Dynamics and Regulation of the Central Carbon Metabolism in *Escherichia coli* During Fed-Batch Cultivations

Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Ingenieurwissenschaften (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von

Timo Hardiman

aus Backnang

Hauptberichter:	Prof. Dr.-Ing. Dr. h. c. Matthias Reuss
Mitberichter:	Prof. Dipl.-Ing. Dr. techn. Elmar Heinzle

Tag der mündlichen Prüfung: 15. Dezember 2009

Institut für Bioverfahrenstechnik der Universität Stuttgart

2009

Für meine Familie

"I never satisfy myself until I can make
a mechanical model of a thing.
If I can make a mechanical model
I can understand it."

Lord Kelvin (1824–1907)

Acknowledgements

First, I want to express my deep gratitude to Prof. Matthias Reuss for giving me the opportunity to perform my PhD thesis on this rewarding topic, for his strong dedication to the project and the academic freedom. I am grateful to Prof. Elmar Heinzle for his interest in this work and for agreeing to referee this thesis. I owe Martin Siemann Herzberg a great debt of gratitude for leading the projects I was involved in, for his professional advice and the motivating support. Many thanks to Prof. Biener for sending excellent students of the HS-ES.

Karin Lemuth deserves special thanks for a great, successful collaboration and for the nice work climate. I want to thank Christina Fritz very much for introducing me into the experimental issues concerning the fed-batch cultivations and for having a really good time. Cordial thanks to all my colleagues at the IBVT: In particular, Dirk Müller, Joachim Schmid, Klaus Maier, Angel Sevilla Camins, Manuel Dietrich, Christoph Hold, Stefan Junne and Corinna Kempter for the great years in the office; Joachim Bucher, Prem Murugan, Wouter Berendsen, Jochen Schaub, Markus Samorski, Alexej Lapin, Michael Klann, Holger Perfahl, Jacek Puchalka, Peter Götz, Florian Kirchner for the cheerful working atmosphere and vital discussions. I am happy to acknowledge the substantial contributions to this work by numerous students as part of their theses or other scientific works, namely Heiko Schwarz, Markus Keller, Volker Windeisen, Jennifer Ewald, Ralph Eckhardt, Christian Reis, Barabara Hörmann, Daniel Schulz, Christophe Martineau, Christian Staub, Aziz Sami, Said Taouli, Simone Tönnessen, Thomas Horn, Tom Schuhmacher, Johannes Hofmann, Daniel Hoffmann and Hannes Meinhold. Petra Schlack, Susanne Zibek, Andreas Freund, Achim Hauck, Manuela Beutner, Thomas Reichert, Daniel Pfeiffer, Oliver Vielhauer, Alexander Müller, Anja Niebel, Lara Bogner, Christina Krämer, Gaby Vacun deserve big thanks for their excellent technical support. I would also like to thank Klaus Mauch, Henning Schmidt, Nikolaus Hansen and Richard Münch and their colleagues for providing the software tools used in this work. I am also indebted to Jochen Rebell, Peter Fischer and Gabriele Beck-Schwadorf for their excellent analytical services. Many thanks to the state of Baden-Württemberg for funding parts of this work in the network research project “Development of systems biology methods and tools for the analysis of complex cellular networks“.

Above all, I am happy to express my deeply-felt thanks to my family for their continuous support, most of all to Michaela for her everlasting tolerance and happiness. This thesis would not have come into being without you. Ingo deserves many thanks for always being ready to help.

Stuttgart, June 2009 Timo Hardiman

Contents

Nomenclature

Symbols

Intracellular Compounds

c_j	intracellular concentration of the compound j [mol·($l_{cytosol}$)$^{-1}$]
$c_{j,t}$	total intracellular concentration of the compound j [mol·($l_{cytosol}$)$^{-1}$]
N_j	number of molecules of compound j [-]
w_j	mass fraction of the compound j [g (g dry weight)$^{-1}$]
X_j	specific intracellular concentration of the compound j [mol·(g dry weight)$^{-1}$]
ρ_j	intracellular mass concentration of the compound j [g·($l_{cytosol}$)$^{-1}$]

Extracellular Compounds

c_{glc}	glucose concentration [g ($l_{reactor}$)$^{-1}$]
c_{ac}	glucose concentration [g ($l_{reactor}$)$^{-1}$]
$X_{cAMP,extra}$	specific extracellular cAMP concentration [mol (g dry weight)$^{-1}$]

Further Symbols

a_j, b_j	parameters for linear approximation functions
C_j^i	control coefficient for species i in response to changes in j [-]
F	feed rate [l h^{-1}]
k_j	individual rate constant concerning the compound j [s^{-1}]

K_k	equilibrium constant of reaction k , [M^{-1}] or [M^{-4}]
m_S	maintenance energy coefficient [h^{-1}]
MW_j	molar weight of compound j [g mol^{-1}]
N_A	Avogadro constant
q_{O_2}	specific O_2 uptake rate [mmol g^{-1} h^{-1}]
q_{CO_2}	specific CO_2 production rate [mmol g^{-1} h^{-1}]
q_S	specific glucose uptake rate [g g^{-1} h^{-1}]
r_i	rate of the intracellular reaction i [mol·($l_{cytosol}$·s)$^{-1}$]
r_m	specific measured uptake or export rates [mmol g^{-1} h^{-1}]
$r_{cAMP,export}$	specific cAMP export rate [µmol g^{-1} h^{-1}]
$score_k$	specificity score for the nucleotide sequence k [-]
SpM	specificity matrix with the entries $a_{m,n}$ [-]
V_{cell}	cell volume [$l_{cytosol}$]
V_R	bioreactor volume [l]
$V_{R,0}$	batch volume [l]
ν_X	specific cell volume [$l_{cytosol}$·(g dry weight)$^{-1}$]
$Y_{CO_2/X}$	rate of CO_2 production relative to biomass production [mol C (mol C)$^{-1}$]
$Y_{X/S}$	biomass yield [g biomass (g glucose)$^{-1}$] = [g g^{-1}]

Greek Symbols

α_k	spacer penalty k [-]
δ_k	enhancement factor k [-]
η_j	efficiency of transcription from promoter j [-]
μ	specific growth rate [h^{-1}]
ρ_j	(see “Intracellular Compounds”)
Φ_k	probability of transcription with respect to binding of a protein to the DNA-binding site k [-]
φ_k	probability of binding to the DNA-binding site k [-]
σ^S	(see “Abbreviations”)
ω	fractional change of the specific cell volume [h^{-1}]

Subscripts

glc	glucose
NH_4	ammonia
ac	acetate
X	biomass
O_2	oxygen
CO_2	carbon dioxide
DNAbs	DNA-binding site

Superscripts

+	reactor feed

Abbreviations

AEC	adenylate energy charge, $AEC = \left(c_{ATP} + 0.5 \cdot c_{ADP}\right) / \left(c_{ATP} + c_{ADP} + c_{AMP}\right)$
AXP pool	$c_{AXP} = \left(c_{ATP} + c_{ADP} + c_{AMP}\right)$
cAMP	cyclic 3',5'-AMP
Cra	Catabolite repression/activation protein (transcriptional dual regulator)
Crp	Catabolite repression protein (transcriptional dual regulator)
EMP	glycolysis
fbp	fructose 1,6-bis(phosphate)
GS	glyoxylate shunt
mRNA	messenger RNA
nt	Nucleotide
pep	phosphoenolpyruvate
pep-GS	phosphoenolpyruvate-glyoxylate cycle
PTS	phosphoenolpyruvate:carbohydrate phosphotransferase system
ppGpp	guanosine 3',5'-bis(diphosphate)
PPP	pentose phosphate pathway
RNAP	RNA polymerase
rRNA	ribosomal RNA
σ^S	sigma S factor (subunit of the RNA polymerase encoded by the *rpoS* gene)
TCA	tricarboxylic acid cycle

Zusammenfassung

In der vorliegenden Arbeit wurde ein systembiologischer Ansatz verfolgt, um die Dynamik des zentralen Kohlenstoffmetabolismus und dessen Regulation in *Escherichia coli* in industrielle wichtigen Fed-Batch Prozessen zu analysieren. Die zwei zentralen Ziele waren (i) die Netzwerkstruktur der globalen Regulation des Zentralstoffwechsels basierend auf experimentellen und mathematischen Methoden zu rekonstruieren und (ii) eine Methode zur dynamischen Modellierung der Transkription der Zentralstoffwechsel-Gene zu entwickeln.

Eine Reihe von Glukose-limitierten Fed-Batch Prozessen mit konstanter Zulaufrate wurde durchgeführt, womit gleiche Bedingungen für alle Experimente geschaffen wurden. Diese Prozessstrategie führt zu einer kontinuierlichen Abnahme der Verfügbarkeit der Glucose. Komplementäre metabolische Fluss- und globale Transkriptionsanalysen zeigten, dass die Stoffflüsse und die meisten Transkript-Level der Glykolyse, des Pentosephosphat-Weges und der Biosynthesen stark abfallen. Die Flüsse im Citrat-Zyklus bleiben konstant und die mRNA Level der Citrat- und Glyoxylat-Zyklus Gene sind erhöht. Weiterhin wurde die Signalbildungsdynamik durch Quantifizierung der intrazellulären Konzentrationen der Alarmone ppGpp und cAMP untersucht. Beim Einsetzen der Kohlenstofflimitation akkumulierten beide Alarmone stark. Ein neuer Befund war die anschließende Rückstellung beider Signale.

Eine Netzwerkstruktur der Regulation des zentralen Kohlenstoffmetabolismus wurde rekonstruiert, um damit die beobachtete Dynamik umfassend beschreiben zu können. Demgemäß bestimmen vorwiegend die *cra* und *crp* Modulons die Transkription der Glykolyse, Citrat- und Glyoxylat-Zyklus Gene. Das *relA*/*spoT* Modulon reguliert hauptsächlich die Proteinbiosynthese und die spezifische Wachstumsrate. Die zentralen, an der Signalbildung und –Rückstellung beteiligten, Komponenten wurden in diese Modellstruktur integriert. Weitere, gut untersuchte regulatorische Phänomene bezüglich der Kohlenhydrat-Transportsysteme und der Chemotaxis wurden beobachtet, während die Stress- und Stationärphasen-Regulation von untergeordneter Rolle waren. Die möglichen Interaktionen dieser Zellfunktionen mit der Versorgung mit Biosynthese-Vorstufen (precursor) und Energie werden diskutiert.

Die rRNA- und die Gesamt-RNA-Gehalte wurden durch eine neu entwickelte Methode quantifiziert. Eine starke Wachstumsraten-abhängige Regulation der rRNA und mRNA Transkription wurde aus diesen Daten abgeleitet. Daher wurde die Netzwerkstruktur um die

Wachstumsraten-abhängige Regulation, über die Verfügbarkeit der RNA Polymerase, erweitert.

Der zweite Schritt hin zur dynamischen Modellierung der Regulation der Zentralstoffwechsel-Gene war die Entwicklung eines neuen Modellierungskonzepts, welches die Berücksichtigung der multiplen Regulation der Transkription (durch Regulatorproteine, die RNA Polymerase Verfügbarkeit und multiple Promotoren) erlaubt. Der Neuigkeitswert liegt insbesondere in der Prädiktion kinetischer Parameter aus der Nukleotid-Sequenz der spezifischen DNA-Bindestellen der Regulatorproteine. Die Vorhersagekraft der Methode wird durch übereinstimmende simulierte und experimentell bestimmte RNA Konzentrationen des *cra* Modulons demonstriert. Darüberhinaus werden die Modellvorhersagen betreffend der Wachstumsraten-abhängigen Regulation durch die experimentellen Daten der Gesamt-RNA gestützt. Das Model sagt eine starke Cra-Regulatorprotein-abhängige Regulation der Zentralstoffwechsel-Gene voraus, welche der Wachstumsraten-abhängigen Regulation überlagert ist. Die Konzentration des Cra-Protein-Inhibitors Fructose 1,6-bis(phosphat) wurde zum ersten Mal im Fed-Batch Prozess quantifiziert. Die Konzentration fiel signifikant ab, was die Hypothese über dessen Schlüsselrolle für die Signalbildung bei Glukose-Limitation bestätigt.

Mit der vorliegenden Arbeit wurde ein verbessertes, quantitatives Verständnis der Dynamik und Regulation des zentralen Kohlenstoffmetabolismus in *E. coli* in Fed-Batch Prozessen entwickelt. Die vorgeschlagene Netzwerkstruktur stellt eine Grundlage für die weitere dynamische Modellierung zellulärer Funktionen, die mit der Versorgung von Biosynthese-Vorstufen und Energie zusammen hängen, zur Verfügung. Das eingeführte Modellierungskonzept eignet sich speziell zur Modellierung großer zellulärer Netzwerke und könnte dazu beitragen, das „Metabolic Engineering“ der Genregulation in Produzentenstämmen weiter voranzutreiben.

Summary

In the current thesis a systems biology approach was chosen for the analysis of the dynamics of the central carbon metabolism and its regulation in *Escherichia coli* during industrially relevant fed-batch processes. The two main goals were (i) to reconstruct the network structure of global regulation of the central carbon metabolism based on experimental and mathematical methods, and, (ii) to develop an approach for dynamic modeling the transcription of the central carbon metabolism genes.

A set of glucose-limited fed-batch processes was performed applying a constant feed rate to provide the same conditions for all examinations. This process strategy leads to a continuous reduction in the glucose availability. Complementary metabolic flux and global transcription analyses revealed that the fluxes and most transcript levels in glycolysis, the pentose phosphate pathway and the biosynthesis strongly decrease. Fluxes in the TCA cycle remained constant and mRNA levels of TCA cycle and glyoxylate shunt genes increased. Moreover, the signaling dynamics were examined by quantification of the intracellular concentrations of the alarmones ppGpp and cAMP. Strong accumulation of both alarmones was observed at the onset of carbon limitation. A new finding was the subsequent resetting of the signals.

A network structure of regulation of the central carbon metabolism was reconstructed to enable a comprehensive explanation of the observed dynamics. Accordingly, the *cra* and *crp* modulons majorly determine the transcription of glycolysis, TCA cycle and glyoxylate shunt genes. The *relA*/*spoT* modulon mainly regulates protein biosynthesis and the specific growth rate. The key cellular components, involved in the signaling and resetting of the alarmone concentrations, were integrated into this model structure. Further, well-known regulatory phenomena concerning the carbohydrate transport systems and chemotaxis were observed, whereas the stress and starvation response were of only minor relevance. These cellular functions are discussed to be interconnected with the precursor and energy supply.

The rRNA and total RNA contents were quantified using a newly developed method. A strong growth rate-dependent regulation of both rRNA and mRNA was concluded from these data. Therefore, the network structure was extended by the growth rate-dependent regulation via RNA polymerase availability.

The second step towards dynamic modeling the regulation of central carbon metabolism genes was the development of a novel modeling framework that enables to consider multiple regulation of transcription (by regulator proteins, RNA polymerase availability and multiple

promoters). The novelty lies particularly in the prediction of kinetic parameters from the nucleotide sequences of the specific DNA-binding sites of the regulator proteins. The predictive power of this approach is demonstrated by the agreement of simulated and experimentally determined mRNA concentrations of the *cra* modulon. Moreover, the experimental data of the total RNA content support the model predictions concerning the growth rate-dependent regulation. The model predicts a strong Cra regulator protein-dependent regulation of the central carbon metabolism genes, which is superimposed by the growth rate-dependent regulation. The concentration of the Cra protein inhibitor fructose 1,6-bis(phosphate) was quantified for the first time in fed-batch cultivations. Its concentration fell significantly, which supports the hypothesis of its key role in signaling glucose availability.

The thesis provides an improved, quantitative understanding of the dynamics and regulation of the central carbon metabolism of *E. coli* in fed-batch processes. The proposed network structure may support further dynamic modeling of cellular functions interrelated with the supply of precursors and energy. The presented modeling framework is especially suitable for modeling large cellular networks and could make an impact on metabolic engineering of gene regulation in producer strains.

1 Introduction

1.1 Introduction

Recognizing that research in industrial biotechnology is the key to cope with many future global challenges the Organization of Economic Co-operation and Development has outlined policies to maximize potential economic and environmental benefits over the next two decades (OECD, 2009). Yet sustainable industrial production based on bio-processes will not be achieved without making headway in the still young field of systems biology and in its application through metabolic engineering. Our understanding of microbial metabolism and regulation will open up new possibilities to modify microorganisms to utilize new substrates, produce new products and – most demanding – increase yields.

Systems biology and metabolic engineering have evolved from natural and engineering sciences into a rich collection of tools for analyzing the behavior and rational optimization of biotechnological producer strains. The majority of the very advanced experimental and mathematical methodologies focus on the relevant metabolic networks, deciphering flux limitations and discovering potentially new metabolic routes. Regulation phenomena are equally complex and a dedicated research community has developed, accelerating progress towards the quantitative analysis of the corresponding cellular subnetworks. As knowledge about the regulation of cellular processes increases, it becomes evident that the microbial environment in the bioreactor and the process design can perturb intracellular processes (through changes in the levels of metabolic enzymes, transport proteins, transcription and protein biosynthesis apparatus, toxins and stress response) and make an impact on process performance.

On the other hand, a number of studies illustrate that altering regulation can improve production rates or yields, using intuitive approaches or rational, modular analyses of the cellular subnetworks. A vital example is the engineering of the phosphoenol-pyruvate:carbohydrate phosphotransferase system (PTS) of *Escherichia coli*, which transports the carbon and energy source (sugars) via the cell membrane and is involved in the global regulatory system catabolite repression (reviewed by Gosset, 2005). A series of further successful strategies for strain development considering the metabolic, regulatory and other cellular subnetworks was portrayed by Lee et al. (2005).

New developments of computational algorithms enable to simulate the microbial metabolism, integrating regulation of gene expression under various stationary conditions, and demonstrate that implementation of regulatory aspects make mathematical models more predictive (Kauffman et al., 2003). In this case, the interactions between metabolic and regulatory networks are assumed to be one-way. However, to comprehend the dynamic

interplay of the cellular subnetworks, and to reinforce the predictive power of mathematical models, the mutual interactions must be incorporated into dynamic, mechanistic representations. In the first instance, existing knowledge must be used, and extended by complementary experiments, to reconstruct and combine the biochemical subnetworks generally relevant for biotechnological processes. Moreover, approaches for dynamic modeling of large networks of gene expression are needed that can be coupled to existing metabolic models.

1.2 Thesis Outline

The present thesis majorly aims at the comprehensive understanding of the metabolic and regulatory behavior of the microorganism *Escherichia coli* during the transition from exponential to carbon-limited growth in industrially important fed-batch cultivations. The applied systems biology approach involved the quantitative experimental observation of the dynamics in signaling, transcription, metabolic fluxes and metabolite concentrations, and was accompanied by stationary and dynamic mathematical modeling. The thesis covers two main steps. First, the network structure of global regulation of the central carbon metabolism is reconstructed. In a second step, an approach for dynamic modeling of multiple regulation of transcription is introduced. The broad nature and the given detail of the current study were taken into account by self-contained chapters that portray the development and application of the experimental and mathematical methods.

In the second chapter, the signaling (alarmones cAMP, ppGpp) and metabolic flux dynamics during fed-batch cultivation applying a constant feed rate are depicted. The reorganization of the central carbon metabolism and the concomitant decrease of the biomass yield are discussed and the network architecture of the global regulation of metabolism, protein biosynthesis and the specific growth rate is proposed. Focus is put on the signaling and negative feedback mechanisms that lead to signal resetting. Thus, the chapter provides a network structure that comprehensively explains the considerable number of regulatory processes that adjust the demand of precursors and energy to the limited nutrient-supply. Chapter 3 details the *in vitro* synthesis, isolation and characterization of the alarmone ppGpp, which was needed as an external standard for the analytical procedure applied in Chapter 2. The proposed network structure is verified by comparison of the relative changes of the metabolic fluxes with complementary time series data of the global mRNA levels in Chapter 4. Detailed analysis of the changes in the transcript levels led to the integration of further cellular processes into this structure. These are interconnected with the global regulation of precursor and energy consumption and are discussed to be relevant for specific

biotechnological processes. The fifth chapter deals with a new method for quantification of rRNA using internal standard RNA and for the quantitative extraction of total RNA from *E. coli* suspensions. The dynamics of the total RNA, 16S and 23S rRNA contents observed during the fed-batch cultivation are depicted in Chapter 6 and indicate a strong growth rate-dependent regulation of transcription, which is superimposed on the regulation via the network of global regulatory systems depicted in Chapter 2.

One of the major components of the proposed regulatory network is the Cra regulator protein that regulates transcription of numerous operons of the central carbon metabolism (the *cra* modulon). The concentration of the metabolite, and, Cra protein inhibitor, fructose 1,6-bis(phosphate) was experimentally determined (Chapter 7). The observed dynamics suggest that the metabolite plays a major role in signaling glucose limitation. In Chapter 7 the multiple regulation of the central carbon metabolism is quantitatively analyzed using a dynamic model of *cra* modulon transcription. The contributions of growth rate- and regulator protein-dependent regulations are dissected. Chapter 7 provides a new approach for dynamic modeling the multiple regulation of transcription, using nucleotide sequences of the regulator DNA-binding sites for the prediction of kinetic parameters. The predictive power of the proposed method is critically evaluated by means of experimental data of the intracellular mRNA concentrations, which were determined by quantitative RT-PCR in a separate work by Schuhmacher et al. (2009), and, the total RNA concentrations described in Chapter 6. The thesis concludes with a summary and an outlook on possible future applications of the presented results.

2 Topology of the Global Regulatory Network of Carbon Limitation in *Escherichia coli*

Published in:

Hardiman, T., Lemuth, K., Keller, M.A., Reuss, M., and Siemann-Herzberg, M. (2007). J Biotechnol 132, 359-374.

Abstract

One fundamental shortcoming of biotechnological processes operating under carbon-limiting conditions is the high energy demand (maintenance) of the cells. Although the function of the central carbon metabolism in supplying precursors and energy for biosynthesis has been thoroughly characterized, its regulation and dynamic behavior during carbon-limited growth has not yet been revealed. The current work demonstrates a time series of metabolic flux distributions during fed-batch cultivation of *Escherichia coli* K-12 W3110 applying a constant feed rate. The fluxes in glycolysis, pentose phosphate pathway and biosynthesis fell significantly, whereas TCA cycle fluxes remained constant. The flux redistribution resulted in an enhanced energy generation in the TCA cycle and consequently, in a 20 % lower biomass yield. Both intracellular alarmones ppGpp and cAMP accumulated in large quantities after the onset of nutrient limitation, subsequently declining to basal levels. The network topology of the regulation of the central metabolic pathways was identified so that the observed metabolic and regulatory behavior can be described. This provides novel aspects of global regulation of the metabolism by the *cra*, *crp* and *relA*/*spoT* modulons. The work constitutes an important step towards dynamic mathematical modeling of regulation and metabolism, which is needed for the rational optimization of biotechnological processes.

2.1 Introduction

In biotechnological processes, the cell growth often is controlled via substrate limitation. This strategy is particularly applied in fed-batch cultivations employing exponential and constant feeding profiles, which not least guarantees respectable high cell densities. Substrate-limited growth, however, results in an excessive energy consumption for the maintenance of cellular functions (Lengeler et al., 1999; Pirt, 1982) as well as in other disadvantageous stress-related effects. Enfors' research group (Bylund et al., 1998; Larsson et al., 1996; Schweder et al., 1999; Teich et al., 1999; Xu et al., 1999a), as well as Hewitt et al. (2001; 2000; 1999) and Lapin et al. (2006) found that in large-scale fed-batch processes the extracellular gradients of substrate concentrations have profound effects on the growth yield, product formation and viability of the cell population. Evidently, such findings are the result of the interaction of the individual cell with its abiotic environment, which determines the cell's regulatory response – and thus, its metabolic state. Consequently, the rational optimization of biotechnological processes (metabolic engineering) requires dynamic mathematical models comprising both regulation and metabolism. Ongoing research in several academic groups and industry is focused on modeling the central carbon metabolism (glycolysis, pentose phosphate pathway and TCA cycle) of *E. coli* and its regulation during glucose-limited growth (Ellison et al., 2006; Heijnen et al., 2006). The first step in modeling is the identification of the model structure (topology), which relies on the knowledge from previous publications (bottom-up approach). Detailed information about the transcriptional regulation of the genes encoding metabolic enzymes in *E. coli* is available from databases (Keseler et al., 2009; Salgado et al., 2006). However, the regulators significantly affecting the central carbon metabolism during carbon limitation and consequently, the model topology of the global genetic regulatory *network*[1] have not yet been identified. In the complementary top-down approach the model structure is deduced from the system's response to an external stimulus, applying proper systems-level experimental tools like proteomics, transcriptomics, metabolomics and metabolic flux analysis (MFA). However, the dynamic metabolic response of the relevant pathways to carbon limitation still remains to be clarified in *E. coli*, which can be studied in fed-batch cultivations. When applying a constant feed rate, the substrate limitation continually increases and a succession of (physiological) quasi-steady states can be achieved (Dunn and Mor, 1975). This allows investigating the metabolic and regulatory response during the transition from exponential to carbon-limited growth.

[1] The term global genetic regulatory *network* is used since it implies that the mathematical model can comprise several subsystems, which are termed global genetic regulatory *systems*.

Bacteria control metabolism and growth rate through global genetic regulatory *systems*[1], i. e. regulons and modulons (Lengeler et al., 1999; Neidhardt and Savageau, 1996). Prominent examples in *E. coli* are the catabolite repression (*crp* modulon) and the stringent response (*relA*/*spoT* modulon), two processes that are active under carbon-limiting conditions. During stringent response (reviewed in Braeken et al. (2006), Cashel et al. (1996) and Lengeler et al. (1999)), the limitation of nutrients leads to the intracellular accumulation of ppGpp (guanosine 3',5'-bis(diphosphate)), which is supposed to bind to the RNA polymerase (Artsimovitch et al., 2004). The transcription of genes involved in the translation process – in particular of ribosomal RNA and ribosomal proteins – is negatively regulated by ppGpp. As a result, the protein biosynthesis rate declines, which in turn also leads to a reduction in growth rate (Cashel et al., 1996; Lengeler et al., 1999). During amino acid limitation, the synthesis of ppGpp or guanosine pentaphosphate (pppGpp), collectively referred to as (p)ppGpp, is mediated by RelA (GDP pyrophosphokinase/GTP pyrophosphokinase). Under amino acid-limiting conditions, the ribosome-bound RelA protein is stimulated by uncharged tRNAs at the A site of ribosomes (Wendrich et al., 2002). However, the accumulation of (p)ppGpp depends also on the dual activity of the SpoT protein as (p)ppGpp-hydrolase or (p)ppGpp-synthetase. Although it is known from a homologous protein of *Streptococcus dysgalactiae subsp. equisimilis* that the opposing activities of SpoT are reciprocally regulated (Hogg et al., 2004; Mechold et al., 2002), the regulation of the SpoT protein in *E. coli* is still hypothetical. The most important issue for understanding growth control is the signaling mechanism, which leads to accumulation of ppGpp under carbon-limiting conditions, an aspect that is still not entirely clarified.

Besides various effects on growth-related functions (Cashel et al., 1996), the alarmone ppGpp is known to be involved in the regulation of the sigma S factor concentration (σ^S; *rpoS* gene) on the transcriptional and posttranscriptional level (Hengge-Aronis, 2002). As an alternative subunit of RNA polymerase, σ^S is involved in the regulation of transcription in the general stress response in *E. coli* (also designated as “stationary phase response”). It is assumed that elevated levels of σ^S negatively regulate σ^D-dependent housekeeping genes, such as the TCA cycle genes (Patten et al., 2004). Moreover, ppGpp influences the competition between different stress-related sigma factors in the binding of the RNA polymerase core enzyme at the expense of the sigma factor σ^D (Jishage et al., 2002) and the RNA polymerase availability (Barker et al., 2001a; Barker et al., 2001b; Cashel et al., 1996; Jensen and Pedersen, 1990; Traxler et al., 2006). The *crp* modulon belongs to a group of global genetic regulatory systems, which can be subsumed under the term catabolite control. One basic feature of these systems is that the presence or absence of an extracellular carbon source is indicated by an intracellular metabolite (catabolite) that serves as a signal

for derepression (catabolite activation) or deactivation (catabolite repression) of catabolic genes (Saier et al., 1996). The *crp* modulon includes catabolic operons for the utilization of various carbon sources and is regulated by the Crp-cAMP complex. The synthesis of the alarmone cAMP (cyclic 3',5'-AMP) by the enzyme adenylate cyclase (CyaA) is stimulated by the phophorylated $EIIA^{Glc}$ protein, a component of the *E. coli* phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) (reviewed in Lengeler et al. (1999) and Postma et al. (1993)). It is assumed that a low glucose uptake rate by the PTS and a high ratio of phosphoenolpyruvate and pyruvate concentrations (c_{PEP} / c_{Pyr}) lead to the phosphorylation of the $EIIA^{Glc}$ protein (Hogema et al., 1998). Consequently, limited glucose availability leads to the synthesis of cAMP and the transcriptional regulator complex Crp-cAMP is formed. Catabolite control is also exerted by the catabolite repressor/activator protein Cra (formerly designated FruR), which regulates numerous genes involved in the carbon and energy metabolism (the *cra* modulon) (reviewed in Ramseier, 1996; Saier and Ramseier, 1996; Saier et al., 1996). The regulator protein Cra is inactivated by the catabolites fructose 1-phosphate and fructose 1,6-bis(phosphate) (Saier and Ramseier, 1996).

The present study demonstrates the signal formation and dynamic metabolic responses of *Escherichia coli* K-12 W3110 exposed to an increasing carbon limitation during fed-batch cultivation applying a constant feed rate. The observed decrease in the biomass yield is shown to result from a substantial carbon flow into the TCA cycle and from the subsequent oxidation of the carbon source. Simultaneously, most of the fluxes in the central carbon metabolism and in the biosynthesis pathways decreased significantly. This rearrangement is supposed to re-establish a balance between anabolism and catabolism after nutrient limitation. The genetic regulatory systems responsible for the illustrated metabolic responses are proposed and assembled to a global regulatory network. A new method for determining the qualitative time course of the intracellular cAMP concentration is presented. For the first time the profile of the intracellular cAMP level is shown in fed-batch cultivations of *E. coli* wild-type cells. Most importantly, the resetting of the cAMP signal could be demonstrated. The suggested network takes account of the observed signal resetting (cAMP and ppGpp) and of a probable stringent response signaling pathway during carbon limitation. The provided network topology is novel inasmuch as it comprehensively explains the obtained systems-level data of the metabolic transition from exponential to carbon-limited growth typical of fed-batch processes.

2.2 Materials and Methods

2.2.1 Bacterial Strain and Fed-Batch Cultivation

The fed-batch cultivations were carried out with the bacterial strain *Escherichia coli* K-12 W3110 (DSM 5911, German Collection of Microorganisms and Cell Cultures) in a 30-l bioreactor (Bioengineering AG, Wald, Switzerland). Minimal medium supplemented with glucose as the carbon source was used. The batch medium (batch volume $V_{R,0}$ = 17 l) consisted of 8.8 g l^{-1} glucose·H_2O, 2.0 g l^{-1} Na_2SO_4·$10H_2O$, 2.68 g l^{-1} $(NH_4)_2SO_4$, 1.0 g l^{-1} NH_4Cl, 14.6 g l^{-1} K_2HPO_4, 4.02 g l^{-1} NaH_2PO_4·$2H_2O$, 0.01 g l^{-1} thiamine·HCl; 0.3 mM $CaCl_2$·$2H_20$, 2 mM $MgSO_4$·$7H_2O$; and 3 ml l^{-1} of trace element solution (TES: 16.7 g l^{-1} $FeCl_3$·$6H_2O$, 20.1 g l^{-1} Na_2EDTA, 0.18 g l^{-1} $ZnSO_4$·$7H_2O$, 0.1 g l^{-1} $MnSO_4$·H_2O, 0.16 g l^{-1} $CuSO_4$·$5H_2O$, 0.18 g l^{-1} $CoCl_2$·$6H_2O$). Fed-batch medium contained 220 g l^{-1} glucose·H_2O, 63.36 g l^{-1} $(NH_4)_2HPO_4$, 0.2 g l^{-1} thiamine·HCl, 10 mM $CaCl_2$·$2H_20$, 32 mM $MgSO_4$·$7H_2O$ and 40 ml l^{-1} TES; 85 % H_3PO_4 was added to resolve any precipitate. The fed-batch process (constant feed rate, F = 0.294 l h^{-1}) was started after batch cultivation at a glucose concentration below 0.15 g l^{-1}. The process parameters that were controlled by a process control system (software developed by the Institute of Biochemical Engineering, Stuttgart, based on Visual Designer 4.0, Intelligent Instrumentation, Stuttgart, Germany) were: dissolved oxygen concentration > 50 % saturation, pH 7, pressure p = 1.5 bar, temperature T = 37 °C. 2 M H_3PO_4 or 4 M NaOH was used to adjust the pH values. The gas inflow ($\dot{V}_G$ = 10 l min^{-1}) was monitored by a mass flow meter (Bioengineering AG, Wald, Switzerland). Foam formation was suppressed by the addition of polypropylene glycol P 2000 (Sigma-Aldrich, Steinheim, Germany) when necessary.

2.2.2 Sampling and Analytical Methods

The sampling of extracellular substrates and products was carried out by filtration with a 0.2 µm-pore-size ceramic membrane (FIPS sampling probe, Flownamics, Madison, WI, USA). The culture medium samples were withdrawn with a capillary sampling probe as developed by Theobald et al. (1997), however, without using membrane-covered glass tubes. The biomass was measured in triplicate by determining the dry weight of the samples. The glucose and acetate concentrations were measured using enzymatic assays (R-Biopharm AG, Darmstadt, Germany). The ammonia concentration was determined using a photometric cuvette test (LCK303, Hach Lange, Düsseldorf, Germany) (all measurements were performed in duplicate). The glucose concentration (between 0.1 and 1.0 g l^{-1}) was also

measured online using a ProcessTrace 1.2MT probe (Trace Analytics GmbH, Braunschweig, Germany). The CO_2 and O_2 mole fractions in the gas outflow were monitored using an S710 gas analyzer, (infrared carbon dioxide analyzer FINOR, paramagnetic oxygen analyzer OXOR-P; Maihak AG, Hamburg, Germany).

2.2.3 Determination of Nucleotide Concentrations

The nucleotides were extracted from 4 ml of culture medium by sampling into 1 ml of perchloric acid (- 20 °C) (Theobald et al., 1997) and mixing for 15 min by end-over-end rotation (4 °C). The pH was neutralized by adding 845 µl 5 M K_2CO_3 (0 °C) (Buchholz et al., 2001). After centrifugation (4000 rpm, 4 °C) the supernatant was collected and stored at – 20 °C until further use. The samples were chilled in an ice bath during all transfer steps. The samples were filtered (0.2 µm) prior to HPLC analysis. A Spectra Physics HPLC system (Farmont, USA) was used to separate and quantify the nucleotides by ion-pair reversed-phase HPLC (Theobald et al., 1997), which comprised a low pressure gradient system (double pump P2000), an auto sampler (AS3000) and an UV detector (UV 1000, λ = 254 nm). A Supelcosil LC-18T column (150 mm x 4.6 mm; particle size 3 µm) and a Supelguard LC18T guard column (20 mm x 4.6 mm; particle size 5 µm) were employed (Supelco, Bellefonte, USA). The analysis was carried out with a flow rate of 1.2 ml min^{-1}. A gradient elution modified from Cserjan-Puschmann et al. (1999) was performed with buffer A (0.1 M KH_2PO_4/K_2HPO_4, pH 6.0; 5 mM tetrabutyl ammonium dihydrogen phosphate; 5 µM EDTA; adjusted to a final pH of 5.3 and purged with helium for 15 min) and buffer B (80 % of buffer A; 20 % acetonitrile, adjusted to a final pH of 5.9 and purged with He for 15 min). Gradient program: 5 % buffer B for 3 min, 5 - 40 % buffer B for 36 min, 40 - 100 % buffer B for 15 min, 100 % buffer B for 5 min, 100 - 5 % buffer B for 5 min and 5 % buffer B for 3 min. Each sample was also run spiked with a nucleotide standard solution in order to identify the respective nucleotide peaks in the chromatograms. For the nucleotide standard solution, ppGpp was synthesized according to a modified protocol from Krohn and Wagner (1995). Ribosome preparations were obtained by ultracentrifugation with a two-step sucrose gradient from extracts of *E. coli* K-12 W3110 cells harvested from 20-l batch cultivations. ppGpp was synthesized *in vitro* using the ribosome preparations and GDP and ATP as substrates. Isolation of ppGpp involved two consecutive column chromatographic steps (ion-exchange and size-exclusion chromatography). The synthesis product was structurally analyzed with ^{13}C-NMR (Que et al., 1973); the molecular weight was confirmed by mass spectrometry.

2.2.4 Metabolite Balances and Concentration Time Courses

The time profiles of substrate and product concentrations, $\hat{c}_j(t)$, from five independent fed-batch experiments were obtained from approximations of measured data (glucose, ammonia, acetate, biomass) by analytical functions. Time variant functions for the specific substrate uptake and product excretion rates (measured rates, $r_m(t)$; Figure in Appendix A) were determined from the time derivatives, $d\hat{c}_j(t)/dt$, and the fed-batch balance Equations. Since all fed-batch balances contain the reaction volume, it was considered as an important variable. Thus, its time profile, $\hat{V}_R(t)$, was also approximated taking into account the constant feed rate, pH adjustment, volume loss when withdrawing the samples and upon the evaporation of water during cultivation (> 12 hours). The specific O_2 uptake rate (q_{O_2}) and CO_2 production rate (q_{CO_2}) were determined from gas analysis and mass balancing of the two components in the gas phase. The specific growth rate (μ) was obtained using the biomass balance [Equation (2-1) in Table 2-1] and the approximation of experimental data ($\hat{c}_X(t)$, Equation in Appendix B). Below 0.1 g l^{-1}, which is below the calibration range of the online glucose measurement, the glucose concentration was calculated according to simple Monod-type kinetics according to Equation (2-2) (Table 2-1). The biomass yield ($\hat{Y}_{X/S}(t)$) was calculated according to Equation (2-3). For this, the specific substrate uptake rate [$\hat{q}_S(t)$, Equation (2-4) in Table 2-1] was calculated from the approximation of experimental data ($\hat{c}_{glc}(t)$, Appendix B). Pirt suggested the use of Equation (2-5) (Table 2-1) defining a growth rate-dependent maintenance energy coefficient (m_S) when calculating the specific substrate uptake rate ($\hat{q}_S$) (Pirt, 1982). The parameters m_1 and m' were estimated by plotting the specific glucose uptake rate as a function of the specific growth rate according to Pirt (1982).

Table 2-1 Equations for the estimation of time courses

(2-1) $$\hat{\mu}(t) = \frac{1}{\hat{c}_X(t)} \frac{d\hat{c}_X(t)}{dt} + \frac{F}{\hat{V}_R(t)}$$ [h^{-1}]

(2-2) $$\hat{c}_{glc}(t>0) = \frac{K_S \hat{\mu}(t)}{\mu_{max} - \hat{\mu}(t)}$$ [$g\ l^{-1}$]

(2-3) $$\hat{Y}_{X/S}(t) = \frac{\hat{\mu}(t)}{\hat{q}_S(t)}$$ [-]

(2-4) $$\hat{q}_S(t) = \frac{F}{\hat{V}_R}\left(c^+_{glc} - \hat{c}_{glc}(t)\right) - \frac{d\hat{c}_{glc}}{dt}$$ [$g\ g^{-1}\ h^{-1}$]

(2-5) $$\hat{q}_S(t) = \frac{\hat{\mu}(t)}{Y_{X/S,max}} + m_S(t) \quad \text{with} \quad m_S(t) = m_1 + m'(1 - \frac{\hat{\mu}(t)}{\mu_{max}})$$ [$g\ g^{-1}\ h^{-1}$]

(2-6) $$\hat{v}_X = \frac{0.4860 \cdot 2^{(1.144\hat{\mu})}}{-0.636 + 0.635 \cdot 2^{(0.718\hat{\mu})}}$$ [ml cytosol (g dry weight)$^{-1}$]

(2-7) $$\hat{r}_{cAMP,export}(t) = \frac{d\hat{X}_{cAMP,extra}(t)}{dt} + \hat{X}_{cAMP,extra}(t)\frac{F}{\hat{V}_R(t)} + \frac{\hat{X}_{cAMP,extra}(t)}{\hat{c}_X(t)}\frac{d\hat{c}_X(t)}{dt}$$

$$\hat{r}_{cAMP,export}(t) = \frac{d\hat{X}_{cAMP,extra}(t)}{dt} + \hat{X}_{cAMP,extra}(t)\hat{\mu}(t)$$ [$\mu mol\ g^{-1}\ h^{-1}$]

(2-8) $$\hat{X}_{cAMP,intra}(t) = \frac{\hat{r}_{cAMP,export}(t)}{k_{cAMP,export}}$$ [$\mu mol\ g^{-1}$]

2.2.5 Metabolic Flux Analysis

The model used for metabolic flux analysis by stoichiometric metabolite balancing was previously published by Chassagnole et al. (2002), but slightly adapted to the present study's requirements (Appendix C). The stoichiometric *E. coli* model (129 reactions and 133

metabolites) represents the central carbon metabolism (glycolysis, pentose phosphate pathway, TCA cycle) and the biosynthesis of monomers and polymers. The model assumes that the macromolecular composition remains constant at different growth rates (cell composition according to (Ingraham et al., 1983; Neidhardt and Umbarger, 1996; Pramanik and Keasling, 1998)), the presence of an NADPH-dependent isocitrate dehydrogenase (according to EcoCyc database (Keseler et al., 2009)), a constant P/O ratio of 2.0 (Calhoun et al., 1993; Gennis and Stewart, 1996) and an ATP demand of 31.5 mmol ATP per g protein for the biosynthesis of proteins from amino acids. The stoichiometric model used for the analysis of the metabolic fluxes is represented as a system of balance equations ($d\mathbf{c}/dt = \mathbf{N}\mathbf{r}$), where $d\mathbf{c}/dt$ is the time derivative of the vector of intracellular metabolite concentrations [mmol g^{-1} h^{-1}]. $\mathbf{N}$ is the stoichiometric matrix, containing stoichiometric coefficients of all reactions in the network and $\mathbf{r}$ is the vector of the specific reaction rates [mmol g^{-1} h^{-1}]. The time derivatives are then equated with zero, assuming metabolic (quasi-)steady states ($\mathbf{0} = \mathbf{N}\mathbf{r}$). After the separation of the measured rates (index m) and the rates to be calculated (index c) ($\mathbf{0} = \mathbf{N}_m\mathbf{r}_m + \mathbf{N}_c\mathbf{r}_c$), the intracellular fluxes can be determined using equation $\mathbf{r}_c = -\mathbf{N}_c^{-1}\mathbf{N}_m\mathbf{r}_m$ (Nielsen and Villadsen, 1994; Stephanopoulos et al., 1998). A time series of metabolic flux distributions was determined using the time courses of the specific substrate uptake and product excretion rates (measured rates). With six measured rates ($r_{glc}(t)$, $r_{ac}(t)$, $r_X(t)$, $r_{NH_4}(t)$, $q_{O_2}(t)$ and $q_{CO_2}(t)$; Appendix A) a three-fold overdetermined system was obtained. This allowed the reconciliation of data and the identification of gross measurement errors according to (Nielsen and Villadsen, 1994; Stephanopoulos et al., 1998). All analyses were performed using the Insilico Discovery 1.2 software (Insilico Biotechnology AG, Stuttgart, Germany).

2.3 Results

2.3.1 Fed-Batch Cultivation

Five independent fed-batch cultivation experiments applying a constant feed rate were carried out to study the response of the central carbon metabolism of *E. coli* concerning glucose limitation. The time profiles of the extracellular concentrations of glucose, acetate and biomass are presented in Figure 2-1a. Acetate was produced during the consumption of glucose in the batch phase and was used up after 1 h of glucose-limited growth (Fig. 2-1a). In response to the limited availability of the carbon source, the specific growth rate, μ,

gradually decreased from 0.7 h^{-1} at t = - 2 h to 0.08 h^{-1} after 8 h of glucose-limited growth (Fig. 2-1b). During the fed-batch phase the glucose concentration decreased from approximately 60 mg l^{-1} to 6 mg l^{-1} (Fig. 2-1b). This time profile was obtained from Equation (2-2) (Table 2-1) assuming a substrate saturation constant of K_S = 0.05 g l^{-1} according to Xu et al. (1999b), which can be expected for non-adapted cells during the batch and in the beginning of the fed-batch phase. After adaptation of the glucose transport systems during fed-batch cultivation, however, the saturation constant should be significantly lower (Senn et al., 1994). Consequently, the glucose concentration given in Figure 2-1b might be lower [see Equation (2-2), Table 2-1] and must be regarded as the qualitative time course for the applied feed rate.

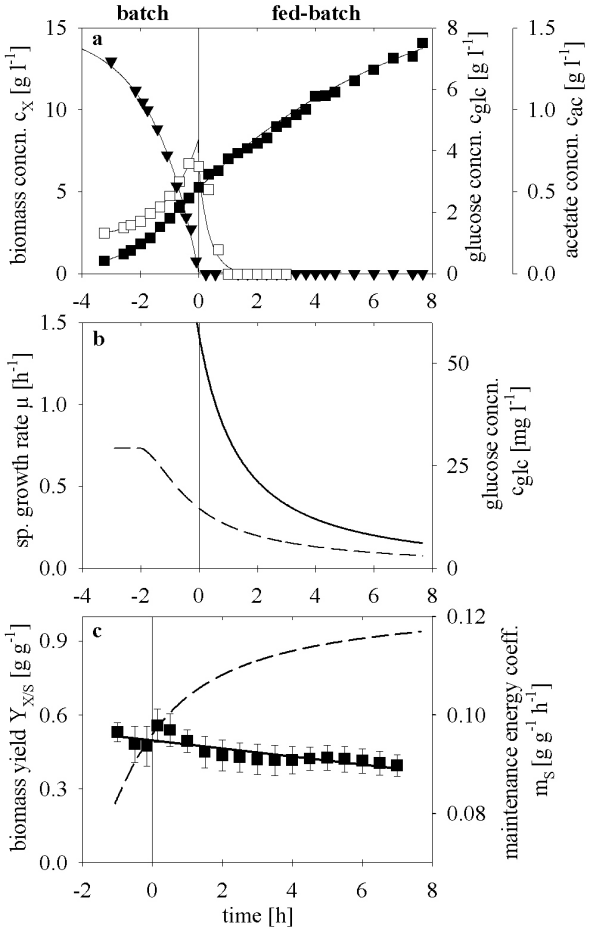


Fig. 2-1 Glucose limited fed-batch cultivation of *E. coli* K-12 W3110 with constant feed rate. Vertical solid lines at t = 0 indicate glucose limitation (as judged from the O_2 consumption rate (q_{O2}) (Appendix A) and dissolved oxygen concentration (data not shown)). (a) measured biomass concentrations (■), glucose (▼) and acetate (□), time profiles (solid lines); (b) approximated time course of specific growth rate (μ) (broken line) and glucose concentration (solid line). (c) biomass yield ($Y_{X/S}$) (symbols and error bars: mean values of five cultivations and standard deviation) and maintenance energy coefficient (m_S) (broken solid line).

Table 2-2 Parameters

Parameter	Value	Source
K_S	0.05 g l^{-1}	Xu et al. (1999b)
μ_{max}	0.7 h^{-1}	This work
$Y_{X/S,max}$	0.6	This work
m_1	0.07 h^{-1}	This work
m'	0.05 h^{-1}	This work
$k_{cAMP,export}$	509 h^{-1}	This work

2.3.2 Biomass Yield and Maintenance

The biomass yield ($Y_{X/S}$) decreased steadily from 0.5 g g^{-1} in the batch phase to 0.4 g g^{-1} at μ = 0.08 h^{-1} during glucose-limited growth (Fig. 2-1c). This represents a significant loss of 20 % in biomass yield. Moreover, the time variant maintenance energy coefficient ($m_S(t)$) was determined according to Pirt (1982) [Equation (2-5), Tables 2-1 and 2-2]. A steep increase was found during periods of glucose limitation (Fig. 2-1c). Consequently, the stronger the limitation, the less substrate was used for biomass production and more energy is generated.

Although this work focuses on the dynamic changes during glucose limitation, two batch phase-related findings are also worth mentioning ($t \leq 0$, Fig. 2-1). First, the growth rate decreased more rapidly than can be explained simply by Monod-type kinetics. The overflow metabolite acetate accumulated in the batch phase, which is known to have many inhibitory effects on growth (summarized in Xu et al., 1999b). A dynamic mathematical model published by Xu et al. (1999b) describes the inhibition of glucose uptake by acetate and proposes a similar reduction of growth rate as found in the current work. Therefore, it must be attributed to the accumulation of acetate in the growth medium. Secondly, the growth

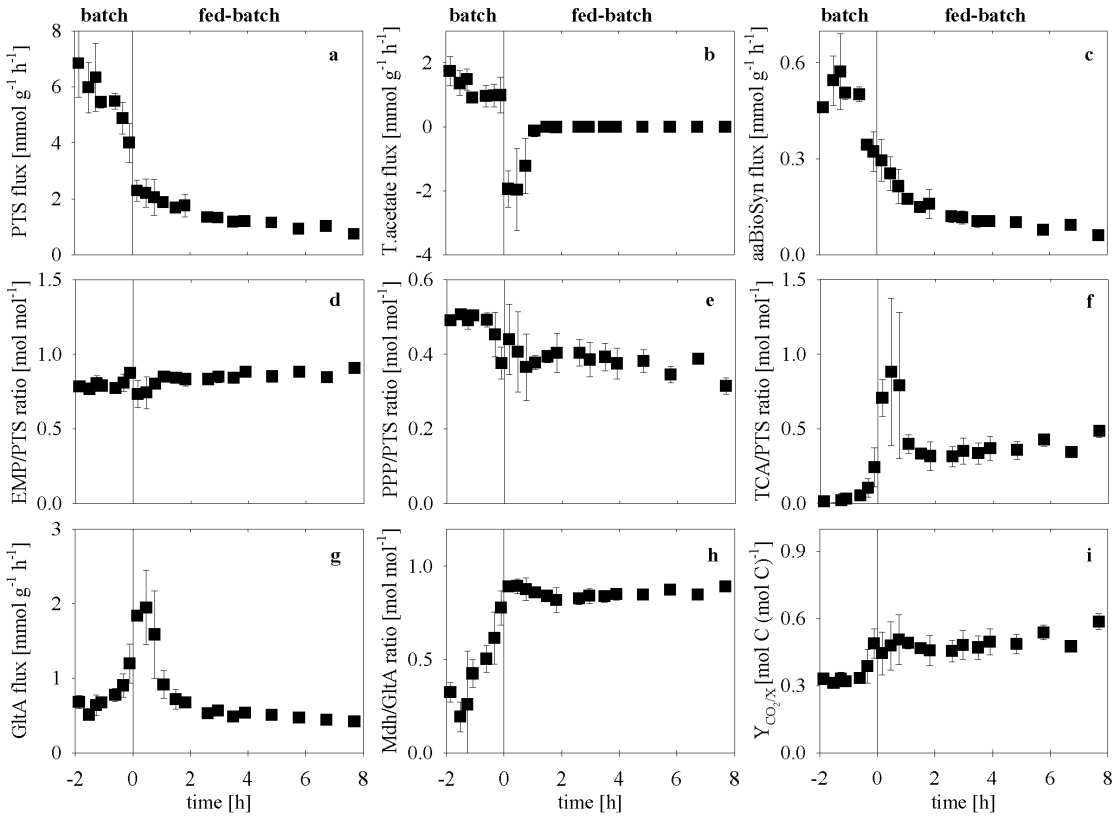


Fig. 2-2 Time courses of metabolic fluxes in intermediary metabolism during the transition to carbon limitation. Fluxes were determined from stoichiometric metabolite balancing. Symbols with standard deviation correspond to mean values from five independent glucose limited fed-batch cultivations of *E. coli* K-12 W3110 with constant feed rate. (a) glucose uptake (PTS) flux. (b) acetate excretion (T.acetate) flux. (c) average flux in amino acid biosynthesis pathways (aaBioSyn). (d) ratio of average glycolysis (EMP) and PTS fluxes. (e) ratio of average PPP and PTS fluxes. (f) ratio of average TCA and PTS fluxes. (g) GltA (citrate synthase) flux. (h) ratio of GltA and Mdh (malate dehydrogenase) fluxes. **i**, ratio of C-fluxes by CO_2 excretion and biomass production.

yield decreased and the maintenance energy coefficient increased in the batch phase (Fig. 2-1c). This is also attributed to the effect of the metabolic overflow (carbon loss).

2.3.3 Central Carbon Metabolism and Biosynthetic Pathways

The dynamic changes in the intermediary metabolism in response to carbon limitation were determined using metabolic flux analysis. The glucose uptake rate during unlimited growth (batch phase) was roughly 6 mmol g^{-1} h^{-1} (PTS, Fig. 2-2a). As a consequence of the constant feed rate, the glucose uptake rate (PTS) was subjected to a gradual decrease and was less than 1 mmol g^{-1} h^{-1} after 8 h of fed-batch cultivation (Fig. 2-2a). Acetate was excreted in the batch phase and used as an additional carbon source during the first hour of the fed-batch cultivation (Fig. 2-2b and Fig. 2-1a). Almost all fluxes were strongly reduced (Fig. 2-3) in

response to the decline in the carbon supply, although some fluxes transiently increased during acetate consumption. The reaction rates of the biosynthesis pathways (monomers and polymers) were reduced uniformly (Fig. 2-3), which is due to the fact that the growth rate declined. The time profile of the average flux in the amino acid biosynthesis pathways is shown in Figure 2-2c. The rates in amino acid biosynthesis were greatly reduced during the initial phase after the onset of glucose limitation (Fig. 2-2c) and continued to decrease slowly along with μ (Fig. 2-1b).

Table 2-3 Reduction of fluxes during glucose limitation

Pathway k [a]	$r_k / r_{k,0}$ [a] [%]			
	NADPH-dep. IcdA[b]		NADH-dep. IcdA[b]	
	μ = 0.13 h^{-1}	μ = 0.08 h^{-1}	μ = 0.13 h^{-1}	μ = 0.08 h^{-1}
PTS	25	15	25	19
EMP	27	17	24	19
PPP	20	10	27	19
TCA	90	75	126	145
bio	24	14	27	19

[a]To obtain the relative average fluxes ($r_k / r_{k,0}$) for each pathway (k) (EMP, glycolysis; PPP, pentose phosphate pathway; TCA cycle; PTS, glucose uptake; bio, growth-related reactions), average fluxes (r_k) of the pathways (k) were determined at the given growth rates in fed-batch cultivation and divided by the corresponding average fluxes during unlimited growth (in batch phase), $r_{k,0}$.

[b]Relative average fluxes are given for the assumption that isocitrate dehydrogenase (IcdA) is NADPH-dependent (average of five cultivations), and additionally, for the opposing assumption of a NADH-dependent IcdA (one cultivation).

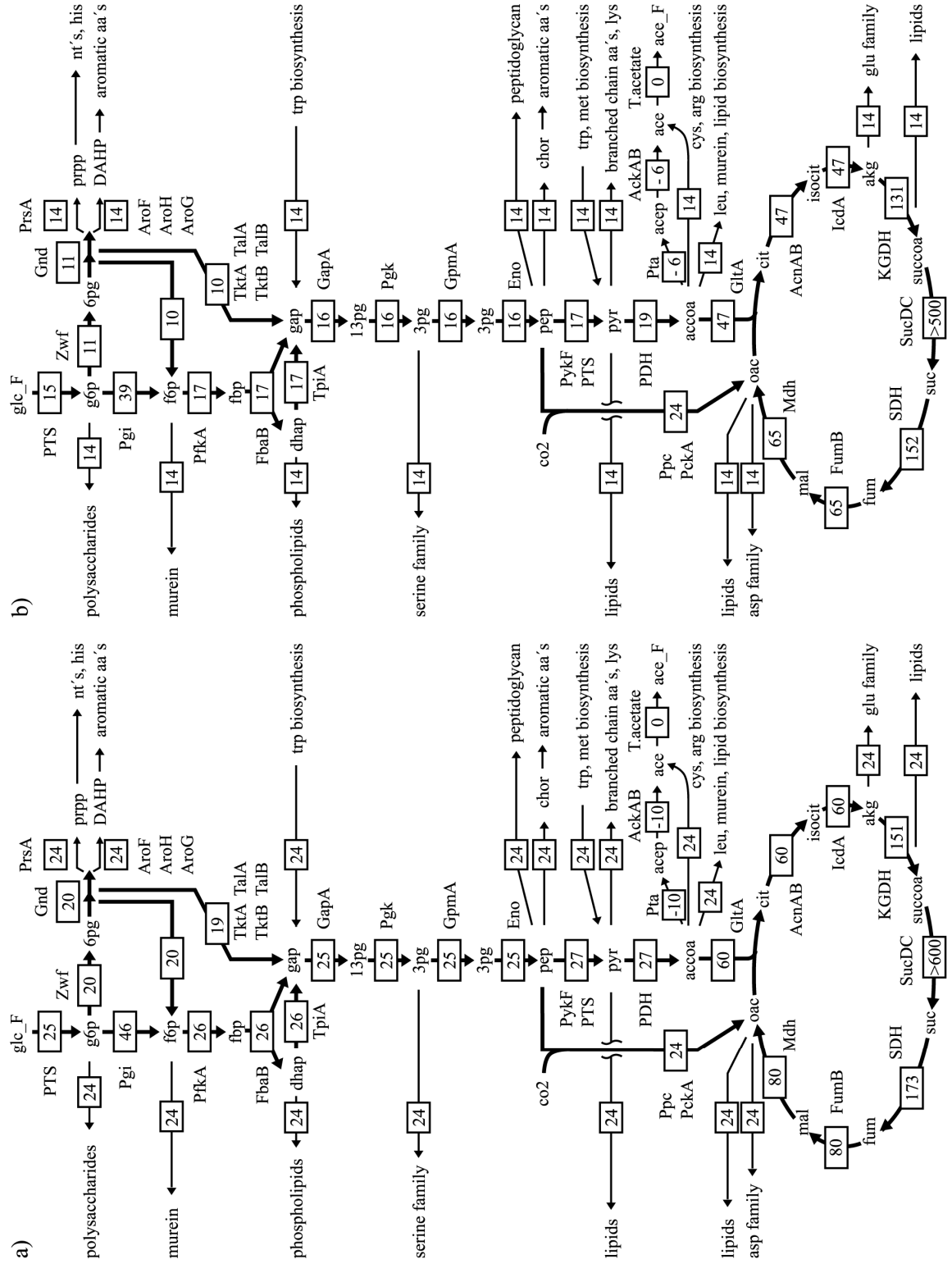


Fig. 2-3 Changes in metabolic flux distributions in central carbon metabolism during fed-batch cultivation. Fluxes are mean values of five independent fed-batch cultivations with constant feed rate of *E. coli* K-12 W3110 determined from stoichiometric metabolite balancing and given as molar percentages of the corresponding fluxes in the reference state during unlimited growth in the batch phase (0.3 h before start of the feed; μ = 0.4 h^{-1}). (a) and (b) glucose limited fed-batch cultivation (μ = 0.13 h^{-1} and μ = 0.08 h^{-1}, respectively). The model used includes also the respiration, biosynthetic pathways and polymerization reactions. Only those reactions involving metabolites from central carbon metabolism (as substrates or products) are shown.

The changes in the central carbon metabolism in response to limited carbon supply are depicted in Figure 2-3. The fluxes in glycolysis (EMP) were reduced to a similar extent as the glucose uptake rate (PTS) during fed-batch cultivation (Fig. 2-2d and Table 2-3). The reaction rates in the pentose phosphate pathway (PPP) were reduced much further than glucose influx (Fig. 2-2e and Table 2-3). It can obviously be assumed that the flux into PPP is reduced because the demand for precursors is lower during slower growth. The TCA cycle flux increased relative to the glucose influx (Fig. 2-2f) and was still 90 % at μ = 0.13 h^{-1} compared to batch growth (Table 2-3). This can be explained as a result of two findings: First, the influx into the TCA cycle via the first reaction, GltA (enzyme: citrate synthase), was maintained at a high level (Fig. 2-2g). Secondly, the precursor efflux from the TCA cycle to biosynthesis decreased substantially (Fig. 2-3). The enhancement of TCA cycle fluxes is also reflected in the flux ratio of the reactions Mdh (malate dehydrogenase) and GltA, which was approximately 0.9 immediately after the onset of limitation compared to roughly 0.3 in the preceding batch phase (Fig. 2-2h). Thus, the available carbon source is primarily oxidized in the TCA cycle to produce energy, and less carbon is available for biosynthesis. This is supported also by the fact that CO_2 production increased relative to biomass production, $Y_{CO_2/X}$ (Fig. 2-2i).

2.3.4 Adenosine Nucleotide and Alarmone Concentrations

The adenylate energy charge (AEC) was 0.8 in the batch phase and decreased slowly to 0.7 during the fed-batch phase (see linear regression in Fig. 2-4c), reflecting a decline in the availability of energy. Surprisingly, an increase in the ATP, ADP and AMP concentrations given in [µmol g^{-1}] was recorded during fed-batch cultivation (data for ATP and the AXP pool are shown in the upper graphs of Fig. 2-4d and 2-4e, respectively). However, it is known that cells become smaller and lighter during the stringent response (Lengeler et al., 1999; Pramanik and Keasling, 1997). Pramanik and Keasling (1997) published equations for the *E. coli* median cell volume and the dry weight per cell as a function of the specific growth rate μ based on experimental results. These equations clearly show that the dry weight decreases more than the cell volume does, when the growth rate declines. Both equations from Pramanik and Keasling (1997) are used for estimation of the *specific* cell volume v_X (reciprocal density ρ_X), i. e. the cell volume per g dry weight, as a function of the specific growth rate μ [Equation (2-6), Table 2-1]. The time course ($\hat{v}_X(t)$) during fed-batch cultivation demonstrates that the specific cell volume increases significantly when the growth rate decreases (Fig. 2-4f). Equation (2-6) (Table 2-1) was now used to calculate the molar

concentrations of the nucleotides. Figures 2-4d and 2-4e (lower graphs) show the results for ATP and AXP, respectively. Based on these approximations, it is assumed that the molar concentrations of AXPs remain constant during fed-batch cultivation (see Fig. 2-4e), which is in accordance with the general assumption of AXP as a conserved moiety. The concentration of the intracellular metabolites generally is related to the biomass concentration and then given in [µmol (g dry weight)$^{-1}$]. This is straightforward and the results can easily be compared with data gained by other laboratories. Nevertheless, the variable specific cell volume (Fig. 2-4f) suggests that it is best to consider molar concentrations (in [mmol (l cytosol)$^{-1}$]), when interpreting experimental results from different growth rates, as these are relevant for the *in vivo* reaction rates.

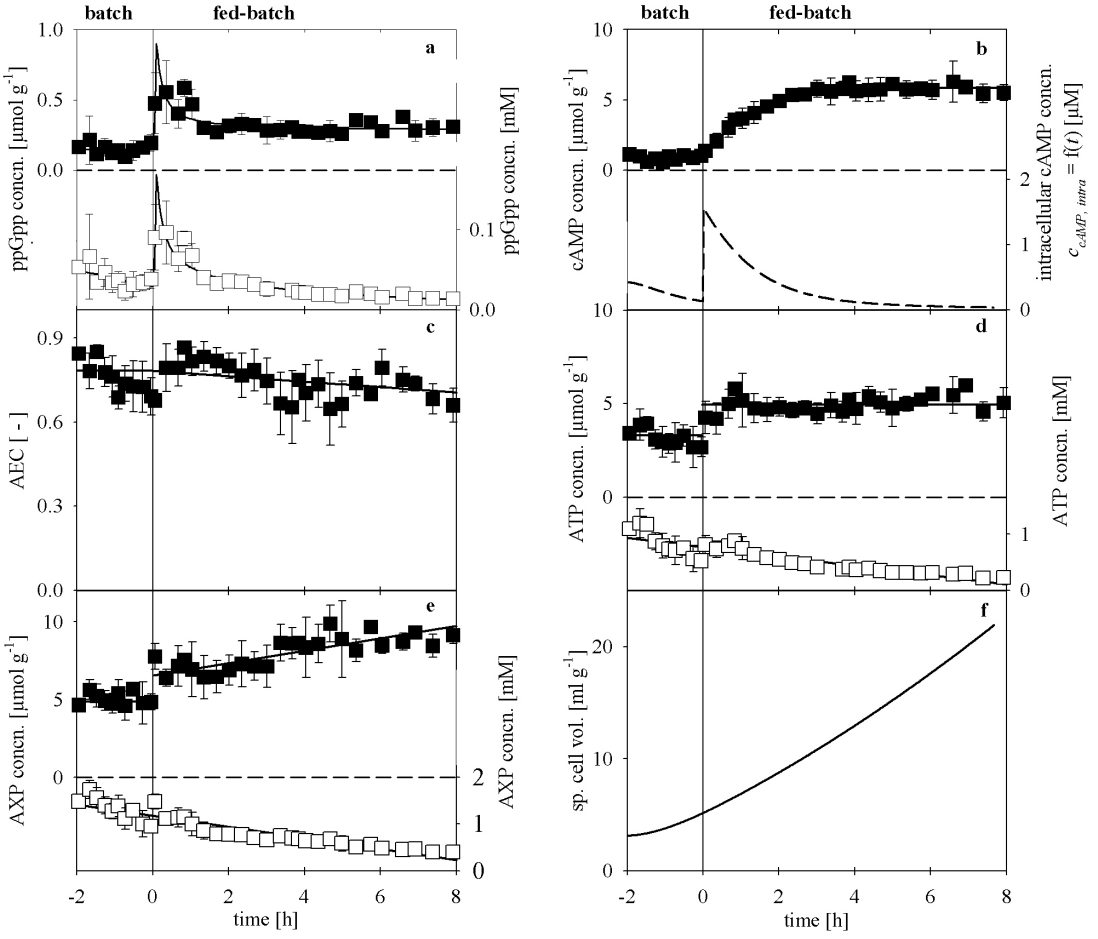


Fig. 2-4 Nucleotide concentrations during glucose limited fed-batch cultivation of *E. coli* K-12 W3110 with constant feed rate. Symbols represent mean values with bars for the standard deviation of measured concentrations from five independent cultivations. (■, upper graphs), concentration in [µmol g^{-1}] = [µmol (g dry weight)$^{-1}$]; (□, lower graphs), concentration in [mM] = [mmol (l cytosol)$^{-1}$]; (solid lines) approximated time courses. (a) ppGpp. (b) total cAMP concentration (intra- plus extracellular) [µmol g^{-1}] (symbols), and estimated intracellular concentration [µM] (solid broken line). (c) adenylate energy charge (AEC).(d) ATP, (e) adenosine nucleotide pool (AXP), and (f), approximated time profile of the specific cell volume.

The dynamics of the intracellular ppGpp and cAMP concentrations were recorded due to the aforementioned effects on regulation of intermediary metabolism. The time profile of the ppGpp concentration is depicted in Figure 2-4a in [µmol g^{-1}] (upper graph) and [mM] (lower graph). With respect to the molar concentration, the alarmone accumulated strongly after the onset of carbon limitation and decreased to a low basal level after approximately 4 h of fed-batch cultivation. The maximum ppGpp level obtained in the five experiments amounted to 0.9 µmol g^{-1} or 0.2 mM (peak values in solid lines in Fig. 2-4a). Although the sampling frequency was high, the peak value could not be resolved in each of the five independent cultivations due to a very fast response in ppGpp accumulation (elevated levels already after 12 s, peak values between 4 min and 20 min). This results in the apparently lower average value after the onset of limitation in Figure 2-4a. It is difficult, however, to determine intracellular cAMP concentrations because it is assumed that 99.9 % of the total cAMP is found in the culture medium (Matin and Matin, 1982). A method to separate *E. coli* cells from the medium and at the same time preserve the metabolic state (quenching) is required. An adequate method has not yet been demonstrated and the results of previous methods differ by three orders of magnitude (from µM to mM), e. g. in (Death and Ferenci, 1994; Epstein et al., 1975; Lin et al., 2004; Matin and Matin, 1982). Therefore, a new method was developed for the estimation of the intracellular concentration of cAMP in *E. coli*. First, the total cAMP concentration (including extracellular and intracellular cAMP) was measured after quenching the metabolism and the extraction of cAMP. A strong increase was observed at the onset of limitation (Fig. 2-4b, upper graph). However, after 4 h the concentration (given in [µmol (g dry weight)$^{-1}$]) remained constant demonstrating that cAMP is continuously synthesized as the cell mass increases (Fig. 2-1a). Additionally, the extracellular concentration was quantified. Essentially the same concentrations were obtained for total and extracellular cAMP (data not shown) in the range of the precision of the analytical procedure. This demonstrates that almost all cAMP is found extracellularly during the fed-batch conditions applied and consequently that newly synthesized cAMP is exported immediately. The time profile of the extracellular cAMP concentration ($\hat{X}_{cAMP,extra}(t)$) was approximated (see Appendix B) and plotted as solid line in the upper graph of Figure 2-4b. Using the balance Equation (2-7) (Table 2-1) for extracellular cAMP, the cAMP export rate ($\hat{r}_{cAMP,export}(t)$) was calculated. Assuming that the export rate is proportional to the intracellular cAMP concentration as suggested from Epstein et al. (1975), Equation (2-8) (Table 2-1) was used to approximate the intracellular cAMP level, $\hat{X}_{cAMP,intra}(t)$ in [µmol g^{-1}]. The result was converted to [µM] by application of Equation (2-6) (Table 2-1). However, using the rate constant given by Epstein et al. (1975), $k_{cAMP,export}$, an unusually high intracellular cAMP

concentration was obtained (> 1 mM). Therefore, $k_{cAMP,export}$ = 509 h^{-1} was estimated from the conditions at the onset of glucose limitation, namely the cAMP export rate [$\hat{r}_{cAMP,export}(t_0 = 0)$ = 4.1 µmol g^{-1} h^{-1}, from Equation (2-7) in Table 2-1], using Equation (2-8) (Table 2-1). Additionally, the concentration of intracellular cAMP was assumed as the highest concentration found by Matin and Matin (1982) ($\hat{X}_{cAMP,intra}(t_0)$ = 8.1 nmol g^{-1}) since the maximum is expected at the beginning of glucose limitation. The qualitative time course of the intracellular cAMP concentration can now be obtained from the Equations (2-8) and (2-6) (Table 2-1) and is depicted in Figure 2-4b in [µM] (lower graph). The alarmone accumulated strongly in the initial phase and decreased to a low basal level after 4 h. It is worth mentioning that the qualitative time profile is congruent with the one determined for ppGpp (Fig. 2-4a, lower graph).

2.4 Discussion

2.4.1 Topology of the Network Regulating the Central Carbon Metabolism

The work presented here demonstrates the formation and resetting of the intracellular signals of carbon limitation, cAMP and ppGpp, and the redistribution of metabolic fluxes during glucose-limited fed-batch growth. In order to explain the observed dynamic behavior the topology of the global genetic regulatory *network* of regulation of the central carbon metabolism is identified. This involves the examination of global genetic regulatory *systems* (regulons and modulons) that are known to be active during carbon limitation and are likely to lead to the observed metabolic responses. The focus will also be put on the structure of the signaling network and the negative feedback regulation mechanisms. These are part of the regulatory response and lend adaptive behavior to the cells as demonstrated in this study (signal reset). Besides the systems-level description of the observed behavior, the proposed network topology shall provide a basis for mathematical modeling.

2.4.2 Indications for Regulation by Global Genetic Regulatory Systems

During fed-batch growth the biomass yield dropped 20 % while the substrate was preferentially used for energy generation (Fig. 2-1c). The energy is used for the maintenance of cellular functions and might also be attributed to the general carbon limitation response, e. g. for making new proteins for sugar transport systems. The substantial energy demand requires the presence of relatively strong fluxes through the (energy-producing) pathways, namely the glycolysis and the TCA cycle. In fact, the observed changes in the glycolysis,

PPP, TCA cycle and biosynthesis (Fig. 2-3) demonstrate that the energy generation is enhanced by high reaction rates in the TCA cycle and that the substrate is less used for biomass synthesis. This comprehensive redistribution of metabolic fluxes indicates that there must be a global response, i. e. a coordinate regulation of the central metabolic pathways and biosynthesis by global genetic regulatory systems such as regulons or modulons (Lengeler et al., 1999; Neidhardt and Savageau, 1996). Of course, besides regulation of enzyme synthesis the enzyme activities additionally might be regulated by protein phosphorylation or metabolic regulation (effectors). However, global genetic regulation is more economic in terms of saving precursors and energy.

2.4.3 Regulation of Precursor and Energy Supply

The fluxes in the TCA cycle were shown to be high during fed-batch growth, which differed notably from the general flux reduction (Fig. 2-3). Provided that this response is accomplished by a coordinative regulation of enzyme levels the Crp-cAMP complex (*crp* modulon) plays possibly a major role. This hypothesis is supported by the fact that cAMP accumulated in large quantities (Fig. 2-4b) and that almost all genes for TCA enzymes are regulated by the Crp-cAMP complex: *gltA*, *acnB*, *acnB*, *acnA*, *sucABCD*, *lpdA*, *sdhCDAB*, *fumB* and *mdh* (several references are given in the EcoCyc database, (Keseler et al., 2009). Furthermore it is the only known coordinative positive regulator of these genes. Finally, an increase of the protein levels of GltA (citrate synthase), SdhA (succinate dehydrogenase subunit) and FumA (fumarase A) was observed by Raman et al. (2005) in a glucose-limited fed-batch cultivation. The proposed regulation of the TCA cycle by the *crp* modulon is illustrated in Figure 2-5. A positive regulation of TCA in response to carbon limitation might counteract the aforementioned high energy consumption and is probably the main reason for the low biomass yield (Fig. 2-1c).

The fluxes in glycolysis decreased strongly in response to carbon limitation (Fig. 2-3). A candidate of choice for a coordinative negative regulation of the glycolytic genes might be the *cra* modulon. Several genes of glycolysis are repressed by the Cra protein: *pfkA*, *fbaA*, *pgk*, *pykF* (references in EcoCyc database (Keseler et al., 2009); *gapA* and *eno* (Shimada et al., 2005). Moreover, the Cra protein is known to repress transcription of the *crr* gene ($EIIA^{Glc}$, a component of the glucose specific PTS) and the *glk* gene (glucokinase). Finally, Siddiquee et al. (2004) determined lower activities of the enzymes PfkA, PykF and the PTS after glucose exhaustion in batch cultivations. Thus, the coordinative repression of glycolysis most probably is accomplished by the *cra* modulon as illustrated in Figure 2-5. This might prevent

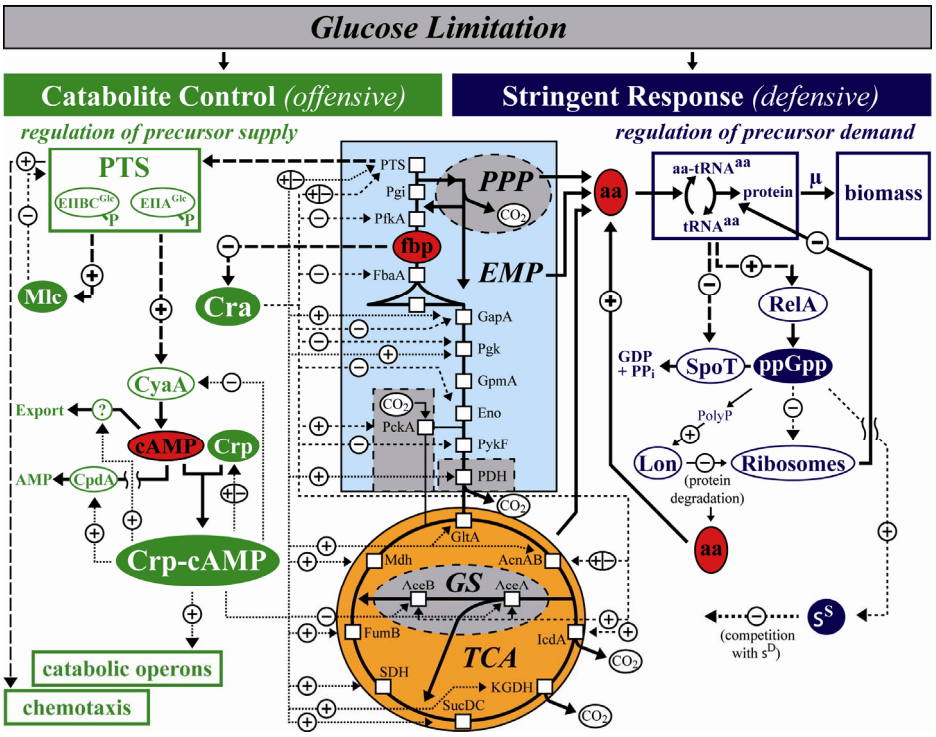


Fig. 2-5 Signaling and global genetic regulation of the central carbon metabolism in *Escherichia coli***.** Filled green and blue ellipses represent transcriptional regulators assigned to catabolite control or stringent response, respectively. Non-filled ellipses correspond to proteins (and ribosome) involved in signal transduction. Red colored ellipses represent intracellular signaling molecules, which are metabolites of the anabolic and catabolic network (*catabolites* and *anabolites*). Short dashed lines, transcriptional regulation; long dashed lines, regulation of enzyme activity. $+$, positive regulation, $-$, negative regulation, $+/-$, positive and negative regulation. **Catabolite Control.** Glucose limitation results in regulation of precursor and energy supply in the central metabolic pathways (EMP, PPP, TCA, GS) by the *cra* and *crp* modulons (*offensive strategy*): (i) The alarmone cAMP is synthesized by the adenylate cyclase (CyaA) in response to glucose limitation and the probable decrease of the metabolite pool of fructose 1,6-bis(phosphate) (fbp) renders the Cra regulator protein to be active. (ii) The Cra protein mainly represses the transcription of glycolysis (EMP) genes. This is proposed to ensure a sufficient carbon flow into the pentose phosphate pathway (PPP). (iii) The Crp-cAMP regulator complex mainly induces the transcription of the TCA cycle genes. Additionally the glyoxylate shunt (GS) genes are regulated by the modulons *cra* (positive) and *crp* (negative). This regulation results in an economic balance in the supply with precursors and energy. Moreover, catabolic operons and chemotaxis are activated in order to exploit potential new carbon sources. (iv) Several negative feedback regulation mechanisms (for details see text) lower the cAMP level after the initial accumulation, while cAMP export is the major one. **Stringent Response**. Carbon limitation also leads to the accumulation of ppGpp, which regulates precursor and energy demand in biosynthesis (*defensive strategy*): (v) It is suggested that carbon limitation results in amino acid limitation and that uncharged tRNAs activate ppGpp synthesis by the RelA protein and inhibit ppGpp hydrolysis by the SpoT protein. (vi) The precursor (particularly for amino acids) and energy demand are reduced following the lowering of protein biosynthesis (repression of ribosome synthesis) and growth-related functions. (vii) The refill of the amino acid pools (for details see text) exerts negative feedback to the central metabolic pathways since equilibrium between anabolism and catabolism is established. Consequently, the ppGpp level decreases. (viii) Further regulators might be affecting gene expression like competition of the sigma factor σ^S (stress-related genes) with σ^D (housekeeping genes).

that the carbon flow into the TCA cycle is too high and therefore, maintain a necessary minimal flux into the pentose phosphate pathway (see Fig. 2-3) providing still sufficient precursors for biomass synthesis. The effectors fructose 1-phosphate and fructose 1,6-bis(phosphate) (fbp) are known to inactivate the Cra protein (Ramseier et al., 1995; Ramseier et al., 1993). A decrease of the fbp level at lower growth rates was found by Jochen Schaub (personal communication) in a series of experiments with different dilution rates in continuous culture of *E. coli* K-12 W3110. Furthermore, an increase in the fbp level was shown during upshift from lower to higher growth rates in continuous culture (Weber et al., 2005b). Thus, the fbp concentration appears to reflect the availability of glucose and to mediate signal transduction to the Cra regulator (Fig. 2-5).

The branch point of the TCA cycle and the glyoxylate shunt (GS) is important for the control of the carbon flux and is both regulated by the Cra protein and the Crp-cAMP complex. The expression of *aceBAK* operon coding for the GS enzymes is activated by the Cra protein and repressed by the Crp-cAMP complex, whereas the *icdA* gene coding for the first TCA enzyme after the branch point is activated by the Cra protein (references in EcoCyc database (Keseler et al., 2009)). Additionally, the enzyme IcdA is regulated by phosphorylation/dephosphorylation (Cozzone and El-Mansi, 2005) and its expression is negatively affected by the sigma factor σ^S (Jung et al., 2006). While the TCA/GS split ratio is not available through stoichiometric metabolite balancing, a whole-cell microarray analysis, which will be extensively treated in a separate publication, showed that mRNA levels for GS enzymes increased. Activation of the GS provides sufficient precursors for biosynthesis and leads to a favorable carbon balance at the expense of the energy balance. Although the energy demand increases (m_S, Fig. 2-1c), this might be advantageous since altogether less energy is required during the slow fed-batch growth. In conclusion, the global regulator protein Cra possibly plays an important role both in regulation of glycolysis and the TCA/GS split ratio (see Fig. 2-5). Ramseier et al. (1996; 1995) suggested that the regulation by the Cra protein depends growth conditions and either affects the glycolysis or parts of the TCA cycle and the GS. However, it is more likely that both the Cra and Crp proteins coordinately regulate large parts of glycolysis, TCA and GS (see Fig. 2-5). This might ensure a good economic balance between precursor and energy supply during carbon-limited growth. As illustrated in Figure 2-5, catabolite control involves also the activation of various catabolic operons (*crp* modulon) and the regulation of chemotaxis. Obviously, the cells pursue an *offensive strategy* using these regulatory mechanisms in order to exploit the low amounts of available glucose and other potentially available carbon sources. There are also other global genetic regulatory systems that additionally influence the gene expression of the mentioned pathways, e. g. the negative regulation of TCA genes by sigma factor competition (Patten et

al., 2004) (Fig. 2-5) or regulation of the RNA polymerase availability by the stringent response (Barker et al., 2001a; Barker et al., 2001b; Cashel et al., 1996; Jensen and Pedersen, 1990; Traxler et al., 2006). However, the ones discussed here obviously affect metabolic flux distribution most significantly.

2.4.4 Regulation of Precursor Demand

A substantial decrease of fluxes in the biosynthesis pathways (Fig. 2-2c) was demonstrated during carbon-limited fed-batch growth concomitantly with the reduction of growth rate (Fig. 2-1b). Particularly, the efflux of precursors from the central metabolic pathways declined (Fig. 2-3). This wide-ranging dynamic metabolic response must be achieved mainly by regulation of gene expression since metabolic resources (precursors and energy) have to be saved during carbon limitation. It is obvious that the stringent response (*relA*/*spoT* modulon) must play the major role. The alarmone ppGpp was shown to accumulate strongly immediately after the onset of limitation and decreased to a low basal level thereafter (Fig. 2-4a). It is well known that the nucleotide ppGpp regulates the ribosome concentration (Cashel et al., 1996) and therefore also the biosynthesis rate of proteins. The reduction of the ribosome concentration might be regarded as a *defensive strategy*, because ppGpp thereby regulates the withdrawal of precursors (and energy) from the central carbon metabolism (see Fig. 2-5). However, although the signaling pathway of the stringent response is well known for amino acid limitation (Cashel et al., 1996), a respective pathway it is not yet clear during carbon limitation. Moreover, the time profile of the ppGpp concentration at low growth rates is not entirely clarified.

2.4.5 Signaling and Negative Feedback Regulation during Stringent Response

An initial drop of the fluxes in all amino acid biosynthesis pathways was determined after glucose limitation (Fig. 2-2c). Thus, for the particular point in time when the carbon source becomes limiting for growth (t = 0, Fig. 2-1a) it can be supposed that the amino acid supply does not meet the demand and consequently that the amino acid pools limit protein biosynthesis. Therefore, it might be concluded that the carbon limitation finally leads to an amino acid limitation. This requires that the signaling pathways (via the RelA and SpoT proteins) during carbon and amino acid limitation are similar. Several previous findings support this hypothesis: First, Murray and Bremer (1996) found that inhibition of the SpoT hydrolase activity is primarily responsible for the accumulation of ppGpp as long as growth was limited by glucose. Second, the SpoT hydrolase activity was found to be inhibited by

uncharged tRNA (An et al., 1979; Richter, 1980). It has also been shown in *relA*$^-$ strains that carbon limitation leads to the accumulation of ppGpp in minimal media only when not supplemented with amino acids (Gentry and Cashel, 1996). From these findings it can be concluded that the SpoT protein is mainly responsible for ppGpp accumulation during carbon limitation and that the signal transduction might occur via amino acid limitation. Finally, although ppGpp accumulated in response to limited amounts of carbon in mutant strains lacking either the RelA, SpoT hydrolase or SpoT synthetase activity (Gentry and Cashel, 1996; Hansen et al., 1975), the respective ppGpp levels were never as high as in wild-type strains. These findings offer the simple model structure given in Figure 2-5. It is suggested that both the functional RelA and SpoT proteins are essential for normal ppGpp accumulation, and that uncharged tRNAs are the main effectors for the regulation of their activities. In this way, carbon limitation could be indicated by low pools of amino acids as the metabolic signals and would finally lead to the stringent response. Following the nomenclature of *catabolites* (catabolite control), the amino acids could be termed *anabolites* since they are metabolites from the anabolism (see Fig. 2-5), which serve as the internal signal of an imbalance between catabolism and anabolism.

Following the initial accumulation of the alarmone ppGpp, its level decreased during fed-batch growth (Fig. 2-4a). Similar time profiles of the ppGpp concentration during the transition from exponential to slow growth were shown in batch (Cashel et al., 1996) and in fed-batch experiments (Teich et al., 1999). However, Teich et al. (1999) also concluded from a series of dilution rate shift experiments in continuous culture – shifting from high to low growth rates – that the ppGpp concentration increases with lower growth rates. Basal concentrations (below 0.2 µmol g^{-1}) in the steady state after the shifts were quoted for this conclusion and the authors mentioned that initial accumulation of ppGpp could not be observed during the shift experiments at low growth rates. Therefore, it is obvious that in the aforementioned shift experiments a very fast response in ppGpp accumulation occurred (as described in the present work) and that the levels found in the steady state are basal levels after the initial accumulation. It can be concluded that after reaching the peak value the resetting of the signal to low basal levels occurs independent of the mentioned experimental designs (batch, fed-batch, continuous). This adaptive behavior might prepare the cells for the response to possible further limitation or stress conditions. However, it requires negative feedback mechanisms. A large network of possible reactions leading to the reset of the signal can be inferred from previous findings (see also Fig. 2-5): It is well known that the ribosome concentration is regulated by ppGpp. The protein biosynthesis rate and the demand for amino acids are thereby reduced (Fig. 2-3 and 2-2c). Additionally, amino acids are supplied through an enhanced protein degradation (St John and Goldberg, 1980).

Particularly the degradation of ribosomal proteins by the Lon protease was suggested to be enhanced via a ppGpp-dependent PolyP (polyphosphate) accumulation (Kuroda, 2006). The resulting higher levels of amino acids and thus of charged tRNAs diminish ppGpp synthesis by the RelA protein (Wendrich et al., 2002) and enhance the SpoT hydrolase activity (An et al., 1979; Richter, 1980). Consequently, the ppGpp level declines as observed during fed-batch cultivation (Fig. 2-4a). Finally, the reduction of the ribosome concentration serves as a negative feedback regulation since the number of the RelA binding sites on the ribosomes (Wendrich et al., 2002) also decreases. The described model structure (Fig. 2-5) allows explaining the resetting the intracellular ppGpp level and the restoration of equilibrium between precursor supply in central carbon metabolism and precursor demand in the protein biosynthesis (see Fig. 2-2c). This is an important (secondary) regulatory effect of the stringent response as it applies feedback to the central metabolic pathways during growth-limiting conditions.

2.4.6 Signaling and Negative Feedback Regulation in Catabolite Repression

The demonstrated accumulation of the intracellular cAMP concentration at the beginning of glucose limitation (Fig. 2-4b, lower graph) obviously results from the well known activation of the adenylate cyclase (CyaA) (Lengeler et al., 1999; Postma et al., 1993). Subsequently, the resetting of the signal to a low basal level was observed within 4 h of fed-batch growth. The importance of the cAMP signal for the regulation of the central carbon metabolism was already pointed out (*crp* modulon). In the following, the main components of the regulatory network contributing to the observed signal reset are discussed. There are at least five negative feedback regulation mechanisms, most of which involve gene products of the *crp* modulon (see also in Fig. 2-5):

(i) cAMP synthesis is negatively regulated by Crp-cAMP-dependent repression of transcription of the *cyaA* gene (adenylate cyclase).

(ii) cAMP degradation is positively regulated through activation of the transcription of the *cpdA* gene (cyclic 3',5'-AMP phosphodiesterase) by the Crp-cAMP complex, even though this is not a strong regulation (Zheng et al., 2004).

(iii) The repressor protein Mlc is bound to the $EIIBC^{Glc}$ protein of the PTS. The phosphorylated $EIIBC^{Glc}$ protein releases the Mlc protein, which represses the transcription of the genes for the PTS (*ptsHI-crr* and *ptsG*) (reviewed in Plumbridge, 2002). This results in a lower concentration of the $EIIA^{Glc}$ protein and therefore in minor activation of the cAMP-synthesizing enzyme CyaA. However, the PTS genes

are regulated by the Cra protein and the Crp-cAMP complex either positively or negatively (Keseler et al., 2009; Ramseier et al., 1993), depending on the conditions.

(iv) Analogously, the transcription of the *crp* gene is regulated either positively or negatively by the Crp-cAMP regulator complex (EcoCyc database) depending on the intracellular cAMP concentration (Kremling and Gilles, 2001).

(v) Most importantly, cAMP is exported in large quantities from the cells (Matin and Matin, 1982). Interestingly, Bhatnagar and Bhattacharya (1984) proposed that the number of cAMP carrier molecules for both uptake and export is regulated by the Crp-cAMP complex.

It has been mentioned, the alarmone cAMP is continuously synthesized and immediately exported. A strong export was observed particularly at the onset of limitation (Fig. 2-4b). It can be concluded that cAMP export is the major negative feedback regulation mechanism since this is a very fast-acting opportunity for resetting the intracellular signal compared to the aforementioned mechanisms that require transcriptional regulation. It might also be conceivable, that the uptake of extracellular cAMP attenuates the feedback regulation. However, the import is negligible in the experiments reported here. This could be ascertained by the dynamic simulation of the intra- and extracellular cAMP concentrations (data not shown). Furthermore, the question arises whether the low intracellular cAMP level (Fig. 2-4b, lower graph) can still be sufficient to activate the transcription of the *crp* modulon. Various additional mechanisms may be involved supporting the expression of the *crp* modulon, e. g. lower chromosome number at lower growth rates (Murray and Bremer, 1996) and an altered availability of RNA polymerase (Barker et al., 2001a; Barker et al., 2001b; Cashel et al., 1996; Jensen and Pedersen, 1990; Traxler et al., 2006).

2.4.7 Stoichiometric Metabolite Balancing

The method of stoichiometric metabolite balancing applied in this work requires assumptions regarding stoichiometry and cofactor usage. It is known that, for instance, the use of NADPH as cofactor in the reaction of the enzyme isocitrate dehydrogenase (IcdA) affects the outcome of the flux split ratio between EMP and PPP (Schmid et al., 2004; Schmidt et al., 1998), which was found to be 42 : 56 at a growth rate of μ = 0.08 h^{-1}. A split ratio of 55 : 44 was determined by Schaub et al. (J. Schaub, personal communication) in continuous culture of *E. coli* K-12 W3110 (μ = 0.1 h^{-1}) and a split ratio of 54 : 44 was found at μ = 0.09 h^{-1} in *E. coli* JM101 by Emmerling et al. (2002). The latter metabolic flux analyses were based on ^{13}C-labelling experiments, which do not rely on the mentioned assumptions. However, the

conclusions drawn in the present study concern the dynamic changes in metabolism. Trends in flux distributions were found to be similar or even more pronounced when using a different cofactor in the IcdA reaction (Table 2-3). Remarkably, the fluxes in the TCA cycle were even higher at low growth rates when using an NADH-dependent IcdA enzyme. This can be attributed to the decoupling of the TCA cycle from the NADPH consuming biosynthesis reactions, which have low rates (Fig. 2-3). Thus, the TCA fluxes should stay constant, or even increase *in vivo*, when the growth rate decreases. Ongoing work is directed at the confirmation of the results by isotopic instationary ^{13}C metabolic flux analysis. This has not yet been applied in fed-batch experiments due to the given metabolic instationarity.

2.5 Conclusions

This work presents a systems-level analysis of the dynamic cellular response to carbon limitation. In accordance with the large experimental data set obtained from metabolic flux analysis the novelty of the work lies not in the discussion of the molecular details of regulatory mechanisms, but in the derivation of a comprehensive model structure comprising the most relevant components – the *cra*, *crp* and *relA*/*spoT* modulons – that can explain the observed dynamic metabolic and regulatory response including the resetting of the signals. According to the model the regulation of glycolysis, the TCA cycle and the glyoxylate shunt by the *cra* and *crp* modulons results in the balancing of the supply with precursors and energy. The *relA*/*spoT* modulon re-establishes equilibrium between the supply and demand of precursors and energy. Moreover, a new method for the qualitative estimation of the intracellular cAMP level was applied and the time course could be determined for the first time in *E. coli* K-12 fed-batch cultivations. A further novel result is the observed resetting of the alarmone levels after adaptation to carbon limitation. It could be clarified that the ppGpp concentration does not increase generally at lower growth rates as concluded by previous authors, but decreases fast to a basal level after the initial accumulation. The resetting of the signals after adaptation can be explained with the model, which reveals the key cell components assembling negative feedback loops. Finally, the signaling pathway of the stringent response was suggested to occur via amino acid limitation and therefore via both the RelA and SpoT protein. The proposed model structure represents the first step towards dynamic mathematical modeling of the central carbon metabolism and its regulation during fed-batch cultivation, which is in the focus of ongoing, challenging systems biology research.

3 *In vitro* Synthesis and Characterization of Guanosine 3',5'-bis(diphosphate) (ppGpp)

Published in:

Hardiman, T., Windeisen, V., Ewald, J.C., Zibek, S., Schlack, P., Rebell, J., Reuss, M., and Siemann-Herzberg, M. (2008). Anal Biochem *383*, 337-339.

Abstract

The intracellular alarmone ppGpp has been thoroughly investigated over the last 40 years and has become one of the best-known effectors in bacterial physiology. ppGpp is also of great importance for biotechnological applications. Systems biology research, involving experimental and mathematical approaches, has contributed a great deal to uncovering the alarmone's complex regulatory effects. HPLC analysis and UV detection are used to quantify intracellular ppGpp. The samples analyzed also contain other phosphorylated guanine nucleotides and are therefore spiked with a standard ppGpp solution. A rapidly growing number of laboratories are turning to synthesizing the nucleotide *in vitro* involving time-consuming protocols and yielding only low amounts of ppGpp. The present work provides a protocol for the preparation of large quantities of a ribosome extract which contains high ppGpp synthesis activity. The demonstrated upscaling from shaking flask to bioreactor cultivation involves the continuous and refrigerated harvest of the biomass. ^{13}C NMR analysis enabled the structural characterization of the synthesis product and was complemented by mass spectrometry and methods that are commonly used to identify ppGpp.

3.1 Introduction

The tetraphosphorylated guanosine ppGpp was initially discovered in *Escherichia coli* where it acts as the intracellular mediator of the global regulatory system "stringent response". It is also important for the physiological adaptation of numerous bacteria and even plants to environmental stress (Braeken et al., 2006; Cashel et al., 1996). Recent research has focused on the dynamics of intracellular ppGpp concentration and the mathematical modeling of the global regulatory effects, which makes important contributions to the understanding of technical production processes, survival under environmental stress as well as pathogenetic and symbiotic aspects (Braeken et al., 2006; Cserjan-Puschmann et al., 1999; Hardiman et al., 2007; Lapin et al., 2006; Neubauer et al., 1995). ppGpp is usually quantified using HPLC analysis coupled with UV detection (Cserjan-Puschmann et al., 1999; Hardiman et al., 2007; Lin et al., 2004; Little and Bremer, 1982; Neubauer et al., 1995). Earlier publications applied 2D-TLC after radioactive labeling (Bochner and Ames, 1982; Cashel et al., 1969) and only one publication reports the structural identification using ^{13}C NMR (Que et al., 1973). Biological samples often contain many other compounds that are as equally well retained by HPLC columns as ppGpp and that also have similar absorption spectra; for example the pentaphosphorylated nucleotide pppGpp. To enable the reliable identification of ppGpp, the samples are usually spiked with a standard ppGpp solution. There is, however, only one commercial source of ppGpp and even low amounts would be very expensive. The nucleotide is therefore synthesized *in vitro* by numerous laboratories (Artsimovitch et al., 2004; Barker et al., 2001b; Cashel, 1974; Cserjan-Puschmann et al., 1999; Hardiman et al., 2007; Haseltine et al., 1972; Krohn and Wagner, 1995; Milon et al., 2006; Oki et al., 1975; Wendrich et al., 2002) involving time-consuming protocols and yielding only low amounts of ppGpp. After that, the ppGpp preparation is often shared with other laboratories (see e. g. in Lin et al., 2004; Neubauer et al., 1995). The synthesized product is subsequently isolated and then characterized with HPLC coupled with UV detection or using 2D-TLC.

The present study describes a protocol that enables the preparation of a large amount of a highly active ribosome extract, which can be used for the *in vitro* synthesis of ppGpp. The ribosome-containing biomass was produced in a 30 liter bioreactor. The upscaling of the preparation step of biomass production – from shaking flask to bioreactor cultivation – must ensure that during the biomass harvest process the ribosomal activity of the biomass is maintained. The work demonstrates that this can be achieved through the continuous and refrigerated harvest. The synthesis product was characterized by ^{13}C NMR analysis and mass spectrometry as well as commonly used methods.

3.2 Results and Discussion

The synthesis of ppGpp is catalyzed by the ribosome-associated RelA protein. High RelA-dependent ppGpp synthesis activity is found in ribosomal extracts isolated from *E. coli* cells (Cashel, 1974; Haseltine et al., 1972; Krohn and Wagner, 1995). The cells are harvested from shaking flask cultures during exponential growth (Cashel, 1974; Haseltine et al., 1972). Due to very low cell densities and small reaction volume available in a shaking flask, only a very small amount of ribosomal extracts can be produced. Therefore, a batch cultivation of the wild-type bacterial strain *Escherichia coli* K-12 W3110 (DSM 5911, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany) was performed in a 30 liter bioreactor (Bioengineering AG, Wald, Switzerland) in 20 l of Luria-Bertani medium and 10 g l^{-1} glucose to produce higher quantities of biomass. The process parameters were adjusted to the following values: dissolved oxygen concentration > 50 % saturation, pressure p = 1.5 bar, temperature T = 37 °C, pH 7, with 2 M H_3PO_4 or 2 M NaOH being used to adjust the pH value. The bioreactor was inoculated with a biomass concentration of c_X = 0.1 g l^{-1}. The cell harvest was started at c_X = 0.8 g l^{-1} and was completed before the end of the exponential growth phase (c_X < 5 g l^{-1}). It should be possible to achieve higher cell densities with higher glucose concentrations. However, it is recommended harvesting the cells during the unlimited growth phase to produce biomass containing a high ppGpp synthesis activity. Otherwise, only low ppGpp synthesis activity will be found in the final ribosomal extract (data not shown). The reason for the higher activity found in the ribosomal extract of rapidly growing cells might be that a large fraction of 70S ribosomes dimerize to 100S ribosomes, when cell growth is limited by the carbon source (Wada et al., 1990). 100S ribosomes show no translational activity (Wada et al., 1995). However, the translational activity of 70S ribosomes appears to be a prerequisite for the complex mechanism that leads to the RelA-mediated ppGpp synthesis (Wendrich et al., 2002) – a process that is not entirely clarified (Potrykus and Cashel, 2008; Srivatsan and Wang, 2008). The final ribosome extract of *E. coli* cells, harvested in the stationary phase, might therefore contain less catalytic activity, whereas information on how a high activity can be achieved is scarce.

Continuous harvesting (flow rate F = 0.22 ml min^{-1}) involved the immediate cooling to 4 °C using a plate heat exchanger (6X-7PI*5, Tranter AG, Hildesheim, Germany), centrifugation (20,000 rpm; CEPA High-Speed centrifuge Z 41/G, Padberg, Lahr, Germany), freezing of aliquots in liquid nitrogen and storage at -70 °C.

Table 3-1 Identification of ppGpp and previously published analytical parameters

Method	Parameter	Value	Reference
HPLC	r [a]	1.19	This work
HPLC	$t_{R,ppGpp} / t_{R,ATP}$	1.19	(Cserjan-Puschmann et al., 1999)
MS [b]	m / z	602	This work
Molecular weight	M_r	603	calculated
UV	$\lambda_{\max}$	253	This work
UV	$\varepsilon_{\lambda=253\,nm}$	12,700 l cm^{-1} mol^{-1}	(Cashel and Kalbacher, 1970)

[a] Relative retention, $r = t'_{R,ppGpp} / t'_{R,ATP} = k_{R,ppGpp} / k_{R,ATP}$, determined from the ratio of the adjusted retention time of ppGpp relative to that of ATP

[b] Fragmentation of the molecule ion (m / z = 602) predominantly resulted in a fragment ion (m / z = 504). $\Delta m / z$ = 98 was found to be specific also for the nucleotides ATP, ADP, GTP and GDP.

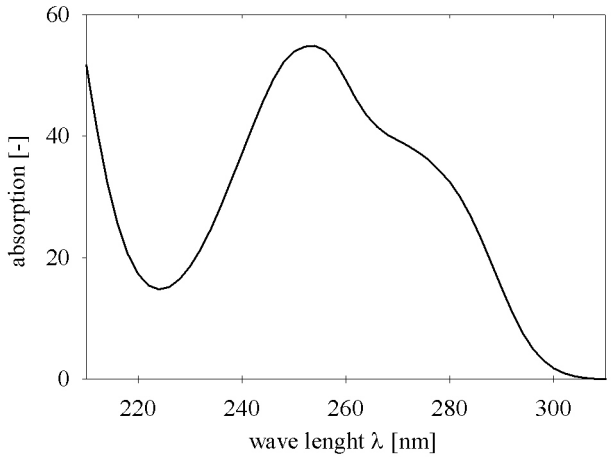


Fig. 3-1 UV spectrum of *in vitro* synthesized ppGpp, recorded during HPLC analysis. Composition of the solvent during analysis: 39.8 % buffer A (0.1 M KH_2PO_4/K_2HPO_4, pH 6.0; 5 mM tetrabutyl ammonium dihydrogen phosphate; 5 M EDTA; pH 5.3) and 60.2 % buffer B (80 % of buffer A; 20 % acetonitrile, pH 5.9).

Table 3-2 ^{13}C chemical shifts of the ribose carbons of guanosine and guanine nucleotides

Compound	δ (ppm)					Reference
	C-1'	C-2'	C-3'	C-4'	C-5'	
ppGpp	89.4	77.4	75.7	86.6	67.8	This work
GDP	89.7	76.5	73.2	86.8	67.8	This work
Guanosine	87.5	74.8	71.4	86.3	62.5	SDBS/RIO-DB (2008)
ppGpp	106.1	119.2	117.9	109.2	127.7	Que et al. (1973)
ppGpp [a]	89.8	76.7	78.0	86.7	68.2	Que et al. (1973)

[a] chemical shifts of Que et al. (1973) were transformed using the equation $\delta = \delta_{CS_2} - \delta_{analyte} + \delta_{medium}$, where the chemical shift given in the table, δ, is calculated from the chemical shift given by Que et al. (1973) ($\delta_{analyte}$, in ppm upfield from CS2) and the chemical shift of carbon disulfide (δ_{CS_2} = 192.58 ppm, SDBS/RIO-DB, (2008)). δ_{medium} = 3.3 ppm was introduced to account for the possible dependence of the chemical shifts on intermolecular interactions in the sample of Que et al. (1973) or in the sample of the current work, respectively.

The cell pellets obtained were extracted according to the protocol of Haseltine *et al.* (1972). This involves the lysis of cells in a French press, removing the cell debris and ultracentrifugation with a two-step sucrose gradient. The resulting ribosomal extract was used for the *in vitro* synthesis of ppGpp from GDP and ATP as previously described by Krohn and Wagner (1995). 90 units of ribosomes/ml (A_{260}) were the optimum amount for achieving a high yield of ppGpp per GDP employed (reaction conditions: 2 mM GDP, 4 mM ATP, 3 h, 27 °C on a shaker). A molar yield of 88 % was obtained, which is about 10 % higher than previously reported by Krohn and Wagner (1995). The new protocol therefore not only leads to high quantities of ribosomal extract, but also to a ppGpp synthesis activity that is at least as high as in the previously published protocols.

Table 3-3 ^{13}C chemical shifts of the guanine carbons of guanosine and guanine nucleotides

Compound	δ (ppm)					Reference
	C-2	C-4	C-5	C-6	C-8	
ppGpp	157.0	154.9	119.2	162.0	140.7	This work
GDP	156.9	154.7	119.2	161.9	140.6	This work
Guanosine	154.7	152.3	117.7	157.8	136.7	SDBS/RIO-DB (2008)
ppGpp	38.9	41.5	76.8	34.2	55.2	Que et al. (1973)
ppGpp [a]	157.0	154.4	119.1	161.7	140.7	Que et al. (1973)

[a] see footnote of Table 3-2.

The guanine nucleotide was purified according to the method of Krohn and Wagner (1995) using ion-exchange chromatography and subsequent gel filtration. The sample was subsequently analyzed using ion-pair reversed-phase HPLC with UV detection as described in Section 2.2.3 and published by Hardiman *et al.* (2007), 2D-TLC according to Bochner and Ames (data not shown) (1982), LC/MS^n (Finnigan LCQ^{Deca}, Thermo Electron Corporation, San Jose, CA) and ^{13}C NMR (ARX 500 spectrometer, Bruker, Rheinstetten, Germany). The results obtained and references are summarized in Table 3-1. The UV spectrum, recorded during HPLC/DAD analysis, shows the typical absorption maximum of guanosine nucleotides at a wave length of 253 nm (Figure 3-1 and Table 3-1). The recorded natural abundance ^{13}C NMR data (250 MHz, D_2O) are listed in Tables 3-2 and 3-3. To enable the comparison of the current analysis with the ^{13}C NMR chemical shifts obtained for a ppGpp preparation by Que *et al.* (1973), the data of Que *et al.* (1973) were transformed (see Tables 3-2, 3-3). The transformed chemical shifts agree with the shifts obtained with the ppGpp preparation of the present work – with the exception of C-2' and C-3' of the ribose carbons. The guanosine 5'-diphosphate (GDP; Fluka, Sigma-Aldrich Chemie GmbH, Buchs, Schweiz) employed as a reactant for *in vitro* synthesis was also analyzed (Tables 3-2, 3-3). The chemical shifts for guanosine are given as additional references (Tables 3-2, 3-3). The ^{13}C chemical shifts of the ppGpp preparation of the current study correspond to the spectra of GDP and guanosine (Tables 3-2, 3-3). In contrast to C-3' of GDP, the diphosphate group attached to carbon 3' of ppGpp leads to a downfield shift of 2.6 ppm (Table 3-2). Similar downfield shifts of 2.5 to 3

ppm were previously reported by Uesugi *et al.* (1978) and Hesbain-Frisque *et al.* (1981) for the phosphorylation of C 3' and 2' of adenine nucleotides and fructose. Mantsch and Smith (1972) observed that the chemical shift of the C-3' of uridine increases by 2.6 ppm after phoshporylation. This suggests that the ^{13}C chemical shifts of δ = 77.4 ppm and 75.7 ppm may be assigned to the carbons C-2' and C-3', respectively, of the chemical structure of guanosine 3',5'-bis(diphosphate) (Tables 3-2, 3-3). It should be mentioned, however, that additional proton resonance spectra recorded for GDP and ppGpp indicated that there were different proportions of the anomers in the reactant and the synthesis product.

3.3 Conclusions

Numerous laboratories studying the regulatory effects of the intracellular alarmone ppGpp in bacteria and plants spend a lot of time and material on the *in vitro* synthesis of the nucleotide. To increase the amount of ppGpp per preparation period, the upscaling from shaking flask to bioreactor cultivation, and consequently, the set-up of a biomass harvesting technique that preserves a high ppGpp synthetic activity in the final ribosomal extract, are needed. The protocol proposed in this work involves the cultivation of cells in a 30 liter bioreactor and the continuous and refrigerated harvest of the biomass, providing a large amount of ribosomal extract with high ppGpp synthesis activity. This increases the productivity of the preparation process, i. e. the amount of ribosomal extract obtained per preparation period. This is an important improvement since the ribosomal extract can be used for only one *in vitro* synthesis reaction (one batch) and cannot be recycled. The structures of the synthesized products are not usually analyzed. However, due to the presence of intracellular compounds with similar physical and chemical properties, it is strongly recommended using NMR spectroscopy or equivalent structural analyses such as LC-MS/MS for the unequivocal identification of ppGpp. LC-MS/MS analysis of other nucleotides has already been shown (e. g. in Luo et al., 2007) and could possibly be used to resolve the signals of the tetra- and pentaphosphorylated nucleotide. Reference data required for the identification of the alarmone are provided in this study, including the assigned ^{13}C chemical shifts and the characteristic fragment ions for the possible LC-MS/MS analysis.

4 Global Transcription and Metabolic Flux Analysis in Glucose-Limited Fed-Batch Cultivations

Parts of this chapter are published in:

Lemuth, K., Hardiman, T., Winter, S., Pfeiffer, D., Keller, M.A., Lange, S., Reuss, M., Schmid, R.D., and Siemann-Herzberg, M. (2008). Appl Environ Microbiol *74*, 7002-7015.

Abstract

A time series of whole-genome transcription profiling of *E. coli* K-12 W3110 was performed during a carbon-limited fed-batch process. The transcription data described in this Chapter are results of a collaboration with the Institute of Technical Biochemistry (University of Stuttgart) and were published by Lemuth et al. (2008). The microarray analysis was performed by Karin Lemuth as part of her PhD thesis (Lemuth, 2006).

The application of a constant feed rate led to the identification of a dynamic sequence of diverse carbon limitation responses (e.g. the hunger response) at the same time as providing a global view of how cellular and extracellular resources are used: the synthesis of high-affinity transporters guarantees maximal glucose influx, thereby preserving the phosphoenolpyruvate pool. Energy-dependent chemotaxis is reduced in order to provide a more economic "work mode". σ^S-mediated stress and starvation responses were both found to be of only minor relevance. Thus, the experimental setup provided access to the hunger response and enabled the differentiation of the hunger from the general starvation response. The topological model of the global regulation of the *E. coli* central carbon metabolism through the *crp*, *cra* and *relA*/*spoT* modulons (proposed in Chapter 2) is supported by correlating transcript levels and metabolic fluxes, and can now be extended. The substrate is extensively oxidized in the TCA cycle to enhance energy generation. However, the general rate of oxidative decarboxylation within the pentose phosphate pathway and the TCA cycle is restricted to a minimum. Fine regulation of the carbon flux through these pathways supplies sufficient precursors for biosyntheses. The pools of at least three precursors are probably regulated through activation of the (phosphoenolpyruvate-)/glyoxylate shunt. The present work shows that the detailed understanding of the genetic regulation of the bacterial metabolism provides useful insights for manipulating the carbon flux in technical production processes.

4.1 Introduction

A wealth of information is available on transcriptome responses in *Escherichia coli,* which are triggered through various stress conditions, including the limitation of energy and carbon sources. However, surprisingly little is known about the dynamic variation of gene expression under the physiological conditions that are required for technical production processes. Knowledge of this, however, is of great importance because *E. coli* has become the most widely used prokaryotic system for the production of heterologous proteins as well as for the industrial production of bacterial metabolites. Batch and fed-batch operations are the major cultivation strategies used for this purpose. For large-scale applications, fed-batch, high cell density *E. coli* K-12 derivative cultivation strategies have proven suitable for considerably increasing the volumetric productivity of these processes (Lee, 1996; Yee and Blanch, 1992). Irrespective of more sophisticated closed-loop strategies, fed-batch cultivations are usually carried out with open-loop control via exponential or constant feeding. Exponential feeding maintains the specific growth rate at a constant level. The maximal biomass concentration that can be achieved with this strategy depends on sufficient oxygen supply and heat transfer capacities. At a constant feed rate, the specific growth rate gradually decreases due to declining carbon and energy source levels. The proceeding carbon limitation also leads to a range of serious starvation phenomena with manifold regulatory responses of the cells. These responses macroscopically manifest themselves in a loss of viability, such as was previously illustrated by Hewitt et al. (2000).

Numerous experimental studies on transcription profiling have been carried out to characterize *E. coli* physiologically, in which a major focus has been put on the high cell density cultivation with an exponential feed rate, growth on different substrates and influence of regulatory proteins, diauxie or starvation [an overview was given by Lemuth et al. (2008)]. In spite of many physiological effects entailed by limited carbon concentrations, little is known about the thorough dynamics of regulatory events occurring in response to the proceeding carbon limitation during constant feeding conditions. Such knowledge, however, is essential for gaining a better understanding of the dynamic adaptation phenomena of *E. coli* during production processes. In the majority of investigations dealing with heterologous protein production, transcriptome data reflect a superposition of the effects of carbon limitation and foreign protein production. Therefore, valuable information on carbon limitation may be masked by other stress responses. A deeper insight into important regulation phenomena and the changes in metabolism in response to carbon limitation can therefore only be gained with investigations involving wild-type strains. Microarray data obtained from such

cultivations should also be directly compared with transient variations of the metabolic fluxes to support hypotheses concerning the regulation of metabolic reactions.

One of the key regulation phenomena related to the onset of carbon limitation is what is known as stringent response, which is mediated by guanosine 3',5'-bis(diphosphate) (ppGpp) (Cashel et al., 1996), a nucleotide derivative that affects the affinity of RNA polymerase to different promoters and hence transcription (Magnusson et al., 2005). Elevated ppGpp concentrations lead to the reduced synthesis of ribosomal proteins, stable RNAs (tRNA and rRNA) and biosynthesis enzymes for fatty acids and lipids, as well as proteins involved in DNA replication. On the other hand, ppGpp is a positive regulator of the alternative sigma factor synthesis (σ^S, encoded by the *rpoS* gene) and – presumably as a secondary effect – of amino acid biosynthesis (Barker et al., 2001a; Barker et al., 2001b; Cashel et al., 1996; Gentry et al., 1993; Loewen and Hengge-Aronis, 1994).

At low glucose concentrations, the limited amount of energy will be exploited as efficiently as possible by activating high-affinity glucose transport systems and by tapping alternative carbon sources (Ferenci, 2001; Ihssen and Egli, 2005). This response is, among others, mediated by the *crp* modulon and occurs at the transition from the exponential growth to the stationary phase (Ferenci, 1996). The *E. coli* fructose repressor, FruR (Cra), modulates the direction of the carbon flow by repressing the genes involved in fermentative carbon flow and by activating the enzymes involved in oxidative and gluconeogenic carbon flow (Ramseier et al., 1995). Submicromolar glucose concentrations (below 0.1 µM (0.02 mg l^{-1})) induce starvation responses that are primarily mediated by σ^S, which binds to RNA polymerase and leads to higher stress tolerance levels (Ferenci, 2001; Hengge-Aronis, 2002), including the resistance to stress factors such as H_2O_2, oxygen radicals, drought, acidic/basic pH, osmotic stress and ethanol as well as heat and cold (Wick and Egli, 2004).

However, little is known about the changes of the central carbon metabolism of *E. coli* K-12 wild-type cells grown under glucose-limiting conditions, nor is the chronological sequence of the aforementioned regulatory responses known in detail. One reason for this is the quick succession of the different glucose limitation stages in batch cultivations. Rapidly declining glucose concentrations lead to temporary alterations in the transport activities which are difficult to investigate under these experimental conditions (Ferenci, 1999b). Continuous cultivations involving wild-type and mutant *E. coli* strains, however, allowed the adjustment of micromolar glucose concentrations and dilution rates and were therefore assumed to generate an exactly defined physiological steady-state that was stable over a long period of time. Nevertheless, subsequent experiments revealed that continuous cultivation conditions led to changes on the transcriptome and proteome level (Kovarova-Kovar and Egli, 1998;

Wick et al., 2001). In addition, the long-term carbon limitation during continuous culture conditions led to genomic mutations (Notley-McRobb and Ferenci, 1999).

To shed further light on the sequence of bacterial responses induced at the transition from millimolar to submillimolar glucose concentrations, fed-batch cultivations of E. coli K-12 W3110 were performed at a constant feed rate. As shown in Section 2.3.2, flux redistribution in the central carbon metabolism during carbon-limited growth results in a significantly lower biomass yield (Hardiman et al., 2007). The coordinate regulation of the expression of many genes encoding enzymes of the central carbon metabolism was proposed as the most relevant process governing the observed behavior. In the current work, DNA microarrays were used to gain a holistic view of the dynamic changes occurring on the transcriptome level. A major focus was not only put on the time sequence of rearrangements of various cellular functions (transport, central carbon metabolism, growth, chemotaxis, stress and starvation response) but also on discussing the effects of these changes on the availability of resources for growth and maintenance. Thus, the work provides a comprehensive overview of the potentially critical responses that can be expected during carbon-limited biotechnological processes, and which may have to be taken into account when rationally (i. e. through dynamic mathematical modeling) improving bacterial producer strains.

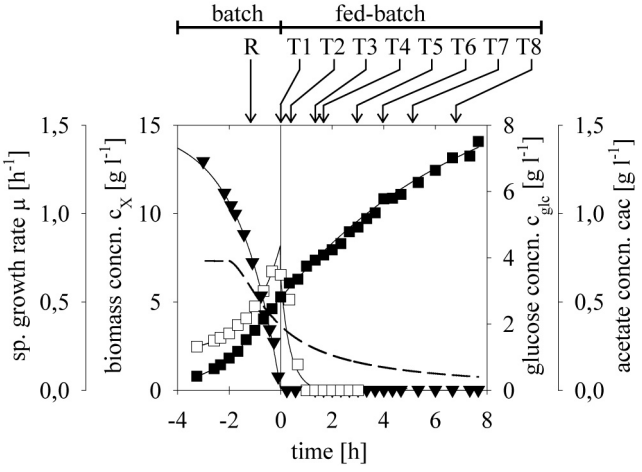


Fig. 4-1 Glucose limited fed-batch cultivation of *E. coli* K-12 W3110 with constant feed rate and sampling for transcriptome analysis. The vertical solid line at t = 0 indicates glucose limitation (as judged from the dissolved oxygen concentration time course (data not shown)). The concentrations of biomass (■), glucose (▼) and acetate (□) are given as well as the time course of the specific growth rate (μ) (broken line). Arrows above the graph indicate the time when the samples were removed for microarray analysis (R, reference; T1 to T8, time series samples).

4.2 Material and Methods

Three independent fed-batch cultivations of *E. coli* K-12 W3110 (DSM 5911) were carried out in a 30-l bioreactor as described in Section 2.2.1 (Hardiman et al., 2007). The transcriptome analysis using DNA microarrays comprising 50mer oligonucleotides of all *E. coli* K-12 ORFs was performed by Lemuth et al. (2008). A sample of the unlimited growth phase from the batch phase of the cultivation process was chosen (see Figure 4-1 for details) as reference (R). This was hybridized on an array together with the sample of interest (T1 – T8) while different labeling agents were used (Cy3 and Cy5). This allowed the direct comparison of the transcripts of a time point of the series taken during fed-batch cultivation with the unlimited reference sample. Further experimental details were published by Lemuth et al. (2008). The data discussed in the present thesis have been deposited in NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through the GEO Series accession number GSE10307 (Barrett et al., 2007; Edgar et al., 2002). Additionally, Quantitative PCR was performed for selected genes (GEO, GSE10307) (Lemuth et al., 2008). Results were in good agreement with the microarray data. The metabolic flux analyses were performed by stoichiometric metabolite balancing of five independent fed-batch cultivations as described in Section 2.2.5 (Hardiman et al., 2007).

4.3 Results and Discussion

4.3.1 Fed-Batch Cultivations

Three independent fed-batch cultivation experiments applying a constant feed rate were performed to study the global physiological response of *E. coli* on gradually decreasing glucose concentrations. Technical details of these cultivations were described in Section 2.2.1 (Hardiman et al., 2007). The extracellular glucose, acetate and biomass concentrations are depicted in Figure 4-1. In the batch phase ($t < 0$, Figure 4-1), acetate accumulated during the consumption of glucose (metabolic overflow). The fed-batch phase was started upon glucose limitation at the end of the batch phase ($t = 0$, Figure 4-1). The accumulated extracellular acetate was consumed at the beginning of the subsequent glucose-limited fed-batch phase ($t > 0$, Figure 4-1). The constant feeding strategy applied led to a gradual decrease in the specific growth rate, μ (Figure 4-1). The qualitative time course of the extracellular glucose concentration, obtained using a Monod-type kinetic, and the experimentally determined growth rate under the assumption of constant substrate affinity (Monod constant, K_S = 0.05 g l^{-1} (Xu et al., 1999b)) throughout the cultivation process, is

also given in Hardiman et al. (2007). The concentration was below 120 mg l^{-1} when the feed was started and decreased to at least 6 mg l^{-1} (Hardiman et al., 2007). However, particular rearrangements in the transport systems, which will be illustrated later [Section 4.3.3, "Transport Systems (I)"], led to changes in substrate affinity. The extracellular glucose concentration was therefore expected to be considerably lower during fed-batch cultivation. The biomass yield, $Y_{X/S}$, decreased by 20 % during fed-batch growth, while the maintenance energy coefficient, m_S, showed a steep increase (Hardiman et al., 2007).

4.3.2 Experimental Design

A time series of a global transcription analysis was carried out for three independent glucose limited fed-batch cultivations. cDNA samples were generated from the phase of unrestricted growth (reference sample R, Figure 4-1; batch) as well as from the carbon-limited growth phase (samples T1 to T8, Figure 4-1; fed-batch). These were hybridized to whole-genome microarrays of *E. coli* K-12, in which each sample was co-hybridized with a reference state sample (R, Figure 4-1). The individual samples corresponded to the following process/physiological conditions: T1, glucose limitation (c_{glc} < 0.05 g l^{-1}); T2, glucose limitation, acetate concentration ≤ 0.35 g l^{-1}; T3, glucose limitation, 30 min after depletion of extracellular acetate; T4, glucose limitation, 50 min after depletion of extracellular acetate; T5 to T7, glucose limitation, one hour intervals corresponding to the specific growth rates of 0.16 h^{-1}, 0.13 h^{-1} and 0.11 h^{-1}; T8, glucose limitation, after seven hours of fed-batch growth (μ = 0.08 h^{-1}).

The transcription profile of 960 genes changed significantly in at least one point in time compared to the reference (GEO, GSE10307). 595 of these transcripts could be assigned to the physiological functions of the central carbon metabolism (glycolysis, Entner-Doudoroff pathway, pentose phosphate pathway, tricarboxylic acid cycle, glyoxylate shunt and respiration (58 transcripts)), transport (121 transcripts), anabolism (128 transcripts), catabolism and macromolecular degradation (51 transcripts), protein biosynthesis (81 transcripts), cell division (16 transcripts), stress response (33 transcripts), flagellar and chemotaxis system (17 transcripts), regulation (48 transcripts), as well as to other proteins (44 transcripts). The remaining 365 transcripts coded for hypothetical proteins or for proteins of unknown function. A general overview of the observed behavior is given in Figure 4-2, which also summarizes the information discussed in the following sections.

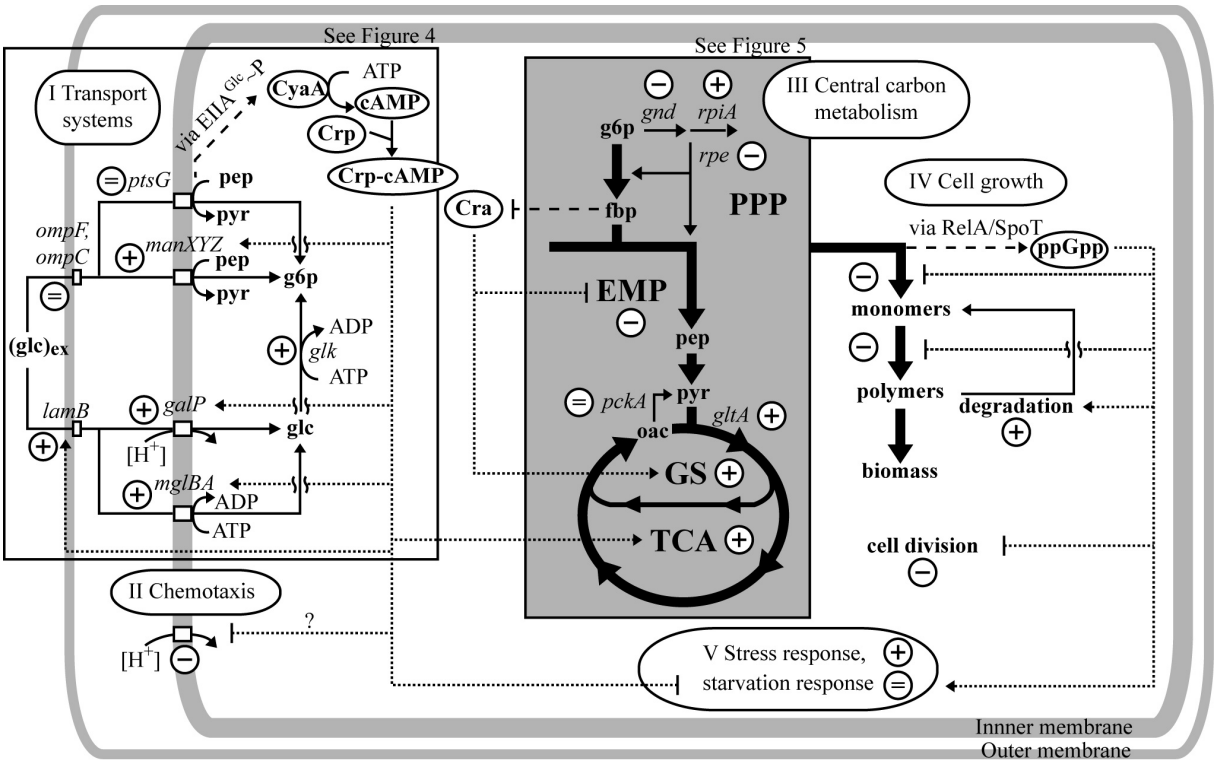


Fig. 4-2 Global regulation of the *E. coli* K-12 W3110 metabolism during carbon-limited growth, derived from a genome-wide transcriptome and metabolic flux analysis. The regulatory processes that are most relevant for utilization of available intra- and extracellular resources are proposed. mRNA/flux levels: +, increase; −, decrease; =, invariable. Short dashed lines, transcriptional regulation; long dashed lines, signaling processes (regulation of protein activities: arrowhead, positive; blunt end, negative). (I) A cluster of high-affinity transporters is synthesized (*mglBA, galP, lamB*), while the activity of medium-affinity transporters is maintained. This is mainly due to their regulation by the Crp-cAMP complex, but also due to the effect of the transcriptional regulatory proteins, MalT, YeeI and Mlc (Figure 4-3b, c). The glucose flux entering the cell is directed via transporters that do not use pep for phosphorylation. This preserves the pool of this metabolite (homeostasis) and affects the $EIIA^{Glc}$~P-dependent activation of cAMP synthesis through the enzyme adenylate cyclase (CyaA). (II) These transport systems in particular depend on a membrane proton gradient for proper function. The expression of the proton gradient-dependent chemotaxis system is reduced, thereby enabling the transport system effectively utilize the energy available. (III) The flux through the upper part of glycolysis is favored whereas the flux through the pentose phosphate pathway is minimized, which is most likely due to the reduced synthesis of *gnd* mRNA. The flux entering the pentose phosphate pathway is used for biosynthesis at the expense of the reflux into the glycolysis pathway, which might be regulated by the RpiA/Rpe split ratio. The reaction rates in the lower glycolysis decrease due to decreasing mRNA levels (*cra* modulon; signaling through fbp), thereby providing a sufficient, though minimal, efflux into the pentose phosphate pathway. The carbon flux entering the TCA cycle (influx is enhanced via *gltA* expression) is split into the glyoxylate shunt (GS), the pep-GS and the full TCA cycle. GS and pep-GS provide a better pep, pyr and oac precursor supply. It is assumed that the global regulation via the *crp* and *cra* modulons is the most relevant in this respect. (IV) Cellular growth is regulated predominantly by the stringent response (alarmone ppGpp, *relA*/*spoT* modulon). (V) No extensive induction of the general *rpoS*-dependent response could be observed (opposing regulation via the *crp* and *relA*/*spoT* modulons). It is expected that slow substrate concentration changes do not trigger a strong starvation response. However, other stress responses were detected. Details are presented in the text.

4.3.3 Transport Systems (I)

Existing transport systems involved in glucose uptake in *E. coli* have been reviewed by several teams of scientists (Ferenci, 1999a; Gosset, 2005; Postma et al., 1993). With respect to their affinity to glucose, these transporters are generally classified as *high*, *medium* and *low.* Moreover, the transporters are sometimes also classified according to their specific location (inner and/or outer membrane) or according to their specific mechanism (PTS-, ABC-, uni-, sym- or antiporter systems). Transport systems of high affinity (ABC-transport systems with low K_M values of < 10 µM), moderate affinity (pep [phophoenolpyruvate]-dependent PTS transport systems, K_M values of 10 – 1000 µM) and of lower affinity (e. g. symporters with low K_M values ranging between 0.1 – 10 mM) are found in the inner bacterial membrane (Ferenci, 1996; Postma et al., 1993). Several porins, with either high or low affinities, are located in the outer cell membrane (Ferenci, 1996).

Under glucose limited steady-state conditions, Ferenci (1996) found that *E. coli* predominantly expresses the high affinity transporters of the outer (e. g. porin LamB*)* and inner membranes (e. g. galactose ABC-transporter encoded by *mglBAC*). Such chemostat cultivations are of limited use for production purposes. Fed-batch strategies (either performed at constant or exponential feed rates) are often the method of choice. Since both strategies are based on limiting bacterial growth by varying the concentration of the substrate, it was important to further investigate whether a constant feed rate, which consequently results in decreasing glucose concentrations, would lead to a transportation behavior that was similar to that of chemostat cultures.

The observed dynamic response caused by this feeding strategy is summarized in Figure 4-3. The majority of transport systems involved in glucose uptake were differentially expressed during the fed-batch cultivation process. This was particularly so for the porin LamB (encoded by *lamB*; Figure 4-3a, b) which is a high–affinity glucose transporter located in the outer membrane. This particular porin was predominantly expressed during the first two hours of glucose limitation whereas the two other outer membrane porins encoded by the *ompF* and *ompC* genes were not affected (Figure 4-3b).

The antiport system, GalP (*galP*), and two genes of the galactose ABC transporter (*mglA* and *mglB*), which transport glucose with high affinity (Ferenci, 2001), were also predominantly expressed during the first two hours of glucose limitation (Figure 4-3a, b). The differential expression of *mglB* was confirmed by real-time PCR (GEO, GSE10307). The mannose PTS system (encoded by *manXYZ*), which transports glucose with moderate affinity (Plumbridge,

1998), was preferentially expressed between the time points T1 and T6. It can be further assumed that the glucose PTS system was active during the entire glucose limitation period, since none of the respective mRNA levels (*crr*, encoding $EIIA^{Glc}$, *ptsG*, encoding $EIIBC^{Glc}$) decreased at all. It is interesting to note that the aforementioned differentially expressed ABC-transport systems and the GalP antiport system are unable to directly phosphorylate glucose during transport (Figure 4-3b). In accordance, the glucokinase-encoding gene (glucokinase is a protein that converts glucose to glucose-6-phosphate (Meyer et al., 1997)) was differentially expressed between T2 and T4.

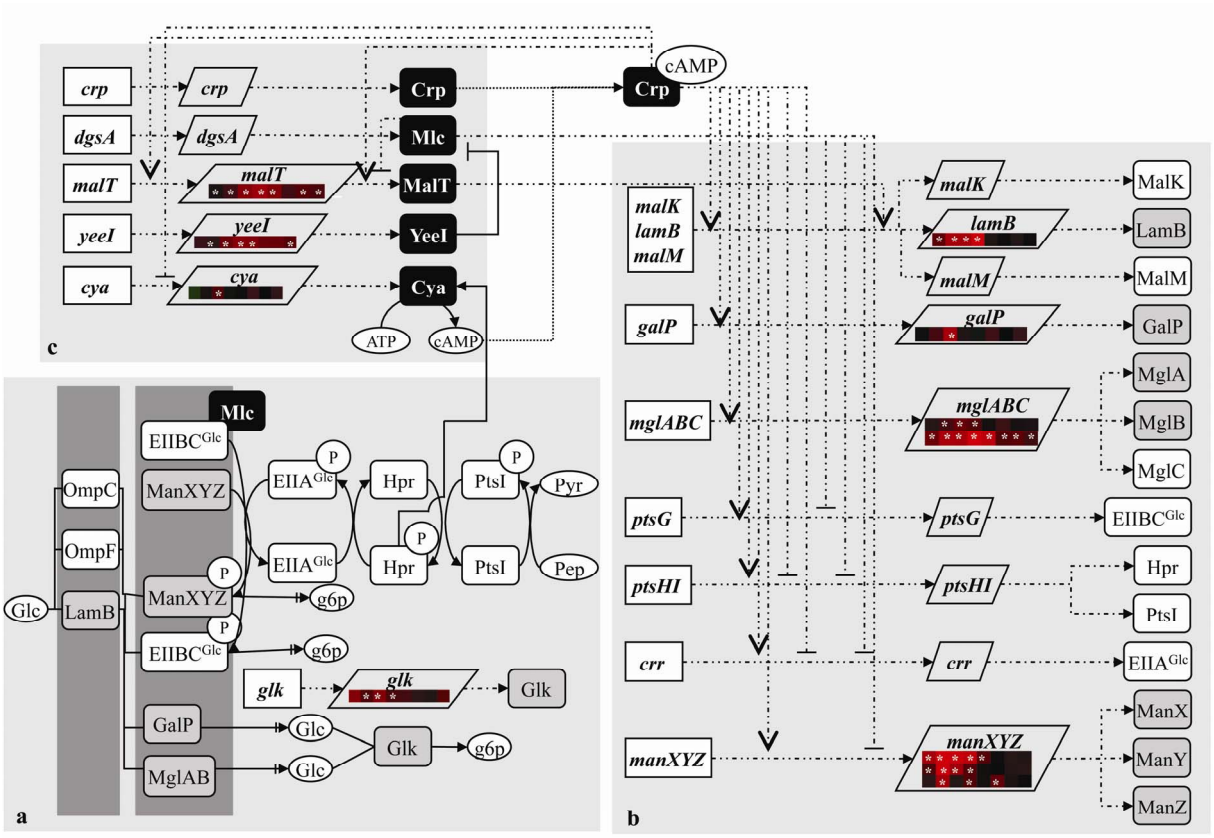


Fig. 4-3 Dynamic changes in transport RNA levels and their regulation in *E. coli* K-12 W3110 during glucose-limited fed-batch growth applying a constant feed rate. A: Sugar transport systems. B: Regulation of transporter gene expression. C: Expression of proteins relevant for the regulation of transporters. The time courses of the transcript levels are given for the samples T1 to T8 relative to the reference sample in the batch phase (R, see Figure 4-1). Green: mRNA level lower compared to the reference state. Red: higher mRNA level. Statistical significance ($p \leq 0.05$) is indicated by asterisks. Glc: glucose; P: phosphoryl group. Node symbols (states): rectangle, gene; parallelogram, RNA; rounded rectangle, protein; filled rounded rectangle (black), regulator protein; filled rounded rectangle (grey), protein with differentially expressed mRNA. Arrow symbols: solid line, regulatory interaction; dash-dot-dot-dash line, transcription; dash-dot-dash-dot line, translation. Arrowhead symbols: filled arrow, transformation; blunt end, inhibition or repression; open arrow, activation; filled arrow with crossbar: transport. [The transcription data of this Figure are results of a collaboration with the Institute of Technical Biochemistry (University of Stuttgart) and were published by Lemuth et al. (2008). The microarray analysis was performed by Karin Lemuth as part of her PhD thesis (Lemuth, 2006).]

The topology of the stimulus-response cascade of the glucose transport system, exhibiting various feed-forward loops and involving the alarmone cAMP as a major internal signal, can be deduced from the observed time course of transcript levels (Figure 4-3b, c). In Section 2.3.4 cAMP was demonstrated to be produced continuously when the external glucose concentration declined from 0.7 mM to at least µM concentrations at time point T8 (Figure 2-4) (Hardiman et al., 2007). All glucose transport systems identified are positively regulated by the Crp-cAMP complex (Figure 4-3b, c) (Boos and Shuman, 1998; Weickert and Adhya, 1993). Although the *crp* gene was not differentially expressed, the genes encoding the regulatory proteins MalT and YeeI (Figure 4-3c) were. MalT generally regulates the expression of transport systems (Cole and Raibaud, 1986) and underwent active transcription (Figure 4-3b, c). The Crp-cAMP complex enhances the expression of this gene (Figure 4-3c). Finally, MalT positively regulates the expression of transport proteins such as MalK, LamB and MalM (Boos and Shuman, 1998). LamB transcripts were differentially expressed (see Figure 4-3b). Mlc regulates several genes that are involved in glycolysis and glucose uptake. In particular, it represses genes of the glucose-specific PTS system (*ptsG*, *ptsHI-crr*), the regulatory protein MalT and the *manXYZ* operon (Bohm and Boos, 2004; Gerber et al., 2005; Plumbridge, 1998) (Figure 4-3b, c). The repressor protein Mlc interacts with the dephosphorylated PtsG protein (which is the $EIIB^{Glc}$ subunit of the glucose-specific PTS-transport system) at non-limited growth conditions and can repress MalT. In the present investigation, the glucose concentration seemed not to be low enough to lead to the inactivation of MalT (Figure 4-3b, c). Thus, the repressor remained inactive, which might explain the absence of further changes in the glucose-specific PTS-transcript levels during carbon-limited conditions at fed-batch cultivation. Recently, Jahreis and colleagues (Becker et al., 2006) identified the YeeI protein (now designated MtfL, Mlc titration factor) which, if present, interacts with Mlc, thereby leading to its inactivation. Elevated levels of YeeI-RNA transcripts were observed over the whole period (Figure 4-3c) analyzed. Therefore, Mlc seems to be under the dual control of YeeI and phosphorylated PtsG, two proteins that make sure that Mlc is inactivated under the examined growth conditions.

It seems that the inferred genetic regulatory network (Figure 4-3b, c) provokes the preferential expression of the high-affinity transporters GalP and Mgl, while, at the same time, medium-affinity transporters such as the glucose-dependent PTS system remain active (Figure 4-2). These physiological refinements are important for sustaining the maximal influx of glucose. In the present investigations, a gradually decreasing glucose flux was observed during the ongoing fed-batch process [see Section 4.3.4, “Central Carbon Metabolism (II)”]. This phenomenon is due to the constant feed rate chosen to technically control the glucose uptake rate.

The specific refinement of the transport system to rather high-affinity systems might also be of relevance in terms of homeostasis: The limited availability of glucose also reduces the metabolic flux through glycolysis (see below). Consequently, the pools of crucial metabolites such as pep, which plays an intermediate role between energy supply and anabolism, might drop to critically low concentrations. The refined (high-affinity) system predominantly consists of ABC-, sym- or antiporter proteins, which are altogether ATP-dependent, rather than of pep-dependent PTS systems of medium affinity (Figure 4-3a). It can therefore be assumed that the glucose flux is directed via ATP-dependent transporter systems and that pep mainly serves as precursor for anabolic purposes (Figure 4-2). Moreover, the pep pool is important for the Crp-cAMP complex-mediated regulation as it determines the phosphorylation state of the PTS system (Figure 4-3a). At high pep levels, cAMP synthesis, which is catalyzed by adenylate cyclase (CyaA), is activated *via* the phosphorylated protein $EIIA^{Glc}$~P (Figure 4-3a). As described in Chapter 2 and by Hardiman et al. (2007), a maintained pep pool, accompanied by the low flux via pep-dependent transporters, might thus lead to continuous cAMP synthesis (and export) during fed-batch growth (Figure 4-2).

In addition to the refined glucose uptake system, the expression of proteins that transport other substances indicates the bacteria's reorganization abilities, which adapt the *E. coli* metabolism to altered environmental conditions: in total, at least 121 transcripts of proteins involved in transport processes of sugars, amino acids, fatty acids and ions were differentially expressed in response to declining glucose concentrations (see supplemental material in Lemuth et al., 2008).

4.3.4 Central Carbon Metabolism (II)

As mentioned above, the dynamic behavior of the central carbon metabolism of *E. coli* K-12 and its regulation at glucose-limiting conditions are not yet understood in detail. Metabolic flux analyses of *Escherichia coli* K-12 W3110 showed that glycolysis and pentose phosphate pathway fluxes decreased strongly while the fluxes in the TCA cycle remained constant (Chapter 2) (Hardiman et al., 2007). This behavior was attributed to the regulation of the expression of many genes of the central carbon metabolism by the *crp*, *cra* and *relA*/*spoT* modulons. In the present investigation, the time courses of the transcript levels could be determined for the genes of the central carbon metabolism (Figure 4-2 and 4-4). These were compared with the metabolic fluxes and provided substantial evidence for the regulatory model structure suggested in Chapter 2 (Hardiman et al., 2007).

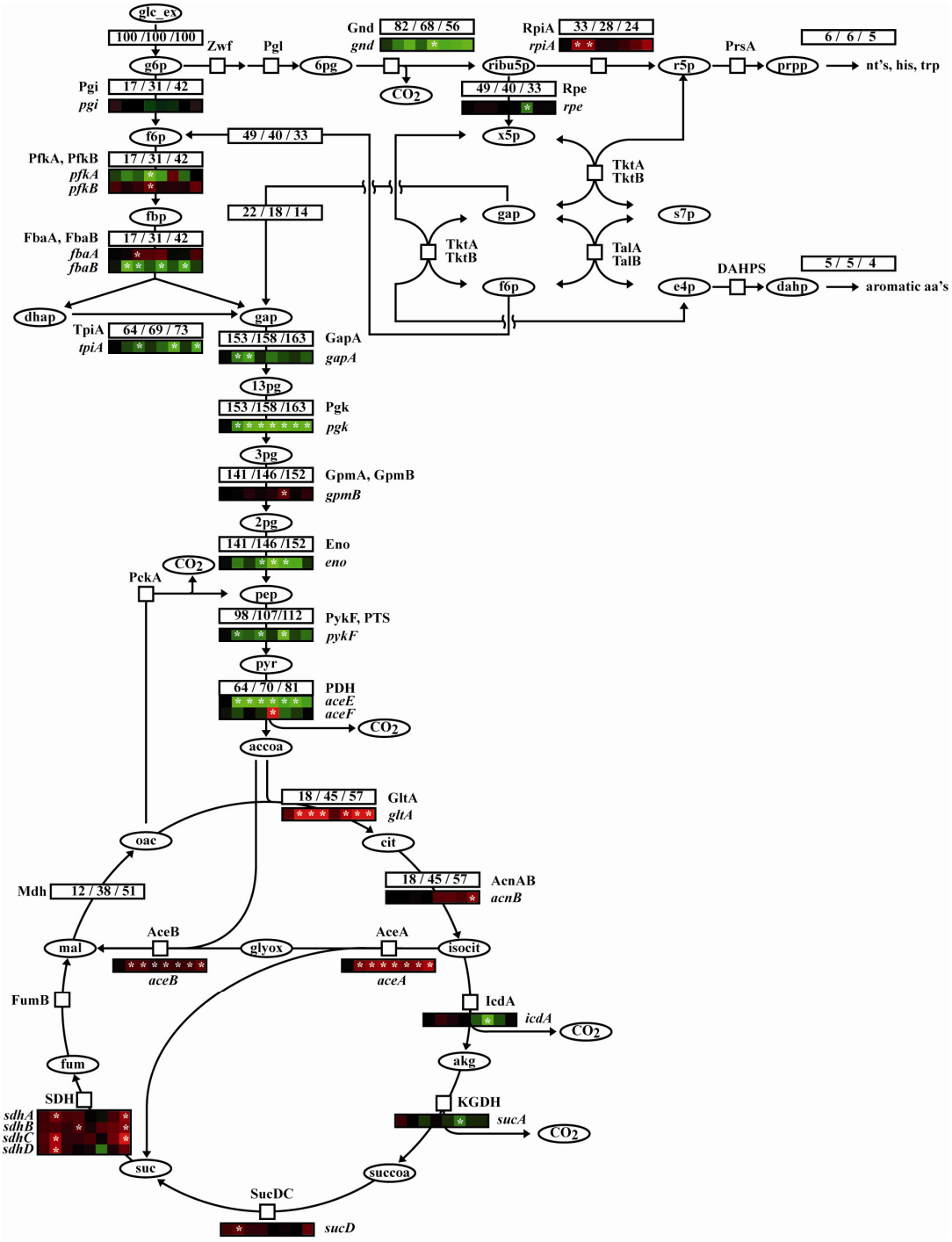


Fig. 4-4 Time series of DNA microarray and metabolic flux analyses of the central carbon metabolism in *E. coli* K-12 W3110 during glucose-limited fed-batch growth applying a constant feed rate. The time courses of the transcript levels are given for the samples T1 to T8 relative to the reference sample in the batch phase (R, see Figure 4-1). Green: mRNA level is lower than that at the reference state. Red: higher mRNA level. Statistical significance ($p \leq 0.05$) is indicated by asterisks. The metabolic fluxes are given for the following times: t = - 0.3 h (batch), t = 3.9 h and t = 7.7 h (fed-batch; see Figure 4-1). Fluxes are mean values from the stoichiometric metabolite balancing of five independent cultivations and are given as molar percentages of the glucose influx. Notation according to Appendix C (Hardiman et al., 2007). [The transcription data of this Figure are results of a collaboration with the Institute of Technical Biochemistry (University of Stuttgart) and were published by Lemuth et al. (2008). The microarray analysis was performed by Karin Lemuth as part of her PhD thesis (Lemuth, 2006).]

Glycolysis (EMP) and pentose phosphate pathway (PPP). Glucose influx decreased dramatically during fed-batch growth as a result of constant glucose feeding (Hardiman et al., 2007). In order to determine the changes in metabolic fluxes that could not be attributed to the general flux decrease in the network, the glucose influx was taken as reference value and set to 100 % at each point in time (Figure 4-4). In response to the limited carbon supply, the expression of the majority of the glycolysis transcripts decreased considerably, whereas the number of isoenzyme transcripts increased (*pfkA*, *fbaB* and *gpmB*, Figure 4-4). Apparently, enzyme levels are downregulated according to the decreased flux caused by constant feeding. Considering the reduction of the reaction rates, $r_i = r_{\max,i} \cdot f\left(c_j\right)$, it can be assumed that the downregulation of the enzyme levels (i. e. maximal reaction rates, $r_{\max}$) leads to constant metabolite concentrations, c_j (homeostasis). The reactions catalyzed by the PTS system, 6-phosphofructokinase (PfkA), pyruvate kinase (PykF) and pyruvate dehydrogenase (PDH) have high flux control coefficients (Chassagnole et al., 2002). Therefore it can be expected that the respective enzyme levels are regulated. It is known that most of the glycolysis genes that are less transcribed (Figure 4) are repressed by the global regulator protein Cra: *pfkA*, *fbaA*, *pgk*, *pykF*, *gapA* and *eno* (cf. EcoCyc database, Keseler et al., 2009; Shimada et al., 2005). Moreover, the enzymes PfkA and PykF were less active during glucose limitation (Siddiquee et al., 2004). The Cra protein is inactivated by the metabolite fructose 1,6-bis(phosphate) (fbp) (Ramseier, 1996), whose level decreases during carbon-limited growth and signals the absence of glucose (Chapters 2 and 7) (Hardiman et al., 2007). This leads to the Cra-dependent repression of glycolysis genes and also to the flux decrease observed (Figure 4-4). It is therefore proposed that the observed behavior is regulated by the *cra* modulon (Figure 4-2). This results in sufficient, though minimal, efflux into the pentose phosphate pathway, thereby maintaining the cell's supply with biosynthesis precursors.

However, the flux into the pentose phosphate pathway decreased more than the fluxes into the glycolysis pathway (EMP/PPP split ratio, Figure 4-4). In accordance, the flux fraction via the upper part of glycolysis (from g6p [glucose 6-phosphate] to gap [glyceraldehdye 3-phosphate]) increased during carbon limited growth at the expense of the flux via the pentose phosphate pathway (g6p to f6p and gap; Figure 4-4). It is expected that the PPP flux is controlled via regulation of a reaction catalyzed by 6-phosphogluconate dehydrogenase (Gnd), i. e. one of the two irreversible reaction in the PPP, which can operate as control point between the oxidative and the non-oxidative branch of the PPP (Sprenger, 1995). It is already known that the amount of Gnd protein is determined primarily by the rate of transcription initiation (Pease and Wolf, 1994; Wolf et al., 1979). It is assumed that the

amount of Gnd correlates with the growth rate (Pease and Wolf, 1994; Wolf et al., 1979). However, this has so far only been shown for balanced growth experiments with e. g. acetate or glucose as carbon and energy source (Pease and Wolf, 1994; Wolf et al., 1979). During fed-batch cultivation, the transcript level of *gnd* decreased considerably (Figure 4-4), which substantiates the assumption of its correlation with the specific growth rate. The underlying regulatory mechanism has, however, not yet been clarified. Even more difficult is the situation with the second irreversible reaction catalyzed by the glucose 6-phosphate-1-dehydrogenase (Zwf). Rowley and colleagues found that the *zwf* transcription rate also varied with the type of substrate used (Rowley et al., 1991). In contrast, the gene was not differentially expressed during fed-batch growth. The lower *gnd* mRNA levels are in accordance with an increase in the flux fraction via the upper glycolysis part (Figure 4-4). It is therefore assumed that the negative regulation of *gnd* transcription is a key parameter for understanding the observed differences in the EMP/PPP split ratio (node g6p, Figure 4-4; Figure 4-2). In addition to this, there are also several factors that fine tune the activities of the PPP enzymes, e. g. the NADP+/NADPH ratio and the fbp, g6p and ribu5p concentrations (Keseler et al., 2009; Sahm et al., 2000). The regulation of PPP enzymes is expected to minimize the oxidation rate of the substrate and reduce the efflux into biosyntheses to a minimum.

Moreover, higher *rpiA* and lower *rpe* mRNA levels could be detected during fed-batch growth (Figure 4-2 and 4-5). The corresponding fluxes were in accordance with the gene expression levels (Figure 4-4): the RpiA flux (from ribu5p [ribulose 5-phosphate] to x5p [xylulose 5-phosphate]; enzyme: ribose 5-phosphate isomerase A) decreased less than the Rpe flux (ribu5p to r5p [ribose 5-phosphate]; enzyme: ribulose phosphate 3-epimerase). The splitting of the flux at the node ribu5p (RpiA/Rpe) is also reflected in the ratio of the efflux into biosyntheses (r5p to prpp [5-phosphoribosyl 1-pyrophosphate]) and the flux re-entering EMP (x5p to f6p and gap), which increases during carbon-limited growth (0.67, 0.70 and 0.73; Figure 4-4). In other words, the flux that enters the pentose phosphate pathway is preferentially directed towards biosynthesis. Although it is assumed that RpiA and Rpe protein level regulation is unimportant for the control of the overall flux through the central carbon metabolism (Chassagnole et al., 2002; Kummel et al., 2006), the regulation of the RpiA/Rpe split ratio might fine tune the efflux into the nucleotide, histidine and tryptophan biosynthesis pathways (Figure 4-2). Simulation of the glycolysis and PPP using the model of Chassagnole *et al.* (Chassagnole et al., 2002) and varying the r_{max}-values of the two reactions confirmed this assumption and suggested that the high *rpiA* mRNA level (Figure 4-4) could play a major role (data not shown). However, little is known about the regulation of the respective genes (Fraenkel, 1996; Keseler et al., 2009; Sprenger, 1995).

TCA cycle and glyoxylate shunt (GS). In Section 2.3.2 it was detailed that the reaction rates in the TCA cycle remained constant during glucose-limited fed-batch growth, and were regarded as a major reason for the substantial decrease in the biomass yield (Hardiman et al., 2007). The existing knowledge about the modular regulation of the enzymes of the central carbon metabolism was integrated into a comprehensive and global structure (Section 2.3.2) (Hardiman et al., 2007). Within this systems oriented picture it is assumed that the Crp-cAMP complex (strong increase in cAMP) activates the expression of TCA cycle genes in a coordinated manner (Chapter 2) (Hardiman et al., 2007). It is supposed further that the Cra regulator protein activates the transcription of GS genes. These hypotheses are supported through findings obtained with microarray analyses (Figure 4-2 and 4-5). The relatively high influx into the TCA cycle is most likely due to a higher GltA enzyme level since the GltA mRNA level is much higher (Figure 4-4). The regulation of the GltA protein level for the major control of the flux into the TCA cycle has also been predicted from thermodynamic analyses (Kummel et al., 2006). In general, the expression of TCA cycle genes was considerably higher during fed-batch conditions, in particular of those genes whose expression is positively regulated by the Crp-cAMP complex (*gltA*, *acnB*, *sucABCD*, *sdhCDAB*, Figure 4-4). A considerable increase in the expression of GS genes was also found (*aceB* and *aceA*, Figure 4-4), suggesting that the products of the *gltA*, *acnB* and *sdhCDAB* (Figure 4-4) genes are also involved in GS. It must therefore be assumed that a large fraction of the flux entering the TCA cycle must be directed into the glyoxylate shunt. The phosphoenolpyruvate(pep)-glyoxylate cycle (pep-GS) is another alternative cycle for substrate oxidation (Fischer and Sauer, 2003). This cycle involves the flux through the GS and the cycling of oxaloacetate (oac) to pep, catalyzed by pep carboxykinase (PckA). Although the *pckA* gene was not differentially expressed in microarray experiments, quantitative PCR analysis nevertheless showed that the gene transcript was more abundant during fed-batch growth (GEO, GSE10307). In *E. coli* MG1655, the flux ratio of TCA : GS : pep-GS was determined as 1.1 : 1.5 : 1.0 at μ = 0.12 h^{-1} (Fischer and Sauer, 2003) or 0.9 : 1.1 : 0.2 in a derivative of MG1655 (Nanchen et al., 2008). These data suggest that the close relative *E. coli* K-12 W3110 also uses these cycles. Considering the cycles' functions, it seems likely that the maintenance of the oac, pep and pyruvate precursor pools are a major goal of the behavior observed, besides supplying the cells with sufficient metabolic energy. The importance of the pep pool for the import of glucose and the synthesis of the signaling molecule cAMP was already highlighted in Section 4.3.3 ["Transport Systems (I)"]. Another benefit for the cells could be seen in the reduced oxidation of the substrate and the reduced production of NADPH (*via* isocitrate dehydrogenase, IcdA), which is not needed in high amounts due to the slower growth of the cells.

4.3.5 Chemotaxis and Flagellar System (III)

Chemotaxis is a phenomenon describing bacterial movement along the concentration gradient of certain chemicals (see Eisenbach, 2007 for further details). Chemotaxis helps the bacteria to detect food sources by swimming towards the highest concentration of food molecules (for example, glucose). E. coli has flagella that principally rotate in two opposing ways, enabling the bacteria to change directions. More than 50 genes are required for the synthesis and function of the E. coli flagellar and chemotaxis system (Chilcott and Hughes, 2000). These genes, which belong to 17 operons, constitute a regulon within which the operons are grouped into three temporally regulated, hierarchically organized transcriptional classes: early, middle and late.

The use of laboratory-scale bioreactors guarantees homogeneously mixed cultivation media at any time. Therefore, different regulatory responses, which are due to local substrate gradients, seem to be irrelevant for these bioprocesses. On the other hand, the chosen feeding strategy leads inevitably to the above mentioned time profile of limiting glucose concentration. Therefore, it was investigated whether this feeding strategy might lead to a global stimulus of the chemotaxis response.

As shown in Table A4-1 (Appendix D), nearly 50% of all genes known to be involved in the chemotaxis and flagellar system were affected during the entire cultivation process. The chemotaxis system seemed to be active during the non-limiting exponential growth phase. Over time, the majority of chemotaxis genes were transcribed to a lesser degree (Table A4-1, Appendix D). The number of transcripts of one of these genes, the dual regulator flhD, which is responsible for initiating the chemotaxis system, was reduced as early as T2, and subsequently followed by middle and late class genes (Table A4-1, Appendix D). It can therefore be assumed that the entire functional flagellar system is displaced from the cell under enhanced carbon limitation (e. g. at T3, when external acetate has been consumed).

As demonstrated in Figure 2-4 (Section 2.3.4), the intracellular cAMP levels rose considerably at the beginning of glucose limitation (Hardiman et al., 2007). This coincided with the time when the transcript levels of the chemotaxis genes were reduced (Table A4-1, Appendix D). The specific role of the alarmone cAMP in bacterial chemotaxis has been under debate for many years. Initially believed to be directly involved in chemotaxis (Black et al., 1983), the role of cAMP in chemotaxis was later refuted or attested barely an “indirect” role (Tribhuwan et al., 1986). Chemotaxis and the glucose-specific PTS system, however, share some common structural elements such as the proteins EI, EII and HPr. Since these proteins are part of the cAMP synthesis pathway, it might be speculated that these elements are

crosslinked through the alarmone cAMP. The presented transcriptome data support previous findings published by Soutourina (1999) who put forward the idea about the multiple control of flagellar biosynthesis. In this context, the transcription control of the early class master operon, *flhDC*, through the global regulator protein H-NS and the Crp-cAMP complex are of particular importance. At T2, H-NS expression was increased while fewer *flhD* transcripts were found (Table A4-1, Appendix D).

A further issue to be addressed in the discussion about downregulation of the chemotaxis refers to the proton gradient. Chemotaxis depends on a steep proton gradient between the periplasmatic space and the cytoplasm [reviewed by Berg (2003)]. Transport phenomena, however, also depend on proton gradients amongst others. The experimental observation of the opposite direction in the gene regulation of chemotaxis and transport systems could be interpreted as the result of a competition, in which the more effective transport of extant sugar or further energy supplying compounds seems to be more important than the chemotaxis.

4.3.6 Cell Growth (IV)

The specific growth rate decreased strongly during fed-batch cultivation (Figure 4-1). 128 genes coding for anabolic enzymes and the salvage pathway were differentially expressed at at least one point in time [see “Supplemental Material” in (Lemuth et al., 2008) for details]. The investigation suggests that the cells carefully utilized the available external and internal resources for their growth. It can therefore be assumed that the capacity of synthesizing recombinant proteins and metabolites decreases under constant feeding conditions. Exponential feeding might have a positive effect although this strategy is also based on the limitation of the carbon source.

Monomer synthesis. Glucose limitation led to the reduced synthesis of mRNAs of genes required for the biosynthesis enzymes of various monomers (amino acids, fatty acids, carbohydrates, nucleotides) as well as of coenzymes and prosthetic groups (Figure 4-2) [“Supplemental Material”, (Lemuth et al., 2008)]. Although much is known about the regulation of the respective genes through the stringent response, further clarification of the mechanisms and their directionality (i. e. positive or negative) is still necessary (Cashel et al., 1996).

Polymer synthesis (cell composition, cell mass). Most of the genes involved in protein biosynthesis (ribosomal proteins, RNA polymerase subunits, ribosomal assembly, protein maturation, RNA modification and aminoacyl-tRNA synthetases) were transcribed to a lesser

extent (Figure 4-2) ["Supplemental Material", (Lemuth et al., 2008)]. 78% of the genes encoding ribosomal proteins were transcribed to a lesser extent. The gene for the 30S ribosomal subunit protein S22 (encoded by *rpsV*) is expressed at higher levels during the stationary phase (Izutsu et al., 2001) and was transcribed more actively during the fed-batch period. Most of the underlying regulatory mechanisms are well known and can be assigned to the stringent response (Cashel et al., 1996). It can be safely assumed that these mechanisms lead to a change in the macromolecular composition of the cell.

Cell division. 16 genes coding for cell division enzymes (Figure 4-2) ["Supplemental Material", (Lemuth et al., 2008)], whose transcript levels decreased (e. g. *ftsAZ, tig, seqA*) were identified; the number of transcripts of cell division inhibition enzymes increased (*minCD* and *cspD*).

Macromolecular degradation. 51 genes coding for enzymes that degrade carbohydrates, amino acids, fatty acids, nucleic acids and nucleotides were differentially expressed at at least one point in time under glucose limiting conditions (Figure 4-2) ["Supplemental Material", (Lemuth et al., 2008)].

4.3.7 Stress and Starvation Response (V)

According to Ferenci (1996), the nutritional state of bacteria can be separated into "feast" (glucose-rich) and "famine" (glucose-starved). The physiological response of *E. coli* to glucose limitation provides evidence for the further separation between "hunger" and "starvation" responses (Ferenci, 2001). Ferenci assumed that the rapid sequence of hunger and starvation responses occurred when the batch cultures were grown until depletion of a nutrient during conventional "starvation" experiments. The fed-batch procedure chosen in the present study enabled the investigation of the dynamic decline of the growth rate due to the decreasing glucose levels. This might provide further insights in support of Ferenci's hypothesis. This is not only of pure academic interest because the majority of production processes (e. g. high cell density cultivations for recombinant protein, amino acid or antibiotics production) are performed under these stress conditions and may hence suffer from stress-related protein turnover.

It is generally assumed that the starvation response occurs during the stationary phase, which is characterized by a complete exhaustion of nutrients (C, N, P, or S) and which is mediated by the stationary phase sigma factor (σ^S, encoded by *rpoS*) (Hengge-Aronis, 1993). Weber et al. (2005a) carried out global transcription analyses and identified 140 genes as a core set of σ^S-regulated genes. In the present fed-batch experiment, only 18 of these genes

showed elevated transcript levels, at least at one point in time (Table A4-2, Appendix D). This number corresponds to 2 % of all genes identified in the present study. This implies that carbon limitation and the concomitant stringent response do not necessarily lead to the σ^S-mediated stress response during glucose-limited growth (Figure 4-2). However, 33 genes associated with stress conditions (heat and cold shock or oxidative stresses) were differentially expressed (Figure 4-2; Table A4-3, Appendix D). Only three of these genes (*hdeA*, *recF* and *yhiO*) are regulated by the sigma S factor (Table A4-4, Appendix D).

The alarmone ppGpp has a positive effect on *rpoS*-transcript levels (Gentry et al., 1993) while the Crp-cAMP complex inhibits *rpoS* transcription. The concentration of these two alarmones is elevated during glucose limitation (Figure 2-4, Section 2.3.4) (Hardiman et al., 2007). It has already been suggested that these opposing regulations could result in the reduction of *rpoS* transcription during glucose-limited growth in bioreactors (Lapin et al., 2006). However, other known regulatory mechanisms may also be worth considering (Hengge-Aronis, 1993).

The global regulator LrhA seems to diminish the σ^S level (Peterson et al., 2006). In the present study, an increase of the *lrhA* mRNA level was observed at T3 until the end of the cultivation (Table A4-1, Appendix D). The effect of LrhA on the σ^S level might be a reason for the lack of a σ^S response. The increased transcription of the gobal regulator protein H-NS from T3 to T6, might be an explanation for the lack of a σ^S-mediated response. H-NS binds to *rpoS* mRNA and enhances its cleavage (Brescia et al., 2004). However, one needs to keep in mind that all of these effects are balanced and lead to constant *rpoS* transcript levels.

The global transcription analysis presented in this study confirms the findings of Teich et al. (Teich et al., 1999) who suggested that slow glucose concentration changes significantly increase the cell's ability to adapt to new physiological states without using the rearrangements mediated by the σ^S-stress response. Therefore, it may be concluded that the strategy of constant nutrient feeding contributes mainly to the hunger state but is less important in terms of the cells' stress-induced starvation.

4.4 Conclusions

The current contribution globally analyses time-dependent transcript and metabolic flux levels in *E. coli* K-12 W3110 fed-batch cultures. It was possible to simultaneously track the carbon limitation responses (in the transport systems, intermediary metabolism, growth-related functions, chemotaxis and stress), which illustrates the power of the applied experimental setup. The constant feeding strategy also provided an appropriate approach for

separating the time-dependent events during the transition from exponential growth to strong carbon limitation. The novelty of this work arises from the integration of the dynamic transcriptional, metabolic and regulatory responses into a comprehensive hypothetical model (Figures 4-2 and 4-3b, c), pinpointing the impact of these variations on the general employment of the available cellular resources.

In Chapter 2 the flux redistribution during carbon-limited growth was shown to result in a significantly lower biomass yield, which is mainly due to the oxidation of the substrate in the TCA cycle for the generation of energy (Hardiman et al., 2007). The current findings led to the hypothesis that the general rate of oxidative decarboxylation can be limited by regulating the EMP/PPP, RpiA/Rpe and TCA/GS/pep-GS split ratios. Accordingly, an optimal carbon and energy balance of the central carbon metabolism (homeostasis) will be achieved. Nevertheless, the split ratios need to be investigated in more detail in order to validate the results from the above-mentioned stoichiometric metabolite balancing method. Flux analyses using isotopic transient ^{13}C-labelling data may be the method of choice as they have become technically possible for fed-batch processes (Maier et al., 2008; Noh et al., 2007; Schaub et al., 2008).

The results obtained in this investigation strongly support the hypothetical regulatory model structure put forward by Hardiman et al. (2007) (Chapter 2). Many other regulatory mechanisms might have minor effects on the *E. coli* metabolism (Chapter 2) (Hardiman et al., 2007), particularly in case of fed-batch processes of recombinant E. coli strains; however, the global genetic regulatory systems discussed are considered as most relevant for the control of the behavior during carbon-limited growth. This model, which is an extension of the previous dynamic metabolic model of Chassagnole et al. (2002), is the basis for the ongoing research relating to the mathematical modeling of the dynamics occurring in the central carbon metabolism of *Escherichia coli* K-12 W3110. The work provides a step forward towards the more detailed understanding of the impacts of carbon limitation on metabolic activities that must be taken into account when optimizing biotechnological processes. For example, the regulation of flux splitting might be an appropriate target for counteracting the excessive loss of carbon and energy in carbon-limited processes via oxidative decarboxylation, and thus for the optimization of the yield of biomass, recombinant proteins and other products.

5 Quantification of rRNA using Capillary Gel Electrophoresis with Laser-Induced Fluorescence Detection

Published in:

Hardiman, T., Ewald, J.C., Lemuth, K., Reuss, M., and Siemann-Herzberg, M. (2008). Anal Biochem *374*, 79-86.

Abstract

Over the last ten years, sophisticated, powerful techniques have been developed for the quantification of mRNA and rRNA, thus enabling researchers in science, industry and molecular medicine to explore gene expression. These techniques require the (reverse) transcription of analyte RNA, hybridization with synthetic oligonucleotides and other additional steps, which makes them costly, time-consuming and quantitatively difficult to perform. The current work demonstrates how 16S and 23S rRNA can be precisely quantified using capillary gel electrophoresis with laser-induced fluorescence detection (CGE-LIF) directly after the extraction of total RNA, without requiring further reactions or calibration. CGE-LIF is normally used for the qualitative examination of RNA preparations. Its quantitative performance could be significantly improved using MS2 bacteriophage RNA as an internal standard. The entire analytical procedure was validated for linearity, repeatability, reproducibility and recovery. This validation also included total RNA extraction from bacterial cells – an aspect examined for the first time in absolute RNA quantification. Recovery is close to one hundred percent and the analytical precision was 10-fold increased (CV < 3 %) as compared with similar approaches. The demonstrated method is simple and opens up new possibilities for the absolute quantification of not only rRNA, but also individual mRNAs.

5.1 Introduction

Scientific and industrial research as well as molecular medicine has always heavily relied on tools for gene expression profiling and transcriptome analysis. Most approaches provide semi-quantitative data (relative expression levels), which is usually sufficient for a wide range of applications. However, the latest challenge to RNA analysis concerns the requirement for absolute, high-precision quantification. Extremely reliable estimates of the concentrations of RNA species such as mRNA, rRNA, or tRNA are a prerequisite for clinical diagnostics or systems biology when quantitatively studying gene expression kinetics based on mathematical modeling (Arnold et al., 2005; Ideker et al., 2001). The concentration of mRNA is an important variable for modeling of a wide range of dynamic processes, e. g. the behavior of producer strain metabolisms during biotechnological processes (Hardiman et al., 2007). Recent results indicate that the dynamic modeling of the central carbon metabolism and its regulation by gene expression also needs to take growth rate control, which is exerted by the regulation of rRNA expression, into account (Chapter 2) (Hardiman et al., 2007). For this purpose, ribosome (or rRNA) concentration is another very important variable. High-throughput technologies and automation become necessary when dealing with large sample quantities so that hands-on time can be reduced to a minimum. However, achieving minimal hands-on time for experimental expression profiling is a challenge in itself, particularly because, up until now, only highly sophisticated methods are available, which produce only reasonably satisfactory results.

For the quantification of ribosomal RNA (rRNA) radioactive labeling of stable RNA (rRNA and tRNA) and uracil starvation (Forchhammer and Kjeldgaard, 1968), or direct quantification of the ribosome concentration using several extremely time consuming ultracentrifugation steps with or without density gradients (Davis et al., 1986; Dong et al., 1995) were previously used, but have proven inappropriate for precise and reasonable high-throughput quantification. The latest developments for this purpose are in the field of diagnostic studies and ecological research, in which either quantitative real-time polymerase chain reaction assays (qPCR) or DNA microarray technology are employed (Collantes-Fernandez et al., 2002; Kuboniwa et al., 2004; Small et al., 2001). The latter are also the most frequently used commercially available automated methods for quantifying mRNA expression. Microarray technology (Butte, 2002; Lemieux et al., 1998) is unbeatable in the parallel determination of thousands of different mRNA molecules on the genome-level in a single run (Epstein and Butow, 2000; Rhodius and LaRossa, 2003; Young, 2000). Although precise semi-quantitative determination is possible, the generation of precise quantitative data (absolute transcript copy numbers per cell) is quite difficult. Even when internal standards are used, it is still only

possible to provide estimations of mRNA concentrations (Carter et al., 2005; Dudley et al., 2002; Frigessi et al., 2005; Kakuhata et al., 2007; Taniguchi et al., 2001; van Bakel and Holstege, 2004). Quantitative real-time PCR is the state-of-the-art method for the quantitative determination of specific mRNA molecules (Bustin, 2000; Kubista et al., 2006; Nolan et al., 2006). However, reproducible quantification depends on careful experimental design, application and validation (Bustin, 2000; Bustin and Nolan, 2004; Smith et al., 2006). Furthermore, due to technical limitations, only a restricted number of mRNA molecules can be detected in parallel (Bustin, 2000). Numerous other biochemical approaches such as Northern blot analysis, ribonuclease protection assay (RPA) and serial analysis of gene expression (SAGE) can also be used for quantification (Reue, 1998; Stanton, 2001). All of the commonly used biochemical approaches deal with samples that contain thousands of individual RNA molecules. The analyses are done in parallel and need to be selective for all individual transcripts, which makes the investigation procedure labor- and cost-intensive (oligonucleotide synthesis, hybridization reactions, cloning and in vitro transcription steps are required). Second, the quantification of RNA usually relies both on a non-linear signal-to-response which necessitates complicated calibration procedures, and also on statistical data analysis. Thus, a high level of experimental expertise is mandatory and the downstream data analysis is also very technical and involves a great deal of work (Bustin, 2000; Butte, 2002; Leung and Cavalieri, 2003).

In more recent, fundamentally different approaches, the analyte RNA molecules are first separated by capillary (gel) electrophoresis (CE or CGE) (Babu et al., 2006) and subsequently detected using spectrometry. Spectrometric detection, e. g. using laser-induced fluorescence (LIF), exhibits a linear signal-to-response behavior and a large dynamic range. Costs and time requirement are exceedingly low and automated microfluidic high-throughput CE-LIF technology is already available (Babu et al., 2006; Krishnan et al., 2001). CGE-LIF is routinely applied for qualitative controlling of RNA integrity prior to microarray or qPCR analyses (Bustin and Nolan, 2004; Nolan et al., 2006). Quantitative CE-based approaches employ either unspecific intercalating dyes or specific spectrometric detection after hybridization with fluorescence-labeled synthetic oligonucleotides (Al-Mahrouki and Krylov, 2005; Goldsmith et al., 2007). Since current CE methods are not yet able to resolve all mRNA molecules, hybridization becomes necessary in order to identify specific RNA molecules in total RNA samples. Unspecific dyes greatly simplify the procedure, but have only been applied for the estimation of relative total RNA and rRNA concentrations (Babu et al., 2006; Ogura et al., 1998). Although present CE systems enable the convenient estimation of RNA concentrations, detailed validation of the degree of precision has so far not been published. This may be attributed to technical limitations in CE

analysis which reduce the analytical precision such as the generally observed injection bias (Gong et al., 2007), changing physico-chemical constitution of samples or the polymer matrix (Mueller et al., 2000), even though the technique is constantly improving (Babu et al., 2006; Krishnan et al., 2001).

A generally disregarded issue – but very important for the absolute quantification of RNA – is the sampling and RNA extraction procedure (Bustin and Nolan, 2004; Smith et al., 2006; van Bakel and Holstege, 2004). In general, equal samples of total RNA amounts are used although many techniques have been developed for quantitative RNA analysis (Nolan et al., 2006). Therefore, the routinely used extraction procedures do not need to be quantitative. Although recommended by the MIAME standard, in many cases modifications of standard protocols are not adequately described (Brazma et al., 2001). Nevertheless, relating transcript numbers to the total RNA content can be critical as total RNA content can vary considerably when cells with different physiological states are being examined (Nolan et al., 2006). For instance, in microorganisms such as Escherichia coli, the total RNA content as well as the proportions of mRNA and rRNA depend on the growth rate (Murray and Bremer, 1996; Pramanik and Keasling, 1997). Consequently, a procedure that guarantees the quantitative extraction of total RNA and the normalization of expression data to biomass (or best possible, to the cell volume (Chapters 2 and 7) (Hardiman et al., 2007) is mandatory.

The current work uses 16S and 23S ribosomal RNA from bacterial cell extracts to show that specific RNA molecules can be precisely quantified using CGE-LIF. Technical limitations inherent to CE can be circumvented by adding an internal standard (IS) to all the samples. The method reported in this work allows the high-precision quantification of rRNA molecules and does not need oligonucleotide synthesis, hybridization reactions and calibration procedures, thereby considerably reducing material costs and working time. This work also presents a procedure that guarantees the quantitative extraction of total RNA from *E. coli*.

5.2 Material and Methods

5.2.1 Capillary Gel Electrophoresis and LIF Detection

All microchip-based capillary gel electrophoresis separations were performed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). The microfluidic glass chips (RNA 6000 Nano LabChip Kit, Agilent) were prepared according to the manufacturer's instructions using the materials supplied in the kit (sieving polymer, intercalating fluorescent dye concentrate, RNA marker (50 b), spin filters) and also included an RNA 6000 ladder

(Ambion, Austin, TX, USA). Injection of the samples into the capillaries, subsequent electrophoretic separation and detection by laser-induced fluorescence (635/685 nm) was fully automated. RNA extracts were diluted in RNase-free water and thoroughly mixed with the internal standard (MS2 bacteriophage RNA). All the samples and the RNA size ladder were incubated at 65 °C for 5 min, centrifuged at maximum speed for 20 s and chilled on ice before application (1 µl) to the microchips. The Bio Sizing software (version A.02.11 SI280, Agilent) was used for data analysis. The rRNA concentrations, c_{rRNA}, were calculated using Equation (5-1) in which A_{rRNA} denotes the peak area of sample rRNA, A_{IS} the peak area of the internal standard and c_{IS} the concentration of the internal standard.

$$c_{rRNA} = \frac{A_{rRNA}}{A_{IS}} \cdot c_{IS} \tag{5-1}$$

In order to demonstrate the improvement in precision (employing the IS), the RNA size ladder solution (Ambion, Austin, TX, USA, 150 mg l^{-1}) was used as an additional external standard for quantification.

5.2.2 Formaldehyde Agarose (FA) Gel Electrophoresis

For FA gel electrophoresis the gel contained 1.2 % (w/v) agarose, 2 mM MOPS, 0.5 mM sodium acetate and 0.1 mM EDTA. 40 ml of gel was mixed with 2.5 µl ethidium bromide and 720 µl formaldehyde (37 % w/v) and equilibrated in running buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, 20 ml l^{-1} formaldehyde (37 % w/v)). FA gel electrophoresis was performed in 400 ml running buffer at 110 V for 50 min. Samples were prepared as follows: RNA samples were mixed 1 : 1 with sample buffer (2x Loading Dye Solution, Fermentas, St. Leon-Rot, Germany), incubated at 65 °C for 5 min and chilled on ice. 3 µl of the preparation was applied to the gel. All solutions were prepared using RNase-free water.

5.2.3 Internal Standard

Each RNA sample was supplemented with RNA from bacteriophage MS2 (3569 nucleotides, Roche Diagnostics, Mannheim, Germany) as an internal standard. The concentration of the MS2 phage RNA was determined using an UV absorption ND-1000 spectrophotometer (ND-1000 nucleic acid software; NanoDrop Technologies, Wilmington, DE, USA). Its final concentration in the RNA samples used for CGE-LIF analysis was 175 mg l^{-1}.

5.2.4 Bacterial Strain, Sampling and Total RNA Extraction

Samples for total RNA extraction were withdrawn from a fed-batch cultivation (30-l bioreactor) of the wild-type bacterial strain *Escherichia coli* K-12 W3110 (DSM 5911, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany) as described in (Section 2.2.1) (Hardiman et al., 2007). Sampling into glass test tubes containing 2 volumes RNAprotect bacteria reagent solution (Qiagen, Hilden, Germany) was performed using a capillary sampling probe (Section 2.2.2) (Hardiman et al., 2007). After mixing thoroughly and incubating for 5 min, aliquots equivalent to 0.1 to 0.5 mg dry weight of cells were transferred to new 2 ml test tubes and centrifuged (10000 g, 10 min, 4 °C). The supernatants were removed and the cell pellets frozen at - 20 °C. The cell pellets were resuspended in 1 ml protoplasting buffer (15 mM Tris/HCl, pH 8.0, 0.45 M sucrose, 8 mM EDTA, stored at 4 °C) and 10 µl lysozyme solution (50 g l^{-1} hen egg white lysozyme; Fluka, Seelze, Germany) was added. After gently mixing and incubating on ice for 15 min, the protoplasts were collected by centrifugation (6000 g, 5 min, 4 °C). After carefully discarding the supernatant, the protoplasts were resuspended in 0.5 ml lyzing buffer (10 mM Tris/HCl pH 8.0, 10 mM NaCl, 1 mM sodium citrate, 1.5 % (w/v) SDS, stored at 37 °C). 15 µl DEPC was added and the resulting preparation was incubated for 5 min at 37 °C, and subsequently chilled on ice. After mixing with 250 µl NaCl (saturated solution) and incubating 10 min on ice, the protein-SDS-DNA precipitate was collected by centrifugation (20000 g, 15 min, 4° C). 500 µl of the supernatant was transferred to a fresh test tube and mixed with 1 ml ice-cold ethanol (100 %). The precipitates thus obtained were incubated at - 80 °C for 30 min prior to centrifugation at 20000 g (20 min, 4 °C). The nucleotide-containing pellet was washed with 1 ml ice-cold ethanol (70 %) and dried (15 min). The remaining liquid was removed in a vacuum centrifuge. The resulting RNA extract was resuspended in 10 µl RNase-free water, heated to 65 °C for 5 min to ensure complete dissolution of the nucleotides and either analyzed immediately with CGE-LIF or frozen at - 80 °C until further required. All solutions were prepared with RNase-free water (0.1 % DEPC in bidest. H_2O, incubated over night at 37 °C, then autoclaved). The RNA extraction procedure applied was modified from Reddy and Gilman (1987) (Reddy and Gilman, 1987).

5.3 Results and Discussion

5.3.1 Quantification Using MS2 Phage RNA as Internal Standard

It is recognized that it is difficult to perform CGE in a highly reproducible manner. It is only recently that the major problems associated with this procedure have been investigated and

summarized (Babu et al., 2006; Gong et al., 2007; Krishnan et al., 2001; Mueller et al., 2000). Internal standard substances are often used in cases where the conditions of an analytical process cannot be guaranteed to be fully reproducible. Therefore, this work examined the suitability of MS2 phage RNA as internal standard (IS) for the quantification of specific RNA molecules by CGE-LIF. An IS of known concentration was added to the RNA samples before analysis and the relative signal area (ratio analyte/IS) was used for quantification [Fig. 5-1 and Equation (5-1)]. A linear signal-to-response behavior of the analytical method was observed (Fig. 5-1). A CGE-LIF electropherogram of a total RNA sample plus IS is depicted in Figure 5-2a. Additionally, the integrity and purity of the MS2 phage RNA was inspected using formaldehyde agarose gel electrophoresis (Fig. 5-2b). It was possible to confirm the size of the IS RNA at 3.6 kb (Fig. 5-2). In accordance with the manufacturer's specifications, the MS2 phage RNA contained 10 % degradation products (Fig. 5-2b). Figure 5-2b also illustrates the separation of the total RNA preparation with and without IS on FA gel electrophoresis.

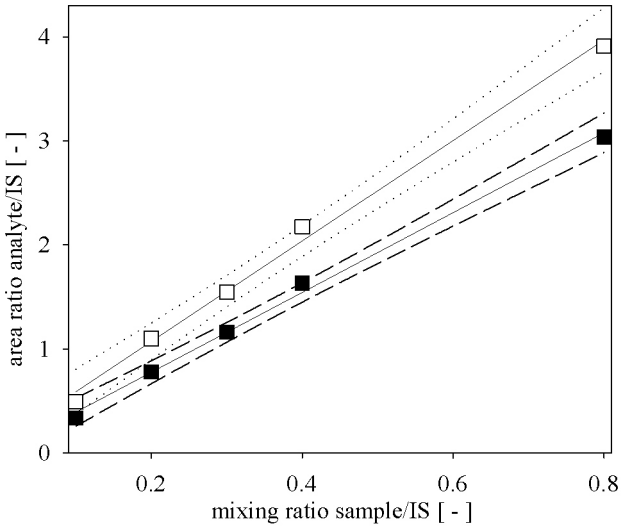


Fig. 5-1 CGE-LIF quantification of specific RNA molecules using an internal standard and linearity plot. Samples of total RNA, isolated from *E. coli* cultures, were mixed in various ratios with MS2 bacteriophage RNA solution as the internal standard (IS) and subjected to CGE-LIF. The signal areas of the analytes (■, 16S and □, 23S ribosomal RNA) were related to the IS area. For a known IS concentration the analyte/IS area ratio can be used for quantification of analyte RNA [see also Equation (5-1)]. Linear regressions were performed (solid lines; R^2 = 0.9971 for 16S rRNA and R^2 = 0.9953 for 23S rRNA), dotted and dashed lines indicate the 95 % confidence intervals.

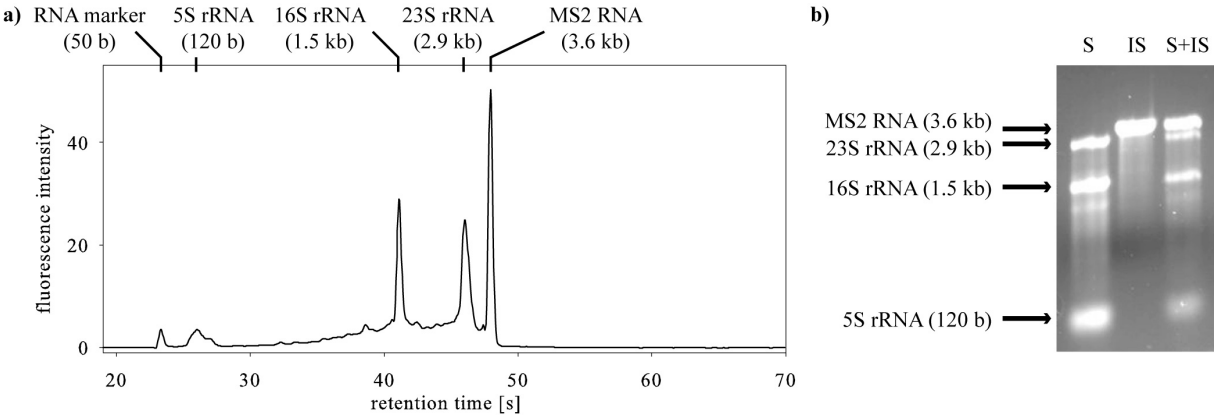


Fig. 5-2 (a) **CGE-LIF electropherogram** of a total RNA extract from *E. coli* supplemented with MS2 phage RNA internal standard solution and RNA marker. (b) **Formaldehyde agarose gel electrophoresis** of 3 µl of the total RNA extract (S), of MS2 phage RNA solution (IS) and of a mixture of both (S+IS).

5.3.2 Improvement of the Analytical Precision

A major aim of the present work was to improve the analytical precision of quantification of RNA molecules using automated CGE-LIF analysis. Three important aspects of the analytical procedure affecting the precision of the microchip-based method used were investigated, including the coefficient of variation using an internal standard (IS, MS2 phage RNA) for quantification. For comparison, an external standard (RNA size ladder) was used (see Fig. 5-3). The intra- and inter-batch precision of the automated microchip-based approach was also determined. This involved the multiple analysis (Fig. 5-3) of sample aliquots using a single microchip (intra-batch precision or repeatability) and several microchips (inter-batch precision or reproducibility). A further important issue is the impact of possible variations caused by manually pipetting small volumes (µl) of required solutions into the wells of the microchips. This was examined by successively applying the sample and analysis solutions and by applying aliquots of a pre-mixed solution for comparison (Fig. 5-3). Best results were obtained by pre-mixing the analysis solutions and by quantification via the IS (CV < 3 % for both 16S and 23S rRNA; Table 5-1). It was possible to achieve a very high level of precision when only a few samples were measured at the same time as using a single microchip (intra-batch precision). Moreover, it was possible to improve the level of precision two-fold in comparison to the quantification using an external standard (Table 5-1). The improvement was even more pronounced when large sets of samples were measured with several chips (inter-batch precision). The inter-batch CVs amounted to 5 to 9 % when the analysis solutions were pre-mixed, and 9 to 13 % when mixed individually (Table 5-1). Again, the use of an external standard only led to an increase of the coefficients of variation by a factor of 2 to 3 (data not shown). CE methods that have been published on the

quantification of specific RNA molecules do not give any detailed information about the precision of the methods (Al-Mahrouki and Krylov, 2005; Goldsmith et al., 2007), which might be due to the aforementioned well-known problems occurring in the quantitative reproduction of CE analyses. The work of Goldsmith *et al.* (2007), who have recently developed a quantitative CE-LIF method for the determination of 28 S rRNA in cardiac tissue, is an exception (2007). Despite using a state-of-the-art method, the latter authors achieved only an unsatisfactory CV exceeding 35 %. A technical note published by Agilent specifies an accuracy of ± 50 % (Agilent, 2003) for estimations of the total RNA content using the CGE-LIF method applied in this work. Accordingly, the proposed introduction of an IS substantially improves the analytical precision by more than 10-fold compared to similar approaches.

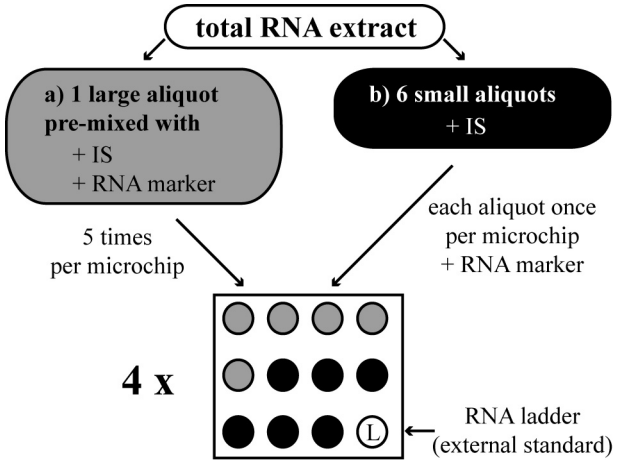


Fig. 5-3 Determination of the precision of quantification via internal standard in microfluidic CGE-LIF analysis. (a) A sample of total RNA extracted from *E. coli* was pre-mixed with IS (MS2 phage RNA) and RNA marker solution and then applied 5 fold to the microchips. (b) 6 total RNA samples were individually supplemented with IS, transferred to the microchips and the RNA marker was added. Additionally, the RNA ladder (L) was loaded into one microchip well, for use both as a size marker and as an external standard for all analyses. In total 4 microchips were prepared in the same manner and analyzed by CGE-LIF.

Table 5-1 Precision of specific RNA quantification using an internal standard in CGE-LIF analysis.

Analyte	Quantification	Handling[c]	Analysis [a]	Sample size (n) [b]	Number of repetitions (m) [b]	CV [%]
16S rRNA	external standard	pre-mixed	intra-batch	4	5	5.0
			inter-batch	5	4	22.1
	internal standard	pre-mixed	intra-batch	4	5	2.6
			inter-batch	5	4	5.0
		individual	intra-batch	4	6	8.7
			inter-batch	6	4	9.0
23S rRNA	external standard	pre-mixed	intra-batch	4	5	7.1
			inter-batch	5	4	20.2
	internal standard	pre-mixed	intra-batch	4	5	2.8
			inter-batch	5	4	9.3
		Individual	intra-batch	4	6	9.5
			inter-batch	6	4	13.3

[a] Intra-batch precision corresponds to repeatability, inter-batch precision to reproducibility.

[b] n refers to the number of repeated measurements used to calculate the CV. m represents the number of repeated experiments for averaging the obtained CVs.

[c] "premixed" indicates that all solutions were mixed prior to analysis. "Individual" means that the solutions were applied one after the other (see also Fig. 5-3).

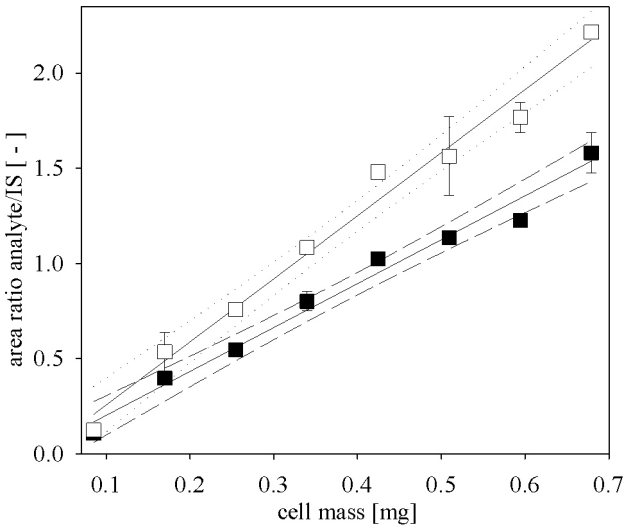


Fig. 5-4 Linearity range of the analytical procedure for specific RNA quantification including RNA extraction from Gram-negative bacterial cell cultures. The total RNA of samples of various dry cell weights of *E. coli* was extracted and the 16S and 23S ribosomal RNA quantified by CGE-LIF after adding an IS (MS2 phage RNA). Symbols with error bars (n = 2) indicate the peak area ratios of the analyte (■, 16S and □, 23S ribosomal RNA) and the IS. Solid line, linear regression (R^2 = 0.9831 for 16S rRNA and R^2 = 0.9845 for 23S rRNA). Dotted and dashed lines denote 95 % confidence intervals.

5.3.3 Linear Range and Recovery of the Total Analytical Procedure

Sample preparation and RNA extraction are often disregarded in the quantification of RNA species (Brazma et al., 2001; Bustin and Nolan, 2004; van Bakel and Holstege, 2004). However, RNA quantification in cell culture samples requires all steps to be quantitative and the intensity of the RNA signal obtained to be directly proportional to the concentration of the RNA species that causes the signal (linear response). First, the linear range of the total analytical procedure was investigated. In order to do this, samples were taken from an *E. coli* fed-batch cultivation, total RNA was extracted and 16S and 23S ribosomal RNA quantified using CGE-LIF. Ten samples of different volumes (equivalent to dry cell weight contents of 0.085 - 0.85 mg) were taken, processed and analyzed in duplicate. A linear response of the total analytical procedure was found for a dry cell mass of 0.08 – 0.70 mg (Fig. 5-4). The analyte/IS ratios were less reproducible when more cell material was used (30 %, n = 2, data not shown). The electropherograms suggest that this can most likely be attributed to the total RNA extract being contaminated with DNA (data not shown). Although bacterial DNA is removed during the extraction procedure, slight contamination cannot be prevented in cases when too much cell material is processed. Cell samples outside of the given linear range lead to insufficient electrophoretic separation, and therefore, to an increased CV. Smaller amounts of cell material often lead to the loss of cell and RNA pellets during sample preparation. It can therefore be concluded that the observed linear range depends primarily

on the amount of cell material used for RNA extraction, and only secondarily on the amount of RNA itself (linear range: 25 – 500 mg l^{-1} (Agilent, 2003)). It is worth noting that further extraction procedures were examined, e. g. commercially available kits for RNA extraction, which are generally recommended by current protocol (Brazma et al., 2001; Nolan et al., 2006). However, these usually use ion-exchange columns that have a limited RNA retention capacity, which makes them unsuitable for quantitative RNA extraction. Secondly, the recovery of the RNA in the extraction procedure was investigated: Protoplasts were isolated from cell culture samples and lyzed, then protein, DNA and cell debris, and finally, RNA was precipitated. Total RNA was extracted from an *E. coli* cultivation sample and the rRNA content was quantified in triplicate (denoted sample 1 to 3 in Table 5-2). Three aliquots (20 µl) of the obtained extract were mixed with 500 µl lyzing buffer and then subjected to a second extraction cycle. The new measurement of the rRNA content (denoted sample 1' to 3' in Table 5-2) enabled rRNA recovery to be determined – 105 % for 16S rRNA and 103 % for 23S rRNA (Table 5-2). Consequently, the extraction procedure is valid for the quantitative determination of concentrations of specific RNA molecules. This concerns all steps after protoplast lysis. In order to estimate the potential losses of whole cells or protoplasts during earlier steps, cell culture samples were withdrawn from an *E. coli* fed-batch cultivation and protoplasts isolated. After separation of cell pellets from the cultivation broth and freezing, thawing and resuspending the cell pellets in protoplasting buffer without, however, adding the lysozyme solution, the obtained cell suspension was examined using light microscopy. Almost 100 % of the cells were found to be physically intact. An insignificant amount of cell debris could be attributed to the cultivation broth. After collecting the protoplasts by centrifugation, the supernatant was also inspected. Only a few, isolated cells could be found and thus, the loss of protoplasts could be estimated to be 1 ppb ($1 : 10^9$). The protoplasts are very fragile and cannot therefore be microscopically analyzed.

However, if the protoplasts broke during the gentle centrifugation step, the debris would most probably be pelleted along with the protoplasts since the supernatants would be free of debris. This means that the early steps in the procedure prior to protoplast lysis cause only negligible losses. Conclusively, the linear response (Fig. 5-4) and the recovery (Table 5-2) prove that the extraction procedure presented in this work is valid, a conclusion that is equally valid for other methods such as microarray analysis and qPCR in cases where the quantity of RNA species has to be determined.

Table 5-2 Recovery of ribosomal RNA after total RNA extraction

Sample [a]	Ratio analyte/IS [–]				Recovery [%]	
	16S rRNA		23S rRNA			
	observation	mean	observation	mean	16S rRNA	23S rRNA
1	2.95		3.56			
2	2.98	2.97 ± 0.02	3.57	3.49 ± 0.13		
3	2.97		3.35			
					105	103
1'	3.14		3.50			
2'	3.08	3.12 ± 0.03	3.37	3.6 ± 0.3		
3'	3.14		3.92			

[a] Three samples were analyzed after total RNA extraction. The sample numbers with a prime correspond to the samples repeatedly subjected to total RNA extraction.

5.4 Conclusions

The proposed method introduces the use of an internal standard in CGE-LIF precluding variations due to irreproducible conditions such as the microfluidic injection bias or ionic strength of the applied solutions. Thus, a highly precise rRNA quantification procedure is provided that can be applied for convenient high-throughput analysis, which is particularly important for systems biology modeling. It can also be employed when further experimental data (transcript, protein, metabolite or parasite concentrations) are supposed to be related to the rRNA content of a (host) cell population (Collantes-Fernandez et al., 2002; Nolan et al., 2006). The results strongly suggest that internal standards should be employed for CE-based analysis of rRNA and mRNA. This will considerably improve the analytical precision, independent of the procedure used. Moreover, the use of unspecific dyes might exceedingly simplify the quantification of RNA. They can be used for the quantification of PCR or *in vitro* transcription products, where only a few mRNA transcripts are present, which can be easily

resolved by 1D-CE. However, 2D-CE has already proven successful for the separation of proteins or peptides (Babu et al., 2006; Chen et al., 2002) and is also conceivable for RNA separation. Careful examination of CGE-LIF analysis demonstrated that extra pipetting steps (small volumes) substantially affect analytical precision. Manual steps should consequently be reduced. Moreover, a detailed experimental validation of total RNA extraction from bacterial cell samples was performed for the first time. The protocol can be used for absolute quantification of individual RNA molecules. Therefore, the work also gives conclusive information for future developments of on-chip sample preparation and RNA extraction, which represents the latest trend in CE and microfluidic technology (Babu et al., 2006; Krishnan et al., 2001; Mueller et al., 2000).

6 Dynamics of the RNA Fractions During Glucose-Limited Fed-Batch Growth

Abstract

The total RNA content of *Escherichia coli* cells is well-known to vary with the specific growth rate. This correlation is majorly attributed to the variations in the rRNA fraction of total RNA, which implies that the total mRNA pool would not change significantly. In the present thesis, the dynamics of the 16S and 23S rRNA fractions were quantified in *E. coli* K-12 W3110 during glucose-limited fed-batch cultivation with constant feeding rate. Additional quantification of the total RNA content enabled the estimation of the dynamics of the total mRNA fraction. Based on the observed dynamics, a strong growth-rate dependent regulation of the global transcription of both rRNA and mRNA is proposed. The RNA polymerase availability is assumed to be the most important factor in growth rate-dependent regulation. Important conclusions for quantifying and modeling the transcription of individual genes are drawn.

6.1 Introduction

The cellular fractions [g (g dry weight)$^{-1}$] of the macromolecules RNA, DNA and protein are well-known to be growth rate-dependent (Bremer and Dennis, 1996; Pramanik and Keasling, 1997) and it is generally assumed that the variations in the total RNA content are due to the strong dynamics in the rRNA fraction (Lengeler et al., 1999). In the present study, the global transcription dynamics of the total RNA and rRNA contents were examined to provide information about the dynamics of the total mRNA fraction. The main goal was to clarify the potential growth rate-dependence of the mRNA content, which is important for dynamic modeling the transcription of mRNA during transient growth conditions.

6.2 Results and Discussion

The 16S and 23S rRNA fractions were quantified during glucose-limited fed-batch cultivation. The applied constant feed rate led to a strong decrease in the specific growth rate as described in Section 2.2.1 (Hardiman et al., 2007). The total RNA was quantitatively extracted from the cell suspensions and the rRNA fractions were determined using capillary gel electrophoresis with laser-induced fluorescence detection (CGE-LIF), applying the protocol described in Sections 5.2.3 and 5.2.4 (Hardiman et al., 2008a). Figure 6-1a and b depict the quantified 16S and 23S rRNA fractions ($w_{16S\,rRNA}$, $w_{23S\,rRNA}$), respectively. The sum of the two fractions, the total rRNA fraction (w_{rRNA}), is illustrated in Figure 6-1c. The dynamics of the total RNA content, $w_{totalRNA}$, were determined using a ND-1000 spectrophotometer (260 nm; ND-1000 nucleic acid software; NanoDrop Technologies) and are given in Figure 6-1c. For the estimation of the mRNA content, analytical functions were fitted to the data of the total RNA and rRNA fractions (Table 6-1). The dynamics of the mRNA content were then approximated by

$$\hat{w}_{mRNA}(t) = \hat{w}_{totalRNA}(t) - \hat{w}_{rRNA}(t), \tag{6-1}$$

assuming that the cellular (total) RNA majorly consists of the rRNA and mRNA (the minor portions of tRNA and 5S rRNA were neglected).

A strong growth rate-dependent variation of the rRNA and mRNA fractions was observed (Figure 6-1), which suggests a significant growth rate-dependent regulation of both rRNA and mRNA transcription. Growth rate-dependent regulation is assumed to be majorly determined by the concentration of free, cytosolic RNA polymerase (RNAP). Jensen and Pedersen (1990) proposed that promoters are generally subsaturated with RNA polymerase.

Accordingly, the *in vivo* rate of transcription is first-order with respect to the RNAP concentration (Jensen and Pedersen, 1990; McClure, 1985). Jensen and Pedersen (1990) suggested that the reduction of the RNAP concentration leads to a significant decrease of the transcription rate at lower growth rates. This hypothesis was confirmed using mathematical models of the partitioning of the RNAP into free, bound and immature fractions by Bremer et al. (2003) and Klumpp and Hwa (2008). The dynamics of the mRNA content observed in the present study suggest that mRNA transcription is globally affected by this growth rate-dependent regulation (Figure 6-1c). However, the strong decrease of the mRNA content in the beginning of the fed-batch process (Figure 6-1c) indicates that adaptation via global regulatory systems, such as the *cra* and *crp* modulons (Chapter 2) (Hardiman et al., 2007), is superimposed on the growth rate-dependent regulation.

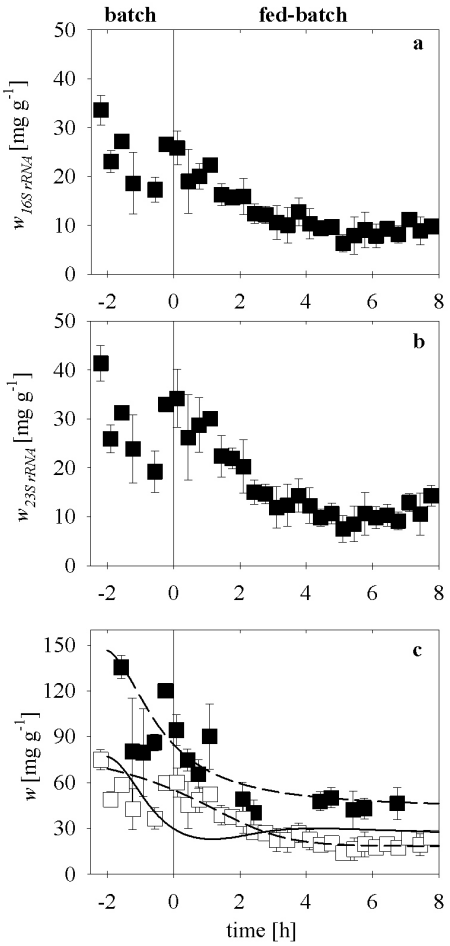


Fig. 6-1 RNA fractions during glucose-limited fed-batch cultivation of *E. coli* K-12 W3110 applying a constant feed rate. (a) experimentally determined 16S rRNA and (b) 23S rRNA fractions. (c) experimentally quantified total RNA content (■) and rRNA content (□) (sum of 16S and 23S rRNA). SDs are indicated by error bars (n = 3). Broken lines depict approximated time courses (see text and Table 1). The estimated total mRNA fraction [Equation (6-1)] is plotted as solid line (c). Vertical solid lines at t = 0 indicate the begin of glucose-limited fed-batch growth.

Table 6-1 Approximation functions of mass fractions [g (g dry weight)$^{-1}$]

$$\hat{w}_{totalRNA}(t) = \frac{(0.08497 - 0.02238 \cdot t + 5.930 \cdot 10^{-3} t^2 + 1.089 \cdot 10^{-3} t^3 + 8.987 \cdot 10^{-4} t^4)}{(1 + 7.870 \cdot 10^{-3} t - 0.01584 t^2 + 0.02197 t^3 + 0.02129 t^4 - 1.758 \cdot 10^{-5} t^5)} \quad (6\text{-}2)$$

$$\hat{w}_{rRNA}(t) = \left(84.52 - \frac{66.06}{\left\{ 1 + e^{\left\{ -\left[t - 0.8710 \cdot \ln\left(2^{1/0.2857} - 1 \right) - 0.3809 \right] / 0.8710 \right\}} \right\}^{0.2857}} \right) \cdot \frac{1}{1000} \quad (6\text{-}3)$$

Data of Figure 6-1 were fitted using Table Curve (Version 3).

In particular, the transcription of rRNA is known to strongly depend on the availability of the RNAP (Barker et al., 2001a; Barker et al., 2001b). This sensitivity is enhanced, when high concentrations of the alarmone ppGpp are present in the cells (Barker et al., 2001a; Barker et al., 2001b). The alarmone accumulates in large quantities in the applied fed-batch cultivation (Figure 2-4, Section 2.3.4) (Hardiman et al., 2007), which presumably contributes to the observed reduction of the rRNA fraction (Figure 6-1). However, it is known that 100S ribosomes are formed by dimerization of 70S ribosomes during carbon limitation and that ribosomal RNA is stored via this mechanism (Wada et al., 1990). It may be assumed that this mechanism reduces the rate of rRNA degradation, which would additionally affect the rRNA pool. This might explain the moderate slope at the onset of the fed-batch cultivation (Figure 6-1).

For the quantitative analysis of global transcription of rRNA and mRNA, such as through dynamic mathematical modeling, it can be concluded that the effect of the RNA polymerase availability on the transcription rate must be taken into consideration. This enables to describe the superimposition of regulation via global regulatory systems and the growth rate-dependent regulation.

Furthermore, the present data support previous studies that are concerned with the normalization of transcript levels quantified by experimental approaches such as DNA microarrays (van de Peppel et al., 2003). The determination of transcript levels relative to total mRNA may lead to artifacts, when the experimental conditions do not assure a constant total mRNA content of the cells. Possible consequences for the quantitative interpretation of such data are detailed in a separate work by Schuhmacher et al. (2009), where absolute concentrations of individual mRNAs were quantified.

6.3 Conclusions

The current work demonstrates significant dynamics of the rRNA and mRNA fractions in *E. coli* K-12 W3110 during fed-batch cultivation. The data supports the hypothesis that RNA availability may play a major role in the global regulation of transcription. It is recommended (i) considering the variable RNA polymerase concentration in dynamic models of gene expression, and, (ii) determining absolute mRNA concentrations for quantitative gene expression studies, when transient growth conditions are investigated.

7 Prediction of Kinetic Parameters from DNA-Binding Site Sequences for Modeling Global Transcription Dynamics

Published in:

Hardiman, T., Meinhold, H., Hofmann, J., Ewald, J.C., Siemann-Herzberg, M., and Reuss, M. (2009). Metab Eng, *doi:10.1016/j.ymben.2009.10.006*.

Abstract

The majority of the current dynamic models of gene regulatory networks (GRNs) comprise few genes and disregard multiple transcription regulation, to minimize the number of kinetic parameters. In the present work a new approach for predicting kinetic parameters from DNA-binding site sequences, by correlating the protein-DNA binding affinities with nucleotide sequence conservation, is proposed. The approach is illustrated by dynamic modeling global regulation of the *cra* modulon in *E. coli* during glucose-limited fed-batch cultivation. The experimentally quantified concentration of the Cra regulator protein inhibitor, fructose 1,6-bis(phosphate), falls strongly and leads to the repression of transcription. Strong regulation via RNA polymerase availability is superimposed. The evidence of this second effect is supported by the experimentally observed total RNA concentration and the critical assessment of the model simulations. Altogether, the work contributes a new method for predicting transcription dynamics, which may make an impact on metabolic engineering of gene regulation in biotechnological processes.

7.1 Introduction

The workflow of dynamic modeling of genetic regulatory networks (GRNs) – structured by network interactions between DNA, RNA, proteins and small molecules – exhibits interesting similarities with corresponding problems in dynamic modeling of metabolic networks. In both cases the initial step comprises the reconstruction of the network model architecture, thereby addressing qualitative aspects of the network performance encoded in the network topology. Hecker et al. (2009) have reviewed the most important models for inferring the network architecture, such as information theory models, Boolean and Bayesian networks, differential and difference equations, to name a few. As far as *Escherichia coli* is concerned, the topology of many of these regulatory networks have been elucidated by fundamental molecular biology studies and structured into hierarchical concepts (Lengeler et al., 1999). The significance of these modular ideas for the qualitative understanding of the key regulation of the central carbon metabolism enzymes, stimulated by limitation of the carbon and energy source glucose, has been verified by complementary microarray and metabolic flux analyses during fed-batch cultivations with constant feeding rates (Chapters 2 and 3) (Hardiman et al., 2007; Lemuth et al., 2008).

After reconstructing the architecture of the network, the next goal in the understanding of the network performance is the dynamic modeling through quantitative details of the molecular interactions, such as those of kinetics. Similar to the corresponding problem in dynamic modeling of metabolic networks, one may distinguish between top-down computational models based on canonical representations of kinetic properties, and, bottom-up approaches, which integrate mechanistic biological knowledge about the individual molecular interactions for modeling the systems behavior through aggregation of the individual reactions. For applications in the field of reconstruction of metabolic networks the two complementary approaches have been juxtaposed by Reuss et al. (2007).

Because of the extensive use of differential and difference equation systems for inferring the network architecture, the two approaches are more difficult to separate in case of GRNs. The models used in these dynamic representation are either linear (Chen et al., 1999) or consider non-linearity through polynomial functions (Sakamoto and Iba, 2001), the S-systems models (Vilela et al., 2008; Voit, 2008), generic sigmoidal functions (Haixin et al., 2007; Weaver et al., 1999) and neural network frameworks (Vohradsky, 2001a, b). Because of the inherent conflict between model quality and complexity, which manifests in the reliability of parameter estimation in higher-connected networks, the puzzle of robust identification of model structure and parameters frequently leads to discrimination problems between different

models. Another drawback of these approaches is the impossibility to interpret the estimated parameters in a physically and biologically meaningful way.

Unlike these constraints for a deterministic interpretation of parameters within the data-driven computational models, bottom-up models initiated from well-known molecular interactions enable the integration of mechanistically meaningful parameters. The use of these deterministic models describing the regulation of gene expression has a long tradition. The majority of these models are centered on the *lac* operon and the phenomena of diauxic growth of *E. coli* (Harder and Roels, 1982; Kremling et al., 2001; Kremling and Gilles, 2001; Lee and Bailey, 1984a, b; Roels, 1978; Schmid et al., 2004; Sevilla et al., 2007; Wong et al., 1997) (Table A7-1, Appendix G). The models are similar in mathematical nature and more or less rest upon the concept suggested by Yagil and Yagil (1971). Based on the operon model of Jacob and Monod (1961) these authors illustrated how to derive the probability of transcription initiation of a gene that is regulated by a repressor or activator protein, assuming equilibrium reactions of the regulator protein, R, and the effector molecule, E, or, the DNA binding site, $DNAbs$:

$$R + E \quad \overset{K_1}{\rightleftharpoons} \quad R.E \tag{7-1}$$

$$R + DNAbs \quad \overset{K_2}{\rightleftharpoons} \quad R.DNAbs \tag{7-2}$$

$$\varphi = \frac{c_{R.DNAbs}}{c_{DNAbs,t}} = \frac{K_2 \cdot c_R\left(K_1, c_E, c_{R,t}\right)}{1 + K_2 \cdot c_R\left(K_1, c_E, c_{R,t}\right)} \tag{7-3}$$

φ is the fraction of DNA-binding sites that are bound by the regulator protein and is used for calculating the probability of transcription (Yagil, 1975; Yagil and Yagil, 1971). c denotes the concentration of the respective model component that is indicated by the subscript, where c_R concerns the free regulator protein, and $c_{R,t}$, the total regulator protein. The major contribution of Yagil and Yagil's approach (1971) to the field is that no more than the two equilibrium constants (K_1, K_2) are needed for mechanistic modeling of gene regulation, where K_2 is the binding constant of the protein-DNA reaction. However, microbial transcription units exhibit many DNA-binding sites and multiple promoters, such as in the case of the *lac* operon (11 DNA-binding sites, 5 of which are bound by the Crp activator protein or the LacI repressor protein; and 5 promoters). Therefore, when applying this concept to large GRNs, the number of parameters (individual DNA-binding constants K_2)

again limits detailed mechanistic modeling. Thus, in previous studies many of the known regulatory interactions were disregarded (Table A7-1, Appendix G). For modeling the expression dynamics under defined environmental conditions, lumping of multiple promoters and regulatory sites may be reasonable to reduce the number of parameters (Table A7-1, Appendix G). For the reconstruction of dynamic GRNs, where researchers desire to dissect the superposition of all potential regulatory interactions and for their quantitative understanding, this might be critical. Therefore, an approach to reduce the number of parameters using *a priori* information is needed (Hecker et al., 2009). The relevant DNA-binding constants may be determined directly by estimation of the model parameters from experimental observations such as from surface plasmon resonance experiments (Majka and Speck, 2007).

An entirely different and straightforward approach rests upon the prediction of parameters from DNA-binding site sequences. The idea of using the information on the frequency of each nucleotide at each position of the DNA-binding sites, embedded in the position weight matrix, PWM (also known as position specific scoring matrix, PSSM) (Wasserman and Sandelin, 2004), for calculation of binding constants, traces back to the works of Mulligan et al. (1984), Schneider et al. (1986), Berg and von Hippel (1987) (reviewed by Stormo, 1990). This approach, however, has never been applied for the purpose of dynamic modeling of GRNs.

In the particular circumstances of a fed-batch operation at constant feeding rate, which is in the focus of the present paper, attention needs to be directed towards an additional factor – growth rate-dependent regulation. In previous models, growth rate-dependent regulation was formulated by simple linear approximation functions assuming that (i) the transcription rate is directly proportional to the specific growth rate, μ, and (ii) that this regulation is kinetically equal for all genes (Gondo et al., 1978; Harder and Roels, 1982; Table A7-1, Appendix G). However, the coupling of transcription initiation to the specific growth rate, μ, is assumed to be majorly determined through the availability of the free, cytosolic RNA polymerase (RNAP) (Barker et al., 2001a; Barker et al., 2001b; Cashel et al., 1996; 2007; Jensen and Pedersen, 1990; McClure, 1985). Bremer et al. (2003) and Klumpp and Hwa (2008) modeled the global partitioning of cellular RNAP into the DNA-bound, free and immature pools – depending on the specific growth rate, μ. These detailed mechanistic models demonstrate the strong non-linear correlation of the number of free RNAPs, and therefore of the transcription rate, and the specific growth rate. Second, dynamics of the growth rate-dependent regulation via RNAP availability are expected to be different for the individual promoter structures known in *E. coli*. Therefore, mechanistic modeling must consider both growth rate- and the

aforementioned regulator protein-dependent regulation. To bring the two ideas together, the RNAP-promoter interaction must be integrated into the transcription rate equation using the binding probability corresponding to Equation (7-3). However, when modeling regulation of the central carbon metabolism by the relevant global regulator proteins (Chapters 2 and 3) (Hardiman et al., 2007; Lemuth et al., 2008) and implementing the individual RNAP-promoter interactions, a large number of additional binding constants [corresponding to K_2 (7-3)] are needed.

In this work kinetic parameters are predicted for the first time from the nucleotide sequences of the DNA-binding sites of regulator proteins and from the promoter sequences that are important for RNA polymerase binding. Moreover, a concept is introduced for modeling the multiple regulation of transcription initiation by regulator proteins and via the RNAP availability. The approach is demonstrated using the predicted parameters for modeling of transcription dynamics of individual genes of the *cra* modulon, encoding central carbon metabolism enzymes, during fed-batch cultivation of *E. coli* K-12 W3110. The contributions of growth rate- and Cra regulator protein-dependent regulation of these genes is discussed.

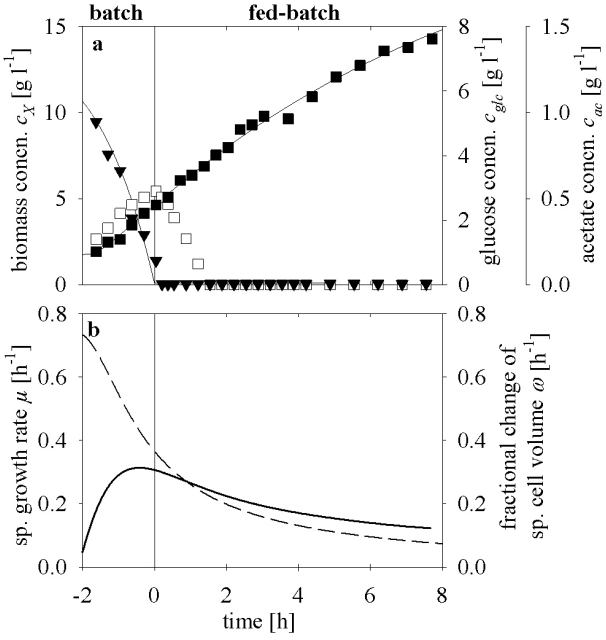


Fig. 7-1 Glucose-limited fed-batch cultivation of *E. coli* K-12 W3110 applying a constant feed rate. (a) measured concentrations of biomass (■), glucose (▼) and acetate (□), (b) Approximated time courses of the specific growth rate (μ , broken line) and the fractional change of the specific cell volume (ω , solid line; see Table 7-4). Vertical solid lines at t = 0 indicate the begin of glucose-limited fed-batch growth. As already detailed in Section 2.3.2, the observed inhibition of the growth of the cells during the batch phase is attributed to the accumulation of acetate (Hardiman et al., 2007).

7.2 Methods

7.2.1 Fed-Batch Cultivation and Analytical Procedures

Two independent fed-batch cultivations of *E. coli* K-12 W3110 were carried out in a 30-l bioreactor as described in Section 2.2.1 and by Hardiman et al. (2007). The application of a constant feed rate led to a continuously decreasing glucose concentration, and therefore, to a strong decrease of the specific growth rate. The concentrations of glucose, acetate and biomass as well as the specific growth rate, μ , are depicted in Figure 7-1. The quantification of fbp was performed using the enzymatic assay of Michal (1984). The experimental and sampling procedures are detailed in the Appendix H. The intracellular mRNA concentrations of the *cra* modulon genes were quantified using quantitative real-time PCR (qPCR) by Schuhmacher et al. (2009). Quantification of the total RNA content at 260 nm was carried out using a ND-1000 spectrophotometer using a quantitative method as detailed in Section 6.2 (Hardiman et al., 2008a). The RNA content in the samples was related to the biomass concentration, which led to the mass fraction, $w_{totalRNA}$ [g RNA (g dry weight)$^{-1}$], and this was used to calculate the total RNA mass concentration by $\rho_{totalRNA} = w_{totalRNA} / \nu_x$ [g RNA (l cytosol)$^{-1}$] using the growth rate-dependent specific cell volume, $\hat{\nu}_x(\mu)$ (which is detailed below).

7.2.2 Model Structure and Balance Equations

In the Chapters 2 and 3 the structure of the global regulatory network regulating the central carbon metabolism in *Escherichia coli* during carbon-limited growth was reconstructed (Hardiman et al., 2007; Lemuth et al., 2008). The *cra* modulon was proposed to be one of the most relevant global regulatory systems, and, is chosen as an example in the present work to demonstrate the new approach for dynamic modeling of gene expression. Figure 7-2 depicts the genes that were included in the model, which are known to be regulated by the Cra regulator protein (negative regulation: *pfkA*, *pykF*, *eno*; positive regulation: *pckA*) (EcoCyc; Keseler et al., 2009). Genes that are known to be regulated by other regulators, besides the Cra-dependent regulation, were not considered in the model (Figure 7-2). Further potential DNA-binding sites of the Cra regulator protein were predicted by pattern matching in the *E. coli* K-12 W3110 genome (see Appendix I). The computational analysis predicted four genes that are not yet known to be regulated by the Cra or other regulator proteins (EcoCyc; Keseler et al., 2009). These genes were implemented in the model (*pgi*, *tpiA*, *gpmM*, *rpiA*; Figure 7-2) to test the hypothesis that they were regulated by the Cra

protein during the fed-batch process. One additional gene, encoding geranyl diphosphate synthase (*ispA*), was chosen to study the transcriptional behavior of a gene that is assumed to be not regulated by any regulator protein (EcoCyc, Keseler et al., 2009) and to examine the sole growth rate-dependent regulation (see below).

The general balance equation (7-4), applied for the individual mRNAs i, is derived in Appendix E and comprises the transcription rate, $r_{tc,mRNA,i}$, the degradation rate, $r_{dg,mRNA,i}$, and a novel expression representing the dilution rate of an intracellular compound, $r_{dl,mRNA,i} = [\mu + \omega] c_{mRNA,i}$. This expression arises from the dynamics of the specific cell volume during the applied fed-batch process, which were illustrated in Section 2.3.4 (Hardiman et al., 2007). The fractional change of the specific cell volume with the specific growth rate, $\omega(\mu)$, must be considered in the mass balances of the intracellular mRNAs (7-4) (Appendix E). The dilution rate of Equation (7-4) denotes the sum of the dilution due to the biomass production ($\mu c_{mRNA,i}$) and the variation of the specific cell volume ($\omega c_{mRNA,i}$).

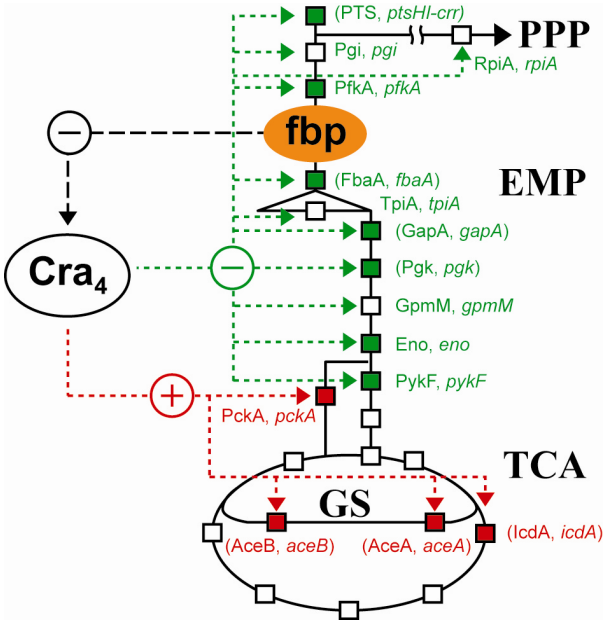


Fig. 7-2 Regulation of the central carbon metabolism by the *cra* modulon. (red) positive regulation and (green) negative regulation by the Cra protein is indicated by arrows. The protein and gene names (e. g. the PfkA protein is encoded by the *pfkA* gene) are indicated next to the corresponding reactions (squares). (red) positive regulation, (green) negative regulation by the Cra protein. Hypothetic negative regulation of a gene (predicted from pattern matching) is illustrated by black squares with white filling for the corresponding metabolic reaction. Genes of the *cra* modulon that are not included in the model (since further regulators are known) are given in brackets. The Cra regulator protein is inhibited by fructose 1,6-bis(phosphate) (fbp) and binds as a tetramer (Cra_4) to its DNA-binding sites. EMP, glycolysis; PPP, pentosephosphate pathway; TCA, tricarboxylic acid cycle; GS, glyoxylate shunt.

Table 7-1 Model parameters and literature data

Symbol	Parameter range	Source	Determination	Optimized parameter value
$k_{dg,mRNA,i}$ [a]	-	Bernstein et al. (2002); Selinger et al., 2003 (2003)	$k_{dg,mRNA,i}\left[\frac{1}{\mathrm{s}}\right]=\frac{\ln 2}{60\cdot t_{1/2,mRNA,i}[\mathrm{min}]}$	-
$\delta_{pckApCra}$ [b]	2 to 7.6	Chin et al. (1989)	CMAES (7-44)	2.2
K_1	$5.0\cdot10^{2}$ to $1.0\cdot10^{8}$ M^{-1}	-	CMAES (7-44)	$7.6\cdot10^{6}$ M^{-1}
$K_{RNAP,tac17}$	$1.7\cdot10^{8}$ to $23.2\cdot10^{8}$ M^{-1}	Mulligan et al. (1985)	CMAES (7-44)	$1.7\cdot10^{8}$
$a_{Cra,t}$	15 to 1500	-	CMAES (7-44)	1300
$b_{Cra,t}$	$1.0\cdot10^{6}$ to $1.0\cdot10^{8}$	-	CMAES (7-44)	$7.638\cdot10^{7}$
$a_{RNAP,t}$	0.1 to 0.5356	-	CMAES (7-44)	0.1004
$b_{RNAP,t}$	0.5 to 1.64	-	CMAES (7-44)	1.619

[a] The individual degradation rate constants, $k_{dg,mRNA,i}$, were calculated from published mRNA half-lifes $t_{1/2,mRNA,i}$.

[b] Chin et al. (1989) demonstrated that *pckA* expression is enhanced by a factor of 4.3 ± 1.3 or 6.7 ± 0.9, respectively, by measuring the enzyme activity of the PckA protein (phosphoenolpyruvate carboxykinase). These numbers include transcription and translation of the *pckA* gene. $\delta_{pckApCra}$ was estimated within the given parameter range, assuming an at least two-fold activation, or, that the maximum enhancement of enzyme activity (6.7 + 0.9 = 7.6) would only dependent on transcription activation, respectively.

$$\frac{dc_{mRNA,i}}{dt} = r_{tc,mRNA,i} - r_{dg,mRNA,i} - \left[\mu + \omega(\mu)\right] c_{mRNA,i} \tag{7-4}$$

The degradation rate is given by

$$r_{dg,mRNA,i} = k_{dg,mRNA,i} \cdot c_{mRNA,i} \,. \tag{7-5}$$

The individual degradation rate constants, $k_{dg,mRNA,i}$, were calculated from data of Selinger et al. (2003) and Bernstein et al. (2002) (Table 7-1), who used DNA microarrays to determine individual mRNA half-lives in *E. coli*. Individual mRNA degradation constants were not considered in several previous works (Table A7-1, Appendix G).

7.2.3 mRNA Transcription Rate

The *in vivo* rate-limiting step of mRNA synthesis is assumed to be the isomerization step from the closed to the open RNAP-promoter complex (McClure, 1980; Walter et al., 1967). The rate of transcription of the individual gene i from the promoter j, $r_{tc,mRNA,ij}$ (7-6), is therefore formulated as the product of the transcription initiation rate constant, $k_{tc,init,j}$ (isomerization step), the growth rate-dependent promoter concentration, $c_{promoter,j}(\mu)$ [$N_{gene} = f(\mu)$ according to Bremer and Dennis (1996)], and the efficiency of transcription initiation from this promoter, η_j, provided that each initiation is productive, i. e. leads to a complete mRNA transcript.

$$r_{tc,mRNA,i,j} = k_{tc,init,j} c_{promoter,j}(\mu) \cdot \eta_j \tag{7-6}$$

The overall efficiency of transcription, considering binding of the RNA polymerase, a repressor and an activator protein, was introduced by Lee and Bailey (1984b) based on Yagil and Yagil's approach (1971) and methods of statistical mechanics:

$$\eta = (1-\varphi_1)(1+(\delta-1)\varphi_2)\varphi_3 \,, \tag{7-7}$$

or,

$$\eta = \Phi_1 \Phi_2 \Phi_3 \,. \tag{7-8}$$

where φ denotes the fractions of DNA-binding sites (indices 1 and 2) or promoters (index 3) that are bound by regulator proteins or RNA polymerases, respectively [cf. Equations (7-1) to

(7-3)]. Moreover, they proposed to include a factor, δ, to characterize the enhancement of promoter activity through the binding of the activator. The multiplicative form of Equation (7-8) is equivalent to logical "AND" operations and reflects the coupling of regulatory interactions. However, the activator-DNA-binding site interaction contributes additively (logical "OR") to the transcription initiation rate, $\Phi_2 = \left(1+\left(\delta-1\right)\varphi_2\right)$ (7-7). These logical operations where applied in the present work to derive a new, extended formulation (Appendix F), considering the binding of the RNA polymerase to multiple promoters j, and multiple repressor and activator proteins, binding to their respective DNA-binding sites (k or l, respectively):

$$\eta_j = \Phi_{RNAP,j}\prod_k \Phi_k^{neg}\cdot\left(1+\sum_l \Phi_l^{pos}\right), \tag{7-9}$$

or,

$$\eta_j = \Phi_{RNAP,j}\prod_k \left(1-\varphi_k\right)\cdot\left(1+\sum_l \left(\delta_l-1\right)\varphi_l\right). \tag{7-10}$$

The fractions of the respectively bound DNA sequences are denoted by a lower case phi [cf. Equations (7-1) to (7-3)], whereas a capital phi is used for the contributions to the efficiency of transcription. When a transcription unit displays multiple promoters, the total transcription rate is obtained by adding up the rates of transcription initiated from the j promoters,

$$r_{tc,mRNA,i} = \sum_j r_{tc,mRNA,i,j}\ , \tag{7-11}$$

and, the total transcription rate of gene i reads:

$$r_{tc,mRNA,i} = k_{tc,init,i}\cdot c_{gene,i}\left(\mu\right)\cdot\sum_j \eta_j\ , \tag{7-12}$$

where the concentrations of the promoter j and the gene i are equal ($c_{promoter,j} = c_{gene,i}$). Individual transcription rate constants were used for each transcription unit, however, for multiple promoters only one rate constant was used ($k_{tc,init,j} = k_{tc,init,i}$). This assumption enables the calculation of $k_{tc,init,i}$ from the quasi-steady state constraints [see below, (7-43)]. Equations (7-10) and (7-12) enable modeling the multiple promoters and DNA-binding sites that are found in the regulatory region of many transcription units (Figure 7-3). As detailed above, multiple regulation was not considered in previous studies to reduce the number of parameters (Table A7-1, Appendix G).

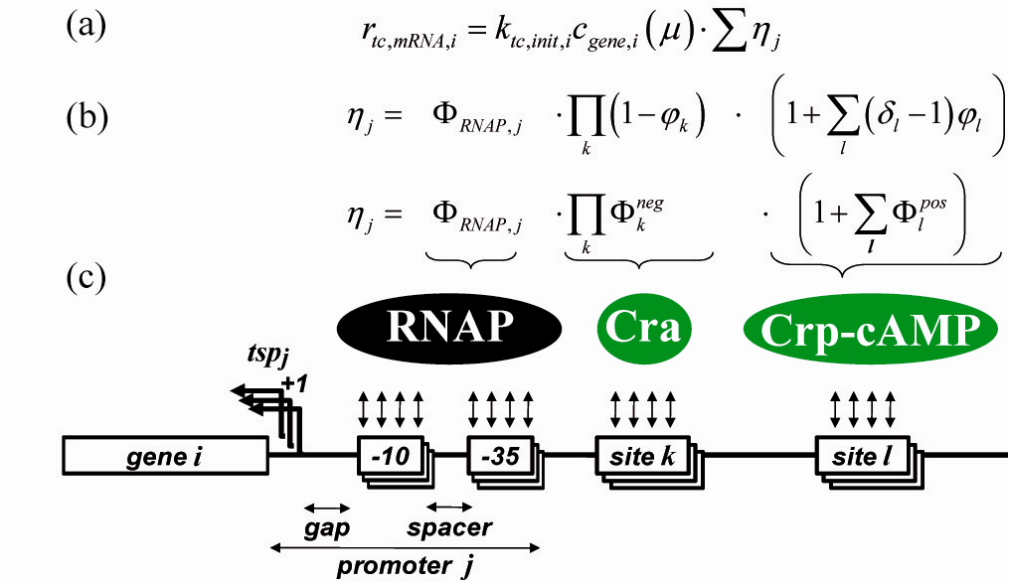


Fig. 7-3 Modeling multiple regulation of transcription initiation. (a) Kinetic rate equation for transcription of gene i from the multiple promoters j with (b) the transcription efficiency from the promoter j, η_j, the product of the probabilities that the RNA polymerase (RNAP) binds to the promoter j, $\Phi_{RNAP,j}$, and, that the repressor DNA-binding sites k are not bound by a repressor protein, $\Phi_k^{neg} = 1-\varphi_k$. The probability, that an activator protein is bound to its DNA-binding sites l is multiplied with an enhancement term ($\delta_l - 1$) and with the former product. The enhancement of transcription contributes additively to the transcription efficiency. (c) Schematic illustration of the regulatory region of gene i with the promoter j, negative (repressor DNA-binding site k; example Cra regulator protein) and positive (activator DNA-binding site l; example Crp regulator protein) regulation. Note, that multiple promoters j and sites k or l may be found in the regulatory region, which is indicated by multiple schematic boxes and transcription start points (tsp_j). Each promoter comprises an individual structure of -10 and -35 hexamer sequences, spacer length (distance between the two hexamers) and the gap (distance of the -10 hexamer from tsp). This substructure mainly determines the promoter strength.

7.2.4 Probability of Transcription Initiation and Growth Rate-Dependent Regulation

The fraction of promoters that are bound by the RNA polymerase are calculated from:

$$\Phi_{RNAP,j}(\mu) = \frac{1}{1+\frac{1}{\left(K_{RNAP,j} \cdot c_{RNAP,f}(\mu)\right)}}, \qquad (7\text{-}13)$$

which is equivalent to (7-3). $K_{RNAP,j}$ is the RNAP-promoter equilibrium constant and $c_{RNAP,f}$ is the concentration of free, cytosolic RNAP. By Equation (7-13) two novel

approaches for dynamic modeling global regulation of expression of individual genes are introduced. First, the aforementioned non-linear growth rate-dependent regulation can be implemented via the availability of RNAP. Second, multiple promoters are considered, using individual RNAP-promoter binding constants ($K_{RNAP,j}$) (Figure 7-3).

7.2.5 Modulation of the Transcription Initiation Rate by Cra-Dependent Regulation

The contribution of transcription regulation to the transcription efficiency is calculated using the binding probabilities, φ_k or φ_l, of the repressor or activator proteins, respectively (7-10) according to Yagil and Yagil (1971). Two equilibria are assumed for the binding reactions of the regulator protein Cra to its DNA-binding sites, $DNAbs_k$, and to the effector molecule fructose 1,6-bis(phosphate), fbp :

$$Cra + fbp \quad \overset{K_1}{\rightleftharpoons} \quad Cra.fbp \tag{7-14}$$

and

$$4\,Cra + DNAbs_k \quad \overset{K_{2,k}}{\rightleftharpoons} \quad Cra_4.DNAbs_k\,, \tag{7-15}$$

with the equilibrium constants

$$K_1 = \frac{c_{Cra.fbp}}{c_{Cra} \cdot c_{fbp}} \text{ and} \tag{7-16}$$

$$K_{2,k} = \frac{c_{Cra_4.DNAbs_k}}{c_{Cra}^4 \cdot c_{DNAbs_k}}\,. \tag{7-17}$$

According to Yagil and Yagil (1971) the probability that the regulator binds to the DNA-binding site reads:

$$\varphi_k = \frac{c_{Cra.DNAbs_k}}{c_{DNAbs_k,t}} = \frac{K_{2,k}\left(\frac{c_{Cra,t}}{1+K_1 c_{fbp}}\right)^4}{1+K_{2,k}\left(\frac{c_{Cra,t}}{1+K_1 c_{fbp}}\right)^4}\,. \tag{7-18}$$

For repression of transcription initiation, Φ_k^{neg} (7-9) is formulated as the fraction of DNA-binding sites that are not bound by the repressor protein Cra:

$$\Phi_k^{neg} = \frac{c_{DNAbs_k}}{c_{DNAbs_k,t}} = 1 - \varphi_k = \frac{1}{1 + K_{2,k}\left(\frac{c_{Cra,t}}{1 + K_1 c_{fbp}}\right)^4}. \qquad (7\text{-}19)$$

Equations (7-14) to (7-18) are given for the DNA-binding sites k of the Cra protein, repressing transcription initiation. To be consistent with (7-9) and (7-10), the subscript k must be replaced by l, when applying Equations (7-14) to (7-18) to activation of transcription initiation by the Cra protein. Thus, the contribution to the transcription efficiency (7-9) is defined by

$$\Phi_l^{pos} = (\delta_l - 1)\varphi_l \ , \qquad (7\text{-}20)$$

where δ_l is the enhancement factor (derivation in Appendix F). It should be noted here that this approach (7-20) separates the enhancement of transcription initiation (δ_l) from the binding probability φ_l – an aspect already considered by Lee and Bailey (1984b) and not included in several later studies (Table A7-1, Appendix G).

7.2.6 Derivation of DNA-Binding Constants

The individual DNA-binding constants of the RNAP-promoter ($K_{RNAP,j}$) and regulator protein-interactions ($K_{2,k}$, $K_{2,l}$) are needed for modeling the mRNA transcription of the *cra* modulon DNA [Equations (7-13), (7-19) and (7-20)]. A novel approach is introduced, to derive the unknown kinetic parameters from the DNA-binding site sequences. First, a set of highly conserved DNA-binding site sequences is chosen, i. e. of binding sites that are known to be bound by the Cra protein. The frequencies, $f_{m,n}$, of the nucleotides m (A, C, G or T) at the positions n of the DNA-binding site sequences are related to the genomic frequency of the nucleotides, p_m, and arranged in a specificity matrix, $\mathbf{SpM}$, with the matrix entries $a_{m,n} = f_{m,n} / p_m$:

$$\mathbf{SpM} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}. \qquad (7\text{-}21)$$

In a second step, the protein-DNA binding reaction is represented by individual reactions of the regulator protein, R, to the isolated nucleotides of the given DNA-binding site sequence:

$$R + nucleotide_{m,n} \quad \overset{K_{2,n}}{\rightleftharpoons} \quad R.nucleotide_{m,n} \tag{7-22}$$

where $nucleotide_{m,n}$ is the nucleotide m (A, C, G or T) at the position n of the given sequence. The equilibrium constant, $K_{2,n}$, is proportional to the ratio of the concentrations of the bound and free nucleotides, and this is equal to the ratio $f_{m,n} / p_{m,n}$ (Stormo, 1990)[2]:

$$K_{2,n} = \frac{c_{R.nucleotide_{m,n}}}{c_R \cdot c_{nucleotide_{m,n}}} \sim \frac{c_{R.nucleotide_{m,n}}}{c_{nucleotide_{m,n}}} = \frac{f_{m,n}}{p_m} \tag{7-23}$$

The (total) equilibrium binding constant for reaction (7-15) is assumed to be the product of the corresponding binding constants $K_{2,n}$ for the nucleotides of the bound DNA-sequence. According to Equation (7-23) this product is proportional to the product of the corresponding matrix entries, $a^k_{m,n}$, of the nucleotides m at positions n of the sequence k:

$$K_{2,k} = \prod_n K_{2,n} \sim \prod_n a^k_{m,n} \,. \tag{7-24}$$

In the current work, the latter product is denoted specificity score, $score_k = \prod_n a^k_{m,n}$. The unknown binding constant $K_{2,k}$ is then calculated by:

$$K_{2,k} = K_{2,ref} \frac{score_k}{score_{ref}}, \tag{7-25}$$

where $K_{2,ref}$ is the binding constant of the protein-DNA reaction of a given reference sequence and $score_{ref}$ is the corresponding specificity score. Equations (7-21) to (7-25) provide a new approach to predict the unknown binding constants for all DNA-binding sites of interest, provided that the value of one (reference) binding constant is known. The illustrated procedure is applied equally to positive or negative regulation and the binding of the RNA polymerase to the promoter ($K_{2,k}$, $K_{2,l}$, $K_{RNAP,j}$), and is summarized in Table A7-2 (Appendix J).

[2] The two main assumptions are: (i) the binding reaction can be represented by independent binding reactions of the regulator protein to the nucleotides. This assumption is also termed “additivity” rule (McClure, 1985) since the logarithmized equilibrium constants may be added to obtain the Gibbs energy of protein-DNA binding. (ii) the binding constants are proportional to the nucleotide frequencies in the set of DNA-binding sites. (Stormo, 1990)

Table 7-2 Cra DNA-binding site sequences considered in the model and predicted binding constants

Site [a]	TU [b]	DNA-sequence [c]	Genomic position [c]	Genomic orientation [c]	Distance from first ORF [c]	Relative score	K_2 [M^{-4}]	Regulation known [b]	Source [c]
aceBpcra [d]	*aceBAK*	GCTGAATCGCTTAACG	4,212,807	-	-242.5	1	1.2E+34 [d]	EcoCyc	Ramseier et al. (1993)
pgipcra	*pgi*	ACTGAAGCGATTCTGC	4,238,341	+	1000.5	2.19E-01	2.63E+33	This work	This work
pfkAp2cra	*pfkA*	CCTGAATCAATTCAGC	3,529,235	-	-113.5	4.77E-01	5.73E+33	EcoCyc	This work
pykFp2cra	*pykF*	CTTGAATGGTTTCAGC	1,757,191	+	-213.5	2.33E-01	2.80E+33	EcoCyc	This work
tpiApcra1	*tpiA*	TCTGAATCGCTTGAAG	3,525,037	+	-129.5	7.50E-02	9.00E+32	This work	This work
tpiApcra2	*tpiA*	GTTGAACCGATTAAGC	3,525,013	+	-153.5	7.78E-01	9.33E+33	This work	This work
gpmMpcra	*gpmM-envC-yibQ*	GCTGAATCGATAAAAT	3,855,219	+	-56.5	5.56E-01	6.67E+33	This work	This work
enop1cra	*pyrG-eno*	TGTGAATCGATCAGTT	2,907,331	+	-741.5	2.90E-03	3.47E+31	EcoCyc	This work
enop2cra	*pyrG-eno*	GCTGAATCGTTTACAA	2,907,528	+	-938.5	1.13	1.35E+34	EcoCyc	This work
enop3cra	*pyrG-eno*	AATGAAACGCTCGTAG	2,908,083	+	-1493.5	1.54E-04	1.85E+30	EcoCyc	This work
rpiApcra	*rpiA*	TCTGAATCGCTTTTTT	3,058,026	-	-52.5	9.37E-02	1.13E+33	This work	This work
pckApcra	*pckA*	GGTGAATCGATACTTT	4,107,869	-	-278.5	4.63E-02	5.56E+32	EcoCyc	This work

[a] Identifier of the sequence in the model according to the notation given in Tables A7-5 to A7-7, Appendix O.
[b] TU = Transcription unit that is regulated by the Cra protein. It is indicated whether this regulation is known [EcoCyc, (Keseler et al., 2009)] or is proposed in the present work (see text).
[c] Genomic information retrieved from the *E. coli* K-12 W3110 genome (see text). For aceBpcra the data refers to the *E. coli* K-12 MG1655 genome. The genomic position of the first nucleotide on the + strand is given.
[d] Reference binding site. The apparent dissociation constant, $K_{d(app),aceBp} = 3\ \mathrm{nM}$, was estimated from in vitro experiments earlier by Ramseier et al. (1993), assuming that the Cra protein binds as a monomer to the aceBpcra DNA-binding site (Ramseier et al., 1993). The desired $K_{2,ref}$ -value is therefore obtained from

$$K_{2,ref} = K_{aceBpcra} = \left(1/K_{d(app)}\right)^4 = 1.2 \cdot 10^{34}\ \mathrm{M}^{-4}.$$

Table 7-3 Structure of promoters considered in the model and predicted binding constants

Site [a]	TU [b]	DNA-sequence [c]		Genomic position [c]		Orien-tation[c]	Spa-cer [c]	Gap [c]	Relative score		K_{RNAP} [M^{-1}]	Curation [b]	Source
		-10	-35	-10	-35				-10	-35			
tac17 [d]	-	TATAAT	TTGACA	-	-	-	17	6	1	1	1.70E+08	-	Parameter estimation [Table 7-3; Mulligan et al.(1985)]
enop	*pyrG-eno*	TAAAAA	GTTCCA	2,906,683	2,906,709	-	20	4	1.59E-01	1.19E-01	2.56E+05	Hypothetical	This work
enop1	*pyrG-eno*	TCGATT	GTGTCA	2,907,340	2,907,364	-	18	4	2.08E-02	1.50E-01	9.79E+04	Curated	This work
enop2	*pyrG-eno*	AAGAAT	TCGAAG	2,907,371	2,907,397	-	20	7	2.07E-02	1.76E-01	4.95E+04	Curated	This work
enop3	*pyrG-eno*	GCGAAT	ATTCAG	2,907,512	2,907,534	-	16	4	4.60E-03	2.71E-02	5.34E+03	Curated	This work
pyrGp	*pyrG-eno*	TATACT	TTGTGG	2,908,383	2,908,406	-	17	8	4.62E-01	8.25E-02	6.48E+06	Hypothetical	This work
pyrGp1	*pyrG-eno*	GCTATT	TCAAAA	2,908,465	2,908,488	-	17	6	4.39E-03	2.41E-02	1.80E+04	Hypothetical	This work
pyrGp2	*pyrG-eno*	TATACT	TTGTGG	2,908,383	2,908,406	-	17	4	4.62E-01	8.25E-02	6.48E+06	Hypothetical	This work
pfkAp2	*pfkA*	TATACT	TCAATT	3,529,221	3,529,244	-	17	8	4.62E-01	1.45E-02	1.14E+06	Hypothetical	This work
pykFp	*pykF*	TATAAT	TTTCCT	1,757,170	1,757,146	+	18	5	1	4.60E-01	1.44E+07	Hypothetical	This work

(Table continued on next page.)

(Table continued from previous page.)

Site [a]	TU [b]	DNA-sequence [c]		Genomic position [c]		Orien-tation[c]	Spa-cer [c]	Gap [c]	Relative score		K_{RNAP} [M^{-1}]	Curation [b]	Source
pykFp1	*pykF*	AAAGCA	TCGCTT	1,757,366	1,757,343	+	17	5	1.40E-03	1.54E-01	3.64E+04	Hypothetical	This work
pykFp2	*pykF*	TATCCT	CTGCAC	1,757,302	1,757,279	+	17	6	1.12E-01	8.25E-02	1.57E+06	Hypothetical	This work
pckAp	*pckA*	GATAAT	ACACCT	4,107,749	4,107,771	-	16	6	6.01E-02	2.35E-03	6.04E+03	Hypothetical	This work
tpiAp1	*tpiA*	TCCTTT	CTGCCC	3,525,141	3,525,119	+	16	10	1.09E-02	8.13E-02	3.78E+04	Hypothetical	This work
tpiAp2	*tpiA*	TATACT	AAGCCT	3,525,099	3,525,076	+	17	7	4.62E-01	5.72E-03	4.49E+05	Hypothetical	This work
pgip	*pgi*	TACAAT	TCACAT	4,237,301	4,237,278	+	17	5	3.69E-01	2.04E-02	1.28E+06	Curated	This work
xseBp	*xseB-ispA-dxs-yajO*	TACCAT	TTTGCT	440,609	440,632	-	17	6	8.95E-02	3.42E-01	5.20E+06	Hypothetical	This work
gpmMp	*gpmM-envC-yibQ*	AATTAC	TTTTAT	3,855,195	3,855,218	-	17	9	1.51E-03	3.22E-01	8.25E+04	Hypothetical	This work
rpiAp	*rpiA*	TATAAT	GGGTCT	3,058,021	3,058,044	-	17	6	1	3.76E-02	6.39E+06	Curated	This work

[a] Identifier of the sequence in the model according to the notation given in Tables A7-5 to A7-7, Appendix O.

[b] TU = Transcription unit that is regulated by the given promoter. It is indicated whether the corresponding promoters in *E. coli* K-12 MG1655 are classified as known or hypothetical according to the RegulonDB (Gama-Castro et al., 2008)

[c] Genomic information for *E. coli* K-12 W3110. The genomic position of the first nucleotide on the + strand is given. The corresponding MG1655 sequences are equal (see text), however, not their genomic positions. The gap and spacer lengths are according to Hawley and McClure's notation (1983).

[d] Reference promoter sequence. The binding constant was determined *in vitro* by Mulligan et al.(1985). The *tac17* promoter was studied in a supercoiled form, to simulate *in vivo* conditions.

A set of twelve highly conserved DNA-binding site sequences (retrieved from the Prodoric Database, see Table A7-3, Appendix K for details), with a length of 16 nts of each binding site, was used for calculating the specificity matrix for the *cra* modulon (7-26).

$$\mathbf{SpM}_{Cra} =$$

$$= \begin{pmatrix} 0.67 & 0 & 0 & 0 & 4 & 4 & 0.67 & 0 & 0.33 & 1 & 0 & 0.67 & 1.67 & 2.67 & 0.33 & 0 \\ 0.67 & 3 & 0 & 0 & 0 & 0 & 0 & 3.33 & 0 & 0.67 & 0 & 0.33 & 1.67 & 0 & 1 & 2 \\ 2.67 & 0.67 & 0 & 4 & 0 & 0 & 0.33 & 0.67 & 3.67 & 0.33 & 0.33 & 0 & 0 & 0.33 & 2.33 & 0.33 \\ 0 & 0.33 & 4 & 0 & 0 & 0 & 3 & 0 & 0 & 2 & 3.67 & 3 & 0.67 & 1 & 0.33 & 1.67 \end{pmatrix} \quad \text{(7-26)}$$

The Cra-binding site of the *aceBAK* operon, *aceBpcra*, was chosen as the reference binding constant, $K_{2,ref}$ (Table 7-2). Relation (7-25) was then used for prediction of the DNA-binding constants of all Cra-binding sites of the model. Table 7-2 lists the calculated individual relative scores and the predicted DNA-binding constants.

The specificity matrix for the RNAP-promoter binding was derived from a list of 401 *E. coli* K-12 MG1655 σ^{70} promoter sequences, published by Shultzaberger et al. (2007). Shultzaberger et al. (2007) examined 684 promoters using information theory, considering the detailed promoter substructure comprising the -10 and -35 hexamers and spacer length. In the present work, specificity matrices were derived for the 401 known sequences of the -10 and -35 hexamers (7-27) (see Appendix L for sequence logos).

$$\mathbf{SpM}_{-10} = \begin{pmatrix} 0.23 & 3.72 & 1.05 & 2.09 & 2.09 & 0 \\ 0.43 & 0 & 0.66 & 0.51 & 0.97 & 0.19 \\ 0.19 & 0.03 & 0.51 & 0.55 & 0.37 & 0.11 \\ 3.15 & 0.25 & 1.79 & 0.85 & 0.57 & 3.70 \end{pmatrix}$$

$$\mathbf{SpM}_{-35} = \begin{pmatrix} 0.42 & 0.18 & 0.18 & 1.73 & 1.42 & 1.34 \\ 0 & 0 & 0.06 & 1.35 & 1.40 & 0.50 \\ 0 & 0.72 & 2.43 & 0 & 0.40 & 0.72 \\ 3.58 & 3.10 & 1.33 & 0.93 & 0.79 & 1.45 \end{pmatrix} \quad \text{(7-27)}$$

However, the substructures of the model-relevant promoters are not known (neither for the MG1655 nor for W3110 strain) and must be reconstructed from the *E. coli* K-12 W3110 genome data. To reconstruct the desired promoter substructures, first, the sequences of the respective promoter regions (comprising 60 nt before and 20 nt after the tsp [cf. Figure 7-3c]; i. e. length = 81-nt) of the MG1655 strain were retrieved from the RegulonDB (Gama-Castro

et al., 2008). Next, the corresponding promoter sequences were located in the genome sequence of the W3110 strain, taken from the Prodoric database (Munch et al., 2003). All examined 81-nt promoter sequences were found to be equal in the W3110 and MG1655 strains. Third, the hexamer location was shuffled within these sequences and the respective specificity scores were calculated. It might be assumed, that the two hexamers leading to the highest score would be the most likely *in vivo* RNAP-binding sites. However, additional information about spacer length and the gap (Figure 7-3c) must be considered to reconstruct the (hypothetically optimal) promoter structure, using

$$K_{RNAP,j} = K_{RNAP,ref} \frac{score_{-10,j} score_{-35,j}}{score_{-10,ref} score_{-35,ref}} \alpha_{spacer,j} \alpha_{gap,j} , \tag{7-28}$$

which is analogous to Equation (7-25). A reference equilibrium constant, $K_{RNAP,j}$, is multiplied by the relative hexamer score ratio of the promoter j and the reference promoter. Note, that a product of the two hexamer scores (-10 and -35) is applied, which is based on the same assumption as in Equation (7-24). Promoter activity changes drastically, when the spacer or gap lengths (Figure 7-3c) are altered – a well-known fact, which was considered by the spacer and gap penalties [$\alpha_{spacer,j}$, $\alpha_{gap,j}$; see Appendix M for details]. Finally, the genomic positions of the wanted hexamer sequences were obtained for the maximal value of Equation (7-28) during the mentioned hexamer shuffling. Table 7-3 depicts the desired (hypothetically optimal) promoter structures for *E. coli* K-12 W3110. The model-relevant RNAP-promoter binding constants were predicted for the *E. coli* K-12 W3110 promoter structures that are given in Table 7-3, using (7-28) and the synthetic *tac17* promoter as the reference binding site. It is worth noting that this promoter structures are also valid for *E. coli* K-12 MG1655 since the nucleotide sequences of these promoters were equal.

7.2.7 Growth Rate-Dependent Model Variables and Model Input Functions

The present work focuses on the transcription during the transient conditions of a fed-batch process with a strong decrease in the specific growth rate (Figure 7-1). As already mentioned the fractional change of the specific cell volume (ω) and the concentrations of the genes ($c_{gene,i}$) and of free, cytosolic RNA polymerase ($c_{RNAP,f}$) are growth rate-dependent variables that have an effect on the transcription rate. Their functional dependence on the specific growth rate is well-known (Section 2.3.4) (Bremer et al., 2003; Bremer and Dennis, 1996; Hardiman et al., 2007; Pramanik and Keasling, 1997), whereas the underlying regulatory mechanisms are not yet understood in detail. Therefore, approximation functions of these

growth rate-dependent model variables, y, were generated in the form $\hat{y} = f(\hat{\mu})$ using the available data (see Table 7-4). The time courses of the specific growth rate and the fractional change of the specific cell volume during fed-batch cultivation are depicted in Figure 7-1b. The number of an individual gene per cell, $N_{gene,i}$, depends on the specific growth rate and can be calculated by the empirical equation formulated by Bremer and Dennis (1996):

$$N_{gene,i}(\mu) = 2^{\left[C(\mu)\cdot[1-m_i']+D(\mu)\right]\cdot\mu/[60\cdot\ln 2]}, \tag{7-29}$$

where $C(\mu)$ (Equation (7-34), Table 7-4) is the time needed for chromosome replication (C period), $D(\mu)$ (Equation (7-35), Table 7-4) is the time period between termination of replication and cell division (D period) and m_i' is the map location of gene i relative to the replication origin. $m_i' = (m_i + 16)/50$ for map locations (m_i) between 0 and 36 min; $m_i' = (84 - m_i)/50$ for map locations between 36 and 84 min; $m_i' = (m_i - 84)/50$ for map locations between 84 and 100 min (Bremer and Dennis, 1996). The gene concentration is calculated from:

$$c_{gene,i}(\mu) = N_{gene,i}(\mu)/(V_{cell}(\mu)\,N_A), \tag{7-30}$$

where N_A is the Avogadro constant and V_{cell} the cell volume. The cell volume varies with growth rate and is given in Equation (7-36) (Table 7-4). The concentration of free, cytosolic RNA polymerase is difficult to quantify directly (Bremer et al., 2003; Klumpp and Hwa, 2008; McClure, 1985). Therefore, it is calculated using the fraction of RNAP that is cytosolic, α_f [Equation (7-37), Table 7-4]:

$$c_{RNAP,f}(\mu) = \frac{N_{RNAP,f}(\mu)}{V_{Cell}(\mu)\,N_A} = \alpha_f(\mu)\cdot\frac{N_{RNAP,t}(\mu)}{V_{Cell}(\mu)\,N_A}, \tag{7-31}$$

where $N_{RNAP,f}$ is the number of free RNAP and $N_{RNAP,t}$ [Equation (7-38), Table 7-4] is the number of total RNAP per cell.

Table 7-4 Model variables

Variable	Approximation [a]	Equation
Specific growth rate [b]	$\hat{\mu}(t) = \frac{\left(0.3656 + 0.1204 \cdot t + 0.02778 \cdot t^2 + 0.002675 \cdot t^3 - 3.280 \cdot 10^{-5} \cdot t^4\right)}{\left(1 + 0.7071 \cdot t + 0.2350 \cdot t^2 + 0.03808 \cdot t^3 + 0.004413 \cdot t^4\right)}$	(7-32)
Fractional change of the specific cell volume [c]	$\hat{\omega}(t) = \frac{1.583 + 0.3312 \cdot t - 0.1163 \cdot t^2 + 0.03041 \cdot t^3 - 0.003743 \cdot t^4 + 1.749 \cdot 10^{-4} \cdot t^5}{5.165 + 1.583 \cdot t + 0.1656 \cdot t^2 - 0.03877 \cdot t^3 + 0.007601 \cdot t^4 - 7.486 \cdot 10^{-4} \cdot t^5 + 2.916 \cdot 10^{-5} \cdot t^5}$	(7-33)
Chromosome replication time (C period) [d]	$\hat{C}(\hat{\mu}) = -136.7 - 87.14\hat{\mu}^2 \cdot \ln\hat{\mu} + 66.93 \cdot e^{\hat{\mu}} - 102 \cdot \ln\hat{\mu}$	(7-34)
Time period between termination of replication and cell division (D period) [d]	$\hat{D}(\hat{\mu}) = 39.1 - 30.86\hat{\mu} + 31.06\hat{\mu}^2 - 14.14\hat{\mu}^{2.5}$	(7-35)
Cell volume [e]	$\hat{V}_{cell}(\hat{\mu}) = 10^{-15} \cdot 0.486 \cdot 2^{1.144\hat{\mu}}$	(7-36)
Fraction of RNAP that is free (unbound) [f]	$\hat{\alpha}_f(\hat{\mu}) = 0.038 \cdot e^{(0.408\hat{\mu})}$	(7-37)
Total number of RNAP per cell [g]	$\hat{N}_{RNAP,t}(\hat{\mu}) = 1000 \cdot (0.1004 + 1.619\hat{\mu})^2$	(7-38)
Concentration of fructose 1,6-bis(phosphate) [h]	$\hat{c}_{fbp} = \frac{1}{1000 \cdot \hat{\nu}_X(\hat{\mu})} \cdot \left[3.575 - 2.152 \frac{1 - 1/\left(e^{-1.532t + 6.043}\right)}{1 + e^{-3.124t - 0.2128}}\right]$ d	(7-39)
Mass concentration of total RNA [h]	$\hat{\rho}_{totalRNA} = \left(0.04338 + 0.7448\hat{\mu}^2 - 0.8896\hat{\mu}^{2.5} + 0.2846\hat{\mu}^3\right) \frac{1}{\hat{\nu}_X(\hat{\mu})}$	(7-40)

(See Footnotes on next page.)

Table 7-4 Footnotes

[a] Approximations were performed using the Software TableCurve 2D (Systat Software Inc.). Time derivates were calculated using Maple (Maplesoft). Dimensions: t [h], μ [h^{-1}], V_{cell} [$l_{cytosol}$], C and D [min], $N_{RNAP,t}$ [molecule RNAP], c_{fbp} [mol $(l_{cytosol})^{-1}$], $\rho_{totalRNA}$ [g RNA (l cytosol)$^{-1}$].

[b] Derived from the approximations $\hat{\mu}(t) = \frac{1}{\hat{c}_X(t)} \frac{d\hat{c}_X(t)}{dt} + \frac{F}{\hat{V}_R(t)}$, $\hat{c}_X = f(t)$ and $\hat{V}_R(t)$ that were obtained from experimental data and were described in Section 2.2.4 (Table 2-1) (Hardiman et al., 2007).

[c] Derived from the growth rate-dependent equations $\hat{\omega}(\hat{\mu}) = \frac{d\hat{v}_X(\hat{\mu})}{\hat{v}_X(\hat{\mu}) \cdot dt}$ (7-56) and $\hat{v}_X = f(\hat{\mu})$ (7-49), defined in Appendix E.

[d] Approximated using experimental data reviewed by Bremer and Dennis (1996) (Figure A7-2, Appendix N).

[e] Taken from (Pramanik and Keasling, 1997).

[f] Approximated using experimental data published by Bremer et al. (2003) (Figure A7-2, Appendix N)

[g] The number of total RNAP molecules per cell, $N_{RNAP,t}$, was approximated by a polynomial function of the form $(a_{RNAP,t} + b_{RNAP,t}\mu)^2$ using simulated data of (Bremer and Dennis, 1996) (Figure A7-2, Appendix N). The parameters obtained from this regression analysis were optimized during the model parameter fitting as described below ($a_{RNAP,t}$ = 0.1004 , $b_{RNAP,t}$ = 1.619; Table 7-1).

[h] Approximation of data from this work. $\hat{v}_X$, see Equation (7-49) in the Appendix E.

Traxler et al. (2006) demonstrated that the mRNA level of the *cra* gene drops, when the growth rate decreases during the diauxic lag and stationary phase in glucose-lactose diauxie experiments. The regulation of the Cra protein concentration is not yet clarified. However, the transcription of the *cra* gene is likely to be attenuated: Merino and Yanofsky (2005) predicted various attenuator RNA hairpin structures and the “phantom” gene *fruL* could function as leader peptide. The *cra* mRNA level could thus be coupled to the growth rate via the limited amino acid availability during glucose-limited growth as detailed in Section 2.4.5 (Hardiman et al., 2007). Since the mechanism of this regulation by attenuation is unknown, it is assumed that the number of Cra proteins per cell, $N_{Cra,t}$, is proportional to the specific growth rate:

$$N_{Cra,t}(\mu) = a_{Cra,t} + b_{Cra,t}\mu \,. \qquad (7\text{-}41)$$

The unknown parameters $a_{Cra,t}$ and $b_{Cra,t}$ were estimated using the optimization algorithm described below (Table 7-1). The total Cra protein concentration, $c_{Cra,t}$, is calculated by

$$c_{Cra,t}(\mu) = \frac{N_{Cra,t}(\mu)}{V_{Cell}(\mu) N_A}. \qquad (7\text{-}42)$$

7.2.8 Quasi-Steady State Conditions

The transcription initiation rate constants were calculated from the initial conditions at an arbitrarily chosen quasi-steady state (t_0 = - 1.5 h):

$$\left.\frac{dc_{mRNA,i}}{dt}\right|_{t=t_0} = 0, \quad \text{or,} \quad k_{tc,init,i} = \frac{r_{dg,mRNA,i}(t=t_0) + r_{dl,mRNA,i}(t=t_0)}{c_{gene,i}(t=t_0)\sum_j \eta_j(t=t_0)}. \qquad (7\text{-}43)$$

The time point t_0 = - 1.5 h was roughly estimated from the time course of fructose 1,6-bis(phosphate) (fbp) during the batch phase (see below in Figure 7-4). The dynamics of the fbp concentration is probably interconnected with the acetate accumulation during the batch process (see Figure 7-1). Acetate is well-known to inhibit the growth of the cells (summarized by Xu et al., 1999b). The growth-inhibiting effect of acetate, which was already detailed in Section 2.3.2 (Hardiman et al., 2007), is attributed to the inhibition of the glucose uptake of the cells (summarized by Xu et al., 1999b). A retarded glucose uptake might lead to the observed decrease of the fbp concentration in the batch phase (see below in Figure 7-4). However, a quasi-steady state at time point t_0 is assumed, as judged from the quantified fbp time profile (see below in Figure 7-4).

7.2.9 SBML Model and Computational Methods

The model is provided in the SBML format (Level 2 Version 1; supplementary material), generated using the Systems Biology Toolbox 2 [SBT2 Version 1.0; (Schmidt and Jirstrand, 2006)] for MATLAB. The model identifiers are given in the Appendix O (Tables A7-5 to A7-7). The SBPD extension package for SBT2 was used for the parameter estimation, with the following modifications: For integration of the ODE system the MATLAB ode15s algorithm was used (MATLAB Version 7.7). For model parameter fitting the CMAES evolution strategy with covariance matrix adaptation by Hansen and Ostermeier (2001) was implemented (cmaes.m, Version 2.55). The objective function

$$\min_{\mathbf{p}} F(\mathbf{p}) = \sum_{time\ h=1}^{H} \sum_{state\ i=1}^{I} \left(\frac{c_{h,i}^{meas} - c_{h,i}^{sim}(\mathbf{p})}{2 \cdot SD_{h,i}} \right)^2 \qquad (7\text{-}44)$$

$$\text{with} \quad \mathbf{p}^{\mathrm{T}} = \left(K_1 \quad K_{RNAP,tac17} \quad a_{Cra,t} \quad b_{Cra,t} \quad a_{RNAP,t} \quad b_{RNAP,t} \quad \delta_{pckApCra} \right),$$

$$\text{subject to} \quad \left. \frac{dc_{mRNA,i}}{dt} \right|_{t=t_0} = 0,$$

was used, where $c_{h,i}^{meas}$ is the mRNA concentration of the gene i that was experimentally determined at the time h, $c_{h,i}^{sim}(\mathbf{p})$ is the corresponding simulated value for the parameter vector $\mathbf{p}$, and $SD_{h,i}$ the standard deviation of the experimental data. The parameters search intervals used for parameter estimation were deduced from literature data and are given in Table 7-1.

7.2.10 Control Coefficients

The effects of concentration changes of the components of the regulatory mechanisms (fbp, Cra and RNAP) on the transcription rate were determined using control coefficients. Note, that in the widely applied metabolic control analysis (Nielsen and Villadsen, 1994; Stephanopoulos et al., 1998) the observed system is linearized, the control coefficients are independent of the size of the perturbation and are thus valid only for small extrapolations. In the current work, however, control coefficients are calculated by Equation (7-45) according to Goldbeter and Koshland (1982). The coefficients, $C_{c_i}^{\Phi_j}$, depend on the size of finite perturbations (Δc_i) and exhibit non-linear dependence of Φ_j on changes in the concentration of a given compound ($\Phi_j = f(c_i)$). The changes were normalized either to the initial conditions (e. g. $c_i^0 = c_i(t = t_0)$, with i = fbp, Cra or RNAP) or to conditions at the end of the fed-batch process ($t = t_{end}$).

$$C_{c_i}^{\Phi_j} = \frac{\Delta\Phi_j / \Phi_j^0}{\Delta c_i / c_i^0} \tag{7-45}$$

7.3 Results and Discussion

7.3.1 Intracellular Concentration of Fructose 1,6-bis(phosphate) (Model Input Signal)

It is the first time that the dynamics of the intracellular concentration of the effector molecule, fructose 1,6-bis(phosphate) (fbp), were quantified during a shift from exponential to glucose-limited growth conditions (Figure 7-4a). A strong initial drop of its cytosolic concentration is seen from a level of 1.10 ± 0.17 during exponential growth to 0.15 ± 0.07 mM at fed-batch growth. This supports the hypothesis that fbp signals the availability of glucose in *Escherichia coli* (Chapters 2 and 3) (Hardiman et al., 2007; Lemuth et al., 2008), and presumably also of other carbon sources. The correlation of the intracellular fbp concentration with growth rate was already demonstrated by Schaub and Reuss (2008). The time series data were approximated in the form $\hat{c}_{fbp} = f(t)$ (Equation (7-39), Table 7-4; Figure 7-4a), to provide a time-dependent function as a model input. Metabolite concentrations are generally related to biomass concentration, to enable direct comparison between experiments with different biomass concentrations. This would lead to masking of the low level of the cytosolic fbp concentration during fed-batch growth (Figure 7-4b). However, Hardiman et al. (2007) already demonstrated that the change of the specific cell volume [Equation (7-49), Appendix E] must be considered, when looking at intracellular concentrations of metabolites determined at different growth rates.

7.3.2 Identification and Prediction of Kinetic Parameters

The present model describes transcription dynamics of nine genes (Figure 7-2) using the balance equation (7-4), the predicted DNA-binding constants (Tables 7-2, 7-3), the growth rate-dependent model variables [Table 7-4; Equations (7-31), (7-42)], and, model parameters derived from literature data (Table 7-1), calculated from the quasi-steady state constraints (7-43) or identified using the optimization algorithm [Table 7-1; Equation (7-44)]. The individual mRNA concentrations quantified by Schuhmacher et al. (2009) were used to estimate seven parameters using the objective function (7-44). Four of these are parameters of the growth rate-dependent functions of the RNAP and Cra protein concentrations ($a_{Cra,t}$, $b_{Cra,t}$, $a_{RNAP,t}$, $b_{RNAP,t}$). These and the two binding constants K_1 and $K_{RNAP,tac17}$ concern the global response of *cra* modulon transcription (Table 7-1). It is important to note that if these parameters are known, the presented approach (Table 7-4) can be used to predict all other kinetic parameters that concern the local features of transcription kinetics of the individual

transcription units. However, when using Equation (7-10), for positive regulation additional (local) kinetic parameters are needed. The enhancement factor for *pckA* transcription ($\delta_{pckApCra}$) is not predicted from the DNA-binding site sequence in this work and only the parameter range could be derived from experimental data (Table 7-1). Nevertheless, in a separate publication a more elaborate approach for prediction of the enhancement factor (δ_l) will be demonstrated for the *crp* modulon, where details about the regulatory mechanism are available.

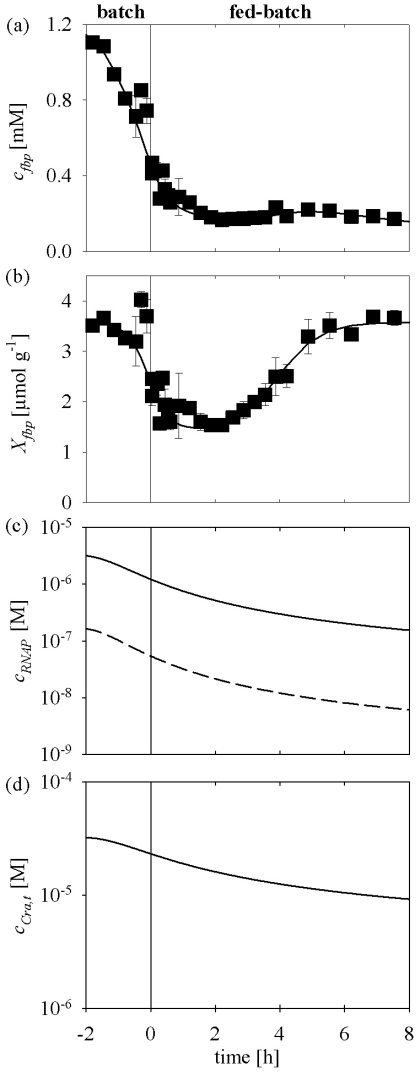


Fig 7-4 Components of the global regulation of the *cra* modulon during fed-batch cultivation of *E. coli* K-12 W3110. The fructose 1, 6-bis(phosphate) (fbp) concentration was quantified using an enzymatic assay (■, error bars indicate SD). (a) cytosolic fbp concentration, c_{fbp} [mM] = [mmol (l cytosol)$^{-1}$]. (b) specific fbp concentration, X_{fbp} [µmol (g dry weight)$^{-1}$]. Approximations of the (a) fbp, (c) total RNA polymerase, and (d) total Cra protein concentrations are given as solid lines; the free RNAP concentration in (c) as broken line.

The predicted parameters $K_{RNAP,j}$ are in the range of 10^3 to 10^7 M^{-1} (Table 7-3), which agrees with previous lists of other kinetically studied promoters (Mulligan et al., 1984; Wagner, 2000). The predicted $K_{2,k}$-values are in the range of 10^{30} to 10^{34} M^{-4} (Table 7-2), which is consistent with the aforementioned reference value for the *aceBpcra* DNA-binding site published by Ramseier et al. (1993). More Cra protein DNA-binding sites were studied using DNA band shift assays by Shimada et al. (2005). These authors estimated an apparent dissociation constant of $K_{d(app)} = 1.91 \cdot 10^{-5}\ \mathrm{M}$ for the *aceBpcra* DNA-binding site. Considering that the Cra protein binds as a tetramer to the DNA, the DNA-binding constant would be $K_2 = \left(1/K_{d(app)}\right)^4 = 7.5 \cdot 10^{18}\ \mathrm{M}^{-4}$. However, the magnitudes of assay data and the estimated parameters are inconsistent [Figure 7-2 in (Shimada et al., 2005)]. Ramseier et al. (1993) titrated an *aceB* DNA fragment with purified Cra protein. They generated a high data density and the estimated binding constant is reasonable since it is consistent with values[3] of other DNA-binding constants of regulator proteins.

For the *eno* transcription unit only the promoters *enop1*, *enop2* and *enop3* were considered in the model. The hypothetical *enop*, *pyrGp*, *pyrGp1* und *pyrGp2* promoters of the *pyrG-eno* transcription unit are not experimentally verified (Gama-Castro et al., 2008) and were not considered. Furthermore, one of the two Cra protein DNA-binding sites predicted by pattern matching for the *tpiA* transcription unit, *tpiAcrp2*, was not considered in the model. Leaving out these hypothetical promoters and DNA-binding site significantly improved the objective function value (7-44), i. e. led to agreement of the simulated and experimental data.

7.3.3 Simulation of mRNA Concentrations during Fed-Batch Growth

Schuhmacher et al. (2009) revealed that, in the fed-batch process, the *in vivo* concentrations of central carbon metabolism mRNAs fall by one to three orders of magnitude (Figure 7-5) – even for genes that are assumed to be constitutively expressed (*ispA*, Figure 7-5i). The model describes this behavior for the *cra* modulon, which is determined by the dynamics of the model input signal, $\hat{c}_{fbp}(t)$ [Equation (7-39), Table 7-4; Figure 7-4a], and the growth-rate dependent variables, $\hat{c}_{Cra,t}(\hat{\mu})$ (7-42) and $\hat{c}_{RNAP,f}(\hat{\mu})$ (7-31) (Figures 7-4c, d). Few

[3] For example, Crp binds as a dimer to a lacZ DNA fragment with a binding constant of $4.8 \cdot 10^7\ \mathrm{M}^{-1}$ [calculated from the monomer concentration; (Pyles and Lee, 1996)]. This corresponds to a value of $5.3 \cdot 10^{30}\ \mathrm{M}^{-4}$ for tetrameric DNA-binding.

parameters concerning the global transcription dynamics were identified using the optimization routine (7-44). These globally affect the behavior of all simulated mRNA concentrations. However, each gene is subjected to the individual regulation by the Cra protein and the RNAP. The kinetic parameters determining the binding probability of those proteins to the respective regulatory regions – and thus the individual behavior of the mRNAs – were predicted from the DNA-sequences. Therefore, the agreement of the simulated and measured individual mRNA concentrations (Figure 7-5) over a dynamic concentration range of three orders of magnitude demonstrates the predictive power of the presented approach.

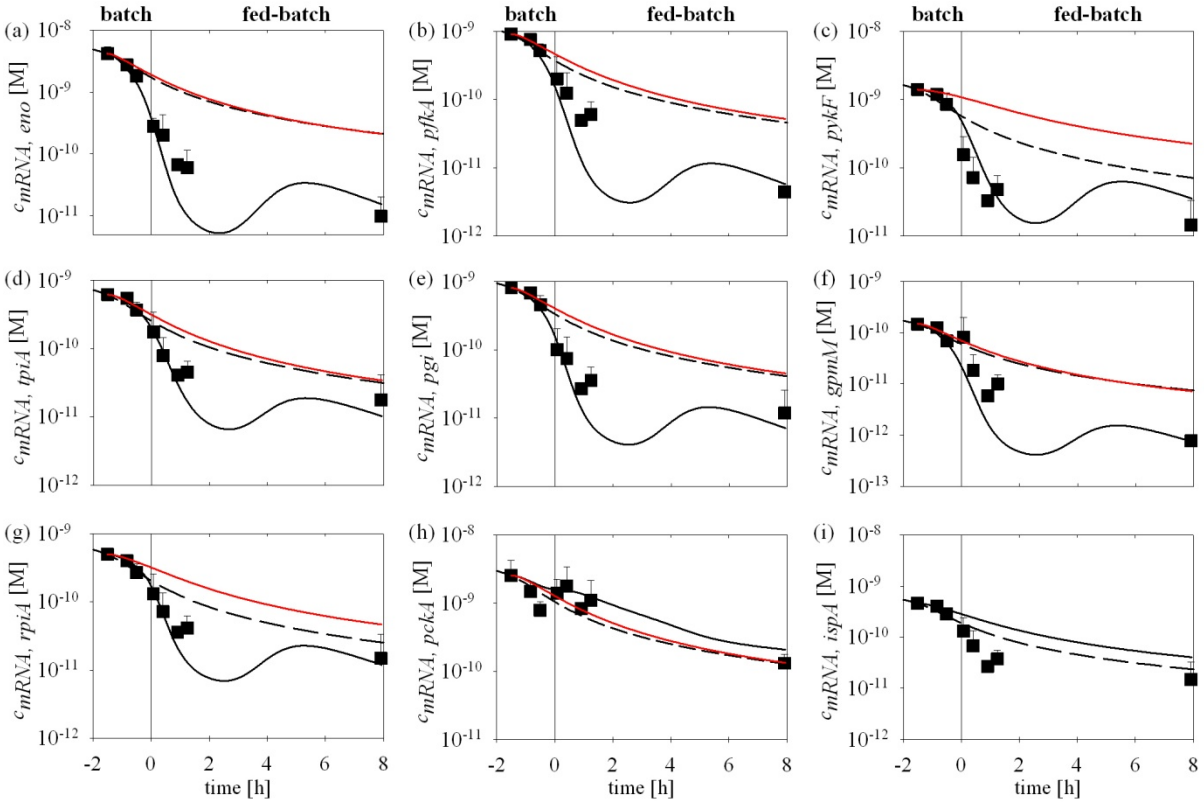


Fig 7-5 Dynamics of the intracellular concentrations of individual mRNAs of *cra* modulon genes during fed-batch cultivation of *E. coli* K-12 W3110. mRNA concentrations were experimentally determined using qPCR (■, error bars indicate SD) by Schuhmacher et al. (2009) and simulated using the presented, optimized model, with the kinetic parameters that were predicted from DNA-binding site sequences (black solid line). (broken line), data-based estimation, and, (red solid line), model-based prediction of the growth rate-dependent regulation ($\Phi_k^{neg} = 1$, $\Phi_k^{pos} = 0$; i. e. without Cra-dependent regulation, see text).

7.3.4 Cra- and Growth-Dependent Regulation of Central Carbon Metabolism Genes

The mRNA concentrations quantified by Schuhmacher et al. (2009) reveal a strong decrease even for constitutively expressed and positively regulated genes (Figures 7-5h, i). It is not possible, however, to quantitatively separate the contributions of Cra- and growth rate-dependent regulation to the changes in the transcription rate from expression data alone. The model presented in the current work was used to dissect this contributions of growth rate-dependent regulation [$\Phi_{RNAP,j}$, Equation (7-9)] and global regulation by the Cra protein [Φ_k^{neg} and Φ_l^{pos}, Equation (7-9)] (Figure 7-5). To simulate the effect of the growth rate-dependent regulation on the transcription dynamics, a new dynamic model – without Cra-dependent regulation [$\Phi_k^{neg}=1$ and $\Phi_k^{pos}=0$ in Equation (7-9)] – was implemented. This model-based, predicted dynamic behavior of the individual mRNA concentrations is depicted in Figure 7-5. To support this prediction by experimental data, the total RNA concentration was quantified during fed-batch cultivation (Figure 7-6). As a first approximation, the dynamic behavior of the individual mRNA concentrations ($c_{mRNA,i}$) should correlate with that of the total RNA mass concentration ($\rho_{totalRNA}$), if no regulator proteins, but only growth rate-dependent regulation, were involved in their transcription. If so, one can assume

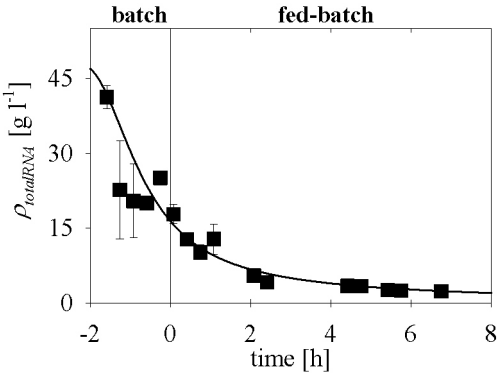


Fig. 7-6 Dynamics of the intracellular total RNA mass concentration, $\rho_{totalRNA}$, during fed-batch growth of *E. coli* K-12 W3110, determined using UV spectroscopy (■, error bars indicate SD) and approximated (solid line) by Equation (7-40), see Table 7-4.

$$\frac{c_{mRNA,i}(t)}{c_{mRNA,i}(t=t_0)} = \frac{\rho_{totalRNA}(t)}{\rho_{totalRNA}(t=t_0)}, \tag{7-46}$$

and thus, this growth rate-dependent dynamic behavior of an individual mRNA concentration can be calculated by

$$c_{mRNA,i}(t) = \frac{\rho_{totalRNA}(t)}{\rho_{totalRNA}(t=t_0)} \cdot c_{mRNA,i}(t=t_0). \tag{7-47}$$

The experimentally determined total RNA concentration was approximated using an analytical function [$\hat{\rho}_{totalRNA}(t)$; Equation (7-40), Table 7-4; and Figure 7-6] and the growth rate-dependent behavior of the individual mRNAs i was then estimated using Equation (7-47), which is depicted for each gene in Figure 7-5. The model-based prediction of growth rate-dependent regulation through the RNAP availability leads to a behavior equal to that estimated using the experimental data (Figure 7-5, red and dashed lines, respectively; *eno*, *pfkA*, *pckA*, *tpiA*, *pgi*, *gpmM*). Individual deviations of the predicted behavior from the (data-based) estimated behavior are due to the individual binding constants $K_{RNAP,j}$ (Table 7-3), degradation constants $k_{dg,mRNA,i}$ (Table 7-1) and transcription rate constants $k_{tc,init,i}$ (7-43). Nevertheless, the general agreement demonstrates that the presented approach is suitable for predicting the growth rate-dependent regulation and dissecting the regulation by superimposed global regulatory mechanisms (Figure 7-5).

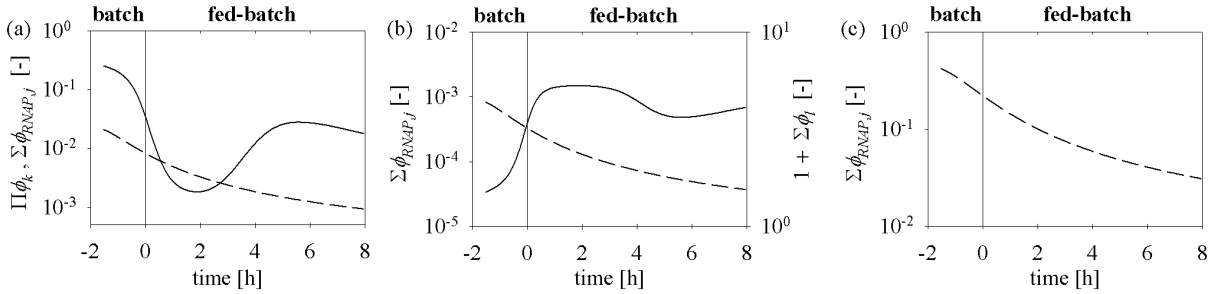


Fig. 7-7 Cra protein- and growth rate-dependent regulation of *cra* modulon transcription. The contributions of repression ($\prod_k \Phi_k^{neg}$) or activation by the Cra regulator protein ($1+\sum_l \Phi_l^{pos}$) (solid lines) and growth rate-dependent regulation ($\sum_j \Phi_{RNAP,j}$) (broken lines) to the transcription efficiency (η_j) were simulated for the genes (a) *eno* (3 repressor sites k; 3 promoters j), (b) *pckA* (1 activation site l; 1 promoter j) and (c) *ispA* (1 promoter j) using the presented, optimized model with the kinetic parameters that were predicted from DNA-binding site sequences.

The contributions of growth rate- and Cra-dependent regulation are depicted in Figure 7-7 for the genes *eno*, *pckA* and *ispA* – exemplary for negative, positive and no regulation by the Cra regulator protein, respectively. The growth rate-dependent regulation [$\Phi_{RNAP,j}$, Equation (7-9)] plays an important role for the transcription dynamics of the three genes (10-fold change in the transcription rate, Figure 7-7). These dynamics apply also for the other genes of the *cra* modulon (data not shown), and probably for most genes in *E. coli*, which are subject to regulation via RNA availability (Bremer et al., 2003; Klumpp and Hwa, 2008). The Cra-dependent regulation contributes a 100-fold change in the transcription rate by repression (e. g. the gene *eno*, Figure 7-7a) and a 4-fold change by positive regulation (*pckA*, Figure 7-7b). Moreover, Figure 7-7a demonstrates that during the first four hours of glucose-limited growth the Cra-dependent repression is stronger than the growth rate-dependent regulation (100- versus 10-fold change, Figure 7-7a). Thereafter, the two regulatory mechanisms contribute equally to the rate of transcription (both 10-fold change). This indicates that global regulatory systems such as the *cra*, *crp* and *relA*/*spoT* modulons majorly control the adaptive behavior during the first four hours, which is supported by the fact that the alarmone concentrations of cAMP and ppGpp increase fast in response to limitation and decrease to low levels only after four hours (Figure 2-4, Section 2.3.4) (Hardiman et al., 2007). RNAP availability is proposed to be one of the most important additional factors determining the mRNA expression level.

Two engaging questions are, which of the components of the regulatory mechanism (fbp, Cra, RNAP) would lead to the largest changes in Φ and therefore in the transcription rate, and, does this situation change after the aforementioned adaptation phase (t > 4 h)? This can be clarified by identification of the component with the largest control coefficient, normalized to the initial conditions (t_0) and to the conditions at the end of the cultivation (t_{end}), respectively. Strong effects of the DNA-binding constants $K_{2,k}$ and of the quasi-steady state chosen are observed (Figure 7-8). Small perturbations in the Cra protein or fbp concentrations generally lead to up to 4-fold amplification, independent of the quasi-steady state that is used for normalization (Figure 7-8). Large perturbations from the initial concentration of the Cra protein set off severe changes in the transcription rate (up to > 20-fold amplification; Figure 7-8), whereas for the fbp concentration the control coefficients decrease to lower than 1.5 (Figure 7-8). When choosing t_{end} for normalization, this relationship is reversed: transitions in the fbp concentration are amplified by a factor of up to > 50 and the control coefficients for the Cra protein decrease to lower than one (Figure 7-8). The reversed behavior at t_{end} supports the above hypothesis of different phases of the

adaptive behavior during the sequence of quasi-steady states in the fed-batch process. The fbp dynamics and the negative feedback regulation through the decreasing Cra protein concentration majorly determined the *cra* modulon transcription during the first four hours. This regulation persists also after 4 hours of cultivation due to the low fbp concentration (Figure 7-4). The low concentration of the Cra protein, which acts as a sensor with respect to glucose availability, leads to a higher sensitivity for changes in the fbp level. This suggests that the cells are poised for a potential new adaptation to a possible up-shift in the glucose concentration.

The value of the control coefficient with respect to changes in the RNAP concentration is approximately one (constant). This can be derived by simply comparing the values of the different terms in (7-13), which leads to $\Phi_{RNAP,j} \approx K_{RNAP,j} \cdot c_{RNAP,f}$, and thus, $C_{c_{RNAP}}^{\Phi_{RNAP,j}} \approx 1.0$. The transcription rate is also proportional to the gene concentration (7-12). Consequently, the respective control coefficient is $C_{c_{gene,i}}^{r_{tc,mRNA,i}} = 1.0 = const$. However, up to twofold changes of the gene concentration (7-30) must be considered during the applied fed-batch conditions (data not shown), which modulates the transcriptions rate by the same factor.

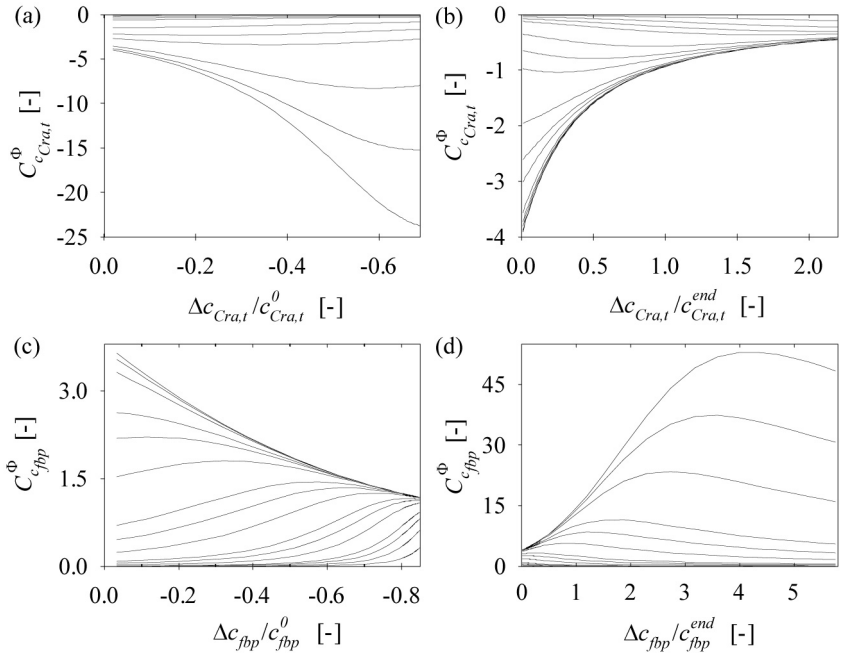


Fig. 7-8 Control coefficients. (a, b) $C_{c_{Cra,t}}^{\Phi_k^{neg}}$ and (c, d) $C_{c_{fbp}}^{\Phi_k^{neg}}$ are given as parameter plots of the DNA-binding constants $K_{2,k}$. (a, c) normalization to the initial conditions ($t = t_0$). (b, d) normalization to the conditions at the end of the fed-batch cultivation ($t = t_{end}$).

7.3.5 Assignment of New Genes to the *cra* Modulon

The genes *tpiA*, *pgi*, *gpmM*, and *rpiA* are not yet known to be regulated by regulator proteins (Figure 7-2). However, the experimental data and the predicted expression dynamics suggest that they should be assigned to the *cra* modulon (Figure 7-5). The new, potential Cra DNA-binding sites have high predicted binding constants (Table 7-2), which however must be verified by dedicated molecular biological experiments.

7.3.6 Fractional Change of Specific Cell Volume

A novel expression for formulating the dilution of intracellular compounds during cellular growth was introduced (7-4). The dilution rate, $r_{dl,mRNA,i} = [\mu + \omega] c_{mRNA,i}$, is important when modeling gene expression since its value may have the same magnitude as the transcription rate (e. g. for *tpiA*, *rpiA* and *pckA*; data not shown). Figure 7-1b demonstrates that at low growth rates $\mu \approx \omega$. Therefore, the fractional change of the specific cell volume, ω, becomes equally important for the dilution rate as the specific growth rate, μ. However, the variable ω must be considered, only when transient conditions are applied, i. e. only when the specific cell volume varies (for $\frac{dv}{dt} \neq 0$; see Appendix E).

7.4 Conclusions

The present work contributes a novel framework for mechanistic dynamic modeling microbial transcription considering the kinetics of multiple transcriptional regulation based on regulatory DNA-sequences. It is the first study to deal with modeling the regulation of central carbon metabolism genes by global regulatory systems in *Escherichia coli* involving the growth rate-dependent regulation via RNAP availability and other variables (cell volume, gene concentration) that are most important for simulation of the transcription rate during transient growth conditions. The kinetic parameters concerning the individual behavior of the mRNAs were predicted from the DNA-sequences (Tables 7-2 and 7-3). Simulation results were critically assessed using experimental data to test the hypotheses that (i) global regulation via the availability of the RNA polymerase can be predicted by evaluating the promoter DNA-sequences, and, (ii) global regulation through regulator proteins can be predicted by evaluating the DNA-binding sites of these regulators. The predictive power of the proposed approach is supported by the evidence of the simulated gene expression that describes the dynamics of the experimentally quantified mRNA concentrations during fed-batch cultivation of *E. coli* K-12 W3110 (Figure 7-5). The agreement of model predictions and

the further experimental data of the total RNA concentration demonstrate that the kinetic parameters derived from DNA-sequences are suitable for dissecting the superimposed regulation by global regulator proteins and the growth rate-dependent regulation via RNA polymerase availability (Figures 7-5).

Dynamic modeling the transcription of the *cra* modulon using the proposed method illustrates the physiological relevance of the central carbon metabolism regulation through the network topology proposed in the Chapters 2 and 3 (Hardiman et al., 2007; Lemuth et al., 2008), and reveals new insights into the dynamics of the underlying regulatory cascade. The model demonstrates the tight regulation of gene expression of the *cra* modulon during the applied *E. coli* fed-batch cultivation by either repression or activation of transcription through the Cra global regulator protein, and its inhibitor fructose 1,6-bis(phosphate) (fbp). The strong dynamics of the fbp concentration, which were quantified for the first time during transition from exponential to glucose-limited growth (Figure 7-4a), are proposed to be a key factor in signaling glucose limitation, and may equally signal potential limitations of other carbon sources under other environmental conditions. Moreover, the global regulation through the RNA polymerase availability was discussed as one of the most important regulatory mechanism that must be considered when modeling transcription dynamics. Analyses of the control strengths of the components involved in the two superimposed regulatory mechanisms indicate that the proposed decrease in the Cra protein concentration might poise the cells for the potential need of a new adaptation to further environmental changes.

The current work provides a predictive tool for reconstructing dynamic GRNs based on mechanistic rate equations, which may be particularly useful for dynamic modeling the interactions of regulatory and metabolic microbial networks such as the demonstrated impact of the dynamics of the metabolite fructose 1,6-bis(phosphate) on the transcription of the central carbon metabolism genes, and, in a second step, the effect of these changes on the metabolic reactions and fluxes. Thus, the approach might be valuable and constructive for metabolic engineering the contributions of regulator proteins to the global regulation of individual enzyme levels during biotechnical processes. The transcription dynamics could be predicted for *in silico* changes in the nucleotide sequence of a DNA-binding site of interest, prior to genetic engineering a producer strain and testing the new behavior under process conditions.

8 Conclusions and Outlook

The main objective of the present thesis was to comprehensively and quantitatively investigate the dynamics in the central carbon metabolism and its regulation in the bacterium *Escherichia coli* in the industrially widely applied fed-batch process using a systems biology approach. The study centers on experimental observations of the dynamics in signaling, transcription, metabolic fluxes and metabolite concentrations and involves stationary and dynamic mathematical modeling approaches.

One major goal of the thesis was the reconstruction of the network structure of the global regulation of the central carbon metabolism. A set of fed-batch processes with constant feeding rate were performed to provide the same conditions for all experimental examinations. The applied metabolic flux analyses and the comparison with complementary global transcription data demonstrated that this process strategy leads to a fundamental reorganization of the central carbon metabolism. In response to the continuous decrease in the supply of the carbon and energy source glucose, the fluxes in glycolysis, pentose phosphate pathway and biosynthesis dropped, whereas TCA cycle fluxes remained constant. Importantly, these changes result in a considerable reduction of biomass yield due to higher oxidation rates of the substrate. It is evident that the high fluxes in the TCA cycle lead to enhanced energy generation at the cost of the biomass yield. However, the complementary analysis of relative changes in transcript levels and fluxes suggests that fine tuning of the carbon flux through the pentose phosphate pathway and also the split of fluxes between the TCA cycle, the glyoxylate shunt and the (only recently discovered cyclic pathway) phosphoenolpyruvate-glyoxylate shunt (Fischer and Sauer, 2003) limits the rate of oxidative decarboxylation, and this supplies sufficient precursors for biosyntheses. Consequently, the carbon and energy balance of the central carbon metabolism will be optimized under the applied conditions.

The reconstruction of the regulatory network structure was supported by the agreement of the observed mRNA dynamics and the changes in the metabolic fluxes. Glycolysis, the TCA cycle and the glyoxylate shunt are proposed to be majorly regulated by the *cra* and *crp* modulons. Protein biosynthesis and the specific growth rate are regulated via the *relA*/*spoT* modulon. Further findings concern the synthesis of high-affinity transporters that guarantees maximal glucose influx and reduction of the energy-dependent chemotaxis, which are probably both driven by regulation through the *crp* modulon. The σ^S-mediated stress and starvation responses were found to be of only minor relevance. The intracellular concentrations of the two alarmones ppGpp and cAMP were demonstrated to accumulate in large quantities. A novel finding is the observed resetting of both alarmone levels. This adaptation was traced back, by means of a short literature review, to the key cell

components assembling several negative feedback loops. The model structure proposed is the first step towards dynamic modeling the regulation of the central carbon metabolism and enables to comprehensively explain the observed dynamic metabolic and regulatory responses.

The further analysis of the global transcription dynamics by quantification of the total RNA and rRNA contents, using a new method based on capillary gel electrophoresis and the quantitative detection using lacer-induced fluorescence, revealed a strong growth rate-dependent regulation of both rRNA and mRNA. This led to the conclusion that the regulatory network structure had to be extended by the growth rate-dependent regulation via the RNA polymerase availability.

The second goal of the current thesis was to provide an approach for dynamic modeling of transcriptional regulation. The *cra* modulon was selected as an example for introducing the method and quantitatively analyzing the transcription dynamics of the central carbon metabolism genes. A novel framework was developed that enables considering growth rate-dependent regulation, multiple regulation by different regulator proteins, multiple promoters and relevant growth rate-dependent variables, such as the cell volume and gene concentration. Most importantly, a new method was proposed, in which the required kinetic parameters are predicted from the DNA-sequences of the binding sites of regulator proteins, and of the RNA polymerase. The agreement of simulated and measured mRNA concentrations, determined by quantitative RT-PCR by Schuhmacher et al. (2009), demonstrates the predictive power of the presented approach. Furthermore, the prediction of the growth rate-dependent regulation was supported by the experimental data of the total RNA content.

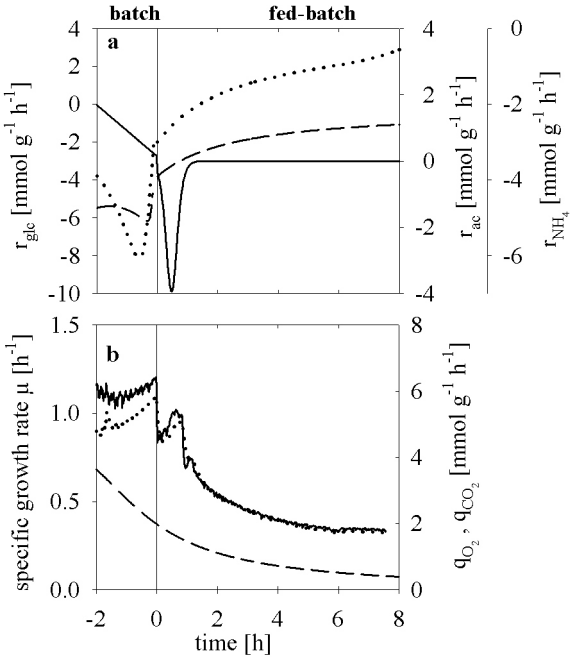
The dynamics of *cra* modulon transcription were analyzed using the developed approach. The Cra regulator protein-dependent regulation majorly determines the behavior of the mRNA concentrations. Yet, a strong growth rate-dependent regulation via the RNA polymerase concentration is superimposed. The metabolite fructose 1,6-bis(phosphate) is an inhibitor of the Cra regulator protein. The concentration of fructose 1,6-bis(phosphate) was quantified and its strong decrease during the transition to glucose-limited growth could be demonstrated for the first time. The model predictions and experimental findings support the hypothesis that fructose 1,6-bis(phosphate) plays a major role in the signaling of glucose availability and that glycolysis is repressed via the *cra* modulon during glucose-limited growth.

The model structure reconstructed in the current thesis may open up new possibilities for strain development concerning the supply and demand of precursors and energy. The relevance of the proposed network topology for biotechnological processes is supported by recent studies of engineered *E. coli* strains. Ow et al. (2007) examined the behavior of a *cra* deletion mutant strain bearing multicopy plasmids. The maintenance of multicopy plasmid DNA has been widely reported to retard cell growth due to the high precursor and energy demand (metabolic burden; Glick, 1995). Ow et al. (2007) demonstrated that in plasmid-bearing *cra* deletion mutants the expression of glycolysis enzymes is upregulated and the specific growth rate is enhanced by 20 %. It can be hypothesized from these observations that the high demand of precursors and energy for plasmid DNA replication leads to a decrease in the level of fructose 1,6-bis(phosphate), and, that the subsequent repression of transcription of glycolysis genes leads to a reduction in growth rate. In the *cra* deletion mutant strain, this regulation is disrupted. The higher enzyme levels may lead to an improved glycolytic flux, and thus, to an enhanced growth rate. In the present work it is proposed that the TCA cycle fluxes are high due to the Crp-dependent regulation during fed-batch growth, and, that this leads to a decrease of the biomass yield. One might hypothesize that altering the Crp-dependent regulation could enhance the biomass yield. Nanchen et al. (2008) recently demonstrated that deletion of the *crp* gene leads to an improvement of the biomass yield by more than 10 %.

However, disruption of a global regulatory system for the purpose of strain development will lead to numerous side effects that could be disadvantageous. Engineering transcription of only selected genes by changing the nucleotide sequences of the respective DNA-binding sites appears to be more purposeful. This approach could be guided by the presented dynamic modeling approach. The desired effects of genetic engineering the nucleotide sequences can be predicted *in silico*, before testing the new behavior under process conditions, which might speed up the development cycles of metabolic engineering.

Appendix

Appendix A – Time Courses of Experimentally Determined Rates



Time courses of experimentally determined rates, $r_m(t)$**.** Example from one individual fed-batch cultivation experiment. **a**, specific rates of glucose uptake $r_{glc}(t)$ (broken line), acetate excretion $r_{ac}(t)$ (solid line) and ammonia uptake $r_{NH_4}(t)$ (dotted line). **b**, specific growth rate $\mu(t)$ (broken line), oxygen consumption rate $q_{O_2}(t)$ (solid line) and carbon dioxide production rate $q_{CO_2}(t)$ (dotted line).

Appendix B – Approximations of Metabolite Concentration Time Courses

---

Biomass concentration [g l^{-1}]

$$\hat{c}_X\left(t<0\right)=0.001816+5.017\cdot\frac{1}{\left(1+0.2913\cdot\left(t-0.52\right)^2\right)e^{0.03027\cdot\left(t-0.52\right)^2}}$$

$$\hat{c}_X\left(t>0\right)=18.45-18.34\cdot\frac{1}{\left(1+e^{\left(1.322\cdot\left(t+10\right)+0.9999\cdot\ln\left(884.1\right)-17.28\right)}\right)^{0.1021}}$$

Glucose concentration [g l^{-1}]

$$\hat{c}_{glc}\left(t<0\right)=8.307-\frac{8.251}{\left(1+\exp\left(-18.72\cdot\left(t+10.07\right)+\ln\left(1794\cdot10^6\right)+165.5\right)\right)^{0.03253}}$$

Extracellular cAMP concentration [µmol g^{-1}]

$$\hat{X}_{cAMP,extra}\left(t<0\right)=0.9$$

$$\hat{X}_{cAMP,extra}\left(t>0\right)=5.841-\left(1.294-0.1624t\right)^{6.175}$$

(examples from individual fed-batch cultivations)

Appendix C – Stoichiometric Model

The stoichiometric model used in this work is based on the one of Chassagnole et al. (2002). Modifications of the reaction stoichiometry and the notation are given below. Reactions (transformers) of the model were named after the respective enzymes (according to genetic nomenclature; e. g. GapA (gene name *gapA*, enzyme glyceraldehyde 3-phosphate dehydrogenase-A). Deviating from this rule, enzyme complexes or polymerizations were named as follows:

- PTS (phosphoenolpyruvate (PEP)-dependent, sugar transporting phosphotransferase system): glc_F + pep → g6p + pyr.
- PDH (pyruvate dehydrogenase complex): pyr + coa + nad → accoa + co2 + nadh
- KGDH (2-oxoglutarate dehydrogenase complex): akg + coa + nad → succoa + nadh + co2
- SDH (succinate dehydrogenase): suc + fad → fum + fadh2
- bio (biomass synthesis according to Chassagnole et al. (2002))

Reactions in addition to the model by Chassagnole et al. (2002):

- Ahyd (ATP hydrolysis): h2o + atp → adp + p + h
- Pta (phosphate acetyltransferase): p + accoa → acep + coa
- AckAB (acetate kinase A): acep + adp → ace + atp
- T.acetate (transport of acetate): ace → ace_F

Modified reactions:

- IcdA (isocitrate dehydrogenase): isocit + nadp → akg + nadph + co2

Modified/additional metabolites:

- acep: acetylphosphate
- ace: acetate
- ace_F: extracellular acetate
- glc_F: extracellular glucose
- fbp: fructose 1,6-bis(phosphate)

Appendix D – Differentially Expressed Genes

Table A4-1 Flagellar / chemotaxis system

Gene	Gene product	T1	T2	T3	T4	T5	T6	T7	T8
Early class genes									
flhD	Transcriptional dual regulator SU	0.18	-0.20	-0.39	-0.05	-0.12	-0.33	-0.37	-0.50
Middle class genes									
fliA	Sigma 28	0.10	-0.26	-0.29	-0.64	-0.55	-0.68	-1.00	-2.42
ycgR	Involved in flagellar motility	0.08	0.02	-0.74	0.15	0.06	0.06	0.22	0.05
flgB	Basal body rod protein	0.07	-0.10	-0.40	-0.29	0.07	-0.20	-0.45	-1.32
flgD	Initiation of hook assembly	-0.22	-0.25	-0.35	-0.59	-0.41	-0.54	-0.98	-1.44
flgE	Flagellar hook protein	0.20	-0.56	-0.68	-0.52	-0.27	-0.58	-0.87	-1.96
flgC	Basal body rod protein	-0.09	-0.47	-0.40	-0.35	-0.50	-0.63	-0.99	-2.04
flgA	Flagellar biosynthesis	-0.08	-0.12	0.03	-0.31	0.09	0.36	-0.05	-0.71
flgJ	Flagellum-specific muramidase	-0.03	-0.45	-0.63	-0.52	-0.08	-0.08	-0.30	-0.61
flgG	Basal body rod protein	-0.09	-0.43	-0.95	-0.36	-0.02	-0.41	-0.41	-0.98
flgH	Flagellar L-ring protein	-0.09	-0.40	-0.48	-0.03	0.33	-0.03	-0.17	-0.59
flgK	Hook filament junction protein 1	-0.08	-0.44	-0.64	0.06	-0.04	-0.31	-0.39	-0.77
fliD	Flagellar cap protein	0.09	-0.08	-0.39	-0.25	-0.14	-0.15	-0.24	-0.25
fliS	Flagellum-biosynthesis protein	-0.04	-0.24	-0.59	-0.63	-0.34	-0.30	-0.63	-0.90
fliM	Motor switch protein	-0.01	-0.32	-0.54	-0.49	-0.33	0.18	-0.29	-1.04
fliE	Basal body protein	-0.36	-0.81	-0.51	-0.84	-0.47	-0.24	-0.39	-1.20
Late class genes									
motA	Flagellar motor complex component	-0.20	-0.19	-0.62	-0.22	-0.21	-0.83	-0.23	-0.27
cheR	Chemotaxis protein methyltransferase	0.09	-0.05	0.42	-0.40	-0.20	0.12	-0.02	0.11
Others									
fhiA	Flagellar system protein	0.03	0.06	0.07	0.47	0.08	0.31	0.21	0.12
Associated regulatory proteins									
csrA	Carbon storage regulator	0.02	-0.48	-0.29	-0.31	-0.50	-0.49	-0.07	-0.15
hns	Transcriptional dual regulator	0.13	0.32	0.42	0.63	0.09	1.21	0.28	0.06
lrhA	Transcriptional repressor	-0.07	0.09	0.36	0.31	0.16	0.32	0.42	□.33
ygiX	Transcriptional activator QseB	-0.20	-0.17	0.08	-0.41	-0.15	-0.26	-0.09	-0.03

Underlining indicates significantly differential expression.

The data of this table are results of a collaboration with the Institute of Technical Biochemistry (University of Stuttgart) and were published by Lemuth et al. (2008). The microarray analysis was performed by Karin Lemuth as part of her PhD thesis (Lemuth, 2006).

Table A4-2 Differentially expressed Sigma S core set genes

Gene	Gene product [a]	T1	T2	T3	T4	T5	T6	T7	T8
bolA	BolA transcriptional regulator, stress regulation	0,21	0,02	-0,65	-0,01	-0,03	-0,08	-0,05	-0,61
ygaM	Predicted protein	-0,10	0,08	-0,55	0,30	0,29	-0,31	0,16	0,12
yhiW	GadW transcriptional repressor	-0,09	-0,03	-0,51	-0,04	0,07	-0,24	-0,05	-0,34
yhiX	GadX transcriptional activator	0,08	-0,15	-0,33	-0,08	-0,12	-0,25	-0,18	-0,09
ychK	Hypothetical protein	0,21	-0,02	-1,47	0,15	0,06	-0,01	0,18	-0,36
ugpC	Glycerol-3-P ABC transporter, SU	-0,28	0,00	-0,56	0,10	-0,14	0,09	-0,05	-0,24
yhhA	Conserved protein	-0,01	-0,26	-0,16	-0,45	-0,36	-0,08	-0,28	-0,15
treA	Trehalase, periplasmic	-0,12	-0,02	-0,36	-0,01	0,02	0,03	0,07	-0,01
yhiO	Ethanol tolerance protein	0,05	0,15	0,62	-0,13	0,01	0,34	0,08	0,35
yedU	Hsp31 molecular chaperone, SU	0,10	0,49	0,27	0,51	0,06	-0,17	0,27	0,16
yceK	Predicted lipoprotein	0,20	0,14	0,42	0,02	0,18	-0,01	0,13	0,12
b1758	Predicted phosphatidyl transferase	-0,13	0,07	0,30	-0,14	-0,20	0,22	0,03	0,31
b2086	Conserved protein	-0,29	0,02	0,51	0,01	0,25	0,40	0,28	0,31
hdhA	7-alpha-hydroxysteroid dehydrogenase, SU	0,11	0,36	0,49	0,11	0,01	-0,10	0,05	0,41
yjeB	NsrR transcriptional repressor	0,10	0,34	0,06	0,11	-0,04	-0,19	0,03	0,11
otsB	Trehalose-6-phosphate phosphatase	0,00	0,30	0,06	-0,04	0,23	0,15	0,32	0,22
rpsV	30S ribosomal subunit protein S22	0,28	0,59	0,51	0,66	1,15	1,13	1,38	1,59
b0753	Putative homeobox protein	0,14	0,57	0,37	0,59	0,34	0,47	0,71	0,61
ygaF	Predicted enzyme	0,24	0,39	0,42	0,44	0,36	0,11	0,27	0,25
yeaG	Conserved protein	0,20	0,08	-0,03	0,47	0,24	0,02	0,04	0,04
ygaE	CsiR transcriptional repressor	0,13	0,27	0,10	0,07	0,11	0,47	0,02	0,14
yjgR	Putative enzyme with P-loop containing nucleotide triphosphate hydrolase domain	0,04	0,03	-0,24	0,27	0,26	0,67	-0,11	-0,04
ygaU	Predicted protein	0,07	0,08	0,26	0,12	0,14	0,30	0,24	0,37
ymgA	Hypothetical protein	-0,19	0,18	0,37	0,12	0,30	0,17	0,29	0,64
b2097	Fructose bisphosphate aldolase class I, SU	-0,03	0,13	0,39	0,39	0,44	-0,05	-0,06	0,36
yjgB	Predicted alcohol dehydrogenase	0,01	0,08	0,64	0,27	0,16	0,14	-0,23	0,14

[a] Weber et al. (2005a)

Underlining indicates significantly differential expression.

The data of this table are results of a collaboration with the Institute of Technical Biochemistry (University of Stuttgart) and were published by Lemuth et al. (2008). The microarray analysis was performed by Karin Lemuth as part of her PhD thesis (Lemuth, 2006).

Table A4-3 Differentially expressed stress-related genes

Gene	Gene product	T1	T2	T3	T4	T5	T6	T7	T8
ppiA	Peptidyl-prolyl cis-trans isomerase A, chaperoning, repair	-0,25	0,01	0,07	0,11	-0,28	-0,56	0,15	-0,04
degS	Inner membrane serine protease (sigmaE response)	-0,73	-0,07	-0,12	0,06	-0,07	-0,3	-0,17	-0,09
narJ	Chaperone subunit (δ subunit) of nitrate reductase 1	0,03	-0,02	-0,77	-0,04	-0,17	0,21	0,13	0,03
hslJ	Heat shock protein	0,07	-0,15	-0,58	-0,09	0,15	-0,27	0,01	-0,22
hdeA	Acid-resistance protein, possible chaperone	0,55	0,11	-0,38	-0,25	-0,34	-0,3	-0,35	-0,37
yabH	Chaperone with DnaK	0,07	-0,2	-0,07	-0,21	-0,57	-0,04	0,13	-0,53
stpA	H-NS-like DNA-binding protein with RNA chaperone activity	-0,04	-0,58	-0,4	-0,31	-0,29	-0,31	0,01	-0,72
bcp	Thiol peroxidase (detoxification)	0,19	-0,17	-0,53	-0,23	-0,42	-0,32	-0,04	-0,55
ydaA	Universal stress protein (resistance to UV irradiation)	0	-0,16	-0,39	-0,37	-0,22	-0,33	-0,14	-0,14
recF	Subunit of RecFOR complex, DNA recombination, rplication, repair	0,09	-0,27	-0,17	-0,45	-0,14	-0,31	-0,27	-0,17
hslV	Peptidase component of the HslVU protease, chaperoning, repair	0,08	-0,13	-0,19	-0,4	-0,2	-0,28	-0,33	-0,29
dnaJ	Chaperone, heat shock protein	0,05	-0,39	-0,31	-0,05	-0,11	-0,12	-0,63	0,1
ymdD	Protein required for succinyl modification of osmoregulated periplasmic glucans	0,03	0,27	0,11	0,04	-0,16	0,14	-0,13	0,55
mutL	Methyl-directed mismatch repair, SU	0,02	-0,02	0,18	0,22	0,03	-0,41	0,43	1,04
yhiO	Ethanol tolerance protein	0,05	0,15	0,62	-0,13	0,01	0,34	0,08	0,35
msrA	Protein-methionine-S-oxide reductase, chaperoning, repair	0,13	0,23	0,41	0,09	0	0,11	0,05	0,13
sodA	Superoxide dismutase, SU	-0,05	-0,07	0,24	0	0,04	0,17	0,01	0,07
ydeB	Inner membrane protein involved in multiple antibiotic resistance	0	-0,03	0,49	-0,1	-0,04	-0,04	0,02	0,13
uvrB	UvrABC Nucleotide Excision Repair Complex, SU	0,19	-0,07	0,94	0,07	-0,11	-0,06	-0,01	-0,24
yedU	Hsp31 molecular chaperone, SU	0,1	0,49	0,27	0,51	0,06	-0,17	0,27	0,16
ydgO	Integral membrane protein of SoxR-reducing complex	0,03	0,28	0,25	0,03	0,04	-0,04	0,02	0,01
cutC	Copper homeostasis protein, detoxification	0,18	0,08	0,33	0,44	0,5	0,25	0,55	0,74
ybeV	Hsc56, co-chaperone of Hsc62	0	0,43	0,55	0,17	0,26	0,43	0,16	0,28
cspI	Qin prophage; cold shock protein	-0,06	0,47	0,36	0,36	0,15	0,47	0,33	0,33
yeaA	Protein-methionine-S-oxide reductase, chaperoning, repair	0,29	0,45	0,69	0,64	0,47	0,44	0,66	0,93
ahpC	Alkylhydroperoxide reductase, SU, detoxification	0,17	0,44	0,6	0,61	0,33	0,27	0,32	0,49
phoH	ATP-binding protein, induced by P starvation	0,09	-0,02	0,06	-0,07	-0,2	0,53	-0,01	0,16
cspF	Qin prophage; cold shock protein	-0,01	-0,02	0,12	0,04	-0,07	0,69	-0,04	0,06
b1631	Member of SoxR-reducing complex	0,04	0,06	0,04	-0,07	0,05	0,36	0,04	-0,03
b0245	Toxin of the YkfI-YafW toxin-antitoxin pair	-0,08	0,16	0,22	-0,21	-0,08	0,51	0,07	0,04
sbmC	DNA gyrase inhibitor	0,02	-0,01	0,09	0,22	0,43	0,16	0,03	-0,15
mutH	MutHLS complex, SU, methyl-directed mismatch repair	-0,01	0,06	0,03	0,48	0,11	-0,28	0,13	-0,23
rpoE	Sigma E factor	0.22	0.09	0.23	0.65	0.79	0.52	0.95	0.83

Underlining indicates significantly differential expression.

The data of this table are results of a collaboration with the Institute of Technical Biochemistry (University of Stuttgart) and were published by Lemuth et al. (2008). The microarray analysis was performed by Karin Lemuth as part of her PhD thesis (Lemuth, 2006).

Table A4-4 Differentially expressed Sigma S-regulated genes

Gene	Gene product [a]	T1	T2	T3	T4	T5	T6	T7	T8
himD	IHF transcriptional dual regulator, SU	0,02	0,20	0,07	0,50	0,22	0,10	-0,01	-0,21
htrE	Putative outer membrane porin protein involved in fimbrial assembly	-0,21	-0,10	0,49	0,09	0,35	0,49	0,14	0,42
proW	Proline ABC transporter UE	-0,07	-0,05	-0,12	0,01	-0,09	-0,11	0,18	0,48
csiE	Stationary phase inducible protein	0,07	0,09	0,01	0,36	0,31	0,15	0,15	0,36
rpsV	30S ribosomal subunit protein S22	0,28	0,59	0,51	0,66	1,15	1,13	1,38	1,59
yehX	YehW/YehX/YehY/YehZ ABC transporter subunit	-0,16	0,00	0,25	0,03	-0,01	0,08	-0,05	0,42
yehY	YehW/YehX/YehY/YehZ ABC transporter subunit	0,10	-0,21	0,10	0,20	0,43	0,73	-0,05	0,01
yeiL	Transcriptional activator	0,00	0,18	0,42	0,15	0,14	0,23	-0,02	0,27
ygaF	Predicted enzyme	0,24	0,39	0,42	0,44	0,36	0,11	0,27	0,25
yhiO	Ethanol tolerance protein	0,05	0,15	0,62	-0,13	0,01	0,34	0,08	0,35
proV	Proline ABC transporter UE	-0,32	-0,40	-0,08	-0,12	-0,11	-0,11	0,36	1,14
hdeA	Acid-resistance protein, possible chaperone	0,55	0,11	-0,38	-0,25	-0,34	-0,30	-0,35	-0,37
treA	Trehalase, periplasmic	-0,12	-0,02	-0,36	-0,01	0,02	0,03	0,07	-0,01
recF	RecFOR complex, SU	0,09	-0,27	-0,17	□0,4	-0,14	-0,31	-0,27	-0,17
bolA	BolA transcriptional regulator, stress regulation	0,21	0,02	-0,65	-0,01	-0,03	-0,08	-0,05	-0,61
yhiX	GadX transcriptional activator	0,08	-0,15	-0,33	-0,08	-0,12	-0,25	-0,18	-0,09
pqiB	Paraquat-inducible protein B	-0,11	-0,36	-0,40	-0,22	-0,27	-0,57	-0,38	-0,42
caiC	Carnitine-CoA ligase / crotonobetaine-CoA ligase	0,07	-0,32	0,04	0,06	-0,33	-0,83	-0,14	0,02
ftsA	Essential cell division protein	0,07	-0,05	-0,09	0,01	-0,45	-0,32	-0,18	0,03
galT	Uridylyltransferase, galactose metabolism	0,01	0,13	-0,56	-0,04	0,13	-0,21	0,06	-0,14

[a] Source: Ecocyc.org database.

Underlining indicates significantly differential expression.

The data of this table are results of a collaboration with the Institute of Technical Biochemistry (University of Stuttgart) and were published by Lemuth et al. (2008). The microarray analysis was performed by Karin Lemuth as part of her PhD thesis (Lemuth, 2006).

Appendix E – General Material Balance Equation for Intracellular Compounds

The balance equation for an intracellular compound j (dimension [mol $(l_{cytosol})^{-1}$]) is derived from the general material balance of the fed-batch process (Equation (7-48); dimension [mol]).

$$\frac{d}{dt}\left(c_j c_X \nu_X V_R\right) = \sum_i \left(\nu_{ij} r_i\right) \cdot c_X \nu_X V_R \tag{7-48}$$

c_j	[mol·$(l_{cytosol})^{-1}$]	intracellular concentration of the compound j
c_X	[(g dry weight)·$(l_{reactor})^{-1}$]	biomass concentration
ν_X	[$l_{cytosol}$·(g dry weight)$^{-1}$]	specific cell volume
V_R	[$l_{reactor}$]	bioreactor volume
$\sum_i \nu_{ij} r_i$	[mol·$(l_{cytosol}$·s$)^{-1}$]	sum of the reactions rates i, where the compounds j are reactants or products; ν_{ij}, stoichiometric coefficient.

The specific cell volume, ν_X, varies with the growth rate (Section 2.3.4) (Hardiman et al., 2007):

$$\hat{\nu}_X(\mu) = \frac{0.4860 \cdot 2^{(1.144 \cdot \mu)}}{\left[-0.636 + 0.635 \cdot 2^{(0.718 \cdot \mu)}\right] \cdot 1000} \tag{7-49}$$

The application of the product rule of differentiation leads to

$$c_X \nu_X V_R \frac{dc_j}{dt} + c_j \nu_X V_R \frac{dc_X}{dt} + c_j c_X V_R \frac{d\nu_X}{dt} + c_j c_X \nu_X \frac{dV_R}{dt} = \sum_i \left(\nu_{ij} r_i\right) \cdot c_X \nu_X V_R . \tag{7-50}$$

$$\frac{dc_j}{dt} + c_j \frac{dc_X}{c_X dt} + c_j \frac{d\nu_X}{\nu_X dt} + c_j \frac{dV_R}{V_R dt} = \sum_i \left(\nu_{ij} r_i\right) \tag{7-51}$$

$$\frac{dc_j}{dt} = \sum_i \left(\nu_{ij} r_i\right) - \left(\frac{dc_X}{c_X dt} + \frac{dV_R}{V_R dt} + \frac{d\nu_X}{\nu_X dt}\right) \cdot c_j \tag{7-52}$$

The latter expression of Equation (7-52) is the dilution rate of the intracellular compound j. After deriving the specific growth rate, μ, from the biomass balance for the fed-batch process,

$$\frac{d}{dt}\left(c_X V_R\right) = \mu c_X V_R \tag{7-53}$$

$$V_R \frac{dc_X}{dt} + c_X \frac{dV_R}{dt} = \mu c_X V_R \tag{7-54}$$

$$\mu = \frac{dc_X}{c_X dt} + \frac{F}{V_R} \text{ [h}^{-1}\text{] with the feed rate } F = \frac{dV_R}{dt}, \tag{7-55}$$

and, defining the fractional change of the specific cell volume,

$$\omega(\mu) = \frac{d\nu_X(\mu)}{\nu_X(\mu) \cdot dt} \text{ [h}^{-1}\text{]}, \tag{7-56}$$

the material balance for the fed-batch cultivation reads:

$$\frac{dc_j}{dt} = \sum_i \left(\nu_{ij} r_i\right) - \left(\mu + \omega(\mu)\right) \cdot c_j . \tag{7-57}$$

Note that $\omega = 0$ for any steady-state conditions, such as for exponential feeding. The application of a constant feed rate, however, leads to a fractional change of the cell volume.

Appendix F – Enhancement of Transcription Initiation

The enhancement of the transcription rate (7-6) by an activator protein is formulated as the raise of the transcription initiation rate constant (7-58), (7-59). When there are multiple DNA-binding sites (l) of activator proteins, the enhancement is assumed to be additive (7-59).

$$k_{tc,init,j}^{app} = k_{tc,init,j} + \sum_{l} \Delta k_{tc,init,j}^{l} \varphi_l \tag{7-58}$$

with

$$\Delta k_{tc,init,j}^{l} = k_{tc,init,j}^{enh,l} - k_{tc,init,j} \tag{7-59}$$

$k_{tc,init,j}^{enh,l}$ is the rate constant, when the activator protein is bound only to the DNA-binding site l, and $k_{tc,init,j}$, when no regulator protein is bound. The difference is multiplied by the probability that the regulator is bound, φ_l, and added up to obtain the apparent rate constant $k_{tc,init,j}^{app}$. Thus, the resulting rate constant is variable and depends on the present concentration of regulators (and effectors).

Equation (7-58) can be converted to

$$k_{tc,init,j}^{app} = k_{tc,init,j} \left[1 + \sum_{l} (\delta_l - 1) \varphi_l \right], \tag{7-60}$$

using

$$\delta_l = \frac{k_{tc,init,j}^{enh,l}}{k_{tc,init,j}}, \tag{7-61}$$

which is the enhancement factor of binding of the regulator protein to the DNA-binding site l.

Appendix G – Dynamic Models of Transcription (Literature Review)

Table A7-1 Formulation of the transcription rate in mechanistic dynamic models (deterministic, non-linear, continuum)

Authors	System studied [a]	Rate constant (number)	c_{gene}	Implementation of regulation (number of sites)				Overlap of sites	Growth rate-dependence	Dimension for mRNA
				RNAP/ promoter	Multiple promoters	Repres-sion	Activation			
Yagil and Yagil (1971)	lacZYA; LacI [b]	-	-	No	-	φ (1) [b]	No	No	-	-
Gondo et al. (1978)	Generally applicable (diauxie)	k_{tc} (1)	No	F [c, d]	No	φ (1) [c]	F [c, d]	No	$r_{tc} \sim (\mu + b)$	Mass fraction [g (g dry weight)$^{-1}$]
Roels (1978)	*lacZYA*; LacI, Crp (diauxie)	$r_{tc}^{\max}$ (1)	No	No	No	φ (1) [c]	φ (1) [c]	No	No	Mass fraction [g (g dry weight)$^{-1}$]
Harder and Roels (1982)	*lacZYA*; LacI, Crp (diauxie)	$r_{tc}^{\max}$ (1)	No	No	▪ ▪	φ (1) [c]	φ (1) [c]	No	$r_{tc} \sim \frac{a\mu + b}{a\mu_{\max} + b}$	Mass fraction [g (g dry weight)$^{-1}$]
Lee and Bailey (1984b)	*lacZYA;* LacI, Crp (variable copy numbers)	k_{tc} (1)	Yes	φ (1) [c]	No	φ (1) [c]	$[1+(\delta-1)\varphi]$ (1) [e]	No	$c_{gene} = f(\mu)$	Intracellular conc. [mol (l cytosol)$^{-1}$]
Wong et al. (1997)	*lacZYA*; *lacI*, LacI, Crp (diauxie)	$k_{tc,i}$ (two individual constants)	Yes	φ (2) [c]	No	φ (3) [c], DNA looping included	φ (1) [c]	No	No	[mol (g dry weight)$^{-1}$]

(Table continued on next page.)

(Table continued from previous page.)

Authors	System studied [a]	Rate constant (number)	c_{gene}	Implementation of regulation (number of sites)				Overlap of sites	Growth rate-dependence	Dimension for mRNA
				RNAP/ promoter	Multiple promoters	Repres-sion	Activation			
Kremling and Gilles (2001), Kremling et al. (2001)	*lacZYA*, *crp*, *cyaA*, *ptsG*, "AT" gene; LacI, Crp (diauxie)	k (one for lumped transcrip-tions and translations)	Yes	φ (3) [c, f]	No	表 EMBED Equation.D SMT4 $\hat{\varphi}$	$\overline{\varphi}$ (4) [c, f]	No	No	[mol (g dry weight)$^{-1}$]
This work	*cra* modulon; Cra (glc-limited fed-batch)	$k_{tc,i}$ [g] (nine individual constants)	Yes	φ_j (14) [c, g]	Yes [c, g]	φ_j (10) [c, g]	$\left[1+\left(\delta-1\right)\varphi\right]$ (1) [c, e]	No	$c_{gene}=f\left(\mu\right)$, $c_{RNAP}=f\left(\mu\right)$	Intracellular conc. [mol (l cytosol)$^{-1}$]

[a] Transcription units studied; regulator proteins considered (perturbation used)
[b] The equilibrium of the repressor-operator binding reaction was studied (not transcription dynamics) and the binding probability, φ, was formulated. The other authors listed in the table, used this approach and applied it to other regulatory systems.
[c] Formulation according to Yagil and Yagil (1971).
[d] F ("catabolite repression index") relates the number of quarternary complexes of Crp-cAMP and RNAP bound at the promoter to the total number of promoter DNA.
[e] Enhancement of transcription through binding of an activator protein was formulated using an enhancement factor, δ, and the binding probability.
[f] To implement hierarchical regulation, the probabilities were nested by assuming that the sequence of binding reactions of the involved proteins is ordered (probability of the one reaction is used as "input" for the next reaction: $\hat{\varphi}=f\left(\overline{\varphi}\right)$ and $\overline{\varphi}=g\left(\varphi\right)$).
[g] The model comprises nine genes (i), three of which are transcribed from multiple promoters. In total, 14 promoters (j) are implemented, considering the interactions of the RNAP with the individual promoters (φ_j). For multiple promoters regulating transcription of the same gene, a mean initiation rate constant, $k_{tc,i}$, is used.

Appendix H - Determination of Fructose 1,6-bis(phosphate)

Enzymatic determination of fbp was performed using the enzymes triosephosphate isomerase (Roche Diagnostics), glycerol-3-phosphate dehydrogenase and aldolase (Sigma-Aldrich) according to Michal (1984). A slow, fructose 1,6-bis(phosphate)-independent zero-order side reaction was observed during the analyses, which was eliminated by recording the absorption time profile at 340 nm and extrapolation according to Henniger (1998). Spiking of the samples with a fructose 1,6-bis(phosphate) standard solution resulted in a decrease of analysis time: The elevated substrate concentration enhanced the rate of the enzymatic conversion, but did not change the side reaction rate. A further modification of the protocol of Michal (1984) concerns the concentration of the NADH solution (30 mM), to obtain a larger measurement range. To quantify the intracellular concentration of fructose 1,6-bis(phosphate) (fbp), samples from the bioreactor were quenched and extracted using perchloric acid according to Section 2.2.3 (Hardiman et al., 2007). Potential metabolite loss due to leakage (Bayer, 1967; Britten and Mc, 1962; Leder, 1972; Wittmann et al., 2004) prior to analysis is avoided using this approach since the intracellular and extracellular compounds are not separated. The procedure may be applied for determination of the intracellular concentration, when the metabolite is not present in the extracellular space. Bolten et al. (2007) and Taymaz-Nikerel et al. (2009) demonstrated that sugar phosphates may be found in the supernatant of *E. coli* suspensions after filtration applying pressure. However, Moses and Sharp (1972) demonstrated that the ratio of the intra- and extracellular fbp is 20,000 : 1 through radiochemical labeling of the sugar phosphates and gently filtering (avoiding clogging of the filters). In the fed-batch cultivation applied in the current work the minimal ratio of the intra- and extracellular volume is roughly $1:150$ ($\rho_x : c_x$). Therefore, the intracellular amount is expected to be at least 100-fold higher than that of extracellular fbp in the samples analyzed in the current work. To confirm that the extracellular amount of fbp is negligible an in-line filtration probe with a 0.2 µm-pore-size ceramic membrane (FIPS sampling probe, Flownamics, Madison, WI, USA) was used. This ensured cell-free sampling, avoiding metabolite leakage that might occur during manual filtration and the concomitant application of pressure. The concentration of extracellular fbp, if any, was below the detection limit, which agrees with previous works using LC-MS analysis (Schaub and Reuss, 2008; Schaub et al., 2006).

Appendix I - Pattern Matching

The computational tool Virtual Footprint and the PRODORIC database (http://www.prodoric.de/vfp; Munch et al., 2005) were used for predicting potential DNA-binding sites of the Cra regulator protein. The pattern matching using default settings (changes of default settings for the *pyrG-eno* operon: sensitivity/threshold, 0.9) and the position weight matrix MX000125 (FruR, synonymous with Cra) resulted in 278 matches of DNA-sequences within the *Escherichia coli* K-12 W3110 genome, 29 of which could be potential Cra-dependent regulatory sites for genes of the central carbon metabolism. Eight genes are not known to be regulated by other regulator proteins (EcoCyc; Keseler et al., 2009). To avoid regulatory interactions, only these genes were implemented in the model (Figure 7-2 of the main text). 11 DNA-binding sites were found in upstream regions and one in the coding region (Table 7-2, main text). The potential binding sites were re-examined by calculating the scores (according to Table A7-2, number 5) of the possible orientations (+/- strand, forward/reverse) and shuffling the sequences. The sequences yielding the maximum score were chosen (Table 7-2, main text).

Appendix J – Workflow Prediction of DNA-binding Constants

Table A7-2 Workflow for prediction of DNA-binding constants and calculation of binding probabilities

1	Choose a reference DNA-binding site sequence, $DNAbs_{ref}$.
2	Determine (experimentally) the binding constant for this reference sequence, $K_{2,ref}$.
3	Choose a set of highly conserved DNA-binding site sequences of a regulon/modulon.
4	Calculate the specificity matrix for the set of DNA-binding sites, $\mathbf{SpM}(m \times n)$; matrix entries: $a_{m,n} = \frac{f_{m,n}}{p_m}$, with $f_{m,n}$, frequency of nucleotide m (A, C, G or T) at position n of the sequence and p_m , genomic frequency of each nucleotide.
5	Calculate the specificity score, $score_k = \prod_n a_{m,n}^k$, of the binding site of interest, $DNAbs_k$, with $a_{m,n}^k$, product of the matrix entries that correspond to the nucleotides m at positions n of the sequence k .
6	Calculate the specificity score of the reference site, $score_{ref} = \prod_n a_{m,n}^{ref}$.
7	For the given binding constant of a reference sequence, $K_{2,ref}$, the unknown individual binding constant, $K_{2,k}$ [a], can now be predicted: $K_{2,k} = K_{2,ref} \frac{score_k}{score_{ref}}$.
8	The predicted binding constant can be used for calculating the binding probability φ_k [a].

[a] The approach can equally applied to calculate DNA-binding constants of repressor and activator proteins ($K_{2,k}$ or $K_{2,l}$, respectively) as well as RNA polymerase-promoter binding constants ($K_{RNAP,j}$). The respective binding probabilities are represented by φ_k , φ_l or $\Phi_{RNAP,j}$.

Appendix K – Cra DNA-Binding Sites for SpM Calculation

Table A7-3 DNA-binding site sequences for calculating the specificity matrix of the *cra* modulon

TU [a]	DNA-sequence [b]	Genomic position [b]	Genomic orientation [b]
edd-eda	ACTGAAACGTTTTTGC	1932726	-
mtlADR	ACTGAATCGGTTAACT	3769785	+
epd-pgk-fbaA	GCTGAAGCGTTTCAGT	3071833	-
pckA	GGTGAATCGATACTTT	3530170	+
ppsA	GGTGAATCGTTCAAGC	1785223	+
aceBAK	GCTGAATCGCTTAACG	4212807	-
ptsHI-crr	GCTGAATCGATTTTAT	2531565	+
icdA	GCTGAATCGCTTAACC	1194100	+
pykF	CTTGAATGGTTTCAGC	1753501	+
adhE	GCTGAAAGGTGTCAGC	1297593	-
pfkA	CCTGAATCAATTCAGC	4105011	+
nirBDC-cysG	GCTGAATCGTTAAGGT	3491601	+

[a] TU = Transcription unit that is regulated by the Cra protein.

[b] Genomic information for *E. coli* K-12 MG1655. The genomic position of the first nucleotide on the + strand is given. 12-nt sequences of the *E. coli* K-12 MG1655 position weight matrix MX000125 for the Cra regulator protein were retrieved from the Prodoric Database (Munch et al., 2003). The sequences were extended by 4 nts to generate 16-nt sequences as proposed by Ramseier et al. (1995), reversed to examine the four possible orientations (+/- strand, forward/reverse), and, shuffled, to obtain optimal matches to the consensus sequence of Ramseier et al. (1995).

Appendix L – Sequence Logos

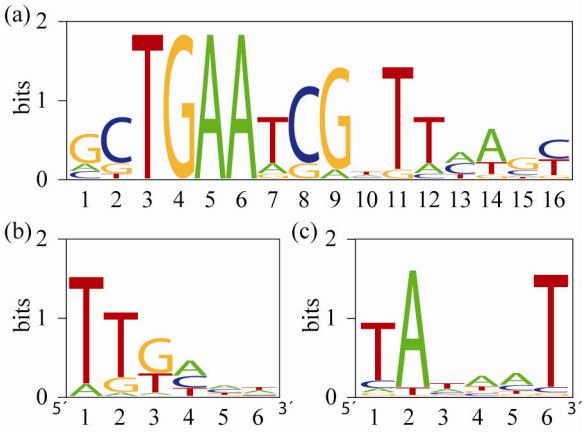


Fig. A7-1 **Sequence logos** for (a) the Cra DNA-binding site sequence (from 12 highly conserved sequences; see Table A7-3), (b) for the promoter -35 and (c) -10 hexamers [each from 401 hexamers derived by Shultzaberger et al. (2007)]. Logos were generated using WebLogo (Crooks et al., 2004).

Appendix M - Spacer and Gap Penalties

The effect of spacer length on *tac* promoter activity was quantified *in vivo* by Typas and Hengge (2006). Therefore, spacer penalties, $\alpha_{spacer,j}$ (between 0 and 1), were used for the different spacer lengths (Table A7-4). Hawley and McClure (1983) and Shultzaberger et al. (2007) demonstrated in their statistical studies that the length of the gap (see Figure 7-3c of the main text) is between 3 and 10 nts (6 nt is optimal). Therefore, a gap penalty was introduced: $\alpha_{gap,j} = 1$, for gaps between 3 and 10 nts; otherwise $\alpha_{gap,j} = 0$.

Table A7-4 Spacer penalties

Spacer length [a]	Spacer penalty (α_{spacer}) [b]
15	0.09
16	0.25
17	1
18	0.18
19	0.04
20	0.08

[a] The spacer lengths are given according to Hawley and McClure's notation (1983).

[b] Spacer penalties = promoter activities of *tac*-promoter derivates relative to the 17-bp spacing promoter *tac17*: *tac15* (0.062/0.688 = 0.09), *tac16* (0.173/0.688 = 0.25), *tac17* (0.688/0.688 = 1), *tac18* (0.127/0.688 = 0.18) and *tac19* (0.027/0.688 = 0.04); data were taken from Typas and Hengge (2006), who determined the *in vivo* activities of chromosomal copies of the synthetic *tac* promoters (Figure 2, therein). The activity for the *tac20* promoter was determined by Aoyama and Takanami (1988) relative to the *tac17* promoter at the natural superhelical density σ = - 0.05.

Appendix N - Growth Rate-Dependent Variables

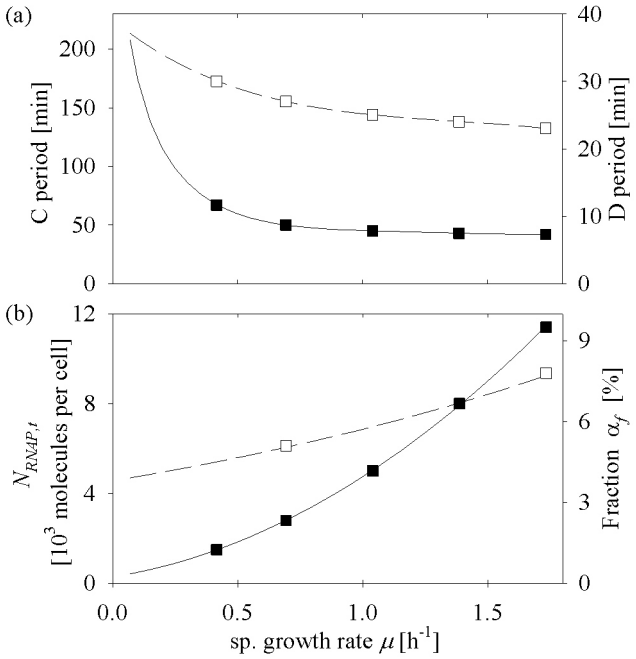


Fig. A7-2 **Growth rate-dependent variables.** (■), C period or chromosome replication time and total number of RNAP molecules, $N_{RNAP,t}$. (□), D period or time period between termination of replication and cell division and fraction of free, cytosolic RNAP. The approximations are depicted as line plots (cf. Table 7-4 of the main text).

Appendix O - Model Indentifiers

The model identifiers follow the microbial notation, such as *lacZp*, *lacZo*, *lacZa* for the promoter, operator DNA-sequences and the activator DNA-binding site of the *lac* operon). However, these cannot be used for modeling multiple regulator proteins regulating the same operon.

Table A7-5 Identifiers for concentrations of model components

Identifier	Example (description)	Denotation
xyzAd	lacZd (*lacZ* gene)	Gene
xyzAp	lacZp1, lacZp2, ... (multiple promoters of the *lac* operon; numbers are left out for single promoters)	Promoter [a]
xyzApabcD	lacZp1crp1, lacZp1crp2, ... (multiple DNA-binding sites of the Crp regulator protein at the *lacZp1* promoter)	Activator or repressor DNA-binding sites [b]
xyzAr	lacZr (*lacZ* mRNA)	mRNA

[a] Nomenclature according to the EcoCyc database (Keseler et al., 2009). Multiple promoters are numbered in the opposite direction of genomic orientation of the operon starting with "1" for the proximal promoter.

[b] In the EcoCyc database (Keseler et al., 2009) each DNA-binding site is assigned to one promoter. This nomenclature is taken on. Note, that the regulator protein name is lower case in the identifier since it specifies a DNA-sequence.

Table A7-6 Identifiers for model reactions

Identifier	Example (description)	Denotation
rtcxyzA	rtclacZ (total *lacZ* transcription rate)	Total transcription rate [a]
rtcxyzBxyzAp	rtclacYlacZp1, rtclacYlacZp2, ... (transcription rates of the *lacY* gene from the *lacZp* promoters)	Individual transcription rate (from multiple promoters) [a]
rdgxyzAr	rdglacZr (degradation rate of *lacZ* mRNA)	Degradation rate
rdlxyzAr	rdllacZr (dilution rate of *lacZ* mRNA)	Dilution rate

[a] The total transcription rate is the sum of the individual transcription rates from multiple promoters.

Table A7-7 Identifiers for model parameters

Identifier	Example (description)	Denotation
KxyzApAbcD	KlacZp1Crp (DNA-binding constant of the reaction of the Crp protein and its binding site at the *lacZp1* promoter)	DNA-binding constant [a]
PhixyzApAbcD	PhilacZp1Crp (probability of finding the *lacZp1* promoter bound by the Crp protein)	Probability of binding (φ) [a]
PhiappxyzApAbcD	PhiapplacZp1Crp (contribution of the Crp-dependent regulation to transcription from the *lacZp1* promoter)	Contribution to the transcription efficiency (Φ) [a] (apparent probability of transcription initiation)
kdgxyzAr	kdglacZr (degradation constant of *lacZ* mRNA)	mRNA degradation constant
deltaxyzApAbcD	deltapckApCra (enhancement factor of Cra-dependent activation of transcription from the *pckAp* promoter)	Enhancenment factor [a]

[a] In the EcoCyc database (Keseler et al., 2009) each DNA-binding site is assigned to one promoter. The identifiers of the parameters are named according to this promoter, even when the same binding site is used to regulate further promoters.

References

Agilent (2003). RNA LabChip kits - Fast quality control of RNA with mnimal sample consumption. In Agilent Technologies, Publication Number: 5989-0229EN (Palo Alto, CA, USA), pp. 2.

Al-Mahrouki, A.A., and Krylov, S.N. (2005). Calibration-free quantitative analysis of mRNA. Anal Chem *77*, 8027-8030.

An, G., Justesen, J., Watson, R.J., and Friesen, J.D. (1979). Cloning the *spoT* gene of *Escherichia coli*: identification of the *spoT* gene product. J Bacteriol *137*, 1100-1110.

Aoyama, T., and Takanami, M. (1988). Supercoiling response of *E. coli* promoters with different spacer lengths. Biochim Biophys Acta *949*, 311-317.

Arnold, S., Siemann-Herzberg, M., Schmid, J., and Reuss, M. (2005). Model-based inference of gene expression dynamics from sequence information. Adv Biochem Eng Biotechnol *100*, 89-179.

Artsimovitch, I., Patlan, V., Sekine, S., Vassylyeva, M.N., Hosaka, T., Ochi, K., Yokoyama, S., and Vassylyev, D.G. (2004). Structural basis for transcription regulation by alarmone ppGpp. Cell *117*, 299-310.

Babu, C.V.S., Song, E.J., Babar, S.M., Wi, M.H., and Yoo, Y.S. (2006). Capillary electrophoresis at the omics level: towards systems biology. Electrophoresis *27*, 97-110.

Barker, M.M., Gaal, T., and Gourse, R.L. (2001a). Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. J Mol Biol *305*, 689-702.

Barker, M.M., Gaal, T., Josaitis, C.A., and Gourse, R.L. (2001b). Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. J Mol Biol *305*, 673-688.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., and Edgar, R. (2007). NCBI GEO: mining tens of millions of expression profiles - database and tools update. Nucleic Acids Res *35*, D760-765.

Bayer, M.E. (1967). Response of cell walls of *Escherichia coli* to a sudden reduction of the environmental osmotic pressure. J Bacteriol *93*, 1104-1112.

Becker, A.K., Zeppenfeld, T., Staab, A., Seitz, S., Boos, W., Morita, T., Aiba, H., Mahr, K., Titgemeyer, F., and Jahreis, K. (2006). YeeI, a novel protein involved in modulation of the activity of the glucose-phosphotransferase system in *Escherichia coli* K-12. J Bacteriol *188*, 5439-5449.

Berg, H.C. (2003). The rotary motor of bacterial flagella. Annu Rev Biochem *72*, 19-54.

Berg, O.G., and von Hippel, P.H. (1987). Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. J Mol Biol *193*, 723-750.

Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S., and Cohen, S.N. (2002). Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci U S A *99*, 9697-9702.

Bhatnagar, D., and Bhattacharya, A.K. (1984). Cyclic AMP transport across membrane vesicles of ultra-violet light irradiated *Escherichia coli*. J Biosci *6*, 173-179.

Black, R.A., Hobson, A.C., and Adler, J. (1983). Adenylate cyclase is required for chemotaxis to phosphotransferase system sugars by *Escherichia coli*. J Bacteriol *153*, 1187-1195.

Bochner, B.R., and Ames, B.N. (1982). Complete analysis of cellular nucleotides by two-dimensional thin layer chromatography. J Biol Chem *257*, 9759-9769.

Bohm, A., and Boos, W. (2004). Gene regulation in prokaryotes by subcellular relocalization of transcription factors. Curr Opin Microbiol *7*, 151-156.

Bolten, C.J., Kiefer, P., Letisse, F., Portais, J.C., and Wittmann, C. (2007). Sampling for metabolome analysis of microorganisms. Anal Chem *79*, 3843-3849.

Boos, W., and Shuman, H. (1998). Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. Microbiol Mol Biol Rev *62*, 204-229.

Braeken, K., Moris, M., Daniels, R., Vanderleyden, J., and Michiels, J. (2006). New horizons for (p)ppGpp in bacterial and plant physiology. Trends Microbiol *14*, 45-54.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C.*, et al.* (2001). Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. Nat Genet *29*, 365-371.

Bremer, H., Dennis, P., and Ehrenberg, M. (2003). Free RNA polymerase and modeling global transcription in *Escherichia coli*. Biochimie *85*, 597-609.

Bremer, H., and Dennis, P.P. (1996). Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli* and *Salmonella*: cellular and molecular biology, F.C. Neidhardt, R. Curtiss, III,, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington DC, American Society for Microbiology Press), pp. 1553-1569.

Brescia, C.C., Kaw, M.K., and Sledjeski, D.D. (2004). The DNA binding protein H-NS binds to and alters the stability of RNA *in vitro* and *in vivo*. J Mol Biol *339*, 505-514.

Britten, R.J., and Mc, C.F. (1962). The amino acid pool in *Escherichia coli*. Bacteriol Rev *26*, 292-335.

Buchholz, A., Takors, R., and Wandrey, C. (2001). Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic-electrospray ionization tandem mass spectrometric techniques. Anal Biochem *295*, 129-137.

Bustin, S.A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. Journal of molecular endocrinology *25*, 169-193.

Bustin, S.A., and Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. J Biomol Tech *15*, 155-166.

Butte, A. (2002). The use and analysis of microarray data. Nat Rev Drug Discov *1*, 951-960.

Bylund, F., Collet, E., Enfors, S.O., and Larsson, G. (1998). Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. Bioprocess Eng *18*, 171-180.

Calhoun, M.W., Oden, K.L., Gennis, R.B., de Mattos, M.J., and Neijssel, O.M. (1993). Energetic efficiency of *Escherichia coli*: effects of mutations in components of the aerobic respiratory chain. J Bacteriol *175*, 3020-3025.

Carter, M.G., Sharov, A.A., VanBuren, V., Dudekula, D.B., Carmack, C.E., Nelson, C., and Ko, M.S. (2005). Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray. Genome biology *6*, R61.

Cashel, M. (1974). Preparation of guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (pppGpp) from *Escherichia coli* ribosomes. Anal Biochem *57*, 100-107.

Cashel, M., Gentry, D.R., Hernandez, V.J., and Vinella, D. (1996). The stringent response. In *Escherichia coli* and *Salmonella*: cellular and molecular biology, F.C. Neidhardt, R. Curtiss, III,, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington DC, American Society for Microbiology Press), pp. 1458-1496.

Cashel, M., and Kalbacher, B. (1970). The control of ribonucleic acid synthesis in *Escherichia coli*. V. Characterization of a nucleotide associated with the stringent response. J Biol Chem *245*, 2309-2318.

Cashel, M., Lazzarini, R.A., and Kalbacher, B. (1969). An improved method for thin-layer chromatography of nucleotide mixtures containing ^{32}P-labelled orthophosphate. J Chromatogr *40*, 103-109.

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., and Reuss, M. (2002). Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnol Bioeng *79*, 53-73.

Chen, T., He, H., and Church, G.M. (1999). Modeling gene expression with differential equations. Paper presented at: Pacific Symposium on Biocomputing.

Chen, X., Wu, H., Mao, C., and Whitesides, G.M. (2002). A prototype two-dimensional capillary electrophoresis system fabricated in poly(dimethylsiloxane). Anal Chem *74*, 1772-1778.

Chilcott, G.S., and Hughes, K.T. (2000). Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica serovar typhimurium* and *Escherichia coli*. Microbiol Mol Biol Rev *64*, 694-708.

Chin, A.M., Feldheim, D.A., and Saier, M.H., Jr. (1989). Altered transcriptional patterns affecting several metabolic pathways in strains of *Salmonella typhimurium* which overexpress the fructose regulon. J Bacteriol *171*, 2424-2434.

Cole, S.T., and Raibaud, O. (1986). The nucleotide sequence of the *malT* gene encoding the positive regulator of the *Escherichia coli* maltose regulon. Gene *42*, 201-208.

Collantes-Fernandez, E., Zaballos, A., Alvarez-Garcia, G., and Ortega-Mora, L.M. (2002). Quantitative detection of *Neospora caninum* in bovine aborted fetuses and experimentally infected mice by real-time PCR. Journal of clinical microbiology *40*, 1194-1198.

Cozzone, A.J., and El-Mansi, M. (2005). Control of isocitrate dehydrogenase catalytic activity by protein phosphorylation in *Escherichia coli*. J Mol Microbiol Biotechnol *9*, 132-146.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Cserjan-Puschmann, M., Kramer, W., Duerrschmid, E., Striedner, G., and Bayer, K. (1999). Metabolic approaches for the optimisation of recombinant fermentation processes. Appl Microbiol Biotechnol *53*, 43-50.

Davis, B.D., Luger, S.M., and Tai, P.C. (1986). Role of ribosome degradation in the death of starved *Escherichia coli* cells. J Bacteriol *166*, 439-445.

Death, A., and Ferenci, T. (1994). Between feast and famine: endogenous inducer synthesis in the adaptation of *Escherichia coli* to growth with limiting carbohydrates. J Bacteriol *176*, 5101-5107.

Dong, H., Nilsson, L., and Kurland, C.G. (1995). Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. J Bacteriol *177*, 1497-1504.

Dudley, A.M., Aach, J., Steffen, M.A., and Church, G.M. (2002). Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. Proc Natl Acad Sci U S A *99*, 7554-7559.

Dunn, I.J., and Mor, J.R. (1975). Variable-volume continuous cultivation. Biotechnol Bioeng *17*, 1805-1822.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res *30*, 207-210.

Eisenbach, M. (2007). A hitchhiker's guide through advances and conceptual changes in chemotaxis. J Cell Physiol *213*, 574-580.

Ellison, M., Goryanin, I., Lee, S.Y., Penn, C., Reuss, M., Tomita, M., Wanner, B.L., and Westerhoff, H.V., eds (2006). The 3rd International *E. coli* Alliance Conference on Systems Biology. Conference Proceedings (October 31 - November 3, 2006. Jeju Island, Republic of Korea).

Epstein, C.B., and Butow, R.A. (2000). Microarray technology - enhanced versatility, persistent challenge. Current opinion in biotechnology *11*, 36-41.

Epstein, W., Rothman-Denes, L.B., and Hesse, J. (1975). Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. Proc Natl Acad Sci U S A *72*, 2300-2304.

Ferenci, T. (1996). Adaptation to life at micromolar nutrient levels: the regulation of *Escherichia coli* glucose transport by endoinduction and cAMP. FEMS Microbiol Rev *18*, 301-317.

Ferenci, T. (1999a). 'Growth of bacterial cultures' 50 years on: towards an uncertainty principle instead of constants in bacterial growth kinetics. Res Microbiol *150*, 431-438.

Ferenci, T. (1999b). Regulation by nutrient limitation. Curr Opin Microbiol *2*, 208-213.

Ferenci, T. (2001). Hungry bacteria - definition and properties of a nutritional state. Environ Microbiol *3*, 605-611.

Fischer, E., and Sauer, U. (2003). A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. J Biol Chem *278*, 46446-46451.

Forchhammer, J., and Kjeldgaard, N.O. (1968). Regulation of messenger RNA synthesis in *Escherichia coli*. J Mol Biol *37*, 245-255.

Fraenkel, D.G. (1996). Glycolysis. In *Escherichia coli* and *Salmonella*: cellular and molecular biology, F.C. Neidhardt, I. R. Curtiss, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington DC, American Society for Microbiology Press), pp. 189-198.

Frigessi, A., van de Wiel, M.A., Holden, M., Svendsrud, D.H., Glad, I.K., and Lyng, H. (2005). Genome-wide estimation of transcript concentrations from spotted cDNA microarray data. Nucleic Acids Res *33*, e143.

Gama-Castro, S., Jimenez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Penaloza-Spinola, M.I., Contreras-Moreira, B., Segura-Salazar, J., Muniz-Rascado, L., Martinez-Flores, I., Salgado, H.*, et al.* (2008). RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. Nucleic Acids Res *36*, D120-124.

Gennis, R.B., and Stewart, V. (1996). Respiration. In *Escherichia coli* and *Salmonella*: cellular and molecular biology, F.C. Neidhardt, R. Curtiss, III,, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington DC, American Society for Microbiology Press), pp. 217-261.

Gentry, D.R., and Cashel, M. (1996). Mutational analysis of the *Escherichia coli spoT* gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation. Mol Microbiol *19*, 1373-1384.

Gentry, D.R., Hernandez, V.J., Nguyen, L.H., Jensen, D.B., and Cashel, M. (1993). Synthesis of the stationary-phase sigma factor sigma S is positively regulated by ppGpp. J Bacteriol *175*, 7982-7989.

Gerber, K., Boos, W., Welte, W., and Schiefner, A. (2005). Crystallization and preliminary X-ray analysis of Mlc from *Escherichia coli*. Acta Crystallograph Sect F Struct Biol Cryst Commun *61*, 183-185.

Glick, B.R. (1995). Metabolic Load and Heterologous Gene-Expression. *13*, 247-261.

Goldbeter, A., and Koshland, D.E., Jr. (1982). Sensitivity amplification in biochemical systems. Q Rev Biophys *15*, 555-591.

Goldsmith, J.G., Ntuen, E.C., and Goldsmith, E.C. (2007). Direct quantification of gene expression using capillary electrophoresis with laser-induced fluorescence. Anal Biochem *360*, 23-29.

Gondo, S., Venkatasubramanian, K., Vieth, W.R., and Constantinides, A. (1978). Modeling the role of cyclic AMP in catabolite repression of inducible enzyme biosynthesis in microbial cells. Biotechnol Bioeng *20*, 1797-1815.

Gong, M., Wehmeyer, K.R., Stalcup, A.M., Limbach, P.A., and Heineman, W.R. (2007). Study of injection bias in a simple hydrodynamic injection in microchip CE. Electrophoresis *28*, 1564-1571.

Gosset, G. (2005). Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. Microb Cell Fact *4*, 14.

Haixin, W., Lijun, Q., and Dougherty, E. (2007). Modeling genetic regulatory networks by sigmoidal functions: A joint genetic algorithm and Kalman filtering approach. Paper presented at: Third International Conference on Natural Computation (ICNC).

Hansen, M.T., Pato, M.L., Molin, S., Fill, N.P., and von Meyenburg, K. (1975). Simple downshift and resulting lack of correlation between ppGpp pool size and ribonucleic acid accumulation. J Bacteriol *122*, 585-591.

Hansen, N., and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. Evol Comput *9*, 159-195.

Harder, A., and Roels, J.A. (1982). Application of Simple Models in Bioengineering. In Advances in biochemical engineering: Microbes and engineering aspects, A. Fiechter, ed. (Berlin Heidelberg New York, Springer-Verlag), pp. 55-107.

Hardiman, T., Ewald, J.C., Lemuth, K., Reuss, M., and Siemann-Herzberg, M. (2008a). Quantification of rRNA in *Escherichia coli* using capillary gel electrophoresis with laser-induced fluorescence detection. Anal Biochem *374*, 79-86.

Hardiman, T., Lemuth, K., Keller, M.A., Reuss, M., and Siemann-Herzberg, M. (2007). Topology of the global regulatory network of carbon limitation in *Escherichia coli*. J Biotechnol *132*, 359-374.

Hardiman, T., Meinhold, H., Hofmann, J., Ewald, J.C., Siemann-Herzberg, M., and Reuss, M. (2009). Prediction of kinetic parameters from DNA-binding site sequences for modeling global transcription dynamics in *Escherichia coli*. Metab Eng *doi:10.1016/j.ymben.2009.10.006*.

Hardiman, T., Windeisen, V., Ewald, J.C., Zibek, S., Schlack, P., Rebell, J., Reuss, M., and Siemann-Herzberg, M. (2008b). *In vitro* synthesis and characterization of guanosine 3',5'-bis(diphosphate) (ppGpp). Anal Biochem *383*, 337-339.

Haseltine, W.A., Block, R., Gilbert, W., and Weber, K. (1972). MSI and MSII made on ribosome in idling step of protein synthesis. Nature *238*, 381-384.

Hawley, D.K., and McClure, W.R. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res *11*, 2237-2255.

Hecker, M., Lambeck, S., Toepfer, S., van Someren, E., and Guthke, R. (2009). Gene regulatory network inference: Data integration in dynamic models - a review. Biosystems *96*, 86-103.

Heijnen, J.J., Bovenberg, R., Hatzimanikatis, V., and Laffend, L., eds (2006). Metabolic Engineering VI: From recDNA towards Engineering Biological Systems. Conference Proceedings. In (October 1 - 5, 2006. Noordwijkerhout, The Netherlands), p. 244.

Hengge-Aronis, R. (1993). Survival of hunger and stress: the role of rpoS in early stationary phase gene regulation in *E. coli*. Cell *72*, 165-168.

Hengge-Aronis, R. (2002). Signal transduction and regulatory mechanisms involved in control of the σ^S (RpoS) subunit of RNA polymerase. Microbiol Mol Biol Rev *66*, 373-395.

Henniger, G. (1998). Enzymatic methods of food analysis. In Analytical methods of food authentication, P.R. Ashurst, and M.J. Dennis, eds. (London, Weinheim Blackie Academic & Professional), pp. 137-181.

Hesbain-Frisque, A.M., van Schaftingen, E., and Hers, H.G. (1981). Structure and configuration of fructose 2,6-bisphosphate by ^{31}P and ^{13}C nuclear magnetic resonance. Eur J Biochem *117*, 325-327.

Hewitt, C.J., and Nebe-Von-Caron, G. (2001). An industrial application of multiparameter flow cytometry: assessment of cell physiological state and its application to the study of microbial fermentations. Cytometry *44*, 179-187.

Hewitt, C.J., Nebe-Von Caron, G., Axelsson, B., McFarlane, C.M., and Nienow, A.W. (2000). Studies related to the scale-up of high-cell-density *E. coli* fed-batch fermentations using multiparameter flow cytometry: effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. Biotechnol Bioeng *70*, 381-390.

Hewitt, C.J., Nebe-Von Caron, G., Nienow, A.W., and McFarlane, C.M. (1999). Use of multi-staining flow cytometry to characterise the physiological state of *Escherichia coli* W3110 in high cell density fed-batch cultures. Biotechnol Bioeng *63*, 705-711.

Hogg, T., Mechold, U., Malke, H., Cashel, M., and Hilgenfeld, R. (2004). Conformational Antagonism between Opposing Active Sites in a Bifunctional RelA/SpoT Homolog Modulates (p)ppGpp Metabolism during the Stringent Response. Cell *117*, 57-68.

Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: systems biology. Annual review of genomics and human genetics *2*, 343-372.

Ihssen, J., and Egli, T. (2005). Global physiological analysis of carbon- and energy-limited growing *Escherichia coli* confirms a high degree of catabolic flexibility and preparedness for mixed substrate utilization. Environ Microbiol *7*, 1568-1581.

Ingraham, J.L., Maaloe, O., and Neidhardt, F.C. (1983). Growth of the bacterial cell. In (Sunderland, MA, Sinauer Associates Inc), p. 435.

Izutsu, K., Wada, C., Komine, Y., Sako, T., Ueguchi, C., Nakura, S., and Wada, A. (2001). *Escherichia coli* ribosome-associated protein SRA, whose copy number increases during stationary phase. J Bacteriol *183*, 2765-2773.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol *3*, 318-356.

Jensen, K.F., and Pedersen, S. (1990). Metabolic growth rate control in Escherichia coli may be a consequence of subsaturation of the macromolecular biosynthetic apparatus with substrates and catalytic components. Microbiol Rev *54*, 89-100.

Jishage, M., Kvint, K., Shingler, V., and Nystrom, T. (2002). Regulation of σ factor competition by the alarmone ppGpp. Genes Dev *16*, 1260-1270.

Jung, I.L., Kim, S.K., and Kim, I.G. (2006). The RpoS-mediated regulation of isocitrate dehydrogenase gene expression in *Escherichia coli*. Curr Microbiol *52*, 21-26.

Kakuhata, R., Watanabe, M., Yamamoto, T., Akamine, R., Yamazaki, N., Kataoka, M., Fukuoka, S., Ishikawa, M., Ooie, T., Baba, Y.*, et al.* (2007). Possible utilization of *in vitro* synthesized mRNAs specifically expressed in certain tissues as standards for quantitative evaluation of the results of microarray analysis. J Biochem Biophys Methods *70*, 755-760.

Kauffman, K.J., Prakash, P., and Edwards, J.S. (2003). Advances in flux balance analysis. Curr Opin Biotechnol *14*, 491-496.

Keseler, I.M., Bonavides-Martinez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R.P., Johnson, D.A., Krummenacker, M., Nolan, L.M., Paley, S., Paulsen, I.T.*, et al.* (2009). EcoCyc: a comprehensive view of *Escherichia coli* biology. Nucleic Acids Res *37*, D464-470.

Klumpp, S., and Hwa, T. (2008). Growth-rate-dependent partitioning of RNA polymerases in bacteria. Proc Natl Acad Sci U S A *105*, 20245-20250.

Kovarova-Kovar, K., and Egli, T. (1998). Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. Microbiol Mol Biol Rev *62*, 646-666.

Kremling, A., Bettenbrock, K., Laube, B., Jahreis, K., Lengeler, J.W., and Gilles, E.D. (2001). The organization of metabolic reaction networks. III. Application for diauxic growth on glucose and lactose. Metab Eng *3*, 362-379.

Kremling, A., and Gilles, E.D. (2001). The organization of metabolic reaction networks. II. Signal processing in hierarchical structured functional units. Metab Eng *3*, 138-150.

Krishnan, M., Namasivayam, V., Lin, R., Pal, R., and Burns, M.A. (2001). Microfabricated reaction and separation systems. Current opinion in biotechnology *12*, 92-98.

Krohn, M., and Wagner, R. (1995). A procedure for the rapid preparation of guanosine tetraphosphate (ppGpp) from *Escherichia coli* ribosomes. Anal Biochem *225*, 188-190.

Kubista, M., Andrade, J.M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjoback, R., Sjogreen, B., Strombom, L.*, et al.* (2006). The real-time polymerase chain reaction. Molecular aspects of medicine *27*, 95-125.

Kuboniwa, M., Amano, A., Kimura, K.R., Sekine, S., Kato, S., Yamamoto, Y., Okahashi, N., Iida, T., and Shizukuishi, S. (2004). Quantitative detection of periodontal pathogens using real-time polymerase chain reaction with TaqMan probes. Oral microbiology and immunology *19*, 168-176.

Kummel, A., Panke, S., and Heinemann, M. (2006). Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. Mol Syst Biol *2*.

Kuroda, A. (2006). A polyphosphate-Lon protease complex in the adaptation of *Escherichia coli* to amino acid starvation. Biosci Biotechnol Biochem *70*, 325-331.

Lapin, A., Schmid, J., and Reuss, M. (2006). Modeling the dynamics of *E. coli* populations in the three-dimensional turbulent field of a stirred-tank bioreactor - A structured-segregated approach. Chem Eng Sci *61*, 4783-4797.

Larsson, G., Tornkvist, M., Wernersson, E.S., Tragardh, C., Noorman, H., and Enfors, S.O. (1996). Substrate gradients in bioreactors: Origin and consequences. Bioprocess Eng *14*, 281-289.

Leder, I.G. (1972). Interrelated effects of cold shock and osmotic pressure on the permeability of the *Escherichia coli* membrane to permease accumulated substrates. J Bacteriol *111*, 211-219.

Lee, S.B., and Bailey, J.E. (1984a). Genetically structured models for *lac* promoter-operator function in the chromosome and in multicopy plasmids: *lac* promoter function. Biotechnol Bioeng *26*, 1383-1389.

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.

Lee, S.Y. (1996). High cell-density culture of *Escherichia coli*. Trends Biotechnol *14*, 98-105.

Lee, S.Y., Lee, D.Y., and Kim, T.Y. (2005). Systems biotechnology for strain improvement. Trends Biotechnol *23*, 349-358.

Lemieux, B., Aharoni, A., and Schena, M. (1998). Overview of DNA chip technology. Mol Breed *4*, 277-289.

Lemuth, K. (2006). Transkriptomanalyse von *Escherichia coli* unter Kohlenhydrat-Limitierung mittels DNA-Microarrays. In Institute of Technical Biochemistry (Stuttgart, University of Stuttgart), pp. 270.

Lemuth, K., Hardiman, T., Winter, S., Pfeiffer, D., Keller, M.A., Lange, S., Reuss, M., Schmid, R.D., and Siemann-Herzberg, M. (2008). Global transcription and metabolic flux analysis of *Escherichia coli* in glucose-limited fed-batch cultivations. Appl Environ Microbiol *74*, 7002-7015.

Lengeler, J.W., Drews, G., and Schlegel, H.G. (1999). Biology of the prokaryotes, 1 edn (Stuttgart, Georg Thieme Verlag).

Leung, Y.F., and Cavalieri, D. (2003). Fundamentals of cDNA microarray data analysis. Trends Genet *19*, 649-659.

Lin, H., Hoffmann, F., Rozkov, A., Enfors, S.O., Rinas, U., and Neubauer, P. (2004). Change of extracellular cAMP concentration is a sensitive reporter for bacterial fitness in high-cell-density cultures of *Escherichia coli*. Biotechnol Bioeng *87*, 602-613.

Little, R., and Bremer, H. (1982). Quantitation of guanosine 5',3'-bisdiphosphate in extracts from bacterial cells by ion-pair reverse-phase high-performance liquid chromatography. Anal Biochem *126*, 381-388.

Loewen, P.C., and Hengge-Aronis, R. (1994). The role of the sigma factor sigma S (KatF) in bacterial global regulation. Annu Rev Microbiol *48*, 53-80.

Luo, B., Groenke, K., Takors, R., Wandrey, C., and Oldiges, M. (2007). Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry. J Chromatogr A *1147*, 153-164.

Magnusson, L.U., Farewell, A., and Nystrom, T. (2005). ppGpp: a global regulator in *Escherichia coli*. Trends Microbiol *13*, 236-242.

Maier, K., Hofmann, U., Reuss, M., and Mauch, K. (2008). Identification of metabolic fluxes in hepatic cells from transient ^{13}C-labeling experiments: Part II. Flux estimation. Biotechnol Bioeng *100*, 355-370.

Majka, J., and Speck, C. (2007). Analysis of protein-DNA interactions using surface plasmon resonance. *104*, 13-36.

Mantsch, H.H., and Smith, I.C. (1972). Fourier-transformed ^{13}C NMR spectra of polyuridylic acid, uridine, and related nucleotides - the use of ^{31}POC ^{13}C couplings for conformational analysis. Biochem Biophys Res Commun *46*, 808-815.

Matin, A., and Matin, M.K. (1982). Cellular levels, excretion, and synthesis rates of cyclic AMP in *Escherichia coli* grown in continuous culture. J Bacteriol *149*, 801-807.

McClure, W.R. (1980). Rate-limiting steps in RNA chain initiation. Proc Natl Acad Sci U S A *77*, 5634-5638.

McClure, W.R. (1985). Mechanism and control of transcription initiation in prokaryotes. Annu Rev Biochem *54*, 171-204.

Mechold, U., Murphy, H., Brown, L., and Cashel, M. (2002). Intramolecular regulation of the opposing (p)ppGpp catalytic activities of Rel_{Seq}, the Rel/Spo enzyme from *Streptococcus equisimilis*. J Bacteriol *184*, 2878-2888.

Merino, E., and Yanofsky, C. (2005). Transcription attenuation: a highly conserved regulatory strategy used by bacteria. Trends Genet *21*, 260-264.

Meyer, D., Schneider-Fresenius, C., Horlacher, R., Peist, R., and Boos, W. (1997). Molecular characterization of glucokinase from *Escherichia coli* K-12. J Bacteriol *179*, 1298-1306.

Michal, G. (1984). Fructose 1,6-bisphosphate, dihydroxyacetone phosphate and glyceraldehyde 3-phosphate. In Methods of enzymatic analysis, H.U. Bergmeyer, J. Bergmeyer, and M. Graßl, eds. (Weinheim, Verlag Chemie), pp. 342–350.

Milon, P., Tischenko, E., Tomsic, J., Caserta, E., Folkers, G., La Teana, A., Rodnina, M.V., Pon, C.L., Boelens, R., and Gualerzi, C.O. (2006). The nucleotide-binding site of bacterial translation initiation factor 2 (IF2) as a metabolic sensor. Proc Natl Acad Sci U S A *103*, 13962-13967.

Moses, V., and Sharp, P.B. (1972). Intermediary metabolite levels in *Escherichia coli*. J Gen Microbiol *71*, 181-190.

Mueller, O., Hahnenberger, K., Dittmann, M., Yee, H., Dubrow, R., Nagle, R., and Ilsley, D. (2000). A microfluidic system for high-speed reproducible DNA sizing and quantitation. Electrophoresis *21*, 128-134.

Mulligan, M.E., Brosius, J., and McClure, W.R. (1985). Characterization *in vitro* of the effect of spacer length on the activity of *Escherichia coli* RNA polymerase at the *TAC* promoter. J Biol Chem *260*, 3529-3538.

Mulligan, M.E., Hawley, D.K., Entriken, R., and McClure, W.R. (1984). *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. Nucleic Acids Res *12*, 789-800.

Munch, R., Hiller, K., Barg, H., Heldt, D., Linz, S., Wingender, E., and Jahn, D. (2003). PRODORIC: prokaryotic database of gene regulation. Nucleic Acids Res *31*, 266-269.

Munch, R., Hiller, K., Grote, A., Scheer, M., Klein, J., Schobert, M., and Jahn, D. (2005). Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. Bioinformatics *21*, 4187-4189.

Murray, K.D., and Bremer, H. (1996). Control of *spoT*-dependent ppGpp synthesis and degradation in *Escherichia coli*. J Mol Biol *259*, 41-57.

Nanchen, A., Schicker, A., Revelles, O., and Sauer, U. (2008). Cyclic AMP-dependent catabolite repression is the dominant control mechanism of metabolic fluxes under glucose limitation in *Escherichia coli*. J Bacteriol *190*, 2323-2330.

Neidhardt, F.C., and Savageau, M.A. (1996). Regulation Beyond the Operon. In *Escherichia coli* and *Salmonella*: cellular and molecular biology, F.C. Neidhardt, R. Curtiss, III,, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington DC, American Society for Microbiology Press), pp. 1310-1324.

Neidhardt, F.C., and Umbarger, H.E. (1996). Chemical composition of *Escherichia coli*. In *Escherichia coli* and *Salmonella*: cellular and molecular biology, F.C. Neidhardt, R. Curtiss, III,, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington DC, American Society for Microbiology Press), pp. 13-16.

Neubauer, P., Ahman, M., Tornkvist, M., Larsson, G., and Enfors, S.O. (1995). Response of guanosine tetraphosphate to glucose fluctuations in fed-batch cultivations of *Escherichia coli*. J Biotechnol *43*, 195-204.

Nielsen, J.H., and Villadsen, J. (1994). Bioreaction engineering principles, 1 edn (New York, NY, Plenum Press).

Noh, K., Gronke, K., Luo, B., Takors, R., Oldiges, M., and Wiechert, W. (2007). Metabolic flux analysis at ultra short time scale: isotopically non-stationary ^{13}C labeling experiments. J Biotechnol *129*, 249-267.

Nolan, T., Hands, R.E., and Bustin, S.A. (2006). Quantification of mRNA using real-time RT-PCR. Nature protocols *1*, 1559-1582.

Notley-McRobb, L., and Ferenci, T. (1999). The generation of multiple co-existing mal-regulatory mutations through polygenic evolution in glucose-limited populations of *Escherichia coli*. Environ Microbiol *1*, 45-52.

OECD (2009). The bioeconomy to 2030: Designing a policy agenda. Main findings and policy conclusions (Paris, Organization of Economic Co-operation and Development, International Futures Programme), pp. 18.

Ogura, M., Agata, Y., Watanabe, K., McCormick, R.M., Hamaguchi, Y., Aso, Y., and Mitsuhashi, M. (1998). RNA chip: quality assessment of RNA by microchannel linear gel electrophoresis in injection-molded plastic chips. Clinical chemistry *44*, 2249-2255.

Oki, T., Yoshimoto, A., Sato, S., and Takamatsu, A. (1975). Purine nucleotide pyrophosphotransferase from *Streptomyces morookaensis*, capable of synthesizing pppApp and pppGpp. Biochim Biophys Acta *410*, 262-272.

Ow, D.S., Lee, R.M., Nissom, P.M., Philp, R., Oh, S.K., and Yap, M.G. (2007). Inactivating FruR global regulator in plasmid-bearing Escherichia coli alters metabolic gene expression and improves growth rate. J Biotechnol *131*, 261-269.

Patten, C.L., Kirchhof, M.G., Schertzberg, M.R., Morton, R.A., and Schellhorn, H.E. (2004). Microarray analysis of RpoS-mediated gene expression in *Escherichia coli* K-12. Mol Genet Genomics *272*, 580-591.

Pease, A.J., and Wolf, R.E., Jr. (1994). Determination of the growth rate-regulated steps in expression of the *Escherichia coli* K-12 *gnd* gene. J Bacteriol *176*, 115-122.

Peterson, C.N., Carabetta, V.J., Chowdhury, T., and Silhavy, T.J. (2006). LrhA regulates *rpoS* translation in response to the Rcs phosphorelay system in *Escherichia coli*. J Bacteriol *188*, 3175-3181.

Pirt, S.J. (1982). Maintenance energy: a general model for energy-limited and energy-sufficient growth. Arch Microbiol *133*, 300-302.

Plumbridge, J. (1998). Control of the expression of the *manXYZ* operon in *Escherichia coli*: Mlc is a negative regulator of the mannose PTS. Mol Microbiol *27*, 369-380.

Plumbridge, J. (2002). Regulation of gene expression in the PTS in *Escherichia coli*: the role and interactions of Mlc. Curr Opin Microbiol *5*, 187-193.

Postma, P.W., Lengeler, J.W., and Jacobson, G.R. (1993). Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol Rev *57*, 543-594.

Potrykus, K., and Cashel, M. (2008). (p)ppGpp: Still Magical? Annu Rev Microbiol *62*, 35–51.

Pramanik, J., and Keasling, J.D. (1997). Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. Biotechnol Bioeng *56*, 398-421.

Pramanik, J., and Keasling, J.D. (1998). Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. Biotechnol Bioeng *60*, 230-238.

Pyles, E.A., and Lee, J.C. (1996). Mode of selectivity in cyclic AMP receptor protein-dependent promoters in *Escherichia coli*. Biochemistry *35*, 1162-1172.

Que, L., Jr., Willie, G.R., Cashel, M., Bodley, J.W., and Gray, G.R. (1973). Guanosine 5'-diphosphate, 3'-diphosphate: assignment of structure by ^{13}C nuclear magnetic resonance spectroscopy. Proc Natl Acad Sci U S A *70*, 2563-2566.

Raman, B., Nandakumar, M.P., Muthuvijayan, V., and Marten, M.R. (2005). Proteome analysis to assess physiological changes in *Escherichia coli* grown under glucose-limited fed-batch conditions. Biotechnol Bioeng *92*, 384-392.

Ramseier, T.M. (1996). Cra and the control of carbon flux via metabolic pathways. Res Microbiol *147*, 489-493.

Ramseier, T.M., Bledig, S., Michotey, V., Feghali, R., and Saier, M.H., Jr. (1995). The global regulatory protein FruR modulates the direction of carbon flow in *Escherichia coli*. Mol Microbiol *16*, 1157-1169.

Ramseier, T.M., Negre, D., Cortay, J.C., Scarabel, M., Cozzone, A.J., and Saier, M.H., Jr. (1993). *In vitro* binding of the pleiotropic transcriptional regulatory protein, FruR, to the *fru*, *pps*, *ace*, *pts* and *icd* operons of *Escherichia coli* and *Salmonella typhimurium*. J Mol Biol *234*, 28-44.

Reddy, K.J., and Gilman, M. (1987). Unit 4.4 Preparation of Bacterial RNA. In Current protocols in molecular biology, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.A. Smith, J.G. Seidman, and K. und Struhl, eds. (New York, Wiley Interscience), pp. 4.4.4-4.4.7.

Reue, K. (1998). mRNA quantitation techniques: considerations for experimental design and application. The Journal of nutrition *128*, 2038-2044.

Reuss, M., Luciano, A.-V., and Mauch, K. (2007). Reconstruction of dynamic network models from metabolite measurements. In Metabolomics, J. Nielsen, and M.C. Jewett, eds. (Berlin, Heidelberg, Springer), pp. 97-127.

Rhodius, V.A., and LaRossa, R.A. (2003). Uses and pitfalls of microarrays for studying transcriptional regulation. Curr Opin Microbiol *6*, 114-119.

Richter, D. (1980). Uncharged tRNA inhibits guanosine 3',5'-bis(diphosphate) 3'-pyrophosphohydrolase [ppGppase], the *spoT* gene product, from *Escherichia coli*. Mol Gen Genet *178*, 325-327.

RIO-DB (2008). Spectral Database for Organic Compounds SDBS (Research Information Database, RIO-DB). In http://riodb01ibaseaistgojp/sdbs/cgi-bin/cre_indexcgi?lang=eng (National Institute of Advanced Industrial Science and Technology (AIST)).

Roels, J.A. (1978). Regulatory mechanisms and the modelling of fermentation processes. Paper presented at: First European Congress on Biotechnology (Verlag Chemie, Weinheim).

Rowley, D.L., Pease, A.J., and Wolf, R.E., Jr. (1991). Genetic and physical analyses of the growth rate-dependent regulation of *Escherichia coli zwf* expression. J Bacteriol *173*, 4660-4667.

Sahm, H., Eggeling, L., and de Graaf, A.A. (2000). Pathway analysis and metabolic engineering in *Corynebacterium glutamicum*. Biol Chem *381*, 899-910.

Saier, M.H., Jr., and Ramseier, T.M. (1996). The catabolite repressor/activator (Cra) protein of enteric bacteria. J Bacteriol *178*, 3411-3417.

Saier, M.H., Jr., Ramseier, T.M., and Reizer, J. (1996). Regulation of carbon utilization. In *Escherichia coli* and *Salmonella*: cellular and molecular biology, F.C. Neidhardt, R. Curtiss, III,, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington DC, American Society for Microbiology Press), pp. 1325-1343.

Sakamoto, E., and Iba, H. (2001). Inferring a system of differential equations for a gene regulatory network by using genetic programming. Paper presented at: IEEE Congress on Evolutionary Computation (IEEE Press).

Salgado, H., Gama-Castro, S., Peralta-Gil, M., Diaz-Peredo, E., Sanchez-Solano, F., Santos-Zavaleta, A., Martinez-Flores, I., Jimenez-Jacinto, V., Bonavides-Martinez, C., Segura-Salazar, J.*, et al.* (2006). RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional

regulatory network, operon organization, and growth conditions. Nucleic Acids Res *34*, D394-397.

Schaub, J., Mauch, K., and Reuss, M. (2008). Metabolic flux analysis in *Escherichia coli* by integrating isotopic dynamic and isotopic stationary ^{13}C labeling data. Biotechnol Bioeng *99*, 1170-1185.

Schaub, J., and Reuss, M. (2008). *In vivo* dynamics of glycolysis in *Escherichia coli* shows need for growth-rate dependent metabolome analysis. Biotechnol Prog *24*, 1402-1407.

Schaub, J., Schiesling, C., Reuss, M., and Dauner, M. (2006). Integrated sampling procedure for metabolome analysis. Biotechnol Prog *22*, 1434-1442.

Schmid, J.W., Mauch, K., Reuss, M., Gilles, E.D., and Kremling, A. (2004). Metabolic design based on a coupled gene expression-metabolic network model of tryptophan production in *Escherichia coli*. Metab Eng *6*, 364-377.

Schmidt, H., and Jirstrand, M. (2006). Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics *22*, 514-515.

Schmidt, K., Marx, A., de Graaf, A.A., Wiechert, W., Sahm, H., Nielsen, J., and Villadsen, J. (1998). ^{13}C tracer experiments and metabolite balancing for metabolic flux analysis: comparing two approaches. Biotechnol Bioeng *58*, 254-257.

Schneider, T.D., Stormo, G.D., Gold, L., and Ehrenfeucht, A. (1986). Information content of binding sites on nucleotide sequences. J Mol Biol *188*, 415-431.

Schuhmacher, T., Lemuth, K., Hardiman, T., Vacun, G., Reuss, M., and Siemann-Herzberg, M. (2009). Quantifying cytosolic messenger RNA concentrations in *Escherichia coli* using real-time polymerase chain reaction for a systems biology approach. Anal Biochem *doi:10.1016/j.ab.2009.11.025*.

Schweder, T., Kruger, E., Xu, B., Jurgen, B., Blomsten, G., Enfors, S.O., and Hecker, M. (1999). Monitoring of genes that respond to process-related stress in large-scale bioprocesses. Biotechnol Bioeng *65*, 151-159.

Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M., and Rosenow, C. (2003). Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. Genome Res *13*, 216-223.

Senn, H., Lendenmann, U., Snozzi, M., Hamer, G., and Egli, T. (1994). The growth of *Escherichia coli* in glucose-limited chemostat cultures: a re-examination of the kinetics. Biochim Biophys Acta *1201*, 424-436.

Sevilla, A., Canovas, M., Keller, D., Reimers, S., and Iborra, J.L. (2007). Impairing and monitoring glucose catabolite repression in L-carnitine biosynthesis. Biotechnol Prog *23*, 1286-1296.

Shimada, T., Fujita, N., Maeda, M., and Ishihama, A. (2005). Systematic search for the Cra-binding promoters using genomic SELEX system. Genes Cells *10*, 907-918.

Shultzaberger, R.K., Chen, Z., Lewis, K.A., and Schneider, T.D. (2007). Anatomy of *Escherichia coli* sigma70 promoters. Nucleic Acids Res *35*, 771-788.

Siddiquee, K.A., Arauzo-Bravo, M.J., and Shimizu, K. (2004). Metabolic flux analysis of pykF gene knockout *Escherichia coli* based on ^{13}C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. Appl Microbiol Biotechnol *63*, 407-417.

Small, J., Call, D.R., Brockman, F.J., Straub, T.M., and Chandler, D.P. (2001). Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. Appl Environ Microbiol *67*, 4708-4716.

Smith, C.J., Nedwell, D.B., Dong, L.F., and Osborn, A.M. (2006). Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples. Environmental microbiology *8*, 804-815.

Soutourina, O., Kolb, A., Krin, E., Laurent-Winter, C., Rimsky, S., Danchin, A., and Bertin, P. (1999). Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon. J Bacteriol *181*, 7500-7508.

Sprenger, G.A. (1995). Genetics of pentose-phosphate pathway enzymes of *Escherichia coli* K-12. Arch Microbiol *164*, 324-330.

Srivatsan, A., and Wang, J.D. (2008). Control of bacterial transcription, translation and replication by (p)ppGpp. Curr Opin Microbiol *11*, 100-105.

St John, A.C., and Goldberg, A.L. (1980). Effects of starvation for potassium and other inorganic ions on protein degradation and ribonucleic acid synthesis in *Escherichia coli*. J Bacteriol *143*, 1223-1233.

Stanton, L.W. (2001). Methods to profile gene expression. Trends in cardiovascular medicine *11*, 49-54.

Stephanopoulos, G.N., Aristidou, A.A., and Nielsen, J. (1998). Metabolic engineering: principles and methodologies, 1 edn (London, Academic Press).

Stormo, G.D. (1990). Consensus patterns in DNA. Methods Enzymol *183*, 211-221.

Taniguchi, M., Miura, K., Iwao, H., and Yamanaka, S. (2001). Quantitative assessment of DNA microarrays - comparison with Northern blot analyses. Genomics *71*, 34-39.

Taymaz-Nikerel, H., de Mey, M., Ras, C., ten Pierick, A., Seifar, R.M., van Dam, J.C., Heijnen, J.J., and van Gulik, W.M. (2009). Development and application of a differential method for reliable metabolome analysis in *Escherichia coli*. Analytical Biochemistry *386*, 9-19.

Teich, A., Meyer, S., Lin, H.Y., Andersson, L., Enfors, S., and Neubauer, P. (1999). Growth rate related concentration changes of the starvation response regulators σ^S and ppGpp in glucose-limited fed-batch and continuous cultures of *Escherichia coli*. Biotechnol Prog *15*, 123-129.

Theobald, U., Mailinger, W., Baltes, M., Rizzi, M., and Reuss, M. (1997). *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*. 1. Experimental observations. Biotechnol Bioeng *55*, 305-316.

Traxler, M.F., Chang, D.E., and Conway, T. (2006). Guanosine 3',5'-bispyrophosphate coordinates global gene expression during glucose-lactose diauxie in *Escherichia coli*. Proc Natl Acad Sci U S A *103*, 2374-2379.

Tribhuwan, R.C., Johnson, M.S., and Taylor, B.L. (1986). Evidence against direct involvement of cyclic GMP or cyclic AMP in bacterial chemotactic signaling. J Bacteriol *168*, 624-630.

Typas, A., and Hengge, R. (2006). Role of the spacer between the -35 and -10 regions in sigma(s) promoter selectivity in *Escherichia coli*. Mol Microbiol *59*, 1037-1051.

Uesugi, S., Tanaka, S., and Ikehara, M. (1978). Carbon-13 nuclear-magnetic-resonance spectra of adenine cyclonucleosides and their phosphates. Effects of neighboring groups for elucidation of fine structure of nucleosides and nucleotides. Eur J Biochem *90*, 205-212.

van Bakel, H., and Holstege, F.C. (2004). In control: systematic assessment of microarray performance. EMBO reports *5*, 964-969.

van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D., and Holstege, F.C. (2003). Monitoring global messenger RNA changes in externally controlled microarray experiments. EMBO reports *4*, 387-393.

Vilela, M., Chou, I.C., Vinga, S., Vasconcelos, A.T., Voit, E.O., and Almeida, J.S. (2008). Parameter optimization in S-system models. BMC Syst Biol *2*, 35.

Vohradsky, J. (2001a). Neural model of the genetic network. J Biol Chem *276*, 36168-36173.

Vohradsky, J. (2001b). Neural network model of gene expression. FASEB J *15*, 846-854.

Voit, E.O. (2008). Modelling metabolic networks using power-laws and S-systems. Essays Biochem *45*, 29-40.

Wada, A., Igarashi, K., Yoshimura, S., Aimoto, S., and Ishihama, A. (1995). Ribosome modulation factor: stationary growth phase-specific inhibitor of ribosome functions from Escherichia coli. Biochem Biophys Res Commun *214*, 410-417.

Wada, A., Yamazaki, Y., Fujita, N., and Ishihama, A. (1990). Structure and probable genetic location of a "ribosome modulation factor" associated with 100S ribosomes in stationary-phase Escherichia coli cells. Proc Natl Acad Sci U S A *87*, 2657-2661.

Wagner, R. (2000). Transcription regulation in prokaryotes (Oxford, Oxford University Press).

Walter, G., Zillig, W., Palm, P., and Fuchs, E. (1967). Initiation of DNA-dependent RNA synthesis and the effect of heparin on RNA polymerase. Eur J Biochem *3*, 194-201.

Wasserman, W.W., and Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet *5*, 276-287.

Weaver, D.C., Workman, C.T., and Stormo, G.D. (1999). Modeling regulatory networks with weight matrices. Pacific Symposium on Biocomputing *4*, 112-123.

Weber, H., Polen, T., Heuveling, J., Wendisch, V.F., and Hengge, R. (2005a). Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. J Bacteriol *187*, 1591-1603.

Weber, J., Kayser, A., and Rinas, U. (2005b). Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. II. Dynamic response to famine and feast, activation of the methylglyoxal pathway and oscillatory behaviour. Microbiology *151*, 707-716.

Weickert, M.J., and Adhya, S. (1993). The galactose regulon of *Escherichia coli*. Mol Microbiol *10*, 245-251.

Wendrich, T.M., Blaha, G., Wilson, D.N., Marahiel, M.A., and Nierhaus, K.H. (2002). Dissection of the mechanism for the stringent factor RelA. Molecular cell *10*, 779-788.

Wick, L.M., and Egli, T. (2004). Molecular components of physiological stress responses in *Escherichia coli*. Adv Biochem Eng Biotechnol *89*, 1-45.

Wick, L.M., Quadroni, M., and Egli, T. (2001). Short- and long-term changes in proteome composition and kinetic properties in a culture of *Escherichia coli* during transition from glucose-excess to glucose-limited growth conditions in continuous culture and vice versa. Environ Microbiol *3*, 588-599.

Wittmann, C., Kromer, J.O., Kiefer, P., Binz, T., and Heinzle, E. (2004). Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria. Anal Biochem *327*, 135-139.

Wolf, R.E., Jr., Prather, D.M., and Shea, F.M. (1979). Growth-rate-dependent alteration of 6-phosphogluconate dehydrogenase and glucose 6-phosphate dehydrogenase levels in *Escherichia coli* K-12. J Bacteriol *139*, 1093-1096.

Wong, P., Gladney, S., and Keasling, J.D. (1997). Mathematical model of the *lac* operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. Biotechnol Prog *13*, 132-143.

Xu, B., Jahic, M., Blomsten, G., and Enfors, S.O. (1999a). Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*. Appl Microbiol Biotechnol *51*, 564-571.

Xu, B., Jahic, M., and Enfors, S.O. (1999b). Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*. Biotechnol Prog *15*, 81-90.

Yagil, G. (1975). Quantitative aspects of protein induction. Curr Top Cell Regul *9*, 183-236.

Yagil, G., and Yagil, E. (1971). On the relation between effector concentration and the rate of induced enzyme synthesis. Biophys J *11*, 11-27.

Yee, L., and Blanch, H.W. (1992). Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*. Biotechnology (N Y) *10*, 1550-1556.

Young, R.A. (2000). Biomedical discovery with DNA arrays. Cell *102*, 9-15.

Zheng, D., Constantinidou, C., Hobman, J.L., and Minchin, S.D. (2004). Identification of the *crp* regulon using *in vitro* and *in vivo* transcriptional profiling. Nucleic Acids Res *32*, 5874-5893.