiating Nucleotide. *Biochemistry*, 28:2760–2762.

Martin, C. T., Muller, D. K., and Coleman, J. E. (1988). Processivity in the Early Stages of Transcription by T7 RNA Polymerase. *Biochemistry*, 27:5755–5762.

Maslak, M. and Martin, C. T. (1994). Effects of Solution Conditions on the Steady-state Kinetics of Initiation of Transcription by T7 RNA Polymerase. *Biochemistry*, 33:6918–6924.

Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999). Expanded Sequence Dependence of Thermodynamic Parameters Provides Robust Prediction of RNA Secondary Structure. *J. Mol. Biol.*, 288:911–940.

Mauch, K., Arnold, S., Posten, C., and Reuss, M. (1997). Computer Algebra Systems in Model-building and Model-analysis for Bioprocesses. *15th IMACS World Congress*, 2:171–178.

Mauch, K., Buziol, S., Schmid, J., and Reuss, M. (2002). Computer-aided Design of Metabolic Networks. *AIChE Symp. Series*, in press.

May, O., Habenicht, A., Mattes, R., Syldatk, C., and Siemann, M. (1998). Molecular Evolution of Hydantoinases. *Biol. Chem.*, 379:743–747.

McCormick, J. R., Zengel, J. M., and Lindahl, L. (1991). Intermediates in the Degradation of mRNA from the Lactose Operon of *Escherichia coli*. *Nucl. Acids Res.*, 19:2767–2776.

McDowall, K. J., Lin-Chao, S., and Cohen, S. N. (1994). A+U Content Rather Than a Particular Nucleotide Order Determines the Specificity of RNase E Cleavage. *J. Biol. Chem.*, 269:10790–10796.

Menninger, J. R. (1983). Computer Simulation of Ribosome Editing. *J. Mol. Biol.*, 171:383–399.

Miczak, A., Kaberdin, V. R., Wei, C. L., and Lin-Chao, S. (1996). Proteins Associated with RNase E in a Multicomponent Ribonucleolytic Complex. *Proc. Natl. Acad. Sci. USA*, 93:3865–9.

Mitchell & Gauthier Associates Inc. (1991). Advanced Continuous Simulation Language (ACSL). Reference Manual.

Moritz, B., Striegel, K., de Graaf, A., and Sahm, H. (2000). Kinetic Properties of the Glucose-6-phosphate and 6-Phosphogluconate Dehydrogenases from *Corynebacterium glutamicum* and Their Application for Predicting Pentose Phosphate Pathway Flux *in vivo*. *Eur. J. Biochem.*, 267:3442–3452.

Morse, D. E. and Guertin, M. (1971). Regulation of mRNA Utilization and Degradation by Amino-acid Starvation. *Nature New Biol.*, 232:165–169.

Naaktgeboren, N., Roobol, K., and Voorma, H. O. (1977). The Effect of Initiation Factor IF-1 on the Dissociation of 70-S Ribosomes of *Escherichia coli*. *Eur. J. Biochem.*, 72:49–56.

Nagaswamy, U., Voss, N., Zhang, Z., and Fox, G. E. (2000). Database of Non-canonical Base Pairs Found in Known RNA Structures. *Nucl. Acids Res.*, 28:375–376.

Neidhardt, F. C., Curtiss III., R., Ingraham, J. L., Lin, E. C. C., Brooks Low, K., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E. (1996). *Escherichia coli* and *Salmonella typhimurium*, Cellular and Molecular Microbiology. American Society for Microbiology, Washington DC.

Neidhardt, F. C. and Umbarger, H. E. (1996). Molecular Architecture and Assembly of Cell Parts. In Neidhardt, F. C., Curtiss III., R., Ingraham, J. L., Lin, E. C. C., Brooks Low, K., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., Editors, *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Microbiology*, pages 13–186. American Society for Microbiology, Washington DC.

Neijssel, O. M., Teixera de Mattos, M. J., and Tempest, D. W. (1996). Growth Yield and Energy Distribution. In Neidhardt, F. C., Curtiss III., R., Ingraham, J. L., Lin, E. C. C., Brooks Low, K., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., Editors, *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Microbiology*, pages 1683–1692. American Society for Microbiology, Washington DC.

Newbury, S. F., Smith, N. H., and Higgins, C. F. (1987). Differential mRNA Stability Controls Relative Gene Expression within a Polycistronic Operon. *Cell*, 51:1131–1143.

Nicholson, A. W. (1999). Function, Mechanism and Regulation of Bacterial Ribonucleases. *FEMS Microbiol. Rev.*, 23:371–390.

Nierhaus, K. H. (1996). Die Tricks der ribosomalen Elongationsfaktoren. *Angew. Chem.*, 108:2342–2345.

Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989). A General Method for Site-specific Incorporation of Unnatural Amino Acids into Proteins. *Science*, 244:182–188.

Oestreich, C. H. and Jones, M. M. (1966). The Effect of Metal Ions on Labile Phosphates. 1. The Hydrolysis of Acetyl Phosphate Dianion. *Biochemistry*, 5:2926–2931.

O'Sullivan, W. J. and Perrin, D. D. (1964). The Stability Constants of Metal-Adenine Nucleotide Complexes. *Biochemistry*, 3:18–26.

Ott, G., Schiesswohl, M., Kiesewetter, S., Förster, C., Arnold, L., Erdmann, V. A., and Sprinzl, M. (1990). Ternary Complexes of *Escherichia coli* Aminoacyl-tRNAs with the Elongation Factor Tu and GTP: Thermodynamic and Structural Studies. *Biochim. Biophys. Acta*, 1050:222–225.

Otto, R. (1996). Charakterisierung von prokaryotischen S30-Extrakten und einem ATP-Regenerierungssystem zum Einsatz in der *in vitro* Proteinsynthese. Studienarbeit am Institut für Bioverfahrenstechnik, Universität Stuttgart.

Pan, T., Artsimovitch, I., Fang, X.-W., and Sosnick, T. R. (1999). Folding of a Large Ribozyme during Transcription and the Effect of the Elongation Factor NusA. *Proc. Natl. Acad. Sci. USA*, 96:9545–9550.

Pape, T., Wintermeyer, W., and Rodnina, M. (1999). Induced Fit in Initial Selection and Proofreading of Aminacyl-tRNA on the Ribosome. *EMBO J.*, 18:38003807.

Pape, T., Wintermeyer, W., and Rodnina, M. V. (1998). Complete Kinetic Mechanism of Elongation Factor Tu-dependent Binding of Aminoacyl-tRNA to the A site of the *E. coli* Ribosome. *EMBO J.*, 17:7490–7497.

Patnaik, R. and Swartz, J. (1998). *E. coli*-based *in vitro* Transcription/Translation: *In vivo*-specific Synthesis Rates and High Yields in a Batch System. *Biotechniques*, 24:862–868.

Pavlov, M. Y. and Ehrenberg, M. (1996). Rate of Translation of Natural mRNAs in an Optimized *in vitro* System. *Arch. Biochem. Biophys.*, 328:9–16.

Pavlov, M. Y., Freistroffer, D. V., Dincbas, V., MacDougall, J., Buckingham, R. H., and Ehrenberg, M. (1998). A Direct Estimation of the Context Effect on the Efficiency of Termination. *J. Mol. Biol.*, 284:579–590.

Pavlov, M. Y., Freistroffer, D. V., MacDougall, J., Buckingham, R. H., and Ehrenberg, M. (1997). Fast Recycling of *Escherichia coli* Ribosomes Requires Both Ribosome Recycling Factor (RRF) and Release Factor RF3. *EMBO J.*, 16:4134–4141.

Pedersen, S. (1984). *Escherichia coli* Ribosomes Translate *in vivo* with Variable Rate. *EMBO J.*, 3:2895–2898.

Pedersen, S., Reeh, S., and Friesen, J. D. (1978). Functional mRNA Half-lives in *E. coli*. *Mol. Gen. Genet.*, 166:329–336.

Petersen, C. (1993). Translation and mRNA Stability in Bacteria: A Complex Relationship. In Belasco, J. G. and Brawerman, G., Editors, *Control of Messenger RNA Stability*, pages 117–145. Academic Press Inc., San Diego.

Pingoud, A., Gast, F.-U., Block, W., and Peters, F. (1983). The Elongation Factor Tu from *Escherichia coli*, Aminoacyl-tRNA, and Guanosine Tetraphosphate Form a Ternary Complex Which Is Bound by Programmed Ribosomes. *J. Biol. Chem.*, 258:14200–14205.

Pingoud, A., Gast, F.-U., and Peters, F. (1990). The Influence of the Concentration of Elongation Factors and tRNAs on the Dynamics and Accuracy of Protein Biosynthesis. *Biochim. Biophys. Acta*, 1050:252–258.

Pingoud, A., Urbanke, C., Krauss, G., Peters, F., and Maass, G. (1977). Ternary Complex Formation between Elongation Factor Tu, GTP and Aminoacyl-tRNA: An Equilibrium Study. *Eur. J. Biochem.*, 78:403–409.

Pipkin, A. C. and Gibbs, J. H. (1966). Kinetics of Synthesis and/or Conformational Changes of Biological Macromolecules. *Biopolymers*, 4:3–15.

Pon, C. L. and Gualerzi, C. O. (1984). Mechanism of Protein Biosynthesis in Procaryotic Cells. Effect of Initiation Factor IF1 on the Initial Rate of 30S Initiation Complex Formation. *FEBS Lett.*, 175:203–206.

Pon, C. L., Paci, M., Pawlik, R. T., and Gualerzi, C. O. (1985). Structure-Function Relationship in *Escherichia coli* Initiation Factors. *J. Biol. Chem.*, 260:8918–8924.

Pozhitkov, A. E., Lavrik, I. N., Sergeev, M. M., and Kochetkov, S. N. (1998). Kinetic Analysis of Reaction Catalyzed by the Phage T7 RNA Polymerase. *Mol. Biol.*, 32:78–82.

Pratt, J. M. (1984). Coupled Transcription-translation in Prokaryotic Cell-free Systems. In Hames, B. D. and Higgins, S. J., Editors, *Transcription and Translation. A Practical Approach*, pages 179–209. IRL Press, Oxford.

Purich, D. L. (1983). *Contemporary Enzyme Kinetics and Mechanism.* Academic Press Inc., Orlando.

Py, B., Higgins, C. F., Krisch, H. M., and Carpousis, A. J. (1996). A DEAD-box RNA Helicase in the *Eschericha coli* RNA Degradosome. *Nature*, 381:169–72.

Régnier, P. and Arraiano, C. M. (2000). Degradation of mRNA in Bacteria: Emergence of Ubiquitous Features. *BioEssays*, 22:235–244.

Rauhut, R. and Klug, G. (1999). mRNA Degradation in Bacteria. *FEMS Microbiol. Rev.*, 23:353–370.

Record, Jr., M. T., Lohman, T. M., and de Haseth, P. (1976). Ion Effects on Ligand-Nucleic Acid Interactions. *J. Mol. Biol.*, 107:145–158.

Reich, J. G. and Sel'kov, E. E. (1981). *Energy Metabolism of the Cell: A Theoretical Treatise.* Academic Press Inc., London.

Rhodes, G. and Chamberlin, M. J. (1974). Ribonucleic Acid Chain Elongation by *Escherichia coli* Ribonucleic Acid Polymerase. *J. Biol. Chem.*, 249:6675–6683.

Rigney, D. R. (1979). Note on the Kinetics and Stochastics of Induced Protein Synthesis as Influenced by Various Models for Messenger RNA Degradation. *J. theor. Biol.*, 79:247–257.

Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., and Gold, L. (1992). Translation Initiation in *Escherichia coli*: Sequences within the Ribosome-binding Site. *Mol. Microbiol.*, 6:1219–1229.

Rivas, E. and Eddy, S. R. (1999). A Dynamic Algorithm for RNA Structure Prediction Including Pseudoknots. *J. Mol. Biol.*, 285:2053–2068.

Rodnina, M. V., Pape, T., Fricke, R., Kuhn, L., and Wintermeyer, W. (1996). Initial Binding of the Elongation Factor Tu·GTP·Aminoacyl-tRNA Complex Preceding Codon Recognition on the Ribosome. *J. Biol. Chem.*, 271:646–652.

Rodnina, M. V. and Wintermeyer, W. (1995). GTP Consumption of Elongation Factor Tu during Translation of Heteropolymeric mRNAs. *Proc. Natl. Acad. Sci. USA*, 92:1945–1949.

Rohrbach, M. S. and Bodley, J. W. (1976). Steady State Kinetic Analysis of the Mechanism of Guanosine Triphosphate Hydrolysis Catalyzed by *Escherichia coli* Elongation Factor G and the Ribosome. *Biochemistry*, 15:4565–4569.

Romero, G., Chau, V., and Biltonen, R. I. (1985). Kinetics and Thermodynamics of the Interaction of Elongation Factor Tu with Elongation Factor Ts, Guanine Nucleotides, and Aminoacyl-tRNA. *J. Biol. Chem.*, 260:6167–6174.

Ron, E. Z., Falk, A., Helberg, D., Horowitz, S., and Zeevi, M. (1978). Preferential Charging of tRNA-Met-f in *Escherichia coli* K12. *Eur. J. Biochem.*, 92:389–395.

Rose, T., Brune, M., Wittinghofer, A., Le Blay, K., Surewicz, W. K., Mantsch, H. H., Barzu, O., and Gilles, A. M. (1991). Structural and Catalytic Properties of a Deletion Derivative (Delta 133-157) of *Escherichia coli* Adenylate Kinase. *J. Biol. Chem.*, 266:10781–10786.

Ruusala, T., Ehrenberg, M., and Kurland, C. G. (1982). Catalytic Effects of Elongation Factor Ts on Polypeptide Synthesis. *EMBO J.*, 1:75–78.

Saenz-Badillos, J., Amin, S. P., and Granstein, R. D. (2001). RNA as a Tumor Vaccine: A Review of the Literature. *Exp. Dermatol.*, 10:143–154.

Saifullin, S. R. and Potapov, A. P. (1995a). Analysis of Stationary Kinetics of Translation Elongation within the Framework Stereospecific Stabilization Hypothesis of Codon-Anticodon Complexes in a Ribosome. I. Kinetic Schemes of Factorless Elongation. *Mol. Biol. (Mosk)*, 29:421–433.

Saifullin, S. R. and Potapov, A. P. (1995b). Analysis of Stationary Kinetics of Translation Elongation within the Framework Stereospecific Stabilization Hypothesis of Codon-Anticodon Complexes in a Ribosome. II. Kinetic Schemes in the Presence of Protein Elongation Factors and GTP. *Mol. Biol. (Mosk)*, 29:434–445.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2nd Edition.

Schilling, C. H., Edwards, J. S., and Palsson, B. O. (1999). Toward Metabolic Phenomics: Analysis of Genomic Data Using Flux Balances. *Biotechnol. Prog.*, 15:288–295.

Schindler, P. (2000). *Proteomics und molekularbiologische Analyse der zellfreien Proteinbiosynthese*. PhD thesis, Universität Stuttgart, Stuttgart, Germany.

Schirmer, F. and Hillen, W. (1998). The *Acinetobacter calcoaceticus* NCIB8250 mop Operon mRNA Is Differentially Degraded, Resulting in a Higher Level of the $3'$ CatA-encoding Segment Than of the $5'$-phenolhydroxylase-encoding Portion. *Mol. Gen. Genet.*, 257:330–337.

Schmid, J. W. (1999). Reaktionskinetische Modellierung der prokaryotischen *in vitro* Translation. Studienarbeit am Institut für Bioverfahrenstechnik, Universität Stuttgart.

Schneider, E., Blundell, M., and Kennell, D. (1978). Translation and mRNA Decay. *Mol. Gen. Genet.*, 160:121–129.

Schulman, L. H. (1991). Recognition of tRNAs by Aminoacyl-tRNA Synthetases. *Proc. Nucleic Acid Res. Mol. Biol.*, 41:23–87.

Schulman, L. H. and Pelka, H. (1988). Anticodon Switching Changes the Identity of Methionine and Valine Transfer RNAs. *Science*, 242:765–768.

Schultzaberger, R. K., Bucheimer, R. E., Rudd, K. E., and Schneider, T. D. (2001). Anatomy of *Escherichia coli* Ribosome Binding Sites. *J. Mol. Biol.*, 313:215–228.

Schulz, V. P. and Reznikoff, W. S. (1990). *In vitro* Secondary Structure Analysis of *mRNA* from *lacZ* Translation Initiation Mutants. *J. Mol. Biol.*, 211:427–445.

Schwarz, A. (1999). Reaktionskinetische Modellierung der tRNA-Beladung und des mRNA-Abbaus in prokaryotischen Zellextrakten. Diplomarbeit am Institut für Bioverfahrenstechnik, Universität Stuttgart.

Segel, I. H. (1993). *Enzyme Kinetics*. Wiley & Sons, New York.

Sen, R. and Dasgupta, D. (1993). Interaction of Ribonucleotides with T7 RNA Polymerase: Probable Role of GTP in Transcription Initiation. *Biochem. Biophys. Res. Commun.*, 195:616–622.

Shen, L., X., Basilion, J. P., and Stanton, Jr., V. P. (1999). Single-nucleotide Polymorphisms Can Cause Different Structural Folds of mRNA. *Proc. Natl. Acad. Sci. USA*, 96:7871–7876.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001). Cell-free Translation Reconstituted with Purified Components. *Nature Biotechnology*, 19:751–755.

Silberberg, A. and Simha, R. (1968). Kinetics of Reversible Reactions on Linear Lattices with Neighbor Effects. *Biopolymers*, 6:479–490.

Simha, R., Zimmerman, J. M., and Moacanin, J. (1963). Polymerization Kinetics of Biological Macromolecules on Templates. *J. Chem. Phys.*, 39:1239–1246.

Singh, U. N. (1969). Polyribosomes and Unstable Messenger RNA: A Stochastic Model of Protein Synthesis. *J. theor. Biol.*, 25:444–460.

Singh, U. N. (1996). Polyribosome Dynamics: Size-distribution as a Function of Attachment, Translocation and Release of Ribosomes. *J. theor. Biol.*, 179:147–159.

Smith, R. M. and Alberty, R. A. (1956). The Apparent Stability Constants of Ionic Complexes of Various Adenosine Phosphates with Divalent Cations. *J. Am. Chem. Soc.*, 78:2376–2380.

Solomovici, J., Lesnik, T., and Reiss, C. (1997). Does *Escherichia coli* Optimize the Economics of the Translation Process? *J. theor. Biol.*, 185:511–521.

Sørensen, M. A. and Pedersen, S. (1991). Absolute *in vivo* Translation Rates of Individual Codons in *Escherichia coli*. The Two Glutamic Acid Codons GAA and GAG Are Translated with a Threefold Difference in Rate. *J. Mol. Biol.*, 222:265–280.

Soultanas, P. and Wigley, D. B. (2001). Unwinding the 'Gordian Knot' of Helicase Action. *Trends Biochem. Sci.*, 26:47–54.

Sousa, R. (1996). Structural and Mechanistic Relationships between Nucleic Acid Polymerases. *Trends Biochem. Sci.*, 21:186–190.

Sousa, R., Patra, D., and Lafer, E. M. (1992). Model for the Mechanism of Bacteriophage T7 RNAP Transcription Initiation and Termination. *J. Mol. Biol.*, 224:319–334.

Spirin, A. S. (1986). *Ribosome Structure and Protein Synthesis*. Benjamin/Cummings, Menlo Park.

Spirin, A. S. and Lishnevskaya, E. B. (1971). Effect of Nonionic Agents on the Stability of Association of Ribosomal Subparticles. *FEBS Lett.*, 14:114–116.

Stahl, S. J. and Zinn, K. (1981). Nucleotide Sequence of the Cloned Gene for Bacteriophage T7 RNA Polymerase. *J. Mol. Biol.*, 148:481–485.

Stephanopoulos, G. N., Aristidou, A. A., and Nielsen, J. (1998). *Metabolic Engineering, Principles and Methodologies*. Academic Press Inc., San Diego.

Stiege, W. and Erdmann, V. A. (1995). The Potentials of the *in vitro* Protein Biosynthesis System. *J. Biotechnol.*, 41:81–90.

Storer, A. C. and Cornish-Bowden, A. (1976). Concentration of $MgATP^{2-}$ and Other Ions in Solution. Calculation of the True Concentrations of Species Present in Mixtures of Associating Ions. *Biochem. J.*, 159:1–5.

Stormo, G. D. (1990). Translation Initiation. In Reznikoff, W. and Gold, L., Editors, *Maximizing Gene Expression*, pages 195–224. Butterworths, Boston.

Straathof, A. J. J. and Heijnen, J. J. (1996). New Constraints Between Kinetic Parameters Explain the (Un)identifiability of Enzymatic Rate Constants. *Biotechnol. Bioeng.*, 52:433–437.

Straathof, A. J. J. and Heijnen, J. J. (1997). Derivation of Enzymatic Rate Equations Using Symbolic Software. *Biocatalysis and Biotransformation*, 14:1–9.

Subbarao, M. N. and Kennell, D. (1988). Evidence for Endonucleolytic Cleavages in Decay of *lacZ* and *lacI* mRNAs. *J. Bacteriol.*, 170:2860–2865.

Talkad, V., Schneider, E., and Kennell, D. (1976). Evidence for Variable Rates of Ribosome Movement in *Escherichia coli*. *J. Mol. Biol.*, 104:299–303.

Thiele, C. (1999). Untersuchungen zur Dimensionierung eines zellfreien Energieregenerierungssystems unter Berücksichtigung metaboler Stoffflußanalysen. Diplomarbeit am Institut für Bioverfahrenstechnik, Universität Stuttgart.

Tomšic, J., Vitali, L. A., Daviter, T., Savelsbergh, A., Spurio, R., Striebeck, P., Wintermeyer, W., Rodnina, M., and Gualerzi, C. O. (2000). Late Events of Translation Initiation in Bacteria: A Kinetic Analysis. *EMBO J.*, 19:2127–2136.

Treiber, D. K. and Williamson, J. R. (1999). Exposing the Kinetic Traps in RNA Folding. *Curr. Opin. Struct. Biol.*, 9:339–345.

Treiber, D. K. and Williamson, J. R. (2001). Beyond Kinetic Traps in RNA Folding. *Curr. Opin. Struct. Biol.*, 11:309–314.

Ujvári, A. and Martin, C. T. (1996). Thermodynamic and Kinetic Measurements of Promoter Binding by T7 RNA Polymerase. *Biochemistry*, 35:14574–14582.

van Meerten, D., Girard, G., and van Duin, J. (2001). Translational Control by Delayed RNA Folding: Identification of the Kinetic Trap. *RNA*, 7:483–494.

Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984). Translation Is a Non-uniform Process. Effect of tRNA Availability on the Rate of Nascent Polypeptide Chains. *J. Mol. Biol.*, 180:549–576.

Vassart, G., Dumont, J. E., and Cantraine, F. R. L. (1971). Translational Control of Protein Synthesis: A Simulation Study. *Biochim. Biophys. Acta*, 247:471–485.

Visser, D., Schmid, J. W., Mauch, K., Reuss, M., and Heijnen, J. J. (2002). Optimal Re-design of Primary Metabolism in *Escherichia coli* Using Linlog Kinetics. Submitted to *Metabolic Engineering*.

Voet, D. and Voet, J. G. (1994). *Biochemie*. VCH Verlags-GmbH, Weinheim, Germany.

von Heijne, G., Nilsson, L., and Blomberg, C. (1977). Translation and Messenger RNA Secondary Structure. *J. theor. Biol.*, 68:321–329.

von Heijne, G., Nilsson, L., and Blomberg, C. (1978). Models for mRNA Translation: Theory versus Experiment. *Eur. J. Biochem.*, 92:397–402.

von Hippel, P. H. and Yager, T. D. (1991). Transcript Elongation and Termination Are Competitive Kinetic Processes. *Proc. Natl. Acad. Sci. USA*, 88:2307–2311.

Wagner, L. A., Gesteland, R. F., Dayhuff, T. J., and Weiss, R. B. (1994). An Efficient Shine-Dalgarno Sequence but Not Translation Is Necessary for *lacZ* mRNA Stability in *Escherichia coli*. *J. Bacteriol.*, 176:1683–1688.

Weickert, M. J., Doherty, D. H., Best, E. A., and Olins, P. O. (1996). Optimization of Heterologous Protein Production in *Escherichia coli*. *Curr. Opin. Biotechnol.*, 7:494–499.

Weiel, J. and Hershey, J. W. B. (1981). Fluorescence Polarization Studies of the Interaction of *Escherichia coli* Protein Synthesis Initiation Factor 3 with 30S Ribosomal Subunits. *Biochemistry*, 20:5859–5865.

Weiel, J. and Hershey, J. W. B. (1982). The Binding of Fluorescein-labeled Protein Synthesis Initiation Factor 2 to *Escherichia coli* 30S Ribosomal Subunits Determined by Fluorescence Polarization. *J. Biol. Chem.*, 257:1215–1220.

Werner, M. (1997). Reaktionskineitsche Analysen zur *in vitro* Transkription. Diplomarbeit am Institut für Bioverfahrenstechnik, Universität Stuttgart.

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Prüß, M., Reuter, I., and Schacherer, F. (2000). TRANSFAC: An Integrated System for Gene Expression Regulation. *Nucl. Acids Res.*, 28:316–319.

Wintermeyer, W. and Gualerzi, C. (1983). Effect of *Escherichia coli* Initiation Factors on the Kinetics of *N*-AcPhe-tRNA^{Phe} Binding to 30S Ribosomal Subunits. A Fluorescence Stopped-flow Study. *Biochemistry*, 22:690–694.

Wolin, S. L. and Walter, P. (1988). Ribosome Pausing and Stacking during Translation of a Eukaryotic mRNA. *EMBO J.*, 7:3559–3569.

Wower, I. K., Zwieb, C. W., Guven, S. A., and Wower, J. (2000). Binding and Cross-linking of tmRNA to Ribosomal Protein S1, on and off the *Escherichia coli* Ribosome. *EMBO J.*, 19:6612–6621.

Yan, W., Augustine, J., and Francklyn, C. (1996). A tRNA Identity Switch Mediated by the Binding Interaction between a tRNA Anticodon and the Accessory Domain of a Class II Aminoacyl-tRNA Synthetase. *Biochemistry*, 35:6559–6568.

Yarchuk, O., Iost, I., and Dreyfus, M. (1991). The Relation between Translation and mRNA Degradation in the *lacZ* Gene. *Biochimie*, 73:1533–1541.

Yarchuk, O., Jacques, N., Guillerez, J., and Dreyfus, M. (1992). Interdependence of Translation, Transcription and mRNA Degradation in the *lacZ* Gene. *J. Mol. Biol.*, 226:581–596.

Young, J. S., Ramirez, W. F., and Davis, R. H. (1997). Modeling and Optimization of a Batch Process for *in vitro* RNA Production. *Biotechnol. Bioeng.*, 56:210–220.

Zhang, S., Goldman, E., and Zubay, G. (1994). Clustering of Low Usage Codons and Ribosome Movement. *J. theor. Biol.*, 170:339–354.

Zimmerman, J. M. and Simha, R. (1965). The Kinetics of Multicenter Macromolecule Growth along a Template. *J. theor. Biol.*, 9:156–185.

Zubay, G. (1973). *In vitro* Synthesis of Protein in Microbial System. *Annu. Rev. Genet.*, 7:267–287.

Zucker, F. H. and Hershey, J. W. B. (1986). Binding of *Escherichia coli* Protein Synthesis Initiation Factor IF1 to 30S Ribosomal Subunits Measured by Flurescence Polarization. *Biochemistry*, 25:3682–3690.

Zuker, M. (2000). Calculating Nucleic Acid Secondary Structure. *Curr. Opin. Struct. Biol.*, 10:303–310.

Zuker, M., Mathews, D. H., and Turner, D. H. (1999). Algorithms and Thermodynamics for RNA Secondary Structer Prediction: A Practical Guide. In Barciszewski, J. and Clark, B. F. C., Editors, *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers, NATO ASI Series.