Emerging *Corynebacterium glutamicum* systems biology

Volker F. Wendisch\textsuperscript{a,b,}\textsuperscript{*}, Michael Bott\textsuperscript{a}, Jörn Kalinowski\textsuperscript{c}, Marco Oldiges\textsuperscript{b}, Wolfgang Wiechert\textsuperscript{d}

\textsuperscript{a} Institute of Biotechnology 1, Research Center Juelich, D-52425 Juelich, Germany
\textsuperscript{b} Institute of Biotechnology 2, Research Center Juelich, Germany
\textsuperscript{c} Institute for Genome Research, University of Bielefeld, Germany
\textsuperscript{d} Department of Simulation, Institute of Systems Engineering, University of Siegen, Germany

Received 8 July 2005; received in revised form 12 October 2005; accepted 1 December 2005

**Abstract**

*Corynebacterium glutamicum* is widely used for the biotechnological production of amino acids. Amino acid producing strains have been improved classically by mutagenesis and screening as well as in a rational manner using recombinant DNA technology. Metabolic flux analysis may be viewed as the first systems approach to *C. glutamicum* physiology since it combines isotop e labeling data with metabolic network models of the biosynthetic and central metabolic pathways. However, only the complete genome sequence of *C. glutamicum* and post-genomics methods such as transcriptomics and proteomics have allowed characterizing metabolic and regulatory properties of this bacterium on a truly global level. Besides transcriptomics and proteomics, metabolomics and modeling approaches have now been established. Systems biology, which uses systematic genomic, proteomic and metabolomic technologies with the final aim of constructing comprehensive and predictive models of complex biological systems, is emerging for *C. glutamicum*. We will present current developments that advanced our insight into fundamental biology of *C. glutamicum* and that in the future will enable novel biotechnological applications for the improvement of amino acid production.

© 2005 Elsevier B.V. All rights reserved.

**Keywords:** Systems biology; White biotechnology; *Corynebacterium glutamicum*; Amino acid production; Strain development; Genomics; Transcriptomics; DNA microarray; Proteomics; Metabolic flux analysis; Metabolite profiling; Metabolomics; Mathematical modeling

1. Introduction

Many fields of biology including biotechnology are experiencing a shift towards systems-level characterizations. Systems-level characterizations are established e.g. in developmental biology, ecology or immunology, but in molecular biology they only became possible by the success of the genome sequencing efforts: the sequencing of whole genomes rather than single genes or operons allowed for the first time unraveling the complete genetic information of an organism. On the other hand, the elucidation of
biochemical pathways, their regulatory principles such as feedback-inhibition and the metabolic network analysis under nonequilibrium conditions are well-established in biochemistry and physiology. Today, the prospect of systems biology based on molecular entities excites researchers from all disciplines of biology.

What is systems biology? Currently, no generally accepted definition of systems biology exists. However, there is general agreement that systems biology (Fig. 1) determines the complete repertoire of genes (genome), mRNAs (transcriptome), proteins (proteome), metabolites (metabolome), fluxes of an organism or cell as well as their interactions across all levels based on global analysis techniques and integrates these data into predictive models (Adlerem, 2005; Bork and Serrano, 2005; Hood et al., 2004; Kirschner, 2005; Liu, 2005; Westerhoff and Palsson, 2004). Systematic perturbations by genetic means or by exposition to environmental stimuli are performed to assess the robustness of a cellular system and its modularity as well as to drive model refinement. Ultimately, emergent systems properties, i.e. those that cannot be extrapolated from the sum of the properties of all individual elements, have to be unraveled.

How do we approach systems biology conceptually and experimentally? Physics has been invoked as an example of how to build on the interdependence and iterative refinement of experimentation, theory and technology (Liu, 2005). Theoretical models based on all available experimental data predicted unforeseen systems properties that could only be verified experimentally after advancing technology. Based on anomalies of a model of the solar system the existence of a missing planet had been predicted in the 18th century before Neptune had been discovered, however, systems analysis needs to mature further, both in physics and in biology, to be successfully applied to very complex systems such as the weather or a living cell (Endy and Brent, 2001). Current systems biology approaches are mostly concentrated on collecting and integrating experimental data and building first models, while the gap between simulation and prediction of unforeseen systems properties to be assessed experimentally still needs to be bridged.

Why Corynebacterium glutamicum systems biology? Systems biology of the bacterium C. glutamicum might be useful more generally for developing and testing concepts and in validating and sharpening experimental systems biology tools as e.g. this relatively simple, single-celled microorganism lacks extensive compartmentalization, thus reducing spatial complexity (Bork and Serrano, 2005). While in molecular
biology tools and concepts had mainly been developed in viruses and bacteria before having been transferred to higher eukaryotes (Brock, 1999), systems biology approaches have been initiated in bacteria and higher eukaryotes in parallel (Dieder et al., 2001) with the possibility of mutual benefit.

Systems biology is of fundamental significance and at the same time holds the promise to lead to practical innovations in biotechnology, medicine and engineering. Mostly the impact of systems biology on medicine e.g. on the drug discovery process is discussed (Hood et al., 2004; Hood and Perlmutt, 2004). Here, we focus on the prospects of systems biology approaches to the development of white biotechnology and in particular on the bacterium C. glutamicum. Since its discovery as l-glutamate producing bacterium in 1957, C. glutamicum has become the work-horse of large-scale production of amino acids by fermentation processes. The annual market growth for amino acids is estimated to be about 10% and currently the production of amino acids in fermentative processes with C. glutamicum amounts to 1,500,000 t of l-glutamate per year, 550,000 t of l-lysine per year while amounts of l-glutamine (2000 t per year) and branched-chain amino acids are lower (Hermann, 2003). Due to its importance in amino acid production, C. glutamicum has been subject of physiological, biochemical and genetic studies for four decades and a number of well-defined fermentation processes based on this bacterium are in place. In addition, C. glutamicum is classified as a ‘generally recognized as safe’ (GRAS) organism. Therefore, it lends itself as an ideal host not only for the established production of amino acids, but also for the production of diverse compounds ranging from organic acids to proteins. To fully exploit the biotechnological potential of C. glutamicum and for its development as a platform microorganism for white biotechnology systems biology approaches may prove helpful and even essential.

2. Genomics

The determination of the whole genome sequence of the C. glutamicum wild-type strain ATCC 13032 marked an important breakthrough. Due to its outstanding industrial interest a number of competing genome projects led to the publication of two complete genome sequences (Ikeda et al., 2003; Kalinowski et al., 2003). These two genome sequences allowed the evaluation of the sequencing quality by sequence comparison and resequencing of all differences found before releasing the second sequence (acc. no. BX927147). None of the point mutations that were detected inside of 61 coding regions was found to be a sequencing error (data not shown) revealing that both genome sequences are of outstanding quality.

However, both genome sequences differ by the presence or absence of insertion element copies and a putative prophage (Kalinowski, 2005) showing that these genetic elements can change the genome within short time. In addition, the ATCC 13032 strain differs from many C. glutamicum strains as it does not possess an S-layer (Hansmeier et al., 2004). Another C. glutamicum strain, named “strain R”, displayed even larger differences and was found to contain a number of additional genetic islands (Suzuki et al., 2005).

A complete genome sequence is brought to life by annotation, supported by powerful bioinformatics tools, either for predicting genes, operons and other signals in DNA, or finding homologous proteins in databases and predicting their cellular localization as well as their function. Especially, gene prediction is critical since a complete gene set is essential for the post-genomic methods transcriptome and proteome analysis. Gene prediction tools should be run in combinations using intrinsic and extrinsic methods (McHardy et al., 2004) and also in this case the remaining “intergenic” regions should be carefully inspected for small genes which often escape prediction. Table 1 lists a collection of genes that were annotated based on strong bioinformatics evidence in the BX927147 genome sequence, but not in the NC003450 genome sequence.

The comparison of genome sequences of related species is of advantage for prediction of coding sequences since mutations accumulate less frequently in coding regions than in non-coding regions. Therefore, the C. glutamicum genome can be compared with that of the closely related amino acid producer, C. efficiens (Nishio et al., 2003), and those of the pathogens C. jejuni (Cerdan-Tarraga et al., 2003) and C. jejuni (Tauch et al., in press). Especially, the comparison to C. eficiens is of high value, since both genomes differ in the G+C content of their DNAs by around 10%, although the encoded protein sequences and the gene order are very highly conserved (Nakamura et
### Table 1

C. glutamicum genes annotated in BX927147 but not in BA000036/NC003450 and their coding evidence

<table>
<thead>
<tr>
<th>Locus tag</th>
<th>Gene name</th>
<th>Gene product</th>
<th>Length (aa)</th>
<th>Blastp hit to(^a)</th>
<th>evalue</th>
<th>Tblastn vs. C. efficiens(^b)</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg0449</td>
<td>Hypothetical protein</td>
<td>210</td>
<td></td>
<td>CE0390; 5.0e−13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg0494</td>
<td>Extremely conserved hypothetical protein</td>
<td>33</td>
<td></td>
<td>CE0433; 1.3e−13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg0757</td>
<td>Conserved hypothetical protein</td>
<td>67</td>
<td></td>
<td>DIP51807; 4.3e−15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg0759</td>
<td>prpD2</td>
<td>Propionate catabolism protein PrpD</td>
<td>504</td>
<td>NCg0627, NCg0628</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg0869</td>
<td>Hypothetical protein predicted by Glimmer/Critica</td>
<td>127</td>
<td></td>
<td>CE0775; 6.1e−37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg0877</td>
<td>extA</td>
<td>Putative anti-sigma factor</td>
<td>89</td>
<td>DIP0710; 2.0e−15 Cd</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg1000</td>
<td>Conserved hypothetical protein</td>
<td>66</td>
<td></td>
<td>CE0952; 3.3e−19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg1070</td>
<td>Conserved hypothetical protein</td>
<td>63</td>
<td></td>
<td>CE1009; 2.3e−28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg1172</td>
<td>Putative predicts maintenance killer protein</td>
<td>93</td>
<td></td>
<td>PIP5186; 1.0e−24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg1241</td>
<td>Hypothetical protein predicted by Glimmer/Critica</td>
<td>87</td>
<td></td>
<td>CE1147; 1.4e−09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg1342</td>
<td>Putative membrane protein</td>
<td>82</td>
<td></td>
<td>CE1479; 4.7e−13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg1704</td>
<td>Putative membrane protein</td>
<td>119</td>
<td></td>
<td>CE0874; 3.6e−34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg1741</td>
<td>Hypothetical protein predicted by Glimmer/Critica</td>
<td>84</td>
<td></td>
<td>CE1667; 8.5e−23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg1943</td>
<td>Putative translation elongation factor (GTPase)</td>
<td>123</td>
<td></td>
<td>CE0505; 2.3e−41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg1972</td>
<td>Hypothetical protein predicted by Glimmer/Critica</td>
<td>111</td>
<td></td>
<td>COG4108</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg2292</td>
<td>Hypothetical protein predicted by Glimmer/Critica</td>
<td>253</td>
<td></td>
<td>COG2340</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg2525</td>
<td>Putative membrane protein</td>
<td>46</td>
<td></td>
<td>COG0456; 2.6e−18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg2546</td>
<td>Putative related or membrane protein</td>
<td>68</td>
<td></td>
<td>CE2202; 3.6e−18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg2555</td>
<td>Conserved hypothetical protein</td>
<td>37</td>
<td></td>
<td>CE2426; 2.4e−17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg2791</td>
<td>rpmJ</td>
<td>Ribosomal protein L36</td>
<td>40</td>
<td>CE2426; 2.4e−17</td>
<td></td>
<td>Lichtinger et al. (1998)</td>
<td></td>
</tr>
<tr>
<td>cg3008</td>
<td>porA</td>
<td>Porin</td>
<td>45</td>
<td>CE2560; 6.0e−20</td>
<td></td>
<td>Hinton et al. (2009)</td>
<td></td>
</tr>
<tr>
<td>cg3009</td>
<td>porD</td>
<td>Porin</td>
<td>57</td>
<td>CE2560; 6.0e−20</td>
<td></td>
<td>Hinton et al. (2009)</td>
<td></td>
</tr>
<tr>
<td>cg3117</td>
<td>cyoX</td>
<td>Ferredoxin-like protein</td>
<td>82</td>
<td>CE2643; 5.0e−43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg3257</td>
<td>Conserved hypothetical protein</td>
<td>46</td>
<td></td>
<td>COG1917</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg3430</td>
<td>Hypothetical protein predicted by Glimmer/Critica</td>
<td>99</td>
<td></td>
<td>CE2945; 9.2e−46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg3433</td>
<td>Hypothetical protein predicted by Glimmer/Critica</td>
<td>73</td>
<td></td>
<td>CE0052; 1.8e−17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Similarity to proteins from *C. glutamicum* NC003450 (NCg), *C. efficiens* (CE), *C. diphtheriae* (DIP), and *Pseudomonas putida* (PP) or to the Clusters of Orthologous Groups of proteins (COG).

\(^b\) Similarity to the translated nucleotide sequence of the *C. efficiens* YS-314 genome (BA000035).

This evolutionary pressure for mutation to higher G+C content very efficiently uncovered the true gene starts as well as putative regulator binding sites and even promoters.

Using a combination of bioinformatics tools, cross-genome comparisons revealed corynebacterial gene families as exemplified for the genes for 127 different DNA-binding transcriptional regulators in the genome sequence of *C. glutamicum* which could be grouped into 25 different families (Brune et al., 2005). It turned out that only 28 regulators were conserved between all four corynebacterial genomes, apparently representing...
essential regulators that are controlling cell division, stress response, metal homeostasis as well as carbohydrate and sulfur metabolism.

Although prediction of gene functions and of regulatory signals is of outstanding importance when analysing thousands of genes within a genome, predictions often are only indicative of a gene’s affiliation to a distinct gene family or of the cellular localization of the encoded protein. In addition, as these annotations are associated with a probability score which often gives no clear prediction, functional analysis of predicted genes by experimentation may become essential. Introducing a defined deletion into the chromosome which removes the gene of interest is often necessary to precisely identify its function, especially when multiple candidates for a given function must be assessed. On the other hand, physiological characterization of strains overexpressing a predicted gene or biochemical characterization of purified proteins may be useful. For these purposes, C. glutamicum is well suited since efficient gene replacement and overexpression techniques are well developed (Kirchner and Tauch, 2003).

By the combination of predictions in silico with functional analyses by gene deletion or overexpression and phenotypic characterizations, complex metabolic and regulatory pathways can be reconstructed, thereby providing a basis for flux and metabolome analyses. Reconstruction of the methionine biosynthesis (Rückert et al., 2003), the prediction and functional analysis of transaminase genes (McHardy et al., 2003) as well as that of the three pathways of trehalose biosynthesis (Wolf et al., 2003) and those of sulfonate and sulfonate ester degradation are examples of such combined analyses.

Another example is the use of metabolic models in conjunction with genome comparisons between a wild-type and a producer strain originally established by random mutagenesis and selection for higher production. The prototype of such experiments was described by Ohnishi et al. (2002), who identified mutations within three genes known to be involved in L-lysine production in a classically derived producer strain. By introducing these mutations into the C. glutamicum wild-type strain a strain was obtained which displayed a high L-lysine yield while retaining a high vitality and thereby leading to an increased productivity (Ohnishi et al., 2002). This approach was extended further by introducing additional mutations (Georgi et al., in press; Ohnishi et al., 2005).

The ultimate goal of these bioinformatic and functional analyses is a fully and deeply annotated genome sequence. Due to the speed with which mutations can be generated and with which high-throughput genomic and post-genomic analyses can be performed, this goal might be in reach within the not so far future. A fully annotated genome will then no longer be a simple “inventory” but might serve as “blueprint” for Systems Biology approaches as well as for rational strain development in industrial production processes.

3. Transcriptomics

DNA microarray technology for C. glutamicum has been established and has been applied to transcriptome analysis (reviewed in Wendisch (2003)), to genome comparisons (Stansen et al., in press) and in medical diagnostics (Roth et al., 2004). Many studies published in 2004 and 2005 employed transcriptome analysis for the characterization of diverse aspects of C. glutamicum physiology: regulation of sulfur (Rey et al., 2005), nitrogen (Silberbach et al., 2005), phosphorus (Rittmann et al., in press; Wendisch and Bott, 2005) and carbon metabolism (Gerstmeir et al., 2004; Netzer et al., 2004a; Stansen et al., in press), glutamate production triggered by ethambutol addition (Radmacher et al., 2005) or by a temperature-shift (Stansen et al., in press), lysine production (Kromer et al., 2004), serine catabolism (Netzer et al., 2004b), magnesium (Pallerla et al., 2005) and iron availability (Krug et al., 2005), regulation by the two-component signal transduction system MtrA-MtrB (Moker et al., 2004), by the regulator of Clp protease gene expression ClgR (Engels et al., 2005; Engels et al., 2004) and by the regulator of aconitase gene expression AcnR (Krug et al., 2005).

Transcriptome analysis of a spontaneous C. glutamicum mutant allowed mapping the mutated gene locus (Netzer et al., 2004a).

While transcriptome analyses helped to advance our knowledge about C. glutamicum in each of these studies, the characterizations of stimulons and regulons are most important from a systems level point of view and are a cornerstone to define the regulatory network topology (Wendisch, 2005). While a stimulon
comprises the group of genes that change expression in response to a stimulus, e.g. nitrogen or phosphate starvation (Ishige et al., 2003; Silberbach et al., 2005), a regulon comprises the group of genes controlled by a transcriptional regulator, e.g. by AcnR or McbR (Krug et al., 2005; Rey et al., 2005).

*C. glutamicum* responds to ammonium availability involving genetic regulation by the transcriptional regulator AmtR (Burkovski, 2003). To identify the ammonium starvation stimulon it was important to distinguish nitrogen-dependent from growth-rate dependent effects as growth ceases when ammonium becomes scarce. This could be achieved by comparing two ammonium-limited continuous cultures differing by the dilution rate with an ammonium-sufficient batch culture (Silberbach et al., 2005). The identification of the ammonium starvation stimulon allowed unraveling the physiological strategy of *C. glutamicum* to respond to the limiting availability of a nitrogen source. Like *E. coli* (Zimmer et al., 2000), *C. glutamicum* scavenges nitrogen from alternative sources when starved for ammonium as deduced from the increased expression of genes for transport and metabolism of amino acids, urea, etc. (Silberbach et al., 2005). In contrast, altered expression of e.g. genes for ribosomal proteins was a consequence of reduced growth rather than nitrogen limitation per se (Silberbach et al., 2005). Continuous cultures are useful to determine long-term adaptation to an environmental stimulus, but the kinetics of a stimulus response can only be determined in shift experiments combined with time-resolved transcriptome analyses as demonstrated for the phosphate starvation response of *C. glutamicum* (Ishige et al., 2003; Wendisch and Bott, 2005). While in *C. glutamicum*, but not in *E. coli*, expression of ushA encoding a secreted 5′-nucleotidase/UDP-sugar hydrolase is induced by phosphate starvation (Rittmann et al., in press), the phosphate starvation stimulon of *C. glutamicum* is in general similar to that of *E. coli* (Wanner, 1996) with the characteristic induction of operons for high affinity uptake of inorganic phosphate and of organophosphate as well as of genes for secreted phosphatases (Ishige et al., 2003; Rittmann et al., in press; Wendisch and Bott, 2005). Taken together, to characterize stimulons as well as the dynamics of a response to a stimulus it will prove useful to combine steady-state analysis of continuous cultures with time-resolved analysis of pulse experiments.

To characterize a regulon, a combination of transcriptome analyses and bioinformatics with biochemical and genetic experiments is often used. By global gene expression analysis, McbR of *C. glutamicum* was shown to control expression of 86 genes. Forty-five of these genes are organized in 22 operons that share a conserved motif in their promoter regions and code for proteins involved in all aspects of transport and metabolism of the macroelement sulfur, i.e. S-adenosylmethionine and cysteine biosynthesis, sulfate reduction, uptake and utilization of sulfur-containing compounds (Rey et al., 2005). Band-shift experiments revealed that S-adenosylhomocysteine, a product of S-adenosylmethionine-dependent transmethylation reactions, prevented the binding of McbR to the conserved sequence motif, thus providing a link between sulfur and CI metabolism. As the McbR regulon also includes two putative regulators and as at least a subset of the McbR regulon might be additionally controlled by other regulators, the hierarchy and network topology of sulfur-dependent regulation remains to be elucidated.

The TetR-type transcriptional regulator AcnR was shown to repress aconitase gene (*acn*) expression in *C. glutamicum* and DNA microarray analyses indicated that *acn* is its primary target gene (Krug et al., 2005). AcnR is active as a homodimer and binds to the *acn* promoter in the region from –11 to –28 relative to the transcription start. This AcnR binding motif and the organization of its gene into an operon together with *acn* is conserved in corynebacteria and mycobacteria (Krug et al., 2005). As the activities of citrate synthase and isocitrate dehydrogenase are relatively constant on various carbon sources while aconitase activities were 2.5–4-fold higher on propionate, citrate, or acetate than on glucose (Krug et al., 2005), it appears that aconitase is a major control point and AcnR an important regulator of tricarboxylic acid cycle activity in *C. glutamicum*. Future studies will reveal whether AcnR control is restricted to *acn* or whether its regulon is more complex. To this end, so-called ChIP-to-chip analyses (Buck and Lieb, 2004) will be helpful as they allow identifying those regions of the genome that are bound by a transcriptional regulator. Thus, these analyses identify the genes and operons directly regulated by a transcriptional regulator and they facilitate consistent model building as direct regulatory effects can be distinguished from indirect effects propagated as consequence of primary regulatory event(s).
4. Proteomics

Proteins fulfill many different functions. They are the major catalysts of metabolism and key to regulatory processes. Due to the fact that proteins are composed of 20 different amino acids, their physicochemical properties are highly diverse. Thus, their comprehensive identification and characterization is much more difficult than that of nucleic acids.

The method currently used most frequently to establish the protein composition of cells is two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and protein identification by mass spectrometry. 2D-PAGE separates proteins according to their isoelectric point (isoelectric focusing) and their apparent molecular mass (sodium dodecylsulfate-polyacrylamide gel electrophoresis). Protocols for 2D-PAGE of C. glutamicum proteins have been established in recent years (for review see Schaffer and Burkovski (2005)) and used to separate the proteins of different cell fractions (Fig. 2), i.e. soluble fraction, membrane fraction and secreted proteins (Hermann et al., 1998, 2000, 2001). In these studies, 42 proteins were identified by N-terminal microsequencing, Western blotting with specific antisera or mass spectrometry. A first high-resolution reference map for cytoplasmic and membrane-associated proteins of glucose-grown C. glutamicum was established by using narrow immobilized pH gradients and tryptic peptide mass fingerprinting (Schaffer et al., 2001). In this work, 152 different proteins of the cytoplasmic fraction were identified as well as 13 proteins of the membrane fraction. Recently, the cytosolic, cell surface and extracellular proteomes of C. efficiens cells grown Luria-Broth were analysed by 2D-PAGE and 164, 49 and 89 proteins of the three fractions were identified by peptide mass fingerprinting, respectively.

Fig. 2. Fractionation of C. glutamicum cells into cellular entities according to Hermann et al. (2001) and Schlüsenner et al. (2005). Abbreviations: UC, ultracentrifugation; ASB-14, amidoxylfolate 14; AIEC, anion exchange chromatography.
For comparison, the corresponding fractions of *C. glutamicum* grown in Luria-Broth were also analysed and compared to that of *C. efficiens*, revealing only minor differences between the two species (Hansmeier et al., in press). The latter study led to the identification of 119 additional *C. glutamicum* proteins.

A major limitation of current 2D-PAGE is the failure to separate integral membrane proteins. For example in the study by Schaffer et al. (2001), only proteins with maximally one transmembrane helix were identified in the membrane fraction. Recently, an alternative method was established (Fig. 2) to separate and identify intrinsic membrane proteins of *C. glutamicum* (Schlüsener et al., 2005). It involves the initial separation of amidosulfobetaine-14-solubilized membrane proteins by anion exchange chromatography (AIEC) and subsequent separation of the individual AIEC fractions by SDS-PAGE followed by the identification of individual protein bands by peptide mass fingerprinting after digestion with trypsin and CNBr cleavage (Schlüsener et al., 2005). Of the 170 different identified proteins, 50 were membrane-integral with up to 13 transmembrane helices. This number corresponds to 7.5% of the predicted membrane proteome. Thus, this method appears to be suitable for the qualitative and possibly also semiquantitative comparison of membrane proteomes.

Phosphorylation of proteins is a central regulatory mechanism both in eukaryotes and prokaryotes. In contrast to phosphohistidine and phosphoaspartate, the residues involved in two-component signal transduction by sensor kinases and response regulators (Moker et al., 2004), phosphoserine, phosphothreonine and phosphotyrosine residues are quite stable and proteins containing groups, e.g. covalently linked flavine dinucleotide. Thirty-one spots were detected both with radiolabeling and immunostaining. By peptide mass fingerprinting, 41 of the detected proteins were identified, most of which act in central metabolic pathways such as glycolysis or the citric acid cycle. This points to an important role of protein phosphorylation on serine and threonine residues (about 20 proteins reacted with phosphothreonine-specific antibodies) in *C. glutamicum*’s regulatory processes. However, the exact residues phosphorylated in the different proteins and the effects of phosphorylation on protein activity remain to be elucidated.

Regulation within the cell may occur on many different levels, i.e. transcription, mRNA stability, translation, protein stability, or protein activity. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability. The first two of these can be analysed at the RNA level with DNA microarrays, whereas the others require analysis of protein stability, protein activity, or protein stability.

Recent studies indicate that many cellular processes are carried out by multiprotein complexes. The identification and analysis of their components provides insight into how the protein repertoire of a cell is organized into functional units. The currently preferred method to identify interactions between individual proteins is large scale two-hybrid systems (Fields and Song, 1989), whereas the composition of multiprotein complexes is elucidated e.g. by tandem affinity purification (Gavin et al., 2002; Ho et al., 2002) of in vivo tagged proteins. In *C. glutamicum*, no large scale studies on protein–protein interaction and the identification of protein complexes have yet been performed. However, in vivo tagging and affinity purification was used to identify a cytochrome-bc1-ao1-supercomplex (Niebisch and Bott, 2003), confirming that this method is applicable to *C. glutamicum*.

There is no doubt that proteins and thus all kind of studies on proteins are of central importance for a
holistic understanding of the cell. As is evident from above, proteome analysis of *C. glutamicum* is in an advanced state, although protein identification has not yet been performed exhaustively for 2D gels or SDS gels of membrane proteins. Some key issues that have to be solved in the future in order to obtain convincing mathematical models of metabolism and its regulation are quantification of intracellular protein concentrations and modifications and a far better knowledge of regulation at the level of enzyme activity. The contribution of in-depth biochemical and biophysical characterizations of individual proteins and enzymes to systems biology often is underestimated. However, as large-scale methods for thorough biochemical and biophysical enzyme analyses are not yet available, the analysis of individual proteins is still an indispensable element of systems biology. Moreover, it appears necessary to focus at first instance on individual subsystems of the cell (e.g., glycolysis, citric acid cycle or biosynthesis pathways) and to obtain reliable data for these, since models can only be as good as the data they are based on. Subsequently, the subsystems can be combined and intertwined to a complete picture of the cell.

5. Metabolomics and fluxomics

Metabolomics function as substrates and products of proteins as the major biocatalysts in the cell, but also have important regulatory functions. Fructose-1,6-bisphosphate, AMP or ATP e.g. modify enzyme activities in *C. glutamicum* (reviewed in Eikmanns (2005)) to allow the cell to adapt to environmental changes very fast. In addition, several metabolites function in genetic regulation by modulating the activities of e.g. transcriptional regulators like *C. glutamicum* McbR (Rey et al., 2005).

Metabolomics deals with the quantification of intracellular metabolite concentrations while fluxomics refers to the reaction velocities between metabolite pools in the biochemical network. Although both values (pool and flux) seem to be connected to each other, no direct quantitative flux data can be derived simply from the concentration of the reacting metabolites because the flux is also determined by the kinetic properties and amount of the corresponding enzyme present in the cell. Hence, the in vitro flux may be determined by genetic regulation from transcription to protein translation to posttranslational modification (e.g. phosphorylation), by product and substrate metabolite concentrations and by allosteric control (inhibition or activation of enzymes by other metabolites). Although the in vitro enzyme properties can be extracted from a database (Schomburg et al., 2002) they do not correctly reflect the enzyme properties in vivo in the presence of the whole metabolic network.

Metabolomics is currently gaining more and more interest because the metabolome is downstream of gene and protein regulation and thus adds valuable information not accessible by other methods, like the in vivo status of allosteric control (Allen et al., 2003; Fiehn et al., 2000a). Metabolomics approaches have been developed relatively late as low amounts of metabolites are present within the cells (3–5% dry cell weight (Neidhardt and Umbarger, 1996)) and their high degree of chemical diversity complicates sample preparation and analytical procedures making it almost impossible to separate and quantify all metabolites using a single experimental technique as is possible for RNA (DNA-chip) and proteins (2D-gel electrophoresis).

The metabolomics approach has been widely applied in plants (Fiehn et al., 2000a,b) but was also developed for microbial systems as *E. coli* (Chassagnole et al., 2002; Oldiges et al., 2004; Schaefer et al., 1999), *S. cerevisiae* (Theobald et al., 1997; Visser et al., 2004) and *B. subtilis* (Soga et al., 2003). In addition to established GC–MS techniques (Schafer et al., 2005; Strelkov et al., 2004) more robust and sensitive analytical devices for LC–MS coupling became available and accelerate the application of metabolomics to diagnostics, functional genomics and systems biology (Fernie et al., 2004). Because of its high impact as an industrial amino acid producing microorganism *C. glutamicum* surprisingly played only a minor role in this area so far, although some relative concentration data of the *C. glutamicum* wild type (Strelkov et al., 2004) and first quantitative data of an amino acid producing *C. glutamicum* strain were presented (Magnus et al., 2005). From a systemic point of view, quantitative metabolite data are advantageous, because kinetic models of the underlying biochemical network can be derived from them, although the experimental effort to achieve absolute quantitative data of metabolites at cytosolic level is much higher even in complex biological sample matrices and a large number of metabolites.
According to Fiehn et al. (2000a), metabolome analysis is multifunctional making use of different analytical approaches depending on the topic of the study and some examples of applications to C. glutamicum have been described. Metabolome analysis can thus be subdivided into (i) target analysis, aiming at measuring distinct substrates or products e.g. to distinguish differences after genetic modification, (ii) metabolic profiling (Magnus et al., 2005), focusing on a complete pathway or linked pathways and (iii) metabolomics (Strelkov et al., 2004), striving for an unbiased, semi-quantitative overview of whole-cell metabolic patterns.

For a more rapid analysis, metabolic fingerprinting (Fiehn, 2001) can be used, which reduces the analytical effort to the analysis of those metabolites with biochemical relevance. Quite recently, another variant of high-throughput classification was presented, called metabolic footprinting (Allen et al., 2003), which differs from the aforementioned methods by analysing extracellular metabolites instead of intracellular pools.

In order to setup kinetic metabolic models, the metabolic profiling approach is preferred, because quantitative metabolite data for a set of pathways (e.g. glycolysis, pentose-phosphate-pathway, TCA-cycle, amino acid pathway) are provided. For the functional genomics approach metabolomics or metabolic fingerprinting is well suited, because data of relative concentration changes for a high number of compounds can contribute to the identification of the function of a gene product in mutant strains (Fernie et al., 2004; Fiehn et al. 2000a; Raamsdonk et al., 2001).

Unlike transcriptome and proteome analysis, metabolome analysis has to be very fast (sub-second scale) to monitor physiological changes in the environment of the cell. Hence, quenching of metabolic activities for metabolomics studies has to be even faster than for transcriptome and proteome analysis (Lange et al., 2001). Several automated techniques allowing reproducible high-frequency sampling (Fig. 3) (Buziol et al., 2002; Schafer et al., 1999; Visser et al., 2004) have been applied mainly for E. coli and S. cerevisiae, but preliminary results for an l-valine producing C. glutamicum strain have been presented recently (Magnus et al., 2005).

Roels (1983) ranked the relaxation times following a particular stimulus for several cellular responses according to their time scale. The fastest responses may be explained by mass action law kinetics followed by the allosteric processes, controlling the activity of enzymes. Changes on the genetic or the proteomic level are slower and the enzyme concentrations can be assumed to be constant for a few minutes after exposition to a stimulus. Keeping this in mind, a stimulus-response experiment (Oldiges and Takors, 2005) using a substrate step increase (glucose pulse), will show no effect for the genetic or proteomic properties of metabolism within an observation window of 1 or 2 min. Hence, the recorded metabolome response will only contain information about the enzyme kinetics and at the same time minimizing the superimposed effects of gene and protein regulation. These stimulus-response experiments can serve as a data source to setup dynamic kinetic models (Wiechert, 2002) and to identify and quantify kinetic enzyme parameters/effectors in order to describe the biochemical network on a systemic level. Current developments aim at the use of 13C-labelled glucose as pulse substrate instead of naturally labelled glucose (Mashego et al., 2004). This will have the advantage that the stimulus shift (step increase) from glucose limitation to glucose saturation will be replaced by a stimulus shift from 13C limitation to 13C saturation, circumventing the physiological switch from glucose limitation to saturation.

The use of 13C-labelled substrates (e.g. glucose) is a general technique for the determination of the flux, i.e. all intracellular metabolic fluxes within the stoichiometric biochemical network. Metabolic flux analysis was the first approach to C. glutamicum strain development with a systems level focus since it requires metabolic network models of the biosynthetic and central carbon metabolic pathways. The investigation of l-lysine production by C. glutamicum was a major driving force for the development of this 13C labelling technology, although it is not limited to a certain biological species. After several hours of feeding a mixture of 13C-labelled and unlabelled glucose to a growing culture, the 13C labelling information has been distributed and equilibrated in the stoichiometric metabolic network and accumulated in the amino acids of the proteome. After protein hydrolysis, 13C information is measured via GC–MS (Christensen and Nielsen, 1999; Dauner and Sauer, 2000) or more detailed using NMR (Marx et al., 1996). The 13C metabolic flux analysis (13C MFA) uses the determined extracellular reaction
rates and the $^{13}$C amino acid data to calculate the intracellular fluxes in the network (Wiechert, 2001a). This method has been intensively used for flux measurements to improve L-lysine production with *C. glutamicum* and revealed the importance of NADPH supply (Marx et al., 1996, 1999, 2003, 1997; Wittmann and Heinzle, 2002), of different anaplerotic pathways and a metabolic futile cycle in anaplerosis (Petersen et al., 2000, 2001; Peters-Wendisch et al., 1996; Wendisch et al., 2000). A series of multiple flux determinations during a fed-batch process for L-lysine production with *C. glutamicum* was described using the sensor-reactor concept (Drysch et al., 2003; El Massaoudi et al., 2003). In combination with a rapid sampling device (sampling frequency < 1 Hz) such concept will provide sample material for intracellular metabolome and transient $^{13}$C metabolome analysis using mass spectrometric methods. This combination of metabolome and $^{13}$C fluxome analysis is no longer limited to growth-coupled systems and will presumably allow faster access to the equilibrated $^{13}$C labelling information in the metabolic pathways of *C. glutamicum* under various growth and production conditions.

Quantitative metabolome and fluxome data are the basis for a quantitative analysis of the metabolic network properties in *C. glutamicum* with mathematical modelling tools. In combination with semi-quantitative metabolomics data to support functional genomics they contribute to the systems-based investigation of *C. glutamicum*. 

---

Fig. 3. Experimental setup for stimulus-response experiments used by (Schaefer et al., 1999). Automated rapid sampling unit in the lower foreground.
6. Mathematical modeling

Due to its integrative nature, modeling has become a key activity of any research project in systems biology. In general, mathematical modeling intends to represent the structure and behavior of a real system by means of mathematical concepts. For example structural graphs, signal flow diagrams, equation systems, ordinary and partial differential equations, logical rules, or probabilistic relations have been used in the past to describe biological systems (Wiechert, 2002).

Mathematics is a universal discipline, and thus model based methods developed for one specific organism can be carried over to a similar system if the underlying structures and data are modified appropriately. This means that tools developed for prokaryotic organisms like *E. coli* or *B. subtilis* will probably also apply to *C. glutamicum*. Likewise, methods from eukaryotic single cells, like *S. cerevisiae*, have good chances to be adapted as long as cellular compartmentation and the details of genetic regulation do not play a central role. On the other hand methods for plant and animal cells typically become more and more different from prokaryotic models because additionally intracellular diffusion and intra/intercellular signaling are becoming important.

It should be distinguished between a mathematical model as an abstract system representation, mathematical methods applicable to a certain class of models and the concrete implementations of these methods as model based software tools. Mathematical tools have been developed for graphical and textual model specification, stationary and instationary simulation, mathematical systems analysis, experimental model validation, model based prediction and in silico systems optimization (Jong, 2002; Wiechert, 2002). The following examples show where *C. glutamicum* has in recent years been the subject of theoretical investigations and tool development on different abstraction levels.

Metabolic flux analysis (MFA) aiming at the quantitative determination of metabolic fluxes certainly is the method that most profited from systemic questions on the *Corynebacterium* metabolism (de Graaf, 2000; Vallino and Stephanopoulos, 1993). The underlying mathematics for the classical stoichiometric MFA have been developed for graphical and textual model specification, stationary and instationary simulation, mathematical systems analysis, experimental model validation, model based prediction and in silico systems optimization (Jong, 2002; Wiechert, 2002). The following examples show where *C. glutamicum* has in recent years been the subject of theoretical investigations and tool development on different abstraction levels.

Metabolic flux analysis (MFA) aiming at the quantitative determination of metabolic fluxes certainly is the method that most profited from systemic questions on the *Corynebacterium* metabolism (de Graaf, 2000; Vallino and Stephanopoulos, 1993). The underlying mathematics for the classical stoichiometric MFA (Heijden et al., 1994) and for the more powerful but also more complicated $^{13}$C MFA (Wiechert and Wurzel, 2001) is now well developed. Designed $^{13}$C labeling experiments (Mölney et al., 1999) could even resolve entangled fluxes like the different anaplerotic as well as reverse fluxes (Petersen et al., 2000, 2001) catalysed by PEP carboxylase (Eikmanns et al., 1989), PEP carboxykinase (Riedel et al., 2001) and pyruvate carboxylase (Peters-Wendisch et al., 1998, 1997). Genetic (Peters-Wendisch et al., 2001; Riedel et al., 2001) and a series of flux analyses of different *Corynebacterium* strains (Petersen et al., 2000, 2001) revealed a futile cycle having a strong influence on lysine production. Current developments in $^{13}$C MFA aim at an application to strain comparison (Shirai et al., 2005), instationary process conditions (Dysch et al., 2004; Wiechert and Noh, 2005), miniaturized reactors (Wittmann et al., 2004) and high throughput methods for metabolic screening (Fischer et al., 2004; Wittmann and Heinzle, 2002).

Flux balance analysis (FBA) (Edwards et al., 2002) and metabolic network analysis (MNA) (Schilling et al., 1999) are also based on stoichiometry and are concerned with the exploration of the space of feasible fluxes that is given by the stoichiometric relations and reaction irreversibility assumptions. MNA tries to obtain a general picture of the metabolic flexibility of an organism to adapt to different external conditions by computing the elementary flux modes (Stelling et al., 2002) as well as to give information concerning the importance of single reactions (Klamt and Gilles, 2004). Although a genome-wide model of *C. glutamicum* is currently not available, it will certainly be ranked in between the very flexible *E. coli* (Reed et al., 2003) and the rather rigidly constrained *B. subtilis* (Fischer and Sauer, 2005; Stelling et al., 2004). FBA is focused on how the metabolic flux map must look like to optimize one single, freely chosen target flux. Unfortunately, the theoretically computed optimum for lysine production by *C. glutamicum* is still far away from the real productivity (Stephanopoulos and Vallino, 1991). Obviously, more constraints like thermodynamic restrictions (Mavrounioti, 1993) must be introduced to obtain better estimates of the productivity limits.

All stoichiometry based methods (MFA, MNA, FBA) are limited because they are assuming a stationary metabolic state whereas the regulation of the pathways is completely left out. However, more
complex mechanistic models for stationary and insta-
tionary states based on enzyme kinetic relations
(Heinrich and Schuster, 1996) allow a detailed anal-
ysis of cellular regulation, e.g. single branch points
of the *Corynebacterium* metabolism (Vallino and
Stephanopoulos, 1994a,b). Sensitivity based methods
like metabolic control theory are widely used estab-
lished tools (Kell and Westerhoff, 1986). However,
mechanistic models must rely on many biological
assumptions and consequently, their prediction must
always be taken with care (Wiechert, 2004). As a com-
promise between stoichiometric and regulatory mod-
els a multi scale model of the anaplerotic reactions
of *Corynebacterium* in the context of global stoi-
chiometry (Petersen et al., 2003) has been developed
and validated with a series of stationary experiments.
A Bayesian approach for parameter estimation and
rigorous error estimation of this model (von Lieres
and Wiechert, 2004) taking all sources of errors into
account showed that the model can predict the effect
of several genetic modifications in the anaplerosis with
a reasonable precision (Petersen et al., 2003; Riedel et
al., 2001; Peters-Wendisch et al., 1998).

The investigation of metabolism under highly
dynamic conditions has been driven forward by the
development of stimulus-response experiments and, in
particular, rapid sampling techniques (see above and
Fig. 3). The goal of stimulus-response experiments
usually is to verify hypotheses on pathway regulation,
to obtain more insight into the regulatory structure
of single pathways, to find hitherto unknown enzyme
effectors, or to determine enzyme kinetic parameters
in vivo. High performance computational methods had
to be developed and many different model variants had
to be considered in the modeling process (Haunschild
et al., 2005). Rapid sampling of a l-valine producing
*C. glutamicum* strain (Magnus et al., 2005) revealed
as a preliminary result of model based data evaluation
that the acetolactate synthase (IlvBN) and the l-valine
export show a high degree of flux control in the l-valine
biosynthesis pathway. Clearly, there cannot be a systemic understanding of a microorganism or even parts of it without taking the genetic and protein regulation into account. One of the goals of *C. glutamicum* systems biology must be the quantitative prediction of the influence of genetic alterations on the metabolic fluxes. However, the quan-
titative modeling of genetic regulation is still an open
field, mainly because the required level of model detail
is not clear (Jong, 2002). It is still under discussion to
what extent gene transcription, RNA degradation, gene
translation, protein phosphorylation or protein turnover
must be taken into account. Also, the available mea-
surements related to genetic and proteomic processes
are at best semi quantitative which strongly hampers
model validation.

As a consequence, phenomenological and statisti-
cal approaches correlating transcriptome, proteome,
fluxome and metabolome data currently dominate the
field (Bro and Nielsen, 2004). DNA microarray data
was interpreted in the context of the metabolic net-
work structure (Oh and Liao, 2000), used to reconstruct
gene regulatory signals by using network component
analysis (Liao et al., 2003), to obtain further con-
straints on metabolic fluxes (Akesson et al., 2004) or
was mapped to physiological states (Stephanopoulos
et al., 2002). For *C. glutamicum*, combinations of tran-
scriptome with proteome analyses have been used to
identify targets to improve valine production (Lange
et al., 2003) and to characterize nitrogen regulation
(Silberbach et al., 2005), acetate regulation (Gerstmeir
et al., 2003) or regulation by the Clp protease (Engels
et al., 2004). While the phase shift from growth to lysine production was studied by combined transcriptome, metabolome and fluxome analyses (Kromer et al., 2004), serine utilizing and
pantothenate producing *C. glutamicum* strains were
analysed using combined transcriptome and fluxome
analyses (Husser et al., 2005; Netzer et al., 2004b).

**7. Conclusion**

Although systems biology approaches have only
recently been established for *C. glutamicum*, impres-
sive progress has since been made especially with
respect to fundamental insight into principles of
metabolic regulation. Current and future efforts focus
on the integration of quantitative data from genomics,
transcriptomics, proteomics, metabolomics and flux
analysis to build and evaluate metabolic and regula-
tory models of *C. glutamicum*. Eventually, the potential
of these approaches for the rational improvement of
*C. glutamicum* strains for amino acid production and,
more broadly, for white biotechnology will fully be
harnessed.
Acknowledgements

The support of the Bundesministerium für Bildung und Forschung (BMBF) through grant 031U13ED031U1213D is gratefully acknowledged. The project was partially funded by the Deutsche Forschungsgemeinschaft (DFG) through grants (TA241/2-1, TA 241/2-2).

References


Eikmanns, B.J., Follettie, M.T., Griot, M.U., Sinskey, A.J., 1989. The phosphoenolpyruvate carboxylase gene of Corynebacterium glutamicum: molecular cloning, nucleotide sequence, and expres-


Eikmanns, B.J., Follettie, M.T., Griot, M.U., Sinskey, A.J., 1989. The phosphoenolpyruvate carboxylase gene of Corynebacterium glutamicum: molecular cloning, nucleotide sequence, and expres-


Eikmanns, B.J., Follettie, M.T., Griot, M.U., Sinskey, A.J., 1989. The phosphoenolpyruvate carboxylase gene of Corynebacterium glutamicum: molecular cloning, nucleotide sequence, and expres-


Fischer, E., Sauer, U., 2005. Large-scale in vivo fluxes reveal rigid-


