Chapter 11
Dynamic Modeling of the Central Metabolism of *E. coli* – Linking Metabolite and Regulatory Networks

Timo Hardiman, Karin Lemuth, Martin Siemann-Herzberg, and Matthias Reuss

Contents

Abstract Coupling complex regulatory and metabolic networks for the purpose of dynamic modeling requires knowledge of the quantitative kinetics of the participating reactions as well as the variation of parameters in the context of the physiological state of the system. This chapter aims at demonstrating the integration of the different networks for *E. coli* exposed to an increasing carbon limitation of a fed-batch process with constant feeding of the carbon and energy source glucose. Starting from a global observation of the response of the bacteria in terms of flux distribution

M. Reuss (✉)
Institute of Biochemical Engineering and Centre Systems Biology, University of Stuttgart, Allmandring 31, 70569, Stuttgart, Germany
e-mail: reuss@ibvt.uni-stuttgart.de

S.Y. Lee (ed.), *Systems Biology and Biotechnology of Escherichia coli*,
DOI 10.1007/978-1-4020-9394-4_11,

and gene expression in the central metabolism, emphasis is given to the dynamic modeling of regulation phenomena in the catabolism. The *cra* regulon which is linked to the dynamic response of the metabolite fructose 1,6- bis(phosphate) serves as an example to introduce a new concept, in which the binding constants are estimated from DNA-binding site sequences of the regulatory proteins. By comparison of the nucleotide frequencies within the DNA-binding sites for the individual target genes of the regulon, it is possible to perform a reasonable estimation of the kinetic parameters. Results of these estimations are compared with experimentally observed transcript concentrations measured with the aid of quantitative PCR. In addition it is shown how these outputs of the regulatory networks can be linked to the maximal rates of the enzymes for the metabolic system of interest. The discussion of this issue is embedded within a critical assessment of different conceptual frameworks for modeling the metabolic network, which covers the spectrum of dynamic modeling at different levels of complexity, such as genome scale, modular approaches and reduced models.

11.1 Introduction

Systems biology as an emerging field of research in bio-, engineering and systems sciences aims at a systems-level understanding of biological processes – and ultimately whole cells and organisms. The grand, and currently unrealistic, hope to even continue these efforts into a whole cell *in silico* model time and again shapes the conceptual framework of this endeavour. There are several reasons that the present state of affairs still falls short of this euphoric expectation. The first is concerned with the fundamental question of a comprehensive definition of a "whole cell model" and closely related to this uncertainty the query about the purpose of such a model. Referring to the fundamental ideas and discussion of Casti (1992a,b) about a model and its intended application, Bailey (1998) reminded us that "mathematical modeling does not make sense without defining, before making the model, what its use is and what problem it is intended to help to solve". The second reason originates from a critical assessment of part of the experimental work in the field of holistic measurements and related top down approaches in inverse engineering for network inference. In spite of spectacular developments in high-throughput technologies such as genome sequencing, transcriptomics, metabolomics, fluxomics etc. – platforms which have monopolized systems biology research in recent years – there is a tendency to fragment the whole into various sub-omes and a great deal of arguments exists about what ome is more important. However, due to multiple border crossings these omes are inseparable parts of a single process – the complex and interwoven dynamics of the living organisms.

Another issue to be addressed in the context of fragmentation is the often observed focus on specific networks and treatment in separated and isolated territories, such as metabolism, regulation and signal transduction. In the course of this partition and kind of downward analysis, levels are reached where the whole meaning of the

system is destroyed because of neglected interactions and missing integration. In order to underline the systemic thinking, the exchanges of material and information between the heuristically isolated modules of a system to be investigated may also be termed “intra-actions”.

In this chapter we will highlight with a few examples the importance of integration of regulatory and metabolic networks in *Escherichia coli* and discuss the framework of how this process of integration can be portrayed dynamically in the structure of the mathematical model.

Taking up the aforementioned attenuation of the importance of defining first the purpose of the mathematical model, the environmental changes triggering the regulation of the metabolism have to be introduced. The example deals with the regulation of the central metabolism of *E. coli* during a fed-batch process with constant feeding rate of the carbon and energy source glucose (Fig. 11.1). This process operation is important for technical processes for production of heterologous proteins as well as bacterial metabolites. For large-scale applications, fed-batch, high cell density 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 maximum biomass concentration that can be achieved with this strategy depends on sufficient

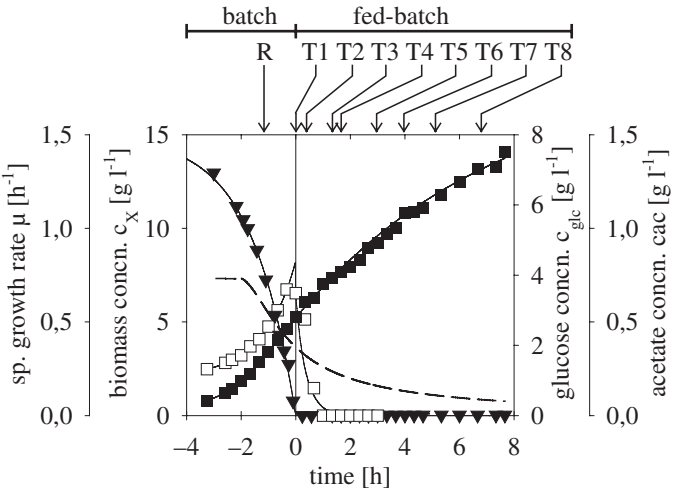

Fig. 11.1 Glucose limited fed-batch cultivation of *E. coli* K-12 W3110 with constant feed rate. The vertical solid line at $t = 0$ indicates glucose limitation. The concentrations of biomass (filled squares), glucose (triangles) and acetate (open squares) are given as well as the time course of the 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)

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 (Dunn and Mor 1975). The proceeding carbon limitation also leads to a range of serious starvation phenomena with manifold regulatory responses of the cells. These processes macroscopically manifest themselves in a loss of viability, such as was illustrated by Hewitt et al. (2000, 1999, Hewitt and Nebe-Von-Caron 2001).

Bacteria control metabolism and growth rate through global genetic regulatory systems, 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 signalling 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,b, 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).

Most of the aforementioned investigations have been performed during the shift from exponential to stationary growth phase in batch cultivations. The dynamic perturbation during these experiments is characterized by a rapid drop of glucose concentration to zero in a short time period. As distinguished from this very fast perturbation the fed-batch cultivation with constant feeding rate prolongs the period of declining glucose concentrations towards a time span of several hours. This prolongation of the proceeding carbon limitation initiates a process of transient adaptation during which the organisms dynamically changes activities of enzymes in the catabolism and regulates the anabolism to adjust the synthesis of macromolecules and reduce growth rate. The result of this concerted action of global regulation differs from the short term regulation during the transient period from batch to stationary phase and the subsequent starvation as well as the behaviour of the organisms during steady state conditions at varying dilution rate during continuous operation.

With the goal to obtain a more in-depth understanding of these complex regulation phenomena and their impact on the flux distribution in the central metabolism experimental work has been initiated which covers three main areas, namely microarray analysis, flux analysis and selected quantitative measurements of metabolites and mRNA via PCR analysis. Results of this work have been already summarized in the papers of Hardiman et al. (2007a) and Lemuth et al. (2008). Part of these results will be presented once more within this chapter to support the tight link between experimental work and computational approach.

11.2 Reconstruction of the Global Regulatory Structure of Carbon Limitation

Current systems biology research in dynamic modeling of the central carbon metabolism of *Escherichia coli* aims at the comprehensive understanding of its global regulation in response to carbon limitation. The long-term goal is of course the support of rational producer strain optimization based on mathematical modeling. Much knowledge about regulatory processes during carbon limitation has accumulated and is available from literature and databases. However, it is not clarified which regulators are dominant under these conditions and thus, which regulators must be considered in a mathematical model. For the assembly of the global regulatory network underlying and for explaining the transient metabolic response to carbon limitation it is necessary to link this *a priori* knowledge with experimental observations in order to identify the relevant components of the network. As mentioned above, observing a single 'ome' alone is not adequate when such complex dynamic processes are being investigated.

The works of Hardiman et al. (2007a) and Lemuth et al. (2008) demonstrate the simultaneous experimental observation of concentrations of signaling molecules (cAMP and ppGpp) and a time series of metabolic flux and transcriptome analyses of *Escherichia coli* K-12 W3110 in a fed-batch cultivation applying a constant feed rate (Fig. 11.1). These omic approaches were employed for the reconstruction of the model structure, focussing on the most relevant parts that must be considered when dynamic modeling the regulatory and metabolic behaviour.

The constant feeding strategy applied, provided an appropriate approach for separating the time-dependent events during the transition from exponential to carbon-limited growth (Fig. 11.1). Both intracellular alarmones ppGpp and cAMP accumulated in large quantities after the onset of nutrient limitation, subsequently declining to basal levels (Hardiman et al. 2007a). The limited supply of the carbon and energy source glucose led to significantly decreasing fluxes in glycolysis, pentose phosphate pathway and biosynthesis, whereas TCA cycle fluxes remained constant (Fig. 11.2a,b). The flux redistribution resulted in an enhanced energy generation in the TCA cycle and consequently, in a 20 % lower biomass yield (Hardiman et al. 2007a). From the correlations of gene expression levels with the metabolic fluxes that were observed (Fig. 11.2), this behaviour can be interpreted as follows and transformed into a model structure (Hardiman et al. 2007a, Lemuth et al. 2008).

The flux through the upper part of glycolysis is favoured 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, thereby providing a sufficient, though minimal, efflux into the pentose phosphate pathway. The regulation of *pfkA*, *fbaA*, *pgk*, *pykF*, *gapA* and *eno* transcription by the Cra regulator protein (*cra* modulon) is suggested to lead to this behaviour (Fig. 11.3). Signalling occurs through

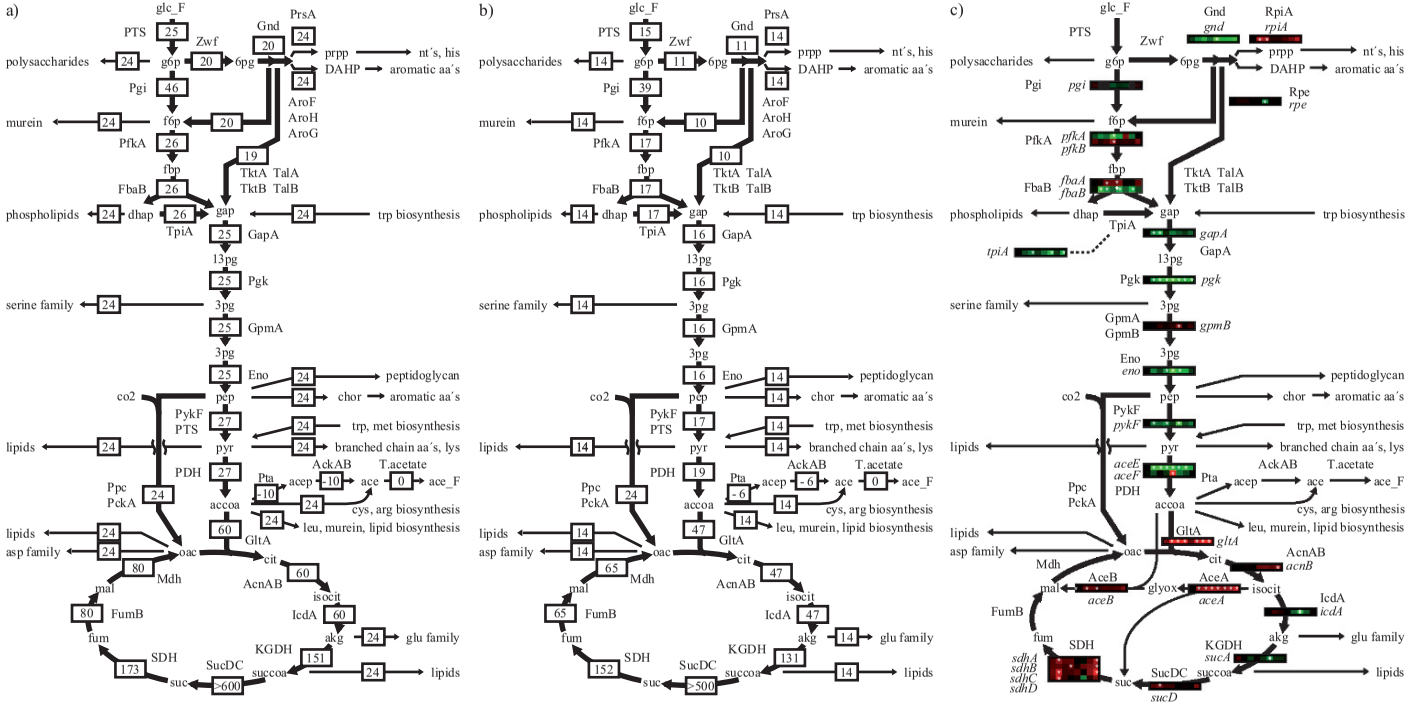

Fig. 11.2 Time series of metabolic flux and DNA microarray analyses of the central carbon metabolism in *E. coli* K-12 W3110. (**a**) and (**b**), metabolic fluxes at $\mu = 0.13\,h^{-1}$ and $\mu = 0.08\,h^{-1}$, respectively, during glucose-limited fed-batch growth applying a constant feed rate. Fluxes are mean values of five independent cultivations 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; $\mu = 0.4\,h^{-1}$; see Fig. 11.1). The model used includes also the respiration, biosynthetic pathways and polymerisation reactions. Only those reactions involving metabolites from central carbon metabolism are shown. (**c**) 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 Fig. 11.1). Green: mRNA level is lower than that at the reference state. Red: higher mRNA level. Statistical significance ($p \leq 0.05$) is indicated by white asterisks

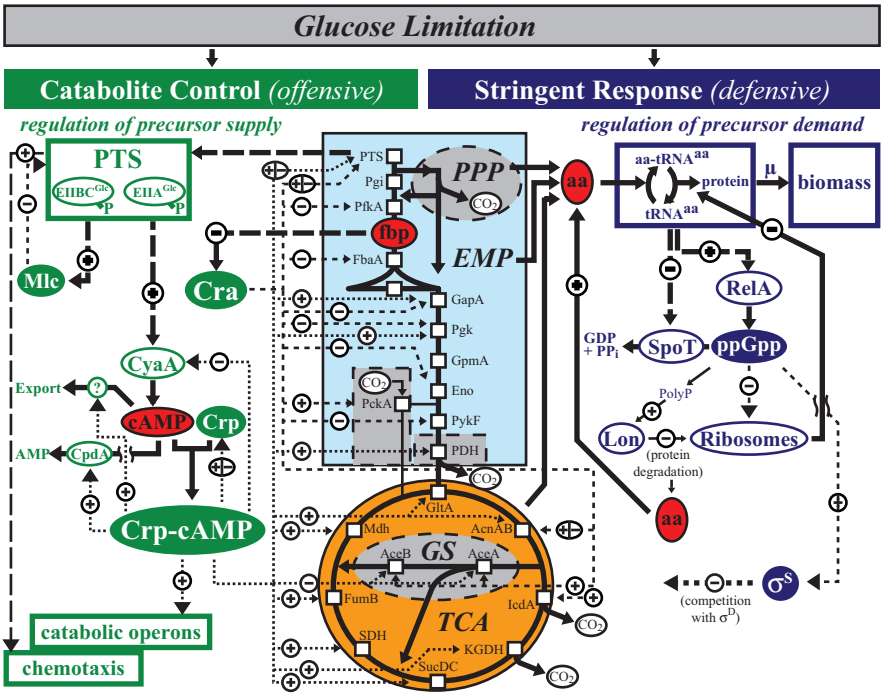

Fig. 11.3 Reconstruction of the global regulatory and metabolic network of carbon limitation. *Left*: Catabolite repression can be seen as an offensive strategy since various catabolic operons are induced, which encode transporters and metabolic pathways for the consumption of sugars other than glucose. Moreover, many genes of the TCA cycle, glyoxylate shunt (GS), PTS system and glycolysis are regulated by the Crp-cAMP regulator complex (*crp* modulon). Additionally, the Cra protein represses genes of glycolysis and activates transcription of GS genes (*cra* modulon). Fbp inactivates the Cra protein. The fbp concentration reflects the availability of extracellular glucose. *Right*: Stringent response is an defensive strategy since it regulates many components of the translational and transcriptional machinery, most prominently, the reduction of rRNA transcription by ppGpp (*relA/spoT* modulon). The dedicated reader is referred to Hardiman et al. (2007a) for a detailed analysis of the major mechanisms that lead to the accumulation of the alarmones cAMP and ppGpp and to the reduction of the fbp concentration during carbon limitation. The major negative feedback regulation mechanisms leading to a resetting of the signals are also discussed therein

the metabolite fructose 1,6-bis(phosphate) (fbp; Fig. 11.3), whiches concentration is proposed to reflect the availability of glucose. A reduction in the enzyme levels of the lower glycolysis concomitantly with the observed decreasing flux levels might be a hint for the control of metabolite concentrations (homeostasis). The carbon flux entering the TCA cycle (influx is enhanced via *gltA* expression) is split into the glyoxylate shunt (GS), the phosphoenolpyruvate(pep)-GS and the full TCA cycle. GS and pep-GS provide a better pep, pyr and oac precursor supply. It is proposed that the global regulation via the *crp* and *cra* modulons is the most relevant in this respect – i.e. the Crp-cAMP regulator complex mainly induces the transcription of

the TCA cycle genes, whereas the glyoxylate shunt (GS) genes are regulated by the Cra regulator protein (positive) and the Crp-cAMP complex (negative) (Fig. 11.3).

In summary, the omic approaches reported by Hardiman et al. (2007a) and Lemuth et al. (2008) demonstrate that 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, i.e. the EMP/PPP, RpiA/Rpe and TCA/GS/pep-GS split ratios, supplies sufficient precursors for biosyntheses. The network topology regulating the central carbon metabolism provided in (Hardiman et al. 2007a) 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 – considering not only signal transduction, transcriptional regulation and metabolic behaviour but also the resetting of the signals (the two intracellular alarmones cAMP and ppGpp) and the effect of the respective feedback mechanisms (ascribed to catabolite repression and stringent response) on the dynamics in the central carbon metabolism (Fig. 11.3).

Besides the reported correlating transcript levels and metabolic fluxes in the central carbon metabolism, a picture of interesting interconnections between enhancement and attenuation of further cellular functions is drawn in (Lemuth et al. 2008), highlighting the importance of this adaptive behaviour for mathematical modeling and optimizing biotechnical production processes. Most of the physiological rearrangements, if not all of them, can clearly be linked to the regulation of the intracellular availability of precursors and energy, i.e., not only the supply and demand rates, but also the (resulting) concentrations of precursors are discussed to be tightly controlled. This physiologically highly important task is exemplified by the tempting proposal that the global regulation of diverse functions such as chemotaxis, transport and flagellar systems as well as glycolysis, TCA cycle and glyoxylate shunt are interconnected in controlling the availability of the precursor phosphenolpyruvate (pep). This and further major findings of Lemuth et al. (2008) are condensed in the following: (i) A cluster of high-affinity transporters is synthesized, while the activity of medium-affinity transporters is maintained. This is mainly due to their regulation by the Crp-cAMP complex. 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 $EIIA^{Glc}\sim P$-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 utilise the energy available. (iii) Cellular growth is regulated predominantly by the stringent response (alarmone ppGpp, *relA/spoT* modulon), however, no extensive induction of the general *rpoS*-dependent response could be observed. This is attributed to the opposing regulation via the *crp* and *relA/spoT* modulons (see also Lapin et al. 2006). It is expected that slow substrate concentration changes do not trigger a strong starvation response Teich et al. (1999). However, other stress responses were detected.

Thus, a model topology has been reconstructed of the global regulation of the *E. coli* central carbon metabolism through the *crp*, *cra* and *relA/spoT* modulons that can be used for mathematical modeling metabolism and regulation (Fig. 11.3). In a second step, physiological functions that are important for precursor and energy availability (transport, chemotaxis, stringent and stress response) are suggested to be implemented as further modules of the mathematical model.

11.3 Basic Principles of Deterministic Modeling the Dynamics of Gene Expression

The development of deterministic models describing the regulation of gene expression (transcription, mRNA degradation and protein biosynthesis) has a long tradition. Already in a 1968 review, Rosen (1968) summarized important methods and approximations essential for modeling and simulation of gene regulatory networks. The majority of the models are similar in mathematical nature and more or less rest upon the concept suggested by Yagil and Yagil (1971) and Yagil (1975). Based on the operon model of Jacob and Monod (1961) these authors illustrated how to derive the probability of transcription initiation if a gene is regulated by a repressor or activation protein.

In the case of negative regulation it is defined as the ratio of the concentration of operators free to be transcribed, c_O, to the total concentration of operators, $c_{O,t}$:

$$\phi_{neg} = \frac{c_O}{c_{O,t}}. \tag{11.1}$$

Accordingly, the ratio of the concentration of activator proteins bound to DNA-binding sites, $c_{A.DNSbs}$, to the total concentration of DNA-binding sites, $c_{DNAbs,t}$, gives the probability:

$$\phi_{pos} = \frac{c_{A.DNSbs}}{c_{DNAbs,t}}. \tag{11.2}$$

The maximal rate of transcription can be achieved for $\phi \rightarrow 1$. In both cases the probability is derived from the equilibrium assumption for the biochemical binding reactions of the regulator protein and its DNA-binding site. This is reasonable because the initiation and the subsequent transcript and peptide elongation occur on different time scales (McClure 1985, Stephanopoulos et al. 1998, Uptain et al. 1997). In case of effectors inhibiting or enhancing the binding activity of regulator proteins (inducers or co-repressors), additional equilibrium reactions can be formulated. Equation (11.3, 11.4) exemplify the inactivation of the repressor protein R by binding the inducer molecule E and binding of the active repressor to the operator DNA sequence O. Equation (11.5) depicts the equilibrium (binding) constants and the derived probability of induction for negative regulation.

$$R + n \cdot E \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} R.E_n \tag{11.3}$$

$$R + O \underset{k_{-2}}{\overset{k_{+2}}{\rightleftharpoons}} R.O \tag{11.4}$$

$$\phi_{neg} = \frac{c_O}{c_{O,t}} = \frac{1 + K_1 c_E^n}{1 + K_1 c_E^n + K_2 c_{R,t}} \tag{11.5}$$

with

$$K_1 = \frac{k_{+1}}{k_{-1}} = \frac{c_{R.E_n}}{c_R \cdot c_E^n} \text{ and } K_2 = \frac{k_{+2}}{k_{-2}} = \frac{c_{R.O}}{c_R \cdot c_O}$$

The transcription rate is then obtained from

$$r_{tc,mRNA_i} = r_{tc,\max} \prod_j \phi_j f(\mu) - k_{Degradation}\, c_{mRNA_i} - \mu c_{mRNA_i} \tag{11.6}$$

and the translation rate of the protein of interest is calculated from

$$r_{TL,\mathrm{Protein}_i} = r_{\max,TL} c_{mRNA_i} - \mu c_{\mathrm{Protein}_i}. \tag{11.7}$$

The term f(μ) considers the impact of the specific growth rate on the transcription rate. Roels (1983) suggested the following form:

$$f(\mu) = \frac{a + b\mu}{a + b\mu_{max}}, \tag{11.8}$$

which reflects the linear dependency between mRNA biosynthesis and the specific growth rate.

The illustrated approach enables modeling of superimposed regulation mechanisms by several regulators and can be extended by the binding of RNA polymerase to the promoter DNA sequence. It is therefore suitable for implementation of gene expression kinetics in large metabolic models.

With increasing amount of knowledge available about the details of catabolite repression (reviewed by Deutscher et al. 2006) more sophisticated models have been developed. Many of these modifications are based on the approach of Lee and Bailey (1984a,b) in which a transcription efficiency is defined as:

$$\eta = \psi_P (1 - \psi_R)(1 + \alpha \psi_A) \tag{11.9}$$

with the fraction of occupied promoters ψ_P, the influence of a repressor $(1 - \psi_R)$ and an activator $(1 + \alpha\psi_A)$. In addition to the comprehensive models suggested by Kremling et al. (2007, 2001, 2000) (Kremling and Saez-Rodriguez 2007, Kremling and Gilles 2001) and Bettenbrock et al. (2006) this approach has been applied by Wong et al. (1997) as well as Van Dien and Keasling (1998) to mention a few.

In a different line of approaches Boolean networks are used for modeling regulatory phenomena. These models have been already introduced in the 1960's by Stuart Kauffman (1969). The conceptual framework of Boolean networks is based on the assumption that binary on/off switches functioning in discrete time steps can describe important aspects of gene regulation (Albert 2004, McAdams and Arkin 1998). In the context with the intended coupling of regulatory and metabolic networks, such a Boolean approach for description of the regulatory network would eventually lead to a hybrid model in which the concentrations of metabolites are expressed as continuous values and connected via enzyme kinetics to describe the dynamics of the metabolic networks described by a system of ODEs.

An alternative option to avoid the computational effort with the hybrid models is to approximate the switch like behavior of the expression with the aid of Hill kinetics. In case of a repression the rate of transcription can be represented by

$$r = r_{\max,transcription} \frac{1}{1 + \left(\frac{c_R}{K_R}\right)^{n_R}}, \tag{11.10}$$

whereas for the event of an activation

$$r = r_{\max,transcription} \frac{1}{1 + \left(\frac{K_A}{c_A}\right)^{n_A}} \tag{11.11}$$

could be an appropriate approximation. A more generic formulation based on the "general" Hill equation suggested by Cornish-Bowden (1995) and Hofmeyr and Cornish-Bowden (1997) for reversible reactions in case of metabolic reactions leads to a very useful rate expression for the concerted action of multiple activators and repressors (Likhoshvai and Ratushny 2007):

$$\frac{dc_{mRNA}^{Target gene(s)}}{dt} = r_{maca,TC} \frac{k + \sum_{si_1}^{c_{As,1}} \left(\frac{R_{si_1}}{K_{si_1}}\right)^{h_{si_1}} + \sum_{si_1,2}^{c_{As,2}} \frac{R_{si_1}^{h_{s1_1}} R_{si_2}^{h_{si_2}}}{K_{si_{1,2}}^{h_{si_1}+h_{si_2}}} + \cdots + \sum_{si_1,\ldots,si_M}^{c_{As,M}} \frac{\prod_{k=1}^{M} R_{si_k}^{h_{si_k}}}{K_{si_{1\ldots M}}^{\sum_{k=1}^{M} h_{i_k}}}}{1 + \sum_{sj_1}^{c_{Is,As,1}} \left(\frac{R_{sj_1}}{K_{sj_1}}\right)^{h_{sj_1}} + \sum_{sj_1,sj_2}^{c_{Is,As,2}} \frac{R_{sj_1}^{h_{sj_1}} R_{sj_2}^{h_{sj_2}}}{K_{sj_{1,2}}^{h_{sj_1}+h_{sj_2}}} + \cdots + \sum_{sj_1,\ldots,sj_N}^{N} \frac{\prod_{w=1}^{N} R_{si_w}^{h_{si_w}}}{K_{sj_{1\ldots N}}^{\sum_{w=1}^{N} h_{sj_w}}}}. \tag{11.12}$$

Here the binding of regulatory proteins R includes inhibition (binding sites Is) and activation (bindig sites As). Figure 11.4 depicts the application of this equation for an example of joint regulation of two genes through two repressors – and one activator molecule. Starting from the framework of statistical mechanics Bintu et al.

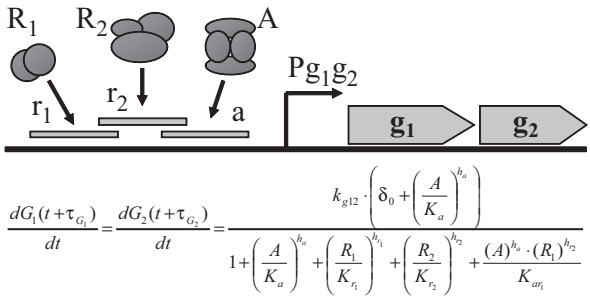

Fig. 11.4 Dynamic modeling of gene expression regulated by two repressors and one activator based on general hill kinetics (Ilya Peshkov, Novosibirsk, Russia: personal communication)

(2005) derived various "regulatory factors" for several different regulatory motifs very similar to the generic structure of Equation (11.12).

For portraying the sigmoid character of the dynamic response alternative approaches are based on generic sigmoidal functions (Weaver et al. 1999), such as

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{11.13}$$

With the aid of additional terms representing system and measurement noise, Haixin et al. (2007) have used this approach in connection with Kalman Filtering for the problem of genetic regulatory network inference from time series microarray data.

Another powerful method in the context of sigmoidal functions is built on the conceptual framework of neural networks (Vohradsky 2001a,b). The model has the form

$$\frac{dz_i}{dt} = r_{\max} \frac{1}{1 + \exp\left[-\left(\sum_j w_{ij} y_j + b_i\right)\right]} - k_{\deg radation} z_i \tag{11.14}$$

with connection weights w_{ij}, delay parameter b_i and rate constant for degradation k. z_i is the target gene regulated by the genes y_j connected to the target (predictor genes).

The focus of application of most of the aforementioned approaches for dynamic modeling of gene regulatory networks is on network inference based on time series "profiles" of microarray data. A crucial point in the evaluation of the majority of these applications is the missing distinction and the rigorous mathematical description of the two processes of transcription and translation. Using nonlinear stability analysis Hatzimanikatis and Lee (1999) have shown that a combination of gene expression information at the mRNA level and at the protein level is required to describe even simple models of gene networks. This issue is all the more important for coupling gene regulatory networks with metabolic networks because at least the output of the regulatory network is linked at the protein level to change enzyme concentrations in the metabolic rate expressions. If balance equations for the

translation process are neglected, the overall dynamics are corrupted by a mixture of characteristic time constants for transcription and translation.

11.4 Dynamic Model for the Intra-actions Between the Regulatory and Central Metabolic Networks of *Escherichia coli*: Translation of Sequence Information into Kinetic Parameters

The focus of this chapter is on the dynamic modeling of the intra-actions between the regulatory and metabolic networks depicted in Figs. 11.2 and 11.3. The ultimate goal of this approach is to quantitatively describe the dynamical changes of traffic patterns and variations in flux distributions in response to the environment changes caused by the diminishing supply of carbon and energy source glucose. The kinetics to describe the dynamics of the regulation phenomena is modeled in terms of probabilities of transcription as described in Section 11.3. The approach is based on a translation of gene sequence information into parameters of binding constants for the individual regulator protein-DNA-binding site interaction of interest. The methodology will be exemplified for the Cra-modulon, illustrated in Fig. 11.3.

The usage of Equations (11.1, 11.2, 11.3, 11.4, and 11.5) for modeling gene expression in large metabolic networks as illustrated in Fig. 11.3 requires the availability of the parameters K_1 and K_2 from literature, data bases or their identification from experimental observations. For estimation of the binding constant K_1 for the reaction between the regulatory protein and its effector E (Equation 11.3) this is of course feasible. However, $K_{2,i}$ has to be determined for each individual gene i coding for the enzymes or regulatory proteins being components of the network. For large networks or large regulons/modulons such an approach is not practicable because of the experimental effort. This is one of the reasons that verification of such models is most often dominated by identification methods for the estimation of large sets of parameters. To circumvent this kind of problems we therefore choose an approach in which the individual binding constants are estimated from the gene sequence information of the DNA-binding side (Hardiman et al. 2007b).

11.4.1 Decomposition of the Binding Reaction

For the purpose of derivation of $K_{2,i}$ from the DNA-binding site sequence, the regulator protein R is first assumed to bind to the mononucleotides, $b \in \{A, C, G, T\}$, of the binding site sequence and that these interactions are independent and additive according to Stormo (1988, 1990):

$$R + b \rightleftharpoons R.b \tag{11.15}$$

Assuming again an equilibrium reaction, the binding constant is proportional to the ratio of the bound pool to the unbound pool of bases:

$$K_b = \frac{c_{R.b}}{c_R \cdot c_b} \propto \frac{c_{R.b}}{c_b} = \frac{f_b}{p_b} \tag{11.16}$$

Equation (11.16) also illustrates that this ratio is equal to the ratio of the frequency at which the base b occurs at the considered position in the DNA-binding site sequence, f_b, to the frequency of this base in the genome of the considered organism, p_b, which was proposed by Stormo (1988, 1990). Considering that the binding to each nucleotide of the sequence is assumed to be independent, the binding constant for the total DNA-binding site, K_2, can be formulated as

$$K_2 = \prod_n K_{b,n} \propto \prod_n \frac{f_{b,n}}{p_b} \tag{11.17}$$

where n corresponds to the position of the nucleotide b in the sequence. Various scientific groups have investigated this relationship and found reasonable correlations between calculated and experimentally determined binding affinities or the equivalent free energy of binding (Equation 11.18). For instance, Berg and von Hippel ((1987)) developed a statistical-mechanical theory based on the assumption that specific DNA sequences have been selected according to their protein binding affinity and that all sequences that show equal affinities are equally likely to occur in the genome. The theory Berg and von Hippel (1987) was able to predict the correlation between the activities ($k_2 K_B$ values) of *E. coli* promoter sequences assuming that nucleotides at different positions in the promoter sites contribute independently to their activities. Many more contributions to the field demonstrated that there is a strong linear relation between base frequency and binding strength (Berg and von Hippel 1988, Fields et al. 1997, Stormo and Fields 1998, Takeda et al. 1989). For an overview the dedicated reader is referred to (Stormo 1990, 2000).

$$\Delta G_b = -RT \ln K_b \propto -\ln\left(\frac{f_b}{p_b}\right) \tag{11.18}$$

The findings of these authors are not surprising, because Equations (11.17, 11.18) simply express that highly conserved DNA sequences are bound stronger than less conserved ones by the respective regulator protein. Therefore, Equations (11.17, 11.18) provide a simple and valuable tool for the quantitative evaluation of any DNA-binding site sequence with respect to a reference sequence.

11.4.2 Application to the cra *Regulon of Escherichia coli*

The regulator protein Cra is a major component of the global regulation of the metabolic fluxes in glycolysis (EMP), the TCA cycle and the glyoxylate shunt (GS) in glucose-limited fed-batch processes of *E. coli* (see Section 11.2 and Fig. 11.3).

Binding of the Cra protein to the DNA-binding site of the transcription units i ($DNAbs_i$; Equation 11.21) of the *cra* modulon is inhibited by high concentrations of fructose 1,6-bis(phosphate) (fbp; Equation 11.19).

$$Cra + fbp \overset{K_1}{\rightleftharpoons} Cra\,.\,fbp \tag{11.19}$$

$$4\,Cra + DNAbs_i \overset{K_{2,i}}{\rightleftharpoons} Cra_4\,.\,DNAbs_i \tag{11.20}$$

$$\phi^{neg}_{Cra.DNAbs,i} = \frac{c_{DNAbs,i}}{\left(c_{DNAbs,i}\right)_{total}} = \frac{1}{1 + K_{2,i}\left(\dfrac{(c_{Cra})_{total}}{1 + K_1\, c_{fbp}}\right)^4} \tag{11.21}$$

$$\phi^{pos}_{Cra.DNAbs,i} = \frac{c_{Cra_4.DNAbs,i}}{\left(c_{DNAbs,i}\right)_{total}} = 1 - \phi^{neg}_{Cra.DNAbs,i} \tag{11.22}$$

$$r_{tc,mRNA_i} = r_{tc,\max} \prod_j \phi_j - k_{Degradation} c_{mRNA_i} - \mu c_{mRNA_i} \tag{11.23}$$

The probability of transcription initiation, ϕ, is determined by the fraction of unbound (Equation 11.21) or bound (Equation 11.22) DNA-binding sites when transcription is repressed or activated, respectively.

11.4.3 Comparison Between Model Prediction and Experimental Observations

Figure 11.5 illustrates the mRNA concentrations of central carbon metabolism genes measured using qPCR analysis during glucose-limited fed-batch cultivation of *E. coli* (see Fig. 11.1) as well as concentrations predicted by the model described by Equations (11.19, 11.20, 11.21, 11.22, and 11.23). The genes *eno* (encoding enolase), *pfkA* (6-phosphofructokinase I) and *pykF* (pyruvate kinase I) are known to be regulated by the Cra regulator protein (see Section 11.2). The repression of their transcription (Fig. 11.5) results in a strong decrease of the respective mRNA concentrations.

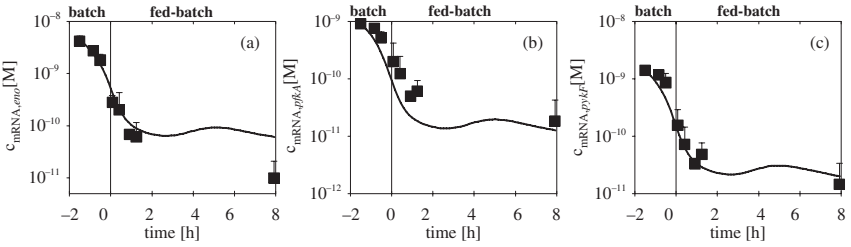

Fig. 11.5 mRNA concentrations during glucose limited fed-batch cultivation of *E. coli* K-12 W3110. The concentrations of mRNA (■) were determined by qPCR analysis (standard deviation, 3 independent samples). Simulation data are indicated by solid lines. (**a**) *eno* mRNA (encoding enolase), (**b**) *pfkA* mRNA (6-phosphofructokinase I) and (**c**) *pykF* mRNA (pyruvate kinase I)

Obviously, the DNA-binding activity of the Cra protein is high due to the low concentration of fructose 1,6-bis(phosphate) (fbp) during the fed-batch process (Fig. 11.6). The strong decrease in fbp concentration (Fig. 11.6a,b) can be attributed to the limited carbon supply (Section 11.2). However, according to Fig. 11.6a the concentration of fbp apparently increases after two hours of fed-batch cultivation, when the experimental data is related to the biomass concentration. Only when

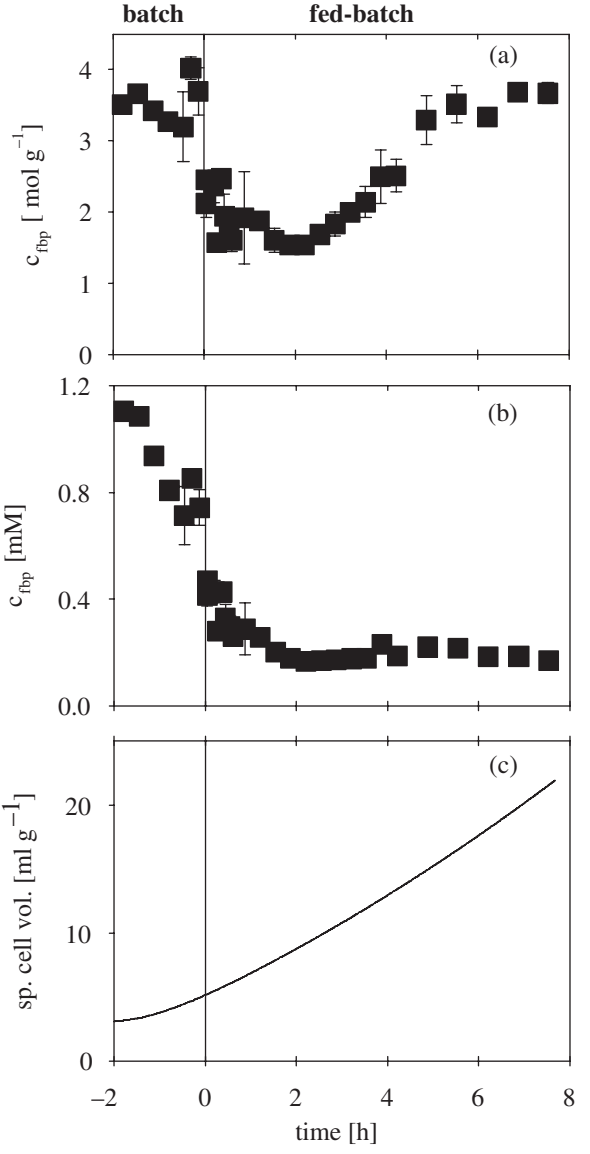

Fig. 11.6 Fructose 1,6-bis(phosphate) concentration during glucose-limited fed-batch cultivation of *E. coli* K-12 W3110. Concentrations were determined after quenching and extraction using perchloric acid as published in Hardiman et al. (2007a). (**a**) Fbp concentration related to biomass [μmol (g dry weight)$^{-1}$]. (**b**) Fbp concentration related to the cell volume [mmol (l cytosol)$^{-1}$] that is obtained by deviding the concentration given in (**a**) by (**c**) the specific cell volume v_x [l cytosol (g dry weight)$^{-1}$], and which is in turn approximated using the growth rate-dependent function $\hat{v}_X = \frac{0.4860 \cdot 2^{(1.144\hat{\mu})}}{-0.636 + 0.635 \cdot 2^{(0.718\hat{\mu})}}$ (Hardiman et al. 2007a)

the growth rate-dependent variation of the cell volume is considered a meaningful result may be obtained from the data (Fig. 11.6b). The time profile of the molar intracellular concentration given in [mmol (l cytosol)$^{-1}$] enables to explain the transcriptome and metabolic flux data as described in Section 11.2. That is, the persisting low concentration of fbp leads to the repression of glycolysis genes by the Cra regulator protein and activation of transcription of glyoxylate shunt genes.

The model predicts the mRNA concentration satisfactorily during the batch and the beginning of the fed-batch process and also at a later process phase where the growth rate is very low (Figs. 11.1 and 11.6). Note, that the model used for the simulations differs from the one introduced in Section 11.3. Equation (11.23) does not take into account the growth rate dependence of transcription initiation, whereas Equation (11.6) considers the impact of the specific growth rate on the transcription rate. Although the Equations (11.19, 11.20, 11.21, 11.22, and 11.23) are sufficient for a rough simulation of the mRNA concentrations (Fig. 11.5), the extension of the model by growth rate dependent variables and further regulons/modulons is needed. This is expected to make an important contribution to the understanding the global regulation of the central carbon metabolism during carbon limitation.

11.5 Conceptual Framework for Dynamic Models of Metabolic Networks of *E. coli* Suitable for Links to Regulatory Networks

A multitude of approaches is available for dynamic modeling of the metabolism of *E. coli*. Here, we shall limit our discussion on continuous and deterministic models, which are derived by considering the balance equations of the individual metabolites and can be represented in the compact form:

$$\frac{d\mathbf{x}}{dt} = \mathbf{N}\mathbf{r}\left(\mathbf{x}\left(t\right), \mathbf{P}\right) - \mu\mathbf{x}. \tag{11.24}$$

N is the *m x n* stoichiometric matrix an **r** is the *n*-dimensional rate vector.

Based on dynamic measurements of intra- and extracellular metabolites in response to a perturbation of a continuous culture with a pulse of glucose Chassagnole et al. (2002) derived a rigorous dynamic model of the central metabolism of *E. coli* (Fig. 11.7). The model is based on kinetic rate expressions for the individual enzymes, the original structures of which have been derived from investigations with isolated enzymes at *in vitro* conditions. The key to afterwards generate the dynamic *in vivo* model is, to extract the kinetic parameters of the biochemical reactions from the *in vivo* metabolite measurements and, as such, considering the reactions in their "systemic" context (Reuss et al. 2007).

To describe the dynamic systems behaviour, deterministic kinetic rate equations of the form

$$r_i = r_{max,i}\, f\left(\mathbf{c}, \mathbf{p}\right) \tag{11.25}$$

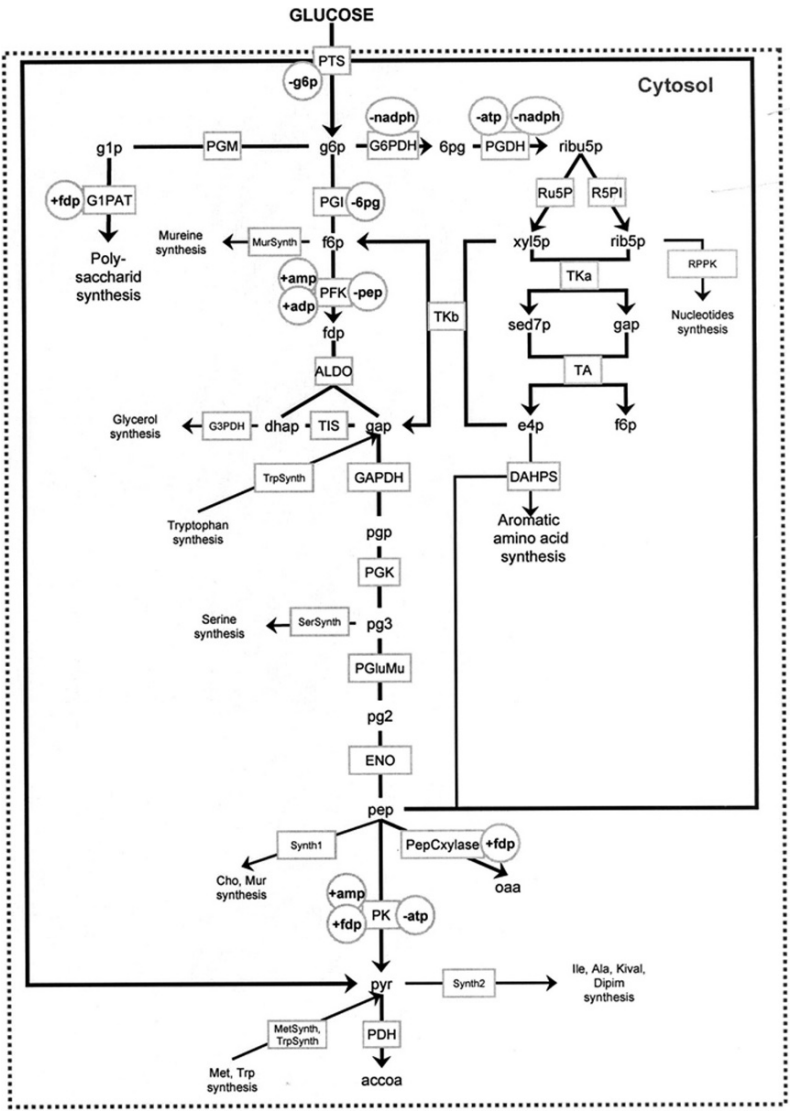

Fig. 11.7 Structure of the metabolic model of glycolysis and pentose phosphate pathway in *Escherichia coli* (Chassagnole et al. 2002)

are formulated, where the capacity of the reaction is characterized by its maximal rate and the kinetic function f represents the kinetic properties of the reaction. Substrates, products and other metabolic effectors influencing the rate of the reaction are represented by the state vector of metabolite concentrations **c**. The parameters of the reaction are summarized in the vector **p**.

If the maximal rate of reaction can be assumed to be proportional to the concentration of the enzyme, Equation (11.25) provides a simple way to integrate the

output of the regulatory network with respect to the concentration of the individual enzymes.

The first step to embed the behaviour of the subsystem into the metabolic network as a whole is provided by the estimation of the maximal rates of the individual reactions. Applying the rate Equation (11.25) to the steady state leads to

$$\tilde{r}_{\max,i} = \frac{r_{i,steady\ state}}{f\left(\mathbf{c}_{steady\ state}, \mathbf{p}\right)}. \tag{11.26}$$

Let us assume that reaction rate $\tilde{r}_i$ at steady has been estimated from metabolic flux analysis. Let us further assume that a first estimate of the structure of the kinetics as well as the parameter vector **p** is available from *in vitro* measurements. If the components of the concentration vector **c** influencing the rate of the reaction have been measured at steady state, the unknown maximal rates are given as depicted in Equation (11.26).

If the stoichiometric model used for metabolic flux analysis has a genome scale or a metabolic submodel in case of ^{13}C analysis is linked to such a model (Schaub et al. 2008), the maximal rates estimated from Equation (11.26) are invariant to the scale of the submodule used for the dynamic model. As such, these rates are intrinsic properties of the system as a whole and in a meaningful way only depend on the physiological state of the system. Further details of the strategy to identify the *in vivo* kinetics from the measured stimulus-response date are discussed in the original papers (Chassagnole et al. 2002, Rizzi et al. 1997) and summarized in a review (Reuss et al. 2007).

The model structure depicted in Fig. 11.7 accounts for the enzymatic rate expressions for the glycolytic enzymes and therefore allows for connection of the most important output signals of the Cra and Crp modulon (Fig. 11.3). Apart of the necessary model extension for incorporation of TCA and glyoxylate shunt reactions, however, interactions between the regulatory and metabolic networks exceed the central metabolism by far. Particularly the precursor demand via e. g. amino acid synthesis and the subsequent polymerisation reactions are regulated through the alarmone ppGpp (Fig. 11.3) and demand further extension of the model structure.

Aside from the possibility to assign the large number of additional reactions with mechanistic enzyme kinetics, which is an excessively laborious and time consuming approach, conceptual frameworks based on canonical formulations of rate expressions leading to less detailed large- scale models may prove to be useful. Such an approach has been introduced by Reuss et al. (2007) and successfully applied for a large-scale dynamic model for *E. coli.* The dynamic model follows from the reaction network model of *Escherichia coli* introduced by Chassagnole et al. (2002). The network comprises both catabolic and anabolic routes with protein, DNA, RNA, polysaccharides, murein, and lipids building up biomass. Sequential reaction steps and parallel routes are lumped. With 129 reactions, 133 balanced metabolites, and seven conserved moieties, the degree of freedom of the null-space of the network is fixed to $129–133 + 7 = 3$. Additional informations regarding inhibition and activation (metabolic regulation) have been gathered from the MetaCyc data base

(www.metacyc.org, (Caspi et al. 2006)). The kinetic behaviour of the individual reactions is assigned according to the universal linlog approach (Visser and Heijnen 2003, Visser et al. 2004, 2000):

$$r = J\frac{c_E}{c_E^0}\left(1+\sum_i \varepsilon_{S,i}\ln\frac{c_{S,i}}{c_{S,i}^0}+\sum_j \varepsilon_{P,j}\ln\frac{c_{P,j}}{c_{P,j}^0}+\sum_k \varepsilon_{A,k}\ln\frac{c_{A,k}}{c_{A,k}^0}+\sum_l \varepsilon_{I,l}\ln\frac{c_{I,l}}{c_{I,l}^0}\right). \tag{11.27}$$

substrates products activators inhibitors

The variables are defined to the relative reference steady state, with concentration levels state c^0, fluxes J^0, and enzyme level c_E^0. The parameters are the elasticity coefficients

$$\varepsilon_M = \frac{c_M}{r}\left(\frac{\partial r}{\partial c_M}\right). \tag{11.28}$$

In total the network holds 921 kinetic parameters (elasticities). The dynamic simulation of the non-linear and stiff system of differential equations was performed with the aid of the extrapolation solver LIMEX from the Konrad-Zuse-Centre for Information Technology in Berlin (Ehrig et al. 1999). For estimation of the parameters the evolutionary algorithm developed by the Computer Science Department of the University of Tuebingen (Streichert and Ulmer 2005) has been applied. Results of the comparisons between model simulations and experimental observation from stimulus response experiments in which a pulse of glucose is added to the steady state of a continuous culture have been presented by Reuss et al. (2007).

One key to understanding how these large scale models do compare with dynamic models based on mechanistic rate expression is to carefully examine the differences between the simulation results of the two approaches. Visser et al. (2000) compared the outcome of the linlog approach with the dynamic model of Chassagnole et al. (2002). These authors noted a reasonable agreement for not to large dynamic perturbations with respect to the external glucose concentration. An important observation from this comparison and associated identification of the elasticity coefficients in Equation (11.27) concerns the expected behaviour of the reversible near-equilibrium reactions in the glycolysis. First, the individual elasticity coefficients of such reversible near-equilibrium reactions are not independent. Furthermore, it can be easily shown that the values of the elacticities must be very high and, in consequence, the flux control coefficient tends to zero. In essence then, these reactions are suited candidates for model reduction.

The issue of this model reduction should be always addressed in the context of the purpose of the model as emphasized in the beginning of this chapter. A first, well-proven concept for model reduction in metabolic engineering is based on the time hierarchy of the metabolism. The kernel of this method is a model analysis, which considers the eigenvalues and eigenvectors of the Jacobian associated to the dynamic model (Heinrich and Schuster 1996). The application of this time-scale separation for the Cassagnole model (Chassagnole et al. 2002) results

in assumptions of quasi-steady state conditions for 11 eigenvectors possessing the highest values.

The result of this reduction, which shows reasonable agreement between the dynamic response of the original and reduced model, yields, however, a differential-algebraic system. Because the algebraic equations do not allow an explicit analytical solution it is necessary to resort to advanced and efficient solver for differential-algebraic systems.

As a promising alternative to the modal analysis we employed a sensitivity analysis based on the flux control coefficients (Lapin et al. 2006). These coefficients relate the fractional change of the steady state fluxes to the infinitesimal changes in the total enzyme concentrations (Heinrich and Schuster 1996). From the hierarchy of these flux control coefficients predicted from the original model reactions with the highest values in relation to the flux control coefficient of the glucose uptake were selected. The resulting network is depicted in Fig. 11.8 Because of low flux control

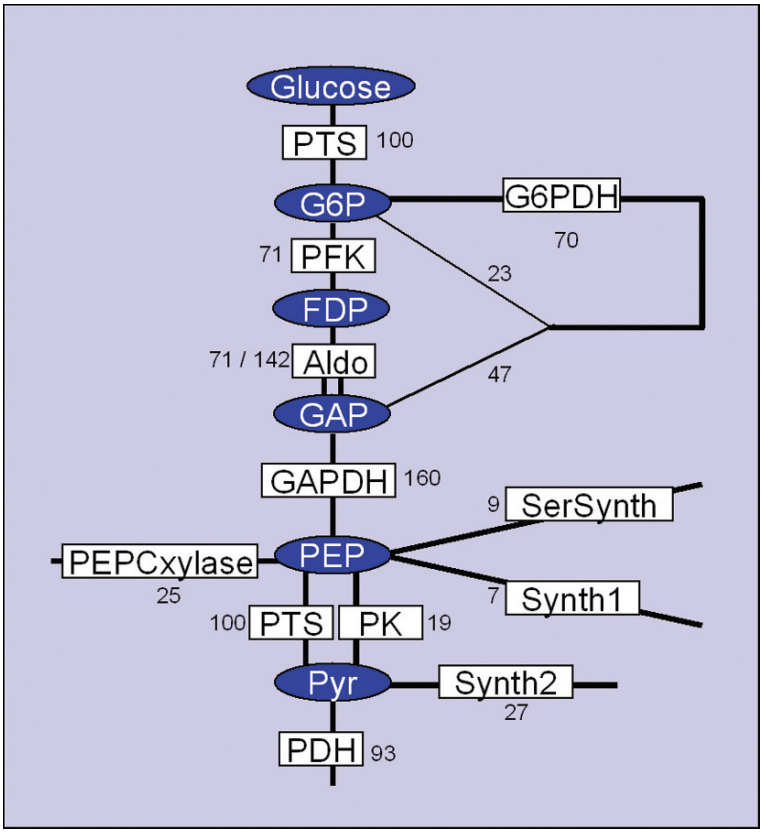

Fig. 11.8 Reduced metabolic network model for the sugar uptake system, glycolysis and pentose phosphate pathway. Reduction of the original model (Chassagnole et al. 2002) is based on the hierarchy of flux control coefficients. The numbers alongside the enzymes depict the metabolic fluxes related to glucose uptake rate 100

coefficients the reactions for the phosphoglucoisomerase, the triose phosphate isomerase, the phosphoglycerate kinase, the phosphoglyceromutase and the enolase could be neglected. The low flux control coefficients result from the reversibility of aforementioned reactions leading to very high values of the elasticity coefficient. For the purpose of model reduction a rapid equilibrium is assumed for these reactions and the dynamics of the metabolites are linked via equilibrium constants.

To summarize the efforts for designing the dynamic model for the central metabolism it is important to emphasize that systems biology modeling of these networks should not be restricted to the task of aggregating and integrating quantitative information on individual enzyme kinetics to "whole-cell models". An equally important challenge is to reduce the complexity and to tailor the model structure for the intended application. Thus, depending on its specific objectives, a model may involve details at different levels.

11.6 Conclusions

The framework for integration of regulatory and metabolic networks provides significant insights on the dynamic response of microorganisms to perturbations of the environmental condition with characteristic times relevant for variations in gene expression. This issue is of particular importance for process operations with dynamic variations in the supply of the carbon and energy source with high relevance for high cell density fermentations. The importance of these regulation phenomena in response to increasing carbon limitation is not restricted to the catabolism of the cell. The strong impact on anabolic reactions (Fig. 11.3) leads to serious variation of the protein expression dynamics with consequences on specific productivities in case of production of recombinant proteins. Future work in our group aims at the extension of integration of regulatory and metabolic networks for these important anabolic phenomena based on dynamic models for protein and ribosome synthesis linked to precursor supply from the central metabolism (Arnold et al. 2005, Elf and Ehrenberg 2005, Elf et al. 2005, Götz and Reuss 1997).

As far as the integration of regulatory networks with modules of the central carbon metabolism is concerned the main contribution of this chapter arises from the fact that a plausible conceptual framework has been developed which enable us to link existing dynamic models for the metabolism with simple models for regulation of transcription and translation of important target enzymes. The approach contains a concise method for the formulation of gene expression. It is demonstrated how the necessary model parameters regarding the gene regulation, i.e. the binding constants of regulator proteins to the DNA-binding site of the individual genes of the regulon, can be derived from the DNA sequence of the sites and minimal literature information.

The overall approach may also serve as an example of how to successfully bridge the top down and bottom up approach for the purpose of modeling and simulation in systems biology. After application of top down analysis for identification of the target genes in the central metabolism, the modeling cycle of the bottom up approach is initiated. This includes quantitative measurements of concentrations of

key compounds such as single mRNA molecules, metabolites and even incorporation of "reductionistic" sequence information. This quantitative information at the compound level is afterwards used for the verification of the dynamic model. The ultimate goal of such a hybrid approach is that the characterization of the behavior of the parts of the system should be consistent with the expected and/or observed behavior of the system as a whole.

References

Albert R (2004) Boolean modeling of genetic regulatory networks. In: (ed) Complex Networks, Springer, Berlin, Heidelberg

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

Artsimovitch I, Patlan V, Sekine S et al. (2004) Structural basis for transcription regulation by alarmone ppGpp. Cell 117(3):299–310

Bailey JE (1998) Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. Biotechnol Prog 14(1):8–20

Barker MM, Gaal T, Gourse RL (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(4):689–702

Barker MM, Gaal T, Josaitis CA et al. (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(4):673–88

Berg OG, von Hippel PH (1987) Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. J Mol Biol 193(4):723–50

Berg OG, von Hippel PH (1988) Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. J Mol Biol 200(4): 709–23

Bettenbrock K, Fischer S, Kremling A et al. (2006) A quantitative approach to catabolite repression in *Escherichia coli*. J Biol Chem 281(5):2578–84

Bintu L, Buchler NE, Garcia HG et al. (2005) Transcriptional regulation by the numbers: models. Curr Opin Genet Dev 15(2):116–24

Braeken K, Moris M, Daniels R et al. (2006) New horizons for (p)ppGpp in bacterial and plant physiology. Trends Microbiol 14(1):45–54

Cashel M, Gentry DR, Hernandez VJ et al. (1996) The stringent response. In: Neidhardt FC, et al. (ed) *Escherichia coli* and *Salmonella*: cellular and molecular biology, American Society for Microbiology Press, Washington DC

Caspi R, Foerster H, Fulcher CA et al. (2006) MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res 34(Database issue):D511–6

Casti JL (1992a) Reality rules. I Picturing the world in mathematics. The fundamentals. John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore

Casti JL (1992b) Reality rules. II Picturing the world in mathematics. The frontier. John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore

Chassagnole C, Noisommit-Rizzi N, Schmid JW et al. (2002) Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnol Bioeng 79(1):53–73

Cornish-Bowden A (1995) Fundamentals of enzyme kinetics. Portland Press Limited, London

Deutscher J, Francke C, Postma PW (2006) How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. Microbiol Mol Biol Rev 70(4):939–1031

Dunn IJ, Mor JR (1975) Variable-volume continuous cultivation. Biotechnol Bioeng 17(12): 1805–22

Ehrig R, Nowak U, Oeverdieck L et al. (1999) Advanced extrapolation method for large scale differential algebraic problems. High performance scientific and engineering computing. In: H.-J. Bungartz et al. (eds) Computational Science and Engineering, Springer, Berlin

Elf J, Ehrenberg M (2005) Near-critical behavior of aminoacyl-tRNA pools in *E. coli* at rate-limiting supply of amino acids. Biophys J 88(1):132–46

Elf J, Paulsson J, Berg O et al. (2005) Mesoscopic kinetics and its applications in protein synthesis. In: Alberghina FA, Westerhoff HV (eds) Systems Biology, Springer, Berlin, Heidelberg

Fields DS, He Y, Al-Uzri AY et al. (1997) Quantitative specificity of the Mnt repressor. J Mol Biol 271(2):178–94

Götz P, Reuss M (1997) Dynamics of microbial growth: modeling time delays by introducing a polymerization reaction. J Biotechnol 58(2):101–114

Haixin W, Lijun Q, Dougherty E (2007) Modeling genetic regulatory networks by sigmoidal functions: a joint genetic algorithm and Kalman filtering approach. In Third International Conference on Natural Computation (ICNC), pp 324–8

Hardiman T, Siemann-Herzberg M, Reuss M (2007b) Derivation of kinetic parameters for coupled regulatory and metabolic network modeling from DNA-binding site sequences. In Foundations in Systems Biology in Engineering (FOSBE), Conference Proceedings, Stuttgart, Germany, pp 255–9

Hardiman T, Lemuth K, Keller MA et al. (2007a) Topology of the global regulatory network of carbon limitation in *Escherichia coli*. J Biotechnol 132(4):359–74

Hatzimanikatis V, Lee KH (1999) Dynamical analysis of gene networks requires both mRNA and protein expression information. Metab Eng 1(4):275–81

Heinrich R, Schuster S (1996) The regulation of cellular systems. Chapman & Hall, New York

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(3):373–95

Hewitt CJ, 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(3):179–87

Hewitt CJ, Nebe-Von Caron G, Axelsson B et al. (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(4):381–90

Hewitt CJ, Nebe-Von Caron G, Nienow AW et al. (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(6):705–11

Hofmeyr JH, Cornish-Bowden A (1997) The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models. Comput Appl Biosci 13(4):377–85

Hogema BM, Arents JC, Bader R et al. (1998) Inducer exclusion in *Escherichia coli* by non-PTS substrates: the role of the PEP to pyruvate ratio in determining the phosphorylation state of enzyme IIA^{Glc}. Mol Microbiol 30(3):487–98

Hogg T, Mechold U, Malke H et al. (2004) Conformational Antagonism between Opposing Active Sites in a Bifunctional RelA/SpoT Homolog Modulates (p)ppGpp Metabolism during the Stringent Response. Cell 117(1):57–68

Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3:318–56

Jensen KF, 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(2):89–100

Jishage M, Kvint K, Shingler V et al. (2002) Regulation of σ factor competition by the alarmone ppGpp. Genes Dev 16(10):1260–70

Kauffman S (1969) Homeostasis and differentiation in random genetic control networks. Nature 224(5215):177–8

Kremling A, Bettenbrock K, Gilles ED (2007) Analysis of global control of *Escherichia coli* carbohydrate uptake. BMC Syst Biol 1(1):42

Kremling A, Bettenbrock K, Laube B et al. (2001) The organization of metabolic reaction networks. III. Application for diauxic growth on glucose and lactose. Metab Eng 3(4):362–79

Kremling A, Gilles ED (2001) The organization of metabolic reaction networks. II. Signal processing in hierarchical structured functional units. Metab Eng 3(2):138–50

Kremling A, Jahreis K, Lengeler JW et al. (2000) The organization of metabolic reaction networks: a signal-oriented approach to cellular models. Metab Eng 2(3):190–200

Kremling A, Saez-Rodriguez J (2007) Systems biology - an engineering perspective. J Biotechnol 129(2):329–51

Lapin A, Schmid J, 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(14):4783–4797

Lee SB, Bailey JE (1984a) Genetically structured models for *lac* promoter-operator function in the chromosome and in multicopy plasmids: *lac* promoter function. Biotechnol Bioeng 26(11):1383–9

Lee SB, Bailey JE (1984b) Genetically structured models for *lac* promoter-operator function in the *Escherichia coli* chromosome and in multicopy plasmids: *lac* operator function. Biotechnol Bioeng 26(11):1372–82

Lee SY (1996) High cell-density culture of *Escherichia coli*. Trends Biotechnol 14(3): 98–105

Lemuth K, Hardiman T, Winter S, Pfeiffer D, Keller MA, Lange S, Reuss M, Schmid RD, Siemann-Herzberg M (2008) Global transcription and metabolic flux analysis of *Escherichia coli* in glucose-limited fed-batch cultivations. Appl Environ Microbiol 74:7002–15

Lengeler JW, Drews G, Schlegel HG (1999) Biology of the prokaryotes. Georg Thieme Verlag, Stuttgart

Likhoshvai V, Ratushny A (2007) Generalized hill function method for modeling molecular processes. J Bioinform Comput Biol 5(2B):521–31

McAdams HH, Arkin A (1998) Simulation of prokaryotic genetic circuits. Annu Rev Biophys Biomol Struct 27:199–224

McClure WR (1985) Mechanism and control of transcription initiation in prokaryotes. Annu Rev Biochem 54:171–204

Mechold U, Murphy H, Brown L et al. (2002) Intramolecular regulation of the opposing (p)ppGpp catalytic activities of Rel_{Seq}, the Rel/Spo enzyme from *Streptococcus equisimilis*. J Bacteriol 184(11):2878–88

Neidhardt FC, Savageau MA (1996) Regulation Beyond the Operon. In: Neidhardt FC, et al. (eds) *Escherichia coli* and *Salmonella*: cellular and molecular biology, American Society for Microbiology Press, Washington DC

Patten CL, Kirchhof MG, Schertzberg MR et al. (2004) Microarray analysis of RpoS-mediated gene expression in *Escherichia coli* K-12. Mol Genet Genomics

Postma PW, Lengeler JW, Jacobson GR (1993) Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. Microbiol Rev 57(3):543–94

Ramseier TM (1996) Cra and the control of carbon flux via metabolic pathways. Res Microbiol 147(6–7):489–93

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

Rizzi M, Baltes M, Theobald U et al. (1997) *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*. 2. Mathematical model. Biotechnol Bioeng 55(4):592–608

Roels JA (1983) Energetics and kinetics in biotechnology. Elsevier Biomedical Press, Amsterdam, New York, Oxford

Rosen R (1968) Recent developments in the theory of control and regulation of cellular processes. 3. Int Rev Cytol 23:25–88

Saier MH, Jr., Ramseier TM (1996) The catabolite repressor/activator (Cra) protein of enteric bacteria. J Bacteriol 178(12):3411–7

Saier MH, Jr., Ramseier TM, Reizer J (1996) Regulation of carbon utilization. In: Neidhardt FC, et al. (ed) *Escherichia coli* and *Salmonella*: cellular and molecular biology, American Society for Microbiology Press, Washington DC

Schaub J, Mauch K, Reuss M (2008) Metabolic flux analysis in *Escherichia coli* by integrating isotopic dynamic and isotopic stationary ^{13}C labeling data. Biotechnol Bioeng 99(5):1170–85

Stephanopoulos GN, Aristidou AA, Nielsen J (1998) Metabolic engineering: principles and methodologies. Academic Press, London

Stormo GD (1988) Computer methods for analyzing sequence recognition of nucleic acids. Annu Rev Biophys Biophys Chem 17:241–63

Stormo GD (1990) Consensus patterns in DNA. Methods Enzymol 183:211–21

Stormo GD (2000) DNA binding sites: representation and discovery. Bioinformatics 16(1):16–23

Stormo GD, Fields DS (1998) Specificity, free energy and information content in protein-DNA interactions. Trends Biochem Sci 23(3):109–13

Streichert F, Ulmer H (2005) JavaEvA: a Java based framework for Evolutionary Algorithms. http://tobias-lib.ub.uni-tuebingen.de/volltexte/2005/1702/

Takeda Y, Sarai A, Rivera VM (1989) Analysis of the sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments. Proc Natl Acad Sci USA 86(2):439–43

Teich A, Meyer S, Lin HY, Andersson L, Enfors S, Neubauer P (1999) Growth rate related concentration changes of the starvation response regulators s^S and ppGpp in glucose-limited fed-batch and continuous cultures of *Escherichia coli*. Biotechnol Prog 15:123–9

Traxler MF, Chang DE, Conway T (2006) Guanosine 3′, 5′-bispyrophosphate coordinates global gene expression during glucose-lactose diauxie in *Escherichia coli*. Proc Natl Acad Sci USA 103(7):2374–2379

Uptain SM, Kane CM, Chamberlin MJ (1997) Basic mechanisms of transcript elongation and its regulation. Annu Rev Biochem 66:117–72

Van Dien SJ, Keasling JD (1998) A dynamic model of the *Escherichia coli* phosphate-starvation response. J Theor Biol 190(1):37–49

Visser D, Heijnen JJ (2003) Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. Metab Eng 5(3):164–76

Visser D, Schmid JW, Mauch K et al. (2004) Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics. Metab Eng 6(4):378–90

Visser D, van der Heijden R, Mauch K et al. (2000) Tendency modeling: a new approach to obtain simplified kinetic models of metabolism applied to *Saccharomyces cerevisiae*. Metab Eng 2(3):252–75

Vohradsky J (2001a) Neural model of the genetic network. J Biol Chem 276(39):36168–73

Vohradsky J (2001b) Neural network model of gene expression. FASEB J 15(3):846–54

Weaver DC, Workman CT, Stormo GD (1999) Modeling regulatory networks with weight matrices. Pacific Symposium on Biocomputing 4:112–123

Wendrich TM, Blaha G, Wilson DN et al. (2002) Dissection of the mechanism for the stringent factor RelA. Mol Cell 10(4):779–88

Wong P, Gladney S, Keasling JD (1997) Mathematical model of the *lac* operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. Biotechnol Prog 13(2): 132–43

Yagil G (1975) Quantitative aspects of protein induction. Curr Top Cell Regul 9:183–236

Yagil G, Yagil E (1971) On the relation between effector concentration and the rate of induced enzyme synthesis. Biophys J 11(1):11–27

Yee L, Blanch HW (1992) Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*. Biotechnology (NY) 10(12):1550–6