Biotechnology and Bioprocess Engineering **2005**, 10: 408-417

Applications of Metabolic Modeling to Drive Bioprocess Development for the Production of Value-added Chemicals

Radhakrishnan Mahadevan, Anthony P. Burgard, Iman Famili, Steve Van Dien, and Christophe H. Schilling*

Genomatica, Inc., 5405 Morehouse Drive, Suite 210, San Diego, CA 92130, USA

Abstract Increasing numbers of value added chemicals are being produced using microbial fermentation strategies. Computational modeling and simulation of microbial metabolism is rapidly becoming an enabling technology that is driving a new paradigm to accelerate the bioprocess development cycle. In particular, constraint-based modeling and the development of genome-scale models of industrial microbes are finding increasing utility across many phases of the bioprocess development workflow. Herein, we review and discuss the requirements and trends in the industrial application of this technology as we build toward integrated computational/experimental platforms for bioprocess engineering. Specifically we cover the following topics: (1) genome-scale models as genetically and biochemically consistent representations of metabolic networks; (2) the ability of these models to predict, assess, and interpret metabolic physiology and flux states of metabolism; (3) the model-guided integrative analysis of high throughput 'omics' data; (4) the reconciliation and analysis of on- and off-line fermentation data as well as flux tracing data; (5) model-aided strain design strategies and the integration of calculated biotransformation routes; and (6) control and optimization of the fermentation processes. Collectively, constraint-based modeling strategies are impacting the iterative characterization of metabolic flux states throughout the bioprocess development cycle, while also driving metabolic engineering strategies and fermentation optimization.

Keywords: bioprocess development, constraint-based modeling, metabolic engineering, SimPheny®

INTRODUCTION

It is estimated that greater than 90% of the chemicals used in industry are made from petroleum products. However, there are increasing pressures being placed on the use of petroleum as a feedstock due to factors associated with the environmental impact, cost stability, as well as geopolitical issues. A result of this pressure is a greater emphasis on the development of biotechnological approaches to produce chemicals from various carbohydrate feedstocks. In addition to reducing the dependency on fossil fuels, industrial biotechnology applications have the potential for value creation in the chemical industry through novel product opportunities, use of renewable resources, and significant reduction in the environmental footprint. A recent McKinsey and Company report estimated the possible value of these opportunities at $160 billion by 2010.

In many respects the future sustainability and growth of the chemical industry may rest on our ability to convert microbes into fine-tuned microscopic chemical factories. This ability relies on a convergence of chemistry, biology and engineering that is becoming more widespread throughout the chemical industry. As a result there is an increasing number of chemical products including commodity chemicals, fine chemicals, and pharmaceutical intermediates now being produced using microbial organisms as biocatalysts.

Further success of bioprocessing rests on the ability to intervene and rapidly engineer microbial metabolism in a wide range of suitable microbes whether they are specialized for the production of a single product or viewed as a platform organism to produce multiple products. Recent high throughput experimental technologies have greatly accelerated our ability to characterize and quantify the components of the metabolic machinery of microbes affordably in an era of increasing cost sensitivity. These technologies include whole genome sequencing, high throughput gene/protein expression profiling, as well as metabolomics and others. In addition there are new tools available to enhance the pace of engineering metabolism that include experimental tools such as directed evolution of enzymes, along with powerful computational approaches to design metabolic networks.

Over the past few years there has been a rapid increase in the use of metabolic modeling and simulation to aid in the design of these host organisms and in the development of accompanying bioprocesses that have improved

***Corresponding author**
Tel: +1-858- 824-1771 Fax: +1-858-824-1772
e-mail: cschilling@genomatica.com

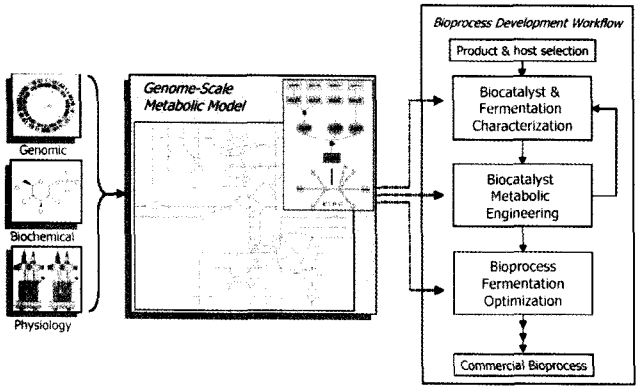

Fig. 1. Applying constraint-based modeling to bioprocess development. the bioprocess development workflow is impacted in many levels by the use of constraint-based modeling and genome-scale models of biocatalysts that are derived from genomic, biochemical, and physiological data.

economic value. In particular the constraint-based modeling (CBM) paradigm [1] is gaining widespread adoption as a preferred approach. This increased adoption is due to the ability of CBM to impact many of the vertical stages involved in bioprocess development from improving our understanding of the host to comprehensively optimizing the bioprocess itself.

Herein we focus on recent developments in CBM and how this modeling and simulation technology combines with experimental technologies to create a powerful integrated model-driven platform to drive bioprocess development (Fig. 1). Several applications of CBM will be discussed including the integrative analysis of 'omics' data, strain design, flux analysis, fermentation data reconciliation, and bioprocess optimization.

Constraint-based Modeling at a Glance

The CBM approach provides a biochemically and genetically consistent framework for the generation of hypotheses and the testing of functions of microbial cells. The approach is essentially an iterative process comprised of network reconstruction followed by the application of biological constraints that lead to the definition of achievable cellular functions. From here a variety of CBM algorithms and methods can be used to probe metabolism and ask a myriad of questions about the fitness and capabilities of the system. While this article focuses on application of CBM to industrial bioprocess development, a recent review article has covered the CBM approach in more detail with particular emphasis on the varying CBM algorithms and methods available [1].

CBM approach relies on implementing a series of physico-chemical and other biological constraints that are derived from phenomenological considerations. These governing constraints limit the range of possible functions and consequently the behavior of cellular networks. These constraints include:

1) Physico-chemical constraints (*e.g.* mass and energy conservation)

2) Topobiological constraints (*e.g.* molecular distribution, enzyme complexing)

3) Environmental constraints (*e.g.* nutrient availability, electron acceptors, pH)

4) Regulatory constraints (*e.g.* enzyme inhibition, gene expression)

After the recognition and definition of these biological constraints, they need to be described mathematically. Once in a mathematical form, they can be used to perform in silico analysis. There are two forms of mathematical constraints that include 1) balances, and 2) bounds. Balances are constraints that are associated with conserved quantities, such as energy, mass, redox potential and momentum, as well as with phenomena such as solvent capacity, electroneutrality and osmotic pressure. Bounds are capacity constraints that limit numerical ranges of individual ranges and parameters such as concentrations, fluxes, or kinetic parameters. Taken together, both bound and balance constraints define the allowable functional states of reconstructed networks. In mathematical terms, the range of allowable network states is described by a solution space that represents the phenotypic potential of an organism. All allowable network states are contained in this solution space.

In recent years, many new *in silico* methods have been developed using the CBM framework. This plethora of methods can be broadly classified into the following categories: finding best or optimal states in the allowable range; investigating flux dependencies; studying all allowable states; altering possible phenotypes as a consequence of genetic variations; and defining and imposing further constraints. All of these methods are applied to genome-scale models of microbial metabolism to help guide metabolism-related research.

Model Development for Industrially Relevant Organisms

Physiological characterization of industrially relevant microbes and gaining a strong understanding of their underlying metabolism is the first step on the path to efficient bioprocess development. Genome-scale models of metabolism have been demonstrated to accelerate metabolic understanding and physiological characterization through their ability to comprehensively predict, assess and interpret metabolic physiology at the genetic, protein, and reaction network levels. For these reasons and others genome-scale models have been developed for several industrially relevant organisms including *Escherichia coli, Saccharomyces cerevisiae, Bacillus subtilis, Pseudomonas fluorescens* and most recently for *Mannheima succiniproducens* [2-6]. Numerous other genome-scale models of other proprietary industrial microorganisms have also been recently developed. It is now becoming standard operating procedure to sequence the genomes of industrial microbes, which enables the rapid development of organism specific genome-scale models.

In order to have industrial efficacy today's models need to be developed with rigorous quality control and review

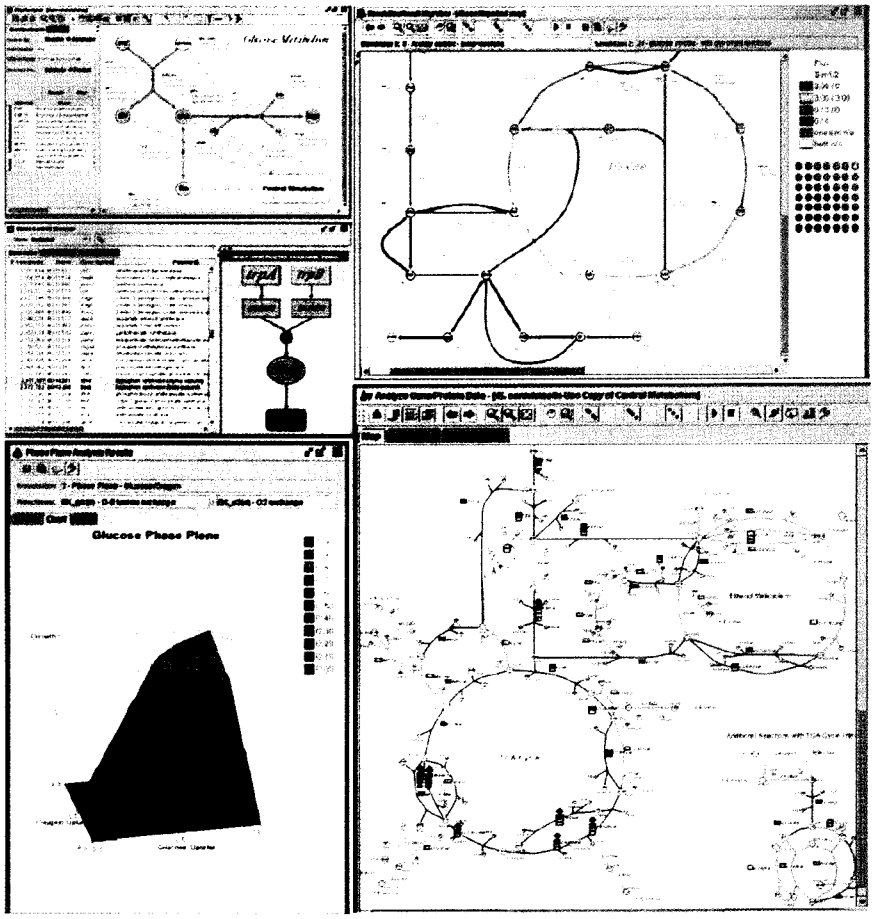

Fig. 2. SimPheny®. Screenshots from the SimPheny® modeling platform, an industrial grade software application developed by Genomatica for the creation and management of genome-scale models, integrative analysis of high-throughput experimental data, and the broad implementation of constraint-based modeling methods and algorithms.

stages, in which the content of both the model and the underlying information database is evaluated by appropriate bioinformatic and cheminformatic procedures. For example, one of the critical issues is the support of iterative model development. Any modeling paradigm must address the situation that gene annotation/ sequence and the functional roles may be continually updated as a result of new experimental discovery in both an intra- and inter-species context [7]. This situation requires careful tracking of the annotation changes and the maintenance of an annotation update history together with automated tools for change detection.

In addition, issues related to proton balancing have been shown to be crucial for the prediction of several physiological properties and by-product secretion rates [3]. Protons that are consumed and generated in metabolic reactions have to be accounted for to obtain an accurate description of the physiology. This need dictates that all of the reactions are elementally and charge balanced and that protons are included in the metabolic reactions accurately. Note that most publicly available metabolic databases do not provide reaction data with such levels of quality control. Sophisticated algorithms for pK_a predictions based on the structure of the metabolite are available (*e.g.*, ACD/pK_a DB, ©Advanced Chemistry Development, Inc., Toronto, Canada; Pipeline Pilot™, Scitegic Inc., San Diego, CA, USA). The ionization state of a metabolite at a physiological pH can be determined using such established algorithms and is important for situations where the cytosolic pH can be very different [8]. Hence, the availability of a high quality compound and reaction database along with an up-to-date gene annotation is the fundamental basis upon which rigorous constraint-based models can be developed and applied for characterizing physiology and for designing optimized bioprocesses. At Genomatica, we have developed these and other control measures for the development of industrial grade genome-scale metabolic models within the SimPheny® modeling platform (Fig. 2), which make today's models a significant step up from initial published models from years past [2,9,10].

In general genome-scale constraint-based models have been shown to be useful in predicting several physiological properties such as growth and by-product secretion

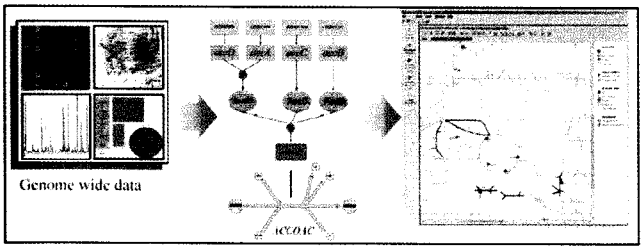

Fig. 3. Integrative data analysis. Genome-scale metabolic models are used to provide the necessary context for both the visualization and comprehensive analysis of high-throughput data sets.

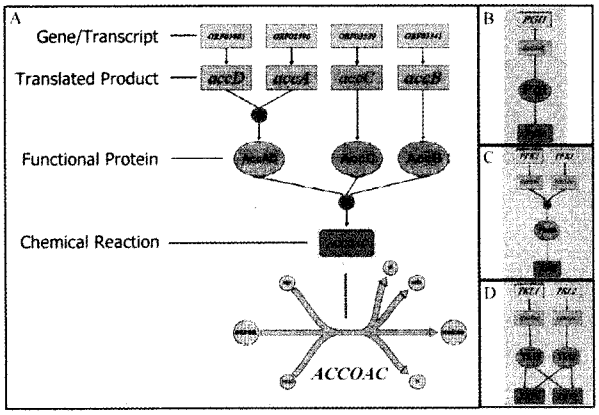

Fig. 4. Gene-protein-reaction associations. Graphical depiction of the logical associations between genes, proteins, and reactions in a genome-scale model as presented in the SimPheny® modeling platform. (A) Example of a multi-protein complex catalyzing acetyl-CoA carboxylase from *Geobacter sulfurreducens*, (B) single gene codes for single protein catalyzing single reaction (phosoglucoisomerase-*S. cerevisae*), (C) multi-subunit protein (phosphofructokinase-*S. cerevisiae*), and (D) isozymes both capable of catalyzing multiple reactions (transketolase-*S. cerevisiae*).

patterns [2,11-13], determining the range of substrate utilization [2], determining the minimal media for growth [10], predicting the outcome of adaptive evolution [14], calculating theoretical product yields [15], predicting knockout phenotypes [16-18] and comparing metabolic capabilities of different organisms [9]. In addition, genome-scale models have also been the basis for several topological studies for understanding network function [19-21]. Furthermore, these models have also been used for the rigorous quantitative analysis of phenotypic data gathered from a wide range of growth/feeding strategies. Based on these predictive capabilities, the models are now used to characterize the metabolic behavior of industrial microbes under laboratory and production scale fermentation conditions.

Integrative Analysis of High-throughput Data

Recent developments in high throughput experimentation have resulted in the ability to generate several genome-wide datasets including gene expression measurements, proteomics, and metabolite concentration measurements [22-24]. These datasets can be used to provide a richer understanding of metabolism. Although several methods for statistical analysis have been proposed to filter these data sets and extract correlations [25,26], there have been relatively few studies that integrate these data in the context of a cellular network [27]. Genome-scale constraint-based models provide a genetic and biochemically consistent representation of the cellular network in terms of biological entities such as genes, proteins, reactions and metabolites. Hence, these models provide a natural framework for the integrative analysis of high-throughput data in the context of the metabolic pathways and the predictions of the model (Fig. 3).

High throughput data quantifying the presence and activity of genes, proteins, metabolites, and reactions can be analyzed based on gene-protein-reaction associations that illustrate the link between genes in the genome, the proteins they encode, and the reactions catalyzed by these proteins. An example of these types of associations is shown for models generated with the SimPheny® platform (Fig. 4).

There are multiple levels of integrative analysis of experimental data that can be carried out using constraint-based models. The first level of analysis involves the visualization and mining of high throughput data in the context of the metabolic pathways and subsystems defined within the SimPheny® platform. The consistency of the data with the model content can be assessed using statistical tools such as Analysis of Variance (ANOVA) to determine conflicts in the datasets. An example of this includes situations where genes for multi-subunit proteins are being differentially expressed in opposite directions according to the data. Additionally, using strategies such as flux coupling analysis [28] the simulation predictions can be compared seamlessly with the large-scale datasets to identify differentially expressed pathways under the conditions examined. Even though there is no quantitative correlation between metabolic fluxes and transcript levels [29,30], qualitative comparison can be carried out to identify changes in the flux distribution that are consistent with predictions.

All of the above described analyses are now well established and have been used in various reports in recent years. Famili *et al.* [4], have compared the model predictions of changes in metabolic flux distribution in *S. cerevisiae* following a metabolic shift (from aerobic to anaerobic glucose limited growth), to gene expression changes under corresponding conditions and found qualitative agreement between the model predictions and the data confirming the predicted mechanism of adaptation. Akesson *et al.* [31] have recently utilized gene expression data to define constraints on simulations and report improvements in the predictive capability of the *S. cerevisiae* model. Patil *et al.* [32] analyze gene expression data based on the metabolic network, by identifying metabolites and subnetworks that are differentially expressed

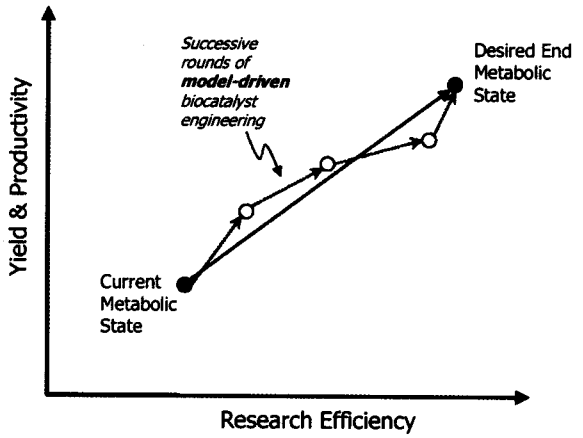

Fig. 5. Model-Driven Research. A systematic approach for enhancing productivity of bioprocesses involves characterizing the current metabolic state of a production strain and then iteratively improving its yield and productivity with corresponding improvements in research efficiency brought on through the use of genome-scale models.

following a perturbation. Interestingly, perturbations in a metabolic subsystem were found to trigger the transcriptional regulation of the genes in the same subsystem to counter the perturbation. Covert *et al.* [7,33] also describe the application of gene expression data to derive regulatory constraints that further improve the predictive ability of the model and form the basis for understanding the changes in *E. coli* metabolism following a shift from aerobic to anerobic conditions. Addtionally, such genome-scale structured representations of the metabolic network can be valuable in interpreting the systems level properties of the metabolic network such as epistatic interactions and evolutionary role of isozymes [19,20].

Reconciliation and Incorporation of Fermentation Data

The objective of most applied research programs in the bioprocess industry is to develop strains and/or processes that produce either an existing cellular metabolite or a novel chemical product at sufficient volumetric productivity to be economically competitive. Model-driven research in this context involves the use of high-quality metabolic models to guide iterative experimentation and design leading towards the improvement of an industrial bioprocess (Fig. 5). The ability to quantify the flux state of the metabolic system throughout each iteration of the biocatalyst engineering phase enables the overall bioprocess development to be carried out in an efficient and rational manner.

As mentioned above, the model can be used to predict the optimal product yield and production rate. These predictions represent the desired endpoint of the biocatalyst design process. In order to determine the appropriate interventions that must be made to engineer the microbe closer and closer to the optimal end goal, it is necessary to characterize the current state of the metabolic system and the fermentation process through experimentation and subsequent model guided interpretation.

Traditionally, fermentation processes are characterized and monitored by a set of macroscopic measurements taken at time points throughout the fermentation, which we henceforth refer to as fermentation data. These measurements can be both on- and off-line, and include feed rates, product concentrations, optical density (as a measure of biomass), residual substrate and byproduct concentrations, temperature, and pH. Often, fermentation data is collected but not utilized to its full potential. In conjunction with a metabolic model and the appropriate analysis platform, such data can be used to gain valuable insight toward both the fermentation process and the cell metabolism.

First, elemental balances evaluate data consistency and determine the likelihood of gross errors in the measurements [34]. Gross errors are inconsistencies in the material balance that cannot be attributed to data variability, and thus indicate a systematic error due to instrument malfunction or improper calibration. Gross errors also result when fermentation byproducts are formed but not accounted for. Statistical techniques also exist to identify the most likely sources of gross error and to reconcile the data so that it satisfies the overall elemental balance [34,35].

Second, fermentation data can be converted to flux units and used to constrain the extracellular fluxes in a CBM framework. In particular, these experimental constraints can be used in combination with flux variability analysis [36] to determine the feasible flux range of all the reactions in the metabolic network. Through *in silico* modeling, we can predict the distribution of metabolic fluxes in the network and identify bottlenecks to increased production [37,38]. The incorporation of constraints reduces the degrees of freedom of the network, thus improving the predictive power of the simulation and our confidence in the results [39]. Guided by the model predictions, research and development projects will lead to improved strains and processes [40]. Finally, fermentation data can be collected and analyzed at intermediate stages to monitor progress as we work toward the target productivity levels.

Flux Analysis Using Isotopomer Tracing Experiments

Fermentation data provides one level of constraints that can be combined with the stoichiometric modeling framework to reduce the size of the possible flux distributions. However, due to the existence of metabolic cycles and parallel pathways, even constraining all extracellular fluxes is often not enough to reduce the network to a fully determined system of equations. Thus we must still rely on linear optimization to predict flux distributions. In many research and biotechnology applications, more precise measurements of metabolic fluxes would be extremely beneficial. To overcome the shortcomings and achieve a single solution, we would need additional measurements to supplement the fermentation data.

Isotopic label tracing is a powerful experimental tech-

nique that can be combined with the CBM framework, providing the additional information required to more-accurately quantify metabolic fluxes in underdetermined systems. Since the most common type of labeling experiments involve ^{13}C-labeled compounds, the subsequent analysis is known as ^{13}C-MFA (for metabolic flux analysis) [41]. The calculation of intracellular fluxes by ^{13}C-MFA is based on the fact that when cells are fed a growth substrate with certain carbon positions labeled with ^{13}C, the distribution of this label in the intracellular metabolites can precisely be determined, based on the known biochemistry of the pathways. Furthermore, due to the high connectivity of metabolic networks, this distribution will depend strongly upon the values of the intracellular fluxes. By measuring the label distribution in the metabolites, these fluxes can therefore be calculated exactly [42]. The distribution of isotopomers, or compounds differing only in label distribution, can be measured using either gas chromatography-mass spectroscopy (GC-MS) or ^{13}C NMR. Performing a ^{13}C-labeling experiment is a complicated procedure that requires careful experimentation and specialized computational tools. A variety of experimental, analytical, and mathematical techniques for ^{13}C-MFA can be found in the literature [41,43-48].

Although isotopomer analysis methods were pioneered 20 years ago to determine how TCA cycle fluxes in *E. coli* differed widely with growth substrate [49], only recently have significant advances in both data acquisition and analysis methods led to use in the biotechnology industry. ^{13}C flux analysis has been used extensively in the study of lysine-producing *Corynebacterium glutamicum* [42,50,51], with particular emphasis on resolving the anapleurotic fluxes [52] or determining the fraction of glucose directed to the pentose phosphate pathway [53]. In one case, a futile cycle was identified as a potential target for directed metabolic engineering [52]. It has also been used with *E. coli* production strains [54]. In a phenylalanine-producing strain, measurement of fluxes in combination with model simulations led to the prediction that increasing expression of the phosphoenolpyruvate synthase enzyme will lead to improved yields [55]. In riboflavin-producing *Bacillus subtilis* [56], measurement of fluxes by ^{13}C-labeling indicated that production was limited by biosynthetic pathways, rather than the supply of precursors through central metabolism. Finally, this approach has been successfully applied to a diverse set of organisms such as *S. cerevisiae* [57], *Penicillium chrysogenum* [58], *Methylobacterium extorquens* AM1 [45], *Lactococcus lactis* [59]. In these cases and others, flux analysis has led to improved understanding of cell physiology in the wild-type, which is the first step toward rational design of improved strains.

The basis of flux analysis using ^{13}C labeled compounds is the availability of an accurate metabolic model. The rapid development of genome-scale models for industrially relevant organisms will enable broader application of flux analysis to bioprocess design and optimization. Fluxes calculated by this method can be incorporated directly into the CBM framework to serve as constraints. Even after performing ^{13}C-MFA, there are still fluxes that are unknown because they do not involve transfer of carbon atoms. Of particular interest may be energy metabolism, oxidative phosphorylation, and redox balance. The use of intracellular fluxes as constraints will lead to an improved quality of prediction of these fluxes, and enable the iterative analysis of modeling results with experimental measurements.

Strain Design

With the model-driven ability to characterize metabolic states of a biocatalyst continuously improving, we now shift focus to direct computational strategies that can guide the biocatalyst design and engineering. Systematic approaches for the rational design of production hosts are being developed and implemented using constraint-based models. Current high quality models of industrial microbes afford the ability to probe metabolism from a systems level in search of promising genetic manipulation candidates for improving biochemical production.

Early constraint-based approaches to strain design focused on pinpointing non-native reactions capable of increasing maximum theoretical yields or network flexibility. For example, mixed-integer programming was used to identify gene additions necessary for generating amino acid production pathways with increased theoretical yields in *E. coli* [60]. Another approach utilized pathway analysis to quantify the increase in maximum theoretical yield and additional flexibility afforded by a non-native transhydrogenase reaction for poly-β-hydroxybutyrate production in *S. cerevisiae* [61]. It is important to note, however, that such approaches are limited because maximum yielding flux routes and network flexibility quantifications rarely mimic flux distributions in vivo where internal biological objectives compete with product formation for key resources.

Central to the goal of applying CBMs in strain design are hypotheses regarding network responses to either genetic modifications or varying external conditions. Two prominent theories have been developed for describing system-wide metabolic responses either immediately following [17] or after an organism has been allotted the chance to adapt to a genetic or environmental perturbation [62]. The bi-level programming framework OptKnock [63,64], for example, was developed on the premise that mutant microbial networks can be evolved towards their computationally predicted maximum-growth phenotypes when subjected to extended periods of growth selection pressures. The approach focuses on exhaustively enumerating gene deletion strategies capable of forcing growth-coupled biochemical production following laboratory adaptive evolution. Knockouts are selected in such a way that the drain towards necessary growth resources (*i.e.*, biomass components, redox potential and energy) must be accompanied, due to stoichiometry, by the production of the desired chemical product. Thus notwithstanding improved production capabilities, it is probable that strains designed *via* the OptKnock approach are likely to possess unique stability advantages as a result of

the compulsory growth-coupled biochemical production.

An equally intriguing CBM approach to strain design builds upon the MOMA methodology [17] which assumes that metabolic networks will undergo a minimal redistribution of fluxes immediately following a genetic perturbation. Following from this premise, a sequential search approach was applied to a genome-scale constraint-based model [3] to identify promising gene knockout targets in *E. coli* for increasing lycopene production [65]. The method yielded numerous gene knockout candidates that were experimentally validated to increase lycopene production. The best approach was a triple knockout construct shown to increase lycopene accumulation 40% (6,600 ppm) over its already engineered, high producing parental counterpart (4,700 ppm). Further work enabled the construction of a strain capable of amassing 20,000 ppm of lycopene, bringing this bioprocess one step closer to industrial feasibility [66]. This work represents a major success story for the use of CBM in strain engineering projects particularly because it resulted in the prediction of non-intuitive mechanisms for enhancing the production of a valuable chemical product. Even more impressive is the fact that lycopene is a secondary metabolite uncoupled to the basal biomass requirements of the organism.

One final area where CBM can impact industrial strain design is in helping to manage and integrate the vast array of synthesis routes achievable to biological systems following enzyme engineering. Traditional model-based approaches to metabolic network redesign have focused only on biotransformation routes to and from chemical compounds already identified in biological systems. However, a novel approach, building upon previous efforts in the petroleum industry [67-69], is capable of suggesting reaction pathways to chemical compounds native and foreign to biological systems [70]. The approach centers upon the idea that enzymes can be engineered to carry out their natural biotransformations on non-natural substrates in turn dramatically expanding the complexity of metabolism beyond conventional metabolic pathway maps [71-73]. The method was recently used to suggest novel biotransformation routes to phenylalanine (~75,000), tyrosine (~350,000), tryptophan [70], and 7-carboxyindole from chorismate [74,75].

Constraint-based models provide the necessary framework to assess the potential system-wide effects, in terms of the tradeoffs between biochemical overproduction and cell growth, of incorporating the various biotransformation routes into a production host. Computational approaches such as OptStrain [76], which combines OptKnock with a selection procedure for finding the minimum number of non-native reactions needed for enabling maximum yielding pathways, may also be of particular importance here for sifting through the plethora of candidate pathways. In any event, the utility of constraint-based models in strain design will eventually reach the point where their development alongside with sequencing the genomes of industrially relevant organisms, as has been done with the succinate producer *Mannheimia succiniciproducens* [6], will become the norm.

Bioprocess Design and Optimization

Once a biocatalyst is well characterized, its metabolism broadly understood, and production capabilities have been engineered, the focus shifts in part to the design and optimization of the actual fermentation process itself. This represents another aspect of the overall bioprocess development that can be impacted by constraint-based models. Bioprocess models typically used for optimization and control are lumped representations of the metabolism in cells [77]. However, detailed representations of cellular metabolism are valuable for designing process parameters and for the control of bioprocesses. Detailed models of cellular metabolism that simulate the dynamics of fed-batch and batch cultures are already available [12,78,79]. There is also well established literature on process optimization and control using both structured and unstructured models of metabolism [80,81]. Such dynamic representations are important as many of the industrial fermentations are carried out in a fed-batch mode.

Metabolic models also have utility for the design of media conditions for maximizing product synthesis. Traditionally media conditions have been designed based on the biomass composition or using empirical methods relying on statistical analysis [82]. The organism-specific models can be used to identify the optimal substrate uptake rates through linear optimization techniques. Even though examples of application to industrial settings is not yet available, similar techniques have been used to identify minimal media composition for less characterized pathogens such as *Helicobacter pylori* [10]. Zhang *et al.* [83] also presented a study where metabolic flux models were used for the optimization of guanosine in *B. subtilis*, and erthyromycin by considering changes in the proton balance associated with metabolism.

CBMs can also impact the design of key process parameters such as the optimal feeding policy and the scheduling of batches for bioprocess applications. As discussed in the previous section, the production of metabolites even when tightly coupled to growth constitutes a loss of carbon that could have been diverted to growth. This aspect is evident in strains designed through these procedures, where the growth rate is decreased compared to the wild type growth. Thus, implementing these optimization-based designs can lead to strains with reduced growth rates. However, in a typical fed-batch bioprocess, the amount of product formed at the end of the batch is the economic objective. Thus, strain designs with reduced growth rate can have an impact on this objective. This trade-off between growth and product formation is considered for determining the optimal induction time in Gadkar *et al.* 2005 [84]. Here, an approach based on the dynamic Flux Balance Analysis [79] was used to identify the optimal profile for manipulating the flux through acetate kinase for ethanol production using *E. coli* in a fed-batch fermentation. Additionally, an extension of this approach was also used to determine the optimal schedule of batch bioprocess to maximize the overall ethanol yield (Gadkar *et al.*, 2004, AIChE Annual Meeting).

SUMMARY

It has been estimated that by 2010 almost 20% of the world's industrial chemicals will be derived from biotechnology. Success in achieving this goal rests on our ability to design and engineer the next generation of microbial biocatalysts to produce value added chemicals. Model-driven approaches that can drive microbially-based bioprocess development are poised to become an enabling technology. This is being achieved today by bringing together the power of high throughput experimentation and computational modeling and simulation in the form of tightly-coupled integrated research platforms that enable model-driven bioprocess development.

CBM is a particularly effective modeling and simulation approach to form the basis of such integrated computational/experimental platforms. In this article, we have reviewed various technical applications of CBM relevant for bioprocess development. The availability of predictive models for metabolic physiology of industrially relevant microbes can be valuable not only for understanding the physiology and characterizing a biocatalyst, but also for designing and manipulating bioprocesses for improved productivity.

Model-driven approaches are now providing a more rapid, cost effective and systematic approach for the design and optimization of bioprocesses. In fact there are several other applications of this technology that exist in both medical, industrial, and environmental biotechnology that are equally significant but are not reviewed here [85,86]. Synergistic development of modeling and simulation technology across all these industries will increase the likelihood that model-driven research approaches may soon revolutionize biotechnology industries just as modeling and simulation has impacted other industries such as the aerospace, automotive and semiconductor industries.

REFERENCES

[1] Price, N. D., J. L. Reed, and B. O. Palsson (2004) Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2: 886-897.

[2] Edwards, J. S. and B. O. Palsson (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* 97: 5528-5533.

[3] Reed, J. L., T. D. Vo, C. H. Schilling, and B. Palsson (2003) *Escherichia coli* iJR904: An expanded genome-scale model of *E. coli* K-12. *Genome Biol.* 4: R54.1-R54.12.

[4] Famili, I., J. Forster, J. Nielsen, and B. O. Palsson (2003) *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl. Acad. Sci. USA* 100: 13134-13139.

[5] Dauner, M. and U. Sauer (2001) Stoichiometric growth model for riboflavin-producing *Bacillus subtilis*. *Biotechnol. Bioeng.* 76: 132-143.

[6] Hong, S. H., J. S. Kim, S. Y. Lee, Y. H. In, S. S. Choi, J. K. Rih, C. H. Kim, H. Jeong, C. G. Hur, and J. J. Kim (2004) The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*. *Nat. Biotechnol.* 22: 1275-1281.

[7] Covert, M. W., E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429: 92-96.

[8] Ciaramella, M., A. Napoli, and M. Rossi (2005) Another extreme genome: How to live at pH 0. *Trends Microbiol.* 13: 49-51.

[9] Forster, J., I. Famili, P. Fu, B. O. Palsson, and J. Nielsen (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 13: 244-253.

[10] Schilling, C. H., M. W. Covert, I. Famili, G. M. Church, J. S. Edwards, and B. O. Palsson (2002) Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* 184: 4582-4593.

[11] Varma, A., B. W. Boesch, and B. O. Palsson (1993) Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol.* 59: 2465-2473.

[12] Varma, A. and B. O. Palsson (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* 60: 3724-3731.

[13] Edwards, J. S., R. U. Ibarra, and B. O. Palsson (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* 19: 125-130.

[14] Ibarra, R. U., J. S. Edwards, and B. O. Palsson (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* 420: 186-189.

[15] Varma, A., B. W. Boesch, and B. O. Palsson (1993) Biochemical production capabilities of *Escherichia coli*. *Biotechnol. Bioeng.* 42: 59-73.

[16] Edwards, J. S. and B. O. Palsson (2000) Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics* 1:1.

[17] Segre, D., D. Vitkup, and G. M. Church (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* 99: 15112-15117.

[18] Shlomi, T., O. Berkman, and E. Ruppin (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl. Acad. Sci. USA* 102: 7695-7700.

[19] Papp, B., C. Pal, and L. D. Hurst (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429: 661-664.

[20] Segre, D., A. Deluna, G. M. Church, and R. Kishony (2005) Modular epistasis in yeast metabolism. *Nat. Genet.* 37: 77-83.

[21] Mahadevan, R. and B. O. Palsson (2005) Properties of metabolic networks: Structure versus function. *Biophys. J.* 88: L07-L09.

[22] DeRisi, J. L., V. R. Iyer, and P. O. Brown (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680-686.

[23] Patterson, S. D. and R. H. Aebersold (2003) Proteomics: The first decade and beyond. *Nat. Genet.* 33 Suppl: 311-323.
[24] Kell, D. B (2004) Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol.* 7: 296-307.
[25] Churchill, G. A. (2004) Using ANOVA to analyze microarray data. *Biotechniques* 37: 173-177.
[26] Sharan, R., R. Elkon, and R. Shamir (2002) Cluster analysis and its applications to gene expression data. *Ernst. Schering. Res Found. Workshop* 83-108.
[27] Ideker, T., V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292: 929-934.
[28] Burgard, A. P., E. V. Nikolaev, C. H. Schilling, and C. D. Maranas (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* 14: 301-312.
[29] Oh, M. K. and J. C. Liao (2000) Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. *Biotechnol. Prog.* 16: 278-286.
[30] Tao, H., R. Gonzalez, A. Martinez, M. Rodriguez, L. O. Ingram, J. F. Preston, and K. T. Shanmugam (2001) Engineering a homo-ethanol pathway in *Escherichia coli*: Increased glycolytic flux and levels of expression of glycolytic genes during xylose fermentation. *J. Bacteriol.* 183: 2979-2988.
[31] Akesson, M., J. Forster, and J. Nielsen (2004) Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.* 6: 285-293.
[32] Patil, K. R. and J. Nielsen (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl. Acad. Sci. USA* 102: 2685-2689.
[33] Covert, M. W., C. H. Schilling, and B. Palsson (2001) Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* 213: 73-88.
[34] van der Heijden, R. T. J. M., J. J. Heijnen, C. Hellinga, B. Romein, and K. C. A. M. Luyben (1994) Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross measurement errors. *Biotechnol. Bioeng.* 43: 11-20.
[35] Raghunathan, A. U., J. R. Perez-Correa, and L. T. Biegler (2003) Data reconciliation and parameter estimation in flux-balance analysis. *Biotechnol. Bioeng.* 84: 700-708.
[36] Mahadevan, R. and C. H. Schilling (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 5: 264-276.
[37] Vallino, J. J. and G. Stephanopoulos (1993) Metabolic fluc distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. *Biotechnol. Bioeng.* 41: 633-646.
[38] van Gulik, W. M., W. T. de Laat, J. L. Vinke, and J. J. Heijnen (2000) Application of metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of penicillin-G. *Biotechnol. Bioeng.* 68: 602-618.
[39] Schilling, C. H., J. S. Edwards, and B. O. Palsson (1999) Toward metabolic phenomics: Analysis of genomic data using flux balances. *Biotechnol. Prog.* 15: 288-295.
[40] Shimizu, H., N. Takiguchi, H. Tanaka, and S. Shioya (1999) A maximum production strategy of lysine based on a simplified model derived from a metabolic reaction network. *Metab. Eng.* 1: 299-308.
[41] Wiechert, W (2001) ^{13}C metabolic flux analysis. *Metab. Eng.* 195-206.
[42] Marx, A., A. A. de Graaf, W. Wiechert, L. Eggeling, and H. Sahm (1996) Determination of the fluxes in central metabolism of *Corynebacterium glutamicum* by NMR spectroscopy combined with metabolite balancing. *Biotechnol. Bioeng.* 49: 111-129.
[43] Dauner, M. and U. Sauer (2000) GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing. *Biotechnol. Prog.* 16: 642-649.
[44] Schmidt, K., M. Carlsen, J. Nielsen, and J. Villadsen (1997) Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol. Bioeng.* 55: 831-840.
[45] van Dien, S. J., T. Strovas, and M. E. Lidstrom (2003) Quantification of central metabolic fluxes in the facultative methylotroph methylobacterium extorquens AM1 using ^{13}C-label tracing and mass spectrometry. *Biotechnol. Bioeng.* 84: 45-55.
[46] Wiechert, W., C. Siefke, A. A. de Graaf, and A. Marx (1997) Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. *Biotechnol. Bioeng.* 55: 118-135.
[47] Wiechert, W. and A. A. de Graaf (1997) Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnol. Bioeng.* 55: 101-117.
[48] Wittmann, C. and E. Heinzle (1999) Mass spectrometry for metabolic flux analysis. *Biotechnol. Bioeng.* 62: 739-750.
[49] Walsh, K. and D. E. Jr. Koshland (1984) Determination of flux through the branch point of two metabolic cycles. The tricarboxylic acid cycle and the glyoxylate shunt. *J. Biol. Chem.* 259: 9646-9654.
[50] Park, S. M., M. I. Klapa, A. J. Sinskey, and G. N. Stephanopoulos (1999) Metabolite and isotopomer balancing in the analysis of metabolic cycles: II. Applications. *Biotechnol. Bioeng.* 62: 392-401.
[51] Wendisch, V. F., A. A. de Graaf, H. Sahm, and B. J. Eikmanns (2000) Quantitative determination of metabolic fluxes during coutilization of two carbon sources: Comparative analyses with *Corynebacterium glutamicum* during growth on acetate and/or glucose. *J. Bacteriol.* 182: 3088-3096.
[52] Petersen, S., A. A. de Graaf, L. Eggeling, M. Mollney, W. Wiechert, and H. Sahm (2000) *In vivo* quantification of parallel and bidirectional fluxes in the anaplerosis of *Corynebacterium glutamicum*. *Metab. Eng.* 3: 195-206.
[53] Wittmann, C., H. M. Kim, and E. Heinzle (2004) Metabolic network analysis of lysine producing *Corynebacterium glutamicum* at a miniaturized scale. *Biotechnol. Bioeng.* 87: 1-6.
[54] Sauer, U., D. R. Lasko, J. Fiaux, M. Hochuli, R. Glaser, T. Szyperski, K. Wuthrich, and J. E. Bailey (1999) Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *J. Bacteriol.* 181: 6679-6688.

[55] Wahl, A., M. El Massaoudi, D. Schipper, W. Wiechert, and R. Takors (2004) Serial ^{13}C-based flux analysis of an L-phenylalanine-producing *E. coli* strain using a sensor reactor. *Biotechnol. Prog.* 20: 706-714.
[56] Sauer, U., V. Hatzimanikatis, J. E. Bailey, M. Hochuli, T. Szyperski, and K. Wuthrich (1997) Metabolic fluxes in riboflavin-producing *Bacillus subtilis*. *Nat. Biotechnol.* 15: 448-452.
[57] Gombert, A. K., S. M. Moreira dos, B. Christensen, and J. Nielsen (2001) Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *J. Bacteriol.* 183: 1441-1451.
[58] Christensen, B. and J. Nielsen (2000) Metabolic network analysis of *Penicillium chrysogenum* using ^{13}C-labeled glucose. *Biotechnol. Bioeng.* 68: 652-659.
[59] Jensen, N. B. S., B. Christensen, J. Nielsen, and J. Villadsen (2002) The simultaneous biosynthesis and uptake of amino acids by *Lactococcus lactis* studied by ^{13}C-labeling experiments. *Biotechnol. Bioeng.* 78: 11-16.
[60] Burgard, A. P. and C. D. Maranas (2001) Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. *Biotechnol. Bioeng.* 74: 364-375.
[61] Carlson, R., D. Fell, and F. Srienc (2002) Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnol. Bioeng.* 79: 121-34.
[62] Fong, S. S. and B. O. Palsson (2004) Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat. Genet.* 36: 1056-1058.
[63] Burgard, A. P., P. Pharkya, and C. D. Maranas (2003) OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84: 647-657.
[64] Pharkya, P., A. P. Burgard, and C. D. Maranas (2003) Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol. Bioeng.* 84: 887-899.
[65] Alper, H., Y. S. Jin, J. F. Moxley, and G. Stephanopoulos (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab. Eng.* 7: 155-164.
[66] Wilson, E. K. (2005) Engineering cell-based factories. *Chem. Eng. News* 83: 41-44.
[67] Broadbelt, L. J., S. M. Stark, and M. T. Klein (1994) Computer-generated pyrolysis modeling-on-the-fly generation of species, reactions, and rates. *Ind. Eng. Chem. Res.* 33: 790-799.
[68] Broadbelt, L. J., S. M. Stark, and M. T. Klein (1995) Termination of computer-generated reaction-mechanisms-species rank-based convergence criterion. *Ind. Eng. Chem. Res.* 34: 2566-2573.
[69] Broadbelt, L. J., S. M. Stark, and M. T. Klein (1996) Computer generated reaction modelling: Decomposition and encoding algorithms for determining species uniqueness. *Comput. Chem. Eng.* 20: 113-129.
[70] Hatzimanikatis, V., C. Li, J. A. Ionita, and L. J. Broadbelt (2004) Metabolic networks: Enzyme function and metabolite structure. *Curr. Opin. Struct. Biol.* 14: 300-306.
[71] Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32 Database issue: D277-D280.
[72] Karp, P. D., M. Riley, M. Saier, I. T. Paulsen, S. M. Paley, and A. Pellegrini-Toole (2000) The EcoCyc and MetaCyc databases. *Nucleic Acids Res.* 28: 56-59.
[73] Krieger, C. J., P. Zhang, L. A. Mueller, A. Wang, S. Paley, M. Arnaud, J. Pick, S. Y. Rhee, and P. D. Karp (2004) MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 32 Database issue: D438-D442.
[74] Li, C., C. S. Henry, M. D. Jankowski, J. A. Ionita, V. Hatzimanikatis, and L. J. Broadbelt (2004) Computational discovery of biochemical routes to specialty chemicals. *Chem. Eng. Sci.* 59: 5051-5060.
[75] Hatzimanikatis, V., C. Li, J. A. Ionita, C. S. Henry, M. D. Jankowski, and L. J. Broadbelt (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics* 21: 1603-1609.
[76] Pharkya, P., A. P. Burgard, and C. D. Maranas (2004) OptStrain: A computational framework for redesign of microbial production systems. *Genome Res.* 14: 2367-2376.
[77] Komives, C. and R. S. Parker (2003) Bioreactor state estimation and control. *Curr. Opin. Biotechnol.* 14: 468-474.
[78] Covert, M. W. and B. O. Palsson (2002) Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* 277: 28058-28064.
[79] Mahadevan, R., J. S. Edwards, and F. J. Doyle (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical J.* 83: 1331-1340.
[80] Gadkar, K. G., F. J. Doyle, III, T. J. Crowley, and J. D. Varner (2003) Cybernetic model predictive control of a continuous bioreactor with cell recycle. *Biotechnol Prog.* 19: 1487-1497.
[81] Mahadevan, R. and F. J. Doyle (2003) On-line optimization of recombinant product in a fed-batch bioreactor. *Biotechnol. Prog.* 19: 639-646.
[82] Parekh, S., V. A. Vinci, and R. J. Strobel (2000) Improvement of microbial strains and fermentation processes. *Appl. Microbiol. Biotechnol.* 54: 287-301.
[83] Zhang, S., J. Chu, and Y. Zhuang (2004) A multi-scale study of industrial fermentation processes and their optimization. *Adv. Biochem. Eng. Biotechnol.* 87: 97-150.
[84] Gadkar, K. G., F. J. Doyle, J. S. Edwards, and R. Mahadevan (2005) Estimating optimal profiles of genetic alterations using constraint-based models. *Biotechnol. Bioeng.* 89: 243-251.
[85] Lovley, D. R. (2003) Cleaning up with genomics: Applying molecular biology to bioremediation. *Nat. Rev. Microbiol.* 1: 35-44.
[86] Beard, D. A. and H. Qian (2005) Thermodynamic-based computational profiling of cellular regulatory control in hepatocyte metabolism. *Am. J. Physiol. Endocrinol. Metab.* 288: E633-E644.

[Received April 8, 2005; accepted July 23, 2005]