Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism

Irina Borodina,[1] Preben Krabben,[2] and Jens Nielsen[1,3]

[1]*Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, DK-2800 Lyngby, Denmark;*
[2]*Department of Biochemical Engineering, University College London, WC1E7JE London, United Kingdom*

Streptomyces are filamentous soil bacteria that produce more than half of the known microbial antibiotics. We present the first genome-scale metabolic model of a representative of this group—*Streptomyces coelicolor* A3(2). The metabolism reconstruction was based on annotated genes, physiological and biochemical information. The stoichiometric model includes 819 biochemical conversions and 152 transport reactions, accounting for a total of 971 reactions. Of the reactions in the network, 700 are unique, while the rest are iso-reactions. The network comprises 500 metabolites. A total of 711 open reading frames (ORFs) were included in the model, which corresponds to 13% of the ORFs with assigned function in the *S. coelicolor* A3(2) genome. In a comparative analysis with the *Streptomyces avermitilis* genome, we showed that the metabolic genes are highly conserved between these species and therefore the model is suitable for use with other Streptomycetes. Flux balance analysis was applied for studies of the reconstructed metabolic network and to assess its metabolic capabilities for growth and polyketides production. The model predictions of wild-type and mutants' growth on different carbon and nitrogen sources agreed with the experimental data in most cases. We estimated the impact of each reaction knockout on the growth of the in silico strain on 62 carbon sources and two nitrogen sources, thereby identifying the "core" of the essential reactions. We also illustrated how reconstruction of a metabolic network at the genome level can be used to fill gaps in genome annotation.

[Supplemental material is available online at www.genome.org.]

Streptomycetes are soil-inhabiting Gram-positive bacteria belonging to the order of *Actinomycetales* (Garrity 2002). The genus is characterized by a high G+C content, large linear chromosomes (>8 Mb), and a complex life cycle, which involves morphological differentiation. Streptomycetes have attracted much attention because of their ability to make secondary metabolites that have a wide range of bioactivities and thereby may find use as antibiotics, immunosuppressants, anti-cancer agents, and the like. It is estimated that *Streptomyces* spp. produce more than 50% of the total of 11,900 known microbial antibiotics (Kieser et al. 2000). *Streptomyces* species also secrete a wide range of extracellular hydrolytic enzymes, which enable them to degrade lignocellulosic materials in soil, and many of these enzymes also have commercial interest.

Streptomyces coelicolor A3(2) is by far the best genetically studied *Streptomyces* strain and has become a model organism for *Streptomyces* species (Hopwood 1999). Release of the *S. coelicolor* A3(2) genome sequence (Bentley et al. 2002) has further expanded the knowledge of this organism and enabled large-scale analyses of transcriptome (Huang et al. 2001; Bucca et al. 2003) and proteome (Hesketh et al. 2002; Novotna et al. 2003). The rapidly growing amount of genome-wide data represents a valuable database for gaining more fundamental insight into the molecular mechanisms governing different processes in *S. coelicolor* A3(2), but particularly integration of these data through the use of mathematical models would enable extraction of more information from this database. In this context it has recently been shown particularly valuable to use genome-scale models for the metabolism (Ideker et al. 2001; Stelling et al. 2002; Covert et al. 2004; Kell 2004).

At the present, metabolic networks have been reconstructed for several bacteria (e.g., *Escherichia coli*) (Edwards and Palsson 2000b; Reed et al. 2003) and for the yeast *Saccharomyces cerevisiae* (Förster et al. 2003a; Duarte et al. 2004). The models have been successfully used to predict organism phenotype for a modified genotype (Edwards and Palsson 2000a; Förster et al. 2003b; Covert et al. 2004), and to predict cellular behavior under different physiological conditions (Edwards et al. 2001; Famili et al. 2003). Furthermore, these models may form the basis for functional characterization (Papp et al. 2004), for extraction of information of coordinated transcriptional responses that cannot be identified through classical statistical methods (Ihmels et al. 2004), and in design of improved strains for production of specific metabolites (Burgard et al. 2003).

Here, the reconstructed metabolic network of *S. coelicolor* A3(2) is applied for detailed study of the organism's metabolism.

Results and Discussion

Reconstruction and characteristics of the metabolic network

Reconstruction of high-quality and well-annotated metabolic networks is laborious as information needs to be acquired from many different sources. The reconstruction can be facilitated by software, for example, PathoLogic (Karp et al. 2002), that assigns reactions to the annotated genes according to E.C. number or enzyme name. We used the *S. coelicolor* A3(2) pathway database in KEGG (ftp://ftp.genome.ad.jp/pub/kegg/pathways/sco/) for extraction of the draft model. Automatically created models need to be manually curated using books, literature, and other avail-

[3]Corresponding author.
E-mail jn@biocentrum.dtu.dk; fax 45 4588 4148.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.3364705.

15:820–829 ©2005 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/05; www.genome.org

able information sources. The most problematic aspects are wrongly or insufficiently defined substrate specificity, reaction reversibility, protein complexes, cofactor specificity, and the missing enzymes. Examples of these problems and the ways of dealing with them are listed in Table 1.

Modeling of *S. coelicolor* A3(2) is a challenging task because of the complexity of this filamentous bacteria. Its genome size of 8,667,507 bp is unusually large for bacteria sequenced to date. Out of 7825 predicted open reading frames (ORFs), a function has been assigned to 5492, which also includes broad definitions such as "putative membrane protein," "putative transmembrane transport protein," and similar ones (S.D. Bentley, pers. comm). It is estimated that 965 proteins have a regulatory function, 819 proteins are secreted (these include numerous hydrolytic enzymes), 614 proteins participate in transport, and 22 gene clusters are involved in secondary metabolites production (Bentley et al. 2002). Only 926 ORFs have an enzyme commission (E.C.) number assigned, and out of these, 711 (77%) were included in the model (Table 2; Supplemental Data Set 1). We considered the pathways, which were known to be operative in *Streptomyces* and/or for which most of the pathway-specific enzymes were present. Genome-scale modeling is a dynamic process, and we expect that the reaction list will be updated as new information is released. The total number of metabolic reactions included in the model is 971, of which 819 are metabolic conversions and the rest are transport reactions (Supplemental Data Set 2). Of the metabolic conversion reactions, 89% have an ORF assigned; the other reactions were included based on physiological evidence. Only 24% of the transport reactions have an assigned ORF, and this low percentage is due to the limited information on transport processes in this organism. The model includes biosynthesis of the polyketide antibiotic actinorhodin (ACT) and a nonribosomal peptide calcium-dependent antibiotic (CDA). The biosynthesis route of undecylprodigiosin (RED) is not yet completely elucidated (Thomas et al. 2002); therefore, only the known steps for its biosynthesis are presented in the model. With the included reactions the total number of metabolites in the reconstructed network is 500 (Supplemental Data Set 3).

Table 1. Typical problems arising from automatic metabolic networks reconstruction and examples of dealing with them

Problem	Example
1) Substrate specificity is insufficiently or wrongly defined.	The ORF SCO4384 encodes an enoyl-CoA hydratase with assigned E.C. number 4.2.1.17, and the enzyme acts on a wide range of 3-hydroxyacyl-CoA compounds. Judging from the neighbor ORF SCO4383, which encodes a 4-coumarate:coenzyme A ligase and is positioned in the same regulon, SCO4384 more likely catalyzes hydration of activated 4-coumarate and hence is involved in the secondary metabolism.
2) Reversibility is not defined.	Pyruvate carboxylase (E.C. 6.4.1.1) is shown in KEGG as acting in both directions even though it is known to be irreversible because of thermodynamic constraints prevailing in most cells, that is, the carboxylation of pyruvate to oxaloacetate is associated with the cleavage of a high-energy phosphate bond in ATP resulting in the formation of AMP and pyrophosphate, whereas the decarboxylation of oxaloacetate to pyruvate does not result in the formation of ATP.
3) Enzyme subunits are shown as catalyzing a reaction independently although they are active only in a complex.	The ORFs SCO0216–0219 form a four-subunit complex with a nitrate reductase activity, while in KEGG each of these ORFs is assigned the given function, basically indicating that each of these four ORFs encodes an isoenzyme.
4) Cofactor requirements are often specific for the given organism and have to be found elsewhere.	Valine dehydrogenase E.C. 1.4.1.8 is given as $NADP^+$-dependent in KEGG; however, it was experimentally determined to use NAD^+ as the preferred hydrogen acceptor (Navarrete et al. 1990).
5) Several reactions that are necessary for making a functional cell have not been assigned a corresponding ORF.	Histidinol phosphatase E.C. 3.1.3.15 is an enzyme in the linear histidine biosynthesis pathway. It was included in the model even though the gene was not found in the genome of *S. coelicolor* A3(2).

Universality of the *S. coelicolor* A3(2) model

We estimated the possibility of extrapolating the use of the *S. coelicolor* A3(2) metabolic network for other *Streptomyces* strains by comparing the genomes of *S. coelicolor* A3(2) and *S. avermitilis* (Supplemental Data Set 4). In total, 53% of the genes in *S. coelicolor* A3(2) are synteny conserved with *Streptomyces avermitilis*, that is, their relative position in the genome has been retained throughout evolution of these two *Streptomyces* strains; 35% of the *S. coelicolor* A3(2) genes have orthologs whose position is not conserved in *S. avermitilis*; only 12% of the genes do not have an ortholog in *S. avermitilis*. Out of 711 ORFs included in the model, 78% are synteny conserved with *S. avermitilis*, which shows that there is a much higher degree of synteny for metabolic genes than for the other genes. A further 22% of the ORFs in the metabolic model are coding for isoenzymes of which at least one synteny-conserved ORF is present in *S. avermitilis*. The high fraction of synteny-conserved ORFs and isoenzymes in the model indicates that the *S. coelicolor* A3(2) metabolic network can be used as a starting point for a rapid reconstruction of the reaction networks for other *Streptomyces* species.

Connectivity

A frequency plot for all the metabolites in the metabolic networks of *S. coelicolor* A3(2), *S. cerevisiae* (Förster et al. 2003a), and *E. coli* (Edwards and Palsson 2000b) is shown in Figure 1. With the exception of proton, which is the most frequent metabolite in yeast because of proton-driven transport reactions, the most connected metabolites are involved in energy metabolism: ATP, ADP, and phosphate. The reducing equivalent NADH is more common for the metabolic network of *S. coelicolor* A3(2), which also has the highest proportion of oxidoreductases. The higher occurrence of glutamate and α-ketoglutarate in *S. coelicolor* A3(2) reflects a higher frequency of the use of aminotransferases. Coenzyme A and acetyl-carrier protein occur more often in the *S. coelicolor* A3(2) model, which directly correlates with extensive propanoate, butanoate metabolism, and the production of polyketide actinorhodin.

Growth energetics

Besides differences in some reactions and the preferred cofactors for some reactions, the core part of microbial metabolism, that is,

Table 2. Comparison of the genome characteristics and the in silico metabolic networks of prokaryotic and eukaryotic species

	Prokaryotes				Eukaryotes
	S. coelicolor A3(2)	*E. coli*[a]	*H. influenzae*[b]	*H. pylori*[c]	*S. cerevisiae*[d]
Genome characteristics					
Genome length (bp)	8,667,507	4,639,221	1,830,135	1,667,867	12,175,026
G+C content	72%	51%	38%	39%	40%
ORF	7825	4288	1743	1590	6281
Assigned function	5492	2656	1011	1091	4379
No assigned function	2333	1632	732	499	1902
In silico metabolic network characteristics					
Total reactions (unique)	971 (700)	931 (931)	461	381	1175 (842)
Biochemical conversions	819	767			826
Transport	152	164			349
Reactions with ORF (% of total reactions)	764 (79%)	873 (94%)	412 (89%)	272 (71%)	1035 (88%)
Metabolites	500	625	367	332	584
ORF (% of ORFs with assigned function)	711 (13%)	904 (34%)	400 (40%)	290 (27%)	708 (16%)

[a]*Escherichia coli* K-12 (Reed et al. 2003).
[b]*Haemophilus influenzae* Rd (Schilling and Palsson 2000).
[c]*Helicobacter pylori* 26695 (Schilling et al. 2002).
[d]*Saccharomyces cerevisiae* (Förster et al. 2003a).

the central carbon metabolism (glycolysis, pentose phosphate pathway, and TCA cycle), biosynthesis of the biomass building blocks, is highly conserved. The biomass composition and the growth energetics are, however, unique properties of each organism. Here, the biomass composition was calculated based on the measurements performed in *S. coelicolor* A3(2) and related species (Supplemental Data Set 5). The growth energetics are described by Stouthamer equation 1:

$$r_{\mathrm{ATP}} = Y_{x\mathrm{ATP}} \cdot \mu + m_{\mathrm{ATP}}, \quad (1)$$

which states that for a pseudo-steady state, the rate of ATP production r_{ATP} is equal to the ATP consumption for growth $Y_{x\mathrm{ATP}} \cdot \mu$ and maintenance m_{ATP}, where μ is the specific growth rate and $Y_{x\mathrm{ATP}}$ is the ATP yield coefficient (Stouthamer and Bettenhaussen 1973). The main source of ATP in aerobically growing organisms is oxidative phosphorylation. *Streptomyces* are known to use menaquinone as an electron transporter in the membrane (Pandya and King 1966; Yassin et al. 1988), even though surprisingly enough the genes for menaquinone synthesis were not found in the sequenced genomes of *S. coelicolor* A3(2) and *S. avermitilis*. The stoichiometry of the electron transport chain was assumed to be the same as recently described in the actinomycete *Corynebacterium glutamicum* (Bott and Niebisch 2003). The values of the ATP yield coefficient $Y_{x\mathrm{ATP}}$ and the maintenance flux m_{ATP} were estimated from the chemostate cultivations data (Melzoch et al. 1997) as described in Methods (Figs. 2, 3).

Reactions activity

In order to investigate the topology of the metabolic network, we created a list of unique reactions, and in this way eliminated the redundancy in the network resulting from the large number of isoenzymes (Supplemental Data Set 7). This model was used for growth simulation in five different media: (1) glucose as carbon source (C-source) and nitrate as nitrogen source (N-source), (2) mixture of carbohydrates, (3) mixture of organic acids and alcohols, (4) mixture of amino acids, and (5) mixture of nucleotides (Supplemental Data Set 8). Except for the first scenario, unlimited ammonia uptake was allowed. Additionally, we simulated secondary metabolites production at slow growth rate on glucose and ammonia. The analysis of the fluxes that computationally provided optimal growth rate or optimal antibiotics production rate at the given conditions (also considering alternative optimal solutions) showed that 88% of reactions were active under some or all of the tested conditions, 6% of reactions were inactive because of the formation of dead ends (that means that one of the reaction's products or reactants cannot be used or produced in the network) (Fig. 4). In this way there was only a very small part (6%) of the non-dead-end reactions that never got involved in the cellular metabolism, perhaps because the conditions that evoke these reactions' activity have not been considered in the simulations.

Interestingly, some of the enzymes that catalyze "dead-end" reactions in the model have been detected on 2D gels in *S. coelicolor* A3(2) grown on a minimal medium supplemented with casamino acids (http://dbkweb.ch.umist.ac.uk/StreptoBASE/s_coeli/referencegel/). One of the enzymes was putative man-

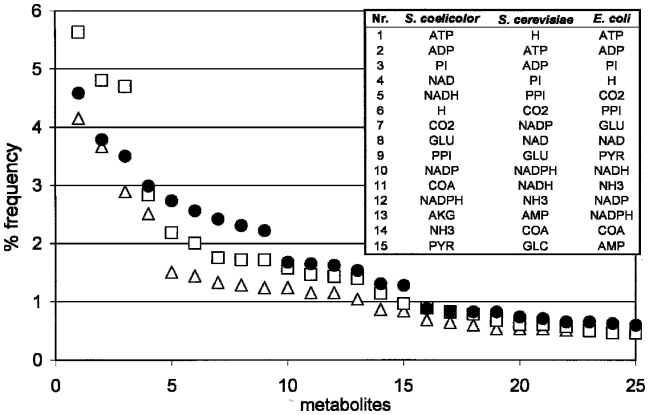

Figure 1. Frequency of the most connected metabolites in the reconstructed metabolic networks of *S. coelicolor* A3(2), *S. cerevisiae* (Förster et al. 2003a), and *E. coli* (Edwards and Palsson 2000b). Frequency is calculated as the number of times a certain metabolite appears in the metabolic network divided by the sum of all metabolite occurrences. *S. coelicolor* A3(2) (●), *S. cerevisiae* (△), *E. coli* (□). (H) External proton; (GLU) glutamate; (PPI) pyrophosphate; (COA) coenzyme A; (AKG) α-ketoglutarate; (PYR) pyruvate; (ACCOA) acetyl-coenzyme A; (ACP) acetyl-carrier protein.

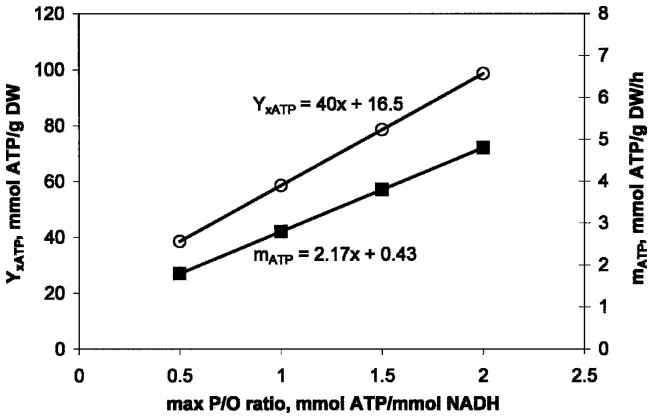

Figure 2. The ATP yield coefficient Y_{xATP} (○) and maintenance energy m_{ATP} (■) as a function of the maximal P/O ratio in *S. coelicolor* A3(2).

nose-1-phosphate guanyltransferase (SCO1388), which is known to be involved in mannosylation of proteins, a common protein modification for Actinomycetes. The reaction appears as a dead end in the model because protein modifications are not included in the reactions list, and therefore the GDP-activated mannose-1-phosphate is not used in any other reactions. Another enzyme is 2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase (SCO0852), which is either involved in Entner-Doudoroff (ED) pathway [this is unlikely as the characteristic ED pathway gene *edd* was not found in the genome sequence of *Streptomyces coelicolor* A3(2)] or is responsible for interconverting 4-hydroxy-α-ketoglutarate into glyoxylate and pyruvate. Experimental evidence has been obtained that the enzyme is important in regulation of glyoxylate levels in the cells of *E. coli* (Cayrol et al. 1995). A hypothesis has been proposed that 4-hydroxy-2-oxoglutarate aldolase action is coupled to α-ketoglutarate dehydrogenase complex and results in pyruvate-catalyzed oxidation of glyoxylate into malyl-CoA (Gupta and Dekker 1984). We judge that it is important to include the reactions in the model even though the complete pathway is not known as the presence of dead ends may give a hint to which enzymes one should look for in future annotation of the genome.

Reactions dispensability

We studied dispensability of the reactions in the network by making single reaction deletions and optimizing for growth on 62 different C-sources and two different inorganic N-sources (ammonia and nitrate, only tested with glucose as C-source). The approach differs from single gene deletion studies, because if isoenzymes are present, then their simultaneous knockout will be simulated by deleting the reaction they catalyze. We find this approach more informative for studying the sensitivity of the metabolic network to perturbations. To quantify the effect of reaction deletion, we defined reaction essentiality as the relative decrease in the specific growth rate with deletion of the reaction in comparison to the specific growth rate with the complete reaction set (2):

$$essentiality_{reaction\ i} = 1 - \frac{\mu_{max}(mutant_{\Delta reaction\ i})}{\mu_{max}(wild\ type)} = [0....1], \quad (2)$$

that is, the reaction essentiality is 1 for an essential reaction, whereas it is 0 for a reaction that upon removal has no growth-retarding effect. The reaction essentialities for the growth on glucose and ammonia are shown in Figure 5A, and Figure 5B illustrates the summed reaction essentialities for growth on the 63 different media (Supplemental Data Set 9). Reactions with a summed reaction essentiality of 63 are required for growth under all the defined conditions, and hence are true essential reactions. The reactions that have a summed reaction essentiality lower than 63 are either necessary only under certain conditions or their deletion leads to growth retardation. The comparison of Figure 5A and Figure 5B shows that the choice of conditions is important for defining the dispensability of reactions. While during growth on glucose and ammonia 64% of the reactions could be eliminated without any consequences for cellular growth, several of these reactions turned out to be essential for growth on other C-sources than glucose, and the number of nonessential deletions was reduced from 64% to 39% when all the growth conditions were considered. The minimal metabolic net, necessary for growth under all the conditions, consists of 146 reactions. Considering that 17% of them have an isoenzyme, there are basically 121 metabolic genes that are truly essential (e.g., essential on the complete medium), which corresponds to 12% of the original metabolic network.

Essential and nonessential reactions were found to have almost equal occurrence of isoenzymes (17% and 15%, respectively), whereas a higher fraction of the growth-retarding or conditionally essential reactions were catalyzed by isoenzymes (26%). This undermines a hypothesis that isoenzymes exist to increase the robustness of the metabolic network to mutations.

Metabolic capabilities of the network

Degradation of carbon and nitrogen sources

In most cases the model correctly predicted growth capability on various C-sources and N-sources for *S. coelicolor* A3(2) wild type and mutant (Table 3). There was a disagreement on usage of aspartate and glutamine as the sole C-source, which was possible according to the model, but has not been observed experimentally. It is known than *S. coelicolor* A3(2) can grow on asparagine, which is presumably degraded via aspartate. In this case, it may be regulatory events that do not allow aspartate utilization rather than the lack of metabolic capacity. *S. coelicolor* A3(2) can use

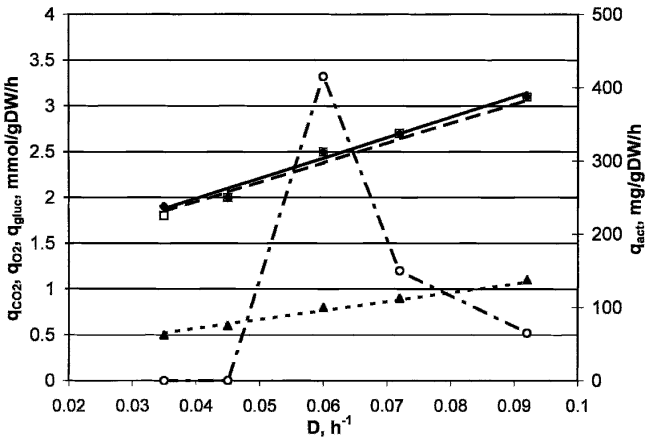

Figure 3. Simulation of experimental chemostates data (Melzoch et al. 1997). Specific glucose uptake rate (experimental ▲, model ----), specific carbon dioxide production rate (experimental ♦, model ——), specific oxygen uptake rate (experimental □, model – – –), and actinorhodin production rate (experimental ○, model — - — - — -).

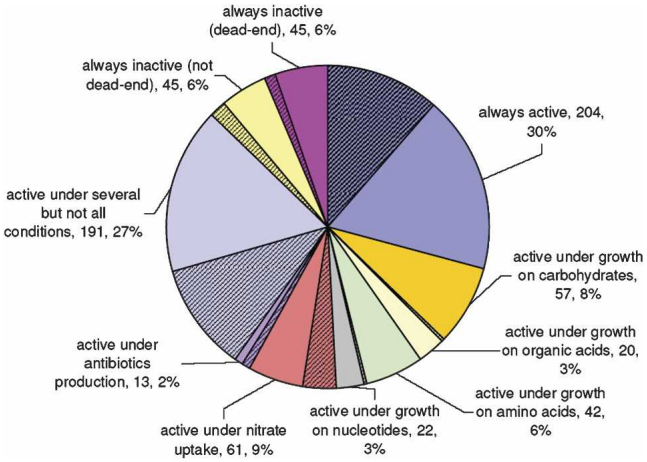

Figure 4. Activity of the reactions in the model of *S. coelicolor* A3(2). All the unique reactions in the model are divided into several categories based on their predicted activity under different growth conditions. The proportion of reactions for which the corresponding enzymes were detected on a 2D gel is shown by the dashed areas.

glutamine as the sole N-source, but not as C-source, while glutamate can be used as both. There is evidence that glutamate is decarboxylated into γ-aminobutanoate upon the uptake (Inbar and Lapidot 1991). Therefore, if glutamine is deaminated by glutaminase intracellularly, the formed glutamate cannot be further degraded. In this way glutamine would only provide nitrogen for growth, but not carbon units.

The deletions of *trpC1* and *trpD1* in *S. coelicolor* A3(2) result in auxotrophy for tryptophan, because the genes with analogical functions—*trpC2* and *trpD2*—are located in the calcium-dependent antibiotic biosynthetic cluster and apparently can be used exclusively in the secondary metabolism (Hu et al. 1999).

The described inconsistencies can be resolved in the future by expanding the model to include regulatory constraints.

Biomass yield

The biomass yields of in silico *S. coelicolor* A3(2) on various C-sources were similar to that of in silico yeast (Fig. 6A; Förster et al. 2003a). The energy requirements were considered for the simulations. The exception was growth on C-2 compounds like acetate and ethanol, where *S. coelicolor* A3(2) was more efficient at using these substrates. This is due to the action of the acetate kinase (encoded by *ackA*) and the phosphate acetyltransferase (encoded by *pta*), which catalyze the conversion of acetate into acetyl-phosphate and further into acetyl-CoA at the expense of only one high-energy phosphate bond hydrolysis in ATP. In eukaryotes these enzymes are not present and acetate is activated by acetyl-CoA synthase (encoded by *acs*) with the cocurrent hydrolysis of ATP to AMP, which is energetically more costly. Comparison of the growth capabilities of in silico *S. coelicolor* and in silico *E. coli* (Edwards and Palsson 2000b) showed that the *S. coelicolor* had higher biomass yields, which was primarily because of higher maintenance energy requirements in *E. coli* (Fig. 6B).

Anaerobic growth

The inability of *Streptomyces* to grow anaerobically has been a puzzle for quite some time (Hodgson 2000). Generally the microbes are capable of growing anaerobically, as they may obtain their Gibbs free energy by substrate-level phosphorylation resulting in fermentative metabolism. The presence of lactate dehydrogenase in the sequenced genome of *S. coelicolor* A3(2) indicates that the organism should be capable of having a fully operational fermentative metabolism with lactate as a key fermentation product. Indeed, excretion of lactate has been observed for *Streptomyces griseus* grown under microaerophilic conditions (Hockenhull et al. 1954). Two reasons were suggested to explain the anaerobic *Streptomyces* paradox: either the organisms are sensitive to the fermentation products or/and there are some essential reactions that require oxygen.

We analyzed the problem from the perspective of the reconstructed metabolic network and did not find any essential reactions that exclusively use oxygen. It seems that the essential dehydroorotate dehydrogenase E.C. 1.3.3.1 (SCO1482) required for pyrimidines biosynthesis and L-aspartate oxidase E.C. 1.4.3.16 (SCO3382) participating in nicotinamide nucleotides biosynthesis can both use oxygen and menaquinone as electron acceptors. Anaerobic growth can simply require that the produced menaquinol can be reoxidized by the reverse action of succinate dehydrogenase. When glucose uptake rate was set to the maximal experimentally observed value (2.2 mmol/g DW/h) (Melzoch et al. 1997) and oxygen uptake rate was fixed to zero, the maximum specific growth rate was predicted to be 0.022 h^{-1} and 0.056 h^{-1} with ammonia and nitrate as the N-source, respectively. The

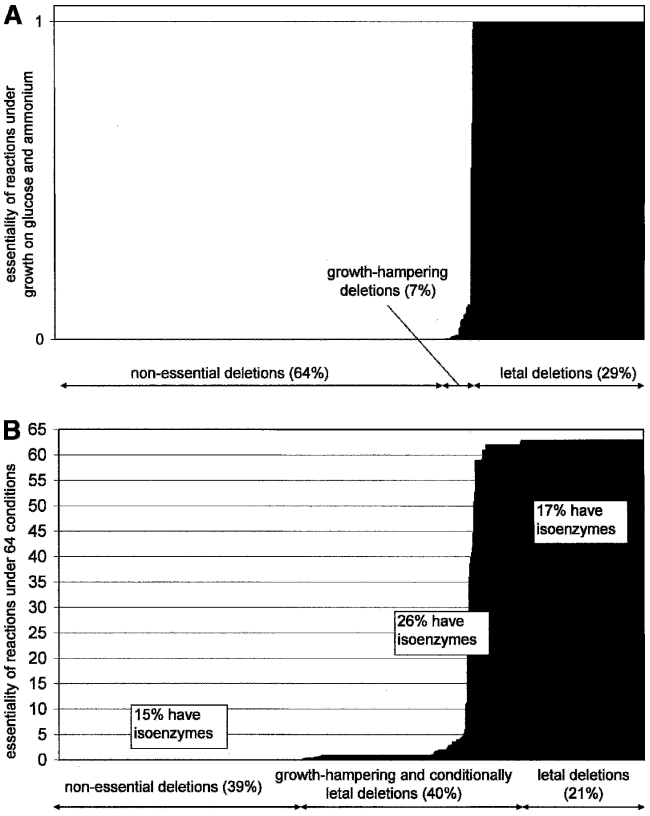

Figure 5. Essentiality of the reactions in the model of *S. coelicolor* A3(2) during growth on glucose (*A*) and during growth under 63 various conditions (*B*). All the reactions were sorted into three categories: essential reactions, conditionally essential reactions, and nonessential reactions depending on their influence on biomass synthesis. The percentage of reactions with isoenzymes among all the reactions in the given category is shown.

Table 3. Comparison of experimentally determined and in silico predicted growth phenotypes

In vivo	In silico	Strain	Substrate/phenotype	Reference
C-source				
+	+	wt	Glycerol, L-arabinose, D-arabitol, D-ribose, D-xylose, L-xylitol, D-fructose, D-galactose, D-gluconate, D-glucose, D-mannitol, D-mannose, L-rhamnose, salicin, cellobiose, lactose, maltose, melibiose, trehalose, acetate, citrate, lactate, malate, pyruvate, succinate, tartrate, propanoate, alanine, asparagine, glutamate, glycine, histidine, isoleucine, leucine, lysine, methionine, proline, serine, threonine, tryptophan, valine	Hodgson 1980; Zhang et al. 1996
		Δ*msdA*	propanoate	Zhang et al. 1996
		Δ*glnA*	no glutamine auxotrophy	Fink et al. 1999
−	−	wt	Sucrose, arginine, methionine, phenylalanine	Hodgson 1980
		Δ*msdA*	Valine	Zhang et al. 1996
		Δ*hisA*	trp^- his^-	Barona-Gomez and Hodgson 2003
		Δ*malE*	Maltose	van Wezel et al. 1997
		Δ*argG*	arg^-	Flett et al. 1987
		Δ*metF*	met^-	Blanco et al. 1998[a]
−	+	wt	Aspartate, glutamine	Hodgson 1980
		Δ*trpC1*Δ*trpD1*	trp^-	Hu et al. 1999
N-source				
+	+	wt	Aspartate, asparagine, glutamate, glutamine, isoleucine, leucine, lysine, praline, valine	Hodgson 1980
−	−	Δ*vdh*	Valine, isoleucine, leucine	Tang and Hutchinson 1993
+	−	wt	Methionine, phenylalanine	Hodgson 1980

(+) Growth; (−) no growth.
[a]The data are for *Streptomyces lividans.*

growth, however, implied conservation of a proton potential over the membrane, that is, the protons exported during secretion of succinate are used for uptake of glucose. This is an unlikely situation in vivo, and then substrate uptake might be the limiting factor for growth without oxygen. Glucose is transported into the *S. coelicolor* A3(2) by proton symport and the proton gradient is created primarily in the respiratory process. Meanwhile, facultative anaerobes like yeast and *E. coli* use facilitated diffusion and the PTS transport system, respectively, which remain active under anaerobic conditions.

Actinorhodin production

Several recent publications are dedicated to expressing bacterial and fungal polyketide synthase (PKS) genes in heterologous hosts. (Kealey et al. 1998; Kennedy et al. 2003; Kinoshita et al. 2003). As an example, we addressed the issue of metabolic capabilities of yeast and *S. coelicolor* A3(2) to produce reducing cofactors and the most common polyketides precursors (Fig. 7). The capabilities of the metabolic networks of yeast and *S. coelicolor* A3(2) to produce acetyl-CoA, malonyl-CoA, and reducing equivalents NADH and NADPH turned out to be very similar. However, *Streptomyces* have a clear advantage in the form of more extensive polyols metabolism, which allows them to make such typical building blocks for polyketides production as methylmalonyl-CoA, propionyl-CoA, and butanoyl-CoA. Development of heterologous production of certain polyketides from *E. coli* or yeast is tedious as it requires not only addition of PKS genes, but also the genes necessary for biosynthesis of precursors.

Filling the gaps in genome annotation

We assumed that the following processes had to be included in order to produce a "viable" model cell:

- transport reactions (nutrients and oxygen uptake, metabolites excretion);
- biosynthesis of the 12 biomass precursors;
- biosynthesis of each biomass component (amino acids, vitamins, cofactors, etc.);
- degradation of the experimentally utilizable substrates.

If any of these conditions were not fulfilled, a missing reaction(s) was added according to the pathway structure (organism specific or general if the former was not available). A total of 205 reactions without assigned ORFs were added to the model: 79 enzymatic reactions, 117 transport reactions, five spontaneous, and four artificial reactions like maintenance and biomass assembly (Supplemental Data Set 10). Out of the 79 enzymatic reactions, 27 were identified as being essential for growth on glucose and salts. In the context of functional genomics, the reconstructed metabolic network hence serves as physiological evidence for the presence of these genes and allows directed search for the ORFs with the necessary function. Each of these 27 reactions was subsequently analyzed on an individual basis. The genome was searched for genes with specific protein motifs (http://www.sanger.ac.uk/Software/Pfam/), and the obtained hits were evaluated using the functions of the neighboring genes and their possible organization in an operon. The previously described Bayesian method (Green and Karp 2004) has essentially the same logics, but it does not account for the protein motifs and uses homology search instead; besides, it requires model implementation in Pathway Tools (Karp et al. 2002).

Phospholipids biosynthesis

One of the enzymes missing in the phospholipids biosynthesis is cardiolipin synthase. No homologs of bacterial cardiopin syn-

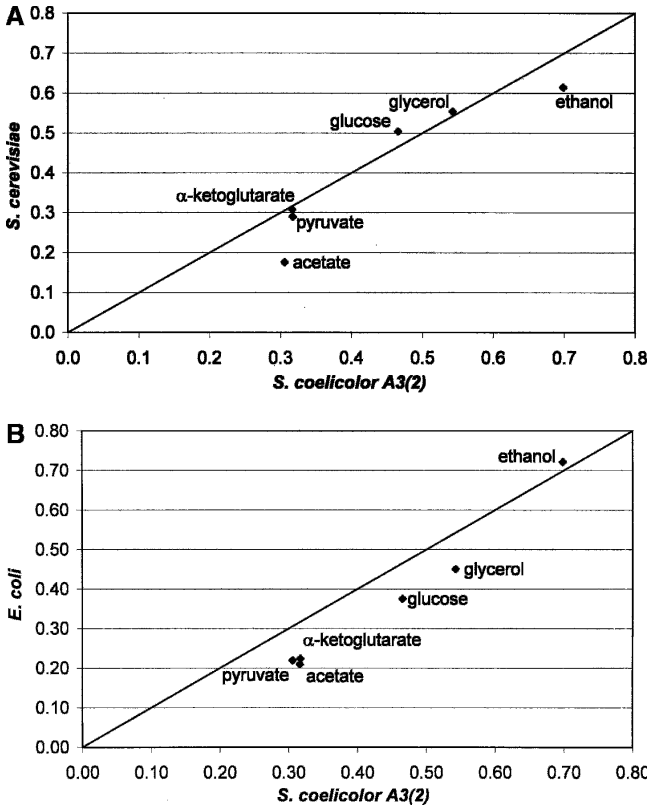

Figure 6. Maximal simulated biomass yield of *S. coelicolor* A3(2) as compared to (*A*) *S. cerevisiae* (Förster et al. 2003a) and (*B*) *E. coli* (Edwards and Palsson 2000b) on different carbon sources (gram/gram substrate). All the calculations were made for substrate uptake rate of 6 C-mmol/g DW/h. The maintenance energy requirement was considered.

thase, which catalyzes the condensation of two phosphatidylglycerols into cardiolipin and glycerol, were found in the genome of *S. coelicolor* A3(2). Presumably the cardiolipin synthesis occurs in the same way as in eukaryotes in the reaction of phosphatidylglycerol with CDP-glycerol accompanied by formation of CMP. In this case, the enzyme should possess a CDP-alcohol phosphatidyltransferase domain. With a Pfam motif search eight ORFs were found to contain such a domain in the *S. coelicolor* A3(2) genome. Three of these ORFs (SCO1527, SCO1389, and SCO5753) were synteny conserved with *Mycobacterium tuberculosis* enzymes coding for phosphatidylinositol synthase (*pgsA*), cardiolipin synthase (*pgsA2*), and phosphatidylglycerol synthase (*pgsA3*), respectively (Jackson et al. 2000). Once a link has been established between gene and enzymatic function, it is possible to transfer the knowledge between different species by homology and synteny information, and we hereby predict that SCO5753 encodes phosphatididylglycerol synthase, SCO1527—a phosphatidylinositol synthase and SCO1389—a cardiolipin synthase.

Polyprenoids biosynthesis

The annotation of the polyprenoid biosynthesis was re-examined because five essential genes were found to be missing in the KEGG database, that is, (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase (IspH/LytB), pentaprenyl diphosphate synthase, hexaprenyl diphosphate synthase, octaprenyl diphosphate synthase, and nonaprenyl diphosphate synthase. The enzymes involved in the polyprenoid biosynthesis all have a prenyltransferase domain IPR001441. *S. coelicolor* A3(2) contains seven genes with a polyprenyltransferase domain, that is, SCO0185, SCO0565, SCO0568, SCO2509, SCO3858, SCO4583, SCO5250, and SCO6763. Two of the genes, SCO2509 and SCO3858, are synteny conserved with *M. tuberculosis* and must therefore have an important function. Furthermore SCO0185, SCO4583, SCO5250, and SCO6763 are synteny conserved with *S. avermitilis*. These six conserved genes are likely to be involved in general polyprenoid biosynthesis. The SCO0185 (*crtB*) is involved in biosynthesis of a secondary carotenoid metabolite (Lee et al. 2001), which is probably connected to the response to UV-radiation. This indicates that SCO0185 might not be true synteny conserved between *S. coelicolor* A3(2) and *S. avermitilis* but is a result of a transfer of a gene cluster. The SCO6763 is placed in another secondary metabolite gene cluster responsible for hopanoid biosynthesis, which is thought to be involved in the stress response in aerial mycelia by decreasing the water permeability of the plasma membrane (Poralla et al. 2000). The last four unaccounted genes must be the genes that are responsible for polyprenoid biosynthesis for general purposes, that is, plasma membrane electron carrier and transmembrane sugar carrier. The homolog of SCO3858 (Rv1086) has been biochemically characterized in *M. tuberculosis* as an ω,E,Z-farnesyl diphosphate synthase (Schulbach et al. 2000).

Conclusions

We have reconstructed the metabolic network of *S. coelicolor* A3(2). Besides resulting in improved annotation of several genes and suggestions for annotation of other genes, the reconstructed network may be used as a model of the metabolism in *Streptomyces*. Earlier attempts to model *Streptomyces* spp. (Avignone et al. 2002; Bruheim et al. 2002; Roubos 2002) show that there is a demand for a high-quality metabolic model of *Streptomyces* both in academic and industrial research. The model presented in this paper has been reconstructed based on the up-to-date knowledge about *Streptomyces* and can readily (or with small adjustments for species specificity) be used for simulations. We therefore expect that the model will be a useful tool for genome-wide analysis of the transcriptome and the proteome as well as for designing strategies for improvement of antibiotic yields in *Streptomyces*.

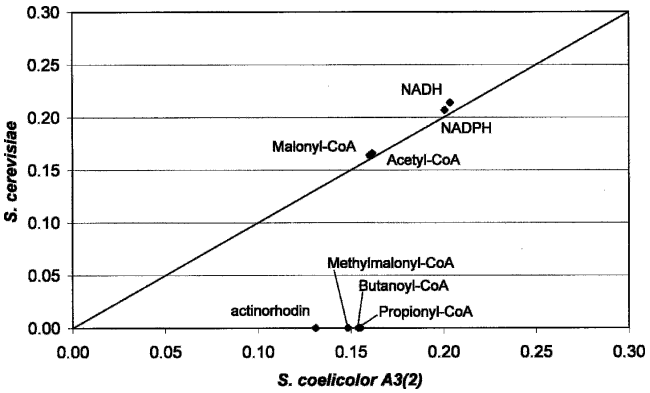

Figure 7. Predicted polyketides production capabilities of *S. coelicolor* A3(2) and *S. cerevisiae* (Förster et al. 2003a) (in moles per mole glucose). All the calculations were made for the glucose uptake rate of 1 mmol/g DW/h.

Methods

Reconstruction process

For reconstruction of the *S. coelicolor* A3(2) metabolic network, we used the annotated genome databases (KEGG PATHWAY database: http://www.genome.ad.jp/dbget-bin/get_htext?S.coelicolor.kegg+-f+T+w+C and The Wellcome Trust Sanger Institute database: http://www.sanger.ac.uk/Projects/S_coelicolor/scheme.shtml), metabolic databases (KEGG Ligand database: http://www.genome.ad.jp/kegg/ligand.html; ExPASy Biochemical Pathways: http://www.expasy.org/cgi-bin/search-biochem-index; ExPASy Enzyme Database: http://www.expasy.org/enzyme; SWISS-PROT database: http://www.expasy.org/sprot/sprot-top.html), biochemistry books (Ingraham et al. 1983; Stanier et al. 1986; Michal 1999), reviews (Hodgson 2000), and journal publications. The reconstruction process was started by downloading the *S. coelicolor* pathway database from the KEGG ftp server (ftp://ftp.genome.ad.jp/pub/kegg/pathways/sco/), and the reactions corresponding to the given E.C. numbers/enzyme names were written using the metabolic databases indicated above. Water, intracellular protons, and hydroxyl ions were not included in the model, assuming that different nonenzymatic processes also use these substrates and therefore don't need to be balanced in the set of enzymatic reactions. The functions of the included genes were compared to the functional assignments in The Wellcome Trust Sanger Institute database and when possible confirmed by information from the literature. In case the reaction was necessary to produce a viable in silico cell but the corresponding gene was not present in the annotated genome, the reaction was included in the model without genomic evidence. Unless the information of irreversibility of a reaction was available, it was set as reversible. A list of the reactions with corresponding references is available as Supplemental material (Supplemental Data Set 2). The biomass equation was made based on the macromolecular composition of *S. coelicolor* A3(2) (Shahab et al. 1996) and related species and is also available as Supplemental material (Supplemental Data Set 5).

Modeling

Flux balance analysis (FBA) was used for quantification of metabolic fluxes (Varma and Palsson 1994; Schilling et al. 1999). The reactions set composed a stoichiometric matrix S with dimensions $m \times n$, where m was the number of metabolites and n was the number of reactions. Assuming pseudo-steady state over intracellular metabolite concentrations, the metabolic model could be defined by the following matrix equation:

$$S \cdot v = 0, \tag{3}$$

where v represented all the metabolic fluxes.

As the number of metabolic fluxes exceeded the number of mass balance constraints, there existed a set of feasible metabolic fluxes distributions. A solution was found using linear programming by introducing an optimization problem: MAXIMIZE $Z = c \cdot v$, where c was a row vector showing the influence of individual fluxes on the objective function. Additional constraints were imposed on fluxes: $a_i \le v_i \le b_i$, indicating the reaction irreversibility ($0 \le v_i < \infty$) or measured flux. The transport fluxes for phosphate, sulfate, ammonia, and oxygen were not restrained ($-\infty \le v_i < \infty$). The uptake rates of metabolites that were not available in the medium were set to zero ($v_i = 0$). The excretion of the major metabolic products (carbon dioxide, pyruvate, acetate, etc.) was allowed ($0 \le v_i < \infty$).

For simulations of growth, the flux to biomass was set to a certain value and the substrate uptake rate was minimized. For simulation of antibiotic production, the substrate uptake rate was constrained and the flux toward the antibiotic was maximized.

In order to find fluxes that were active during alternative optimal solutions, the following algorithm was used:

1. Substrate uptake rates were constrained and the biomass growth rate was maximized.
2. The growth rate was fixed to the found maximal value, and thus an additional constraint was introduced.
3. By changing vector c, each reaction in the network was maximized and then minimized. If the flux through the reaction was different from zero in at least one of these optimizations, the reaction was considered to be active in alternative optimal solutions.

The calculations were performed using the commercially available linear programming package LINDO (Lindo Systems Inc.). The algorithm for finding fluxes that are active during alternative optimal solutions was implemented in Matlab (The MathWorks Inc.), but the LINDO package was used as the solver for the linear programming problems.

Estimation of energetic parameters

Knowing one of the energetic parameters (P/O ratio, $Y_{x\mathrm{ATP}}$ or m_{ATP}) allows calculation of the other two parameters from continuous cultivation data. Because none of the parameters was known, they were estimated for P/O ratios in the range of 0.5 to 2 (Fig. 2).

The $Y_{x\mathrm{ATP}}$ is composed of three parts:

- ATP costs for the synthesis of biomass building blocks (amino acids, nucleotides, etc.), which were accounted for through reactions stoichiometry;
- ATP costs for polymerization of monomers into biological polymers (proteins, DNA, etc.), which were assumed to be the same as for *E. coli* (Ingraham et al. 1983) and were included in the reactions of macromolecules biosynthesis; and
- ATP costs for growth-associated maintenance $Y_{x\mathrm{ATP\_growth\_maintenance}}$, added to the growth equation.

The last composite $Y_{x\mathrm{ATP\_growth\_maintenance}}$ as well as maintenance ATP (m_{ATP}) were not known and were estimated from the experimental chemostate data (Melzoch et al. 1997).

Normally the glucose uptake rate q_{gluc}, carbon dioxide production rate $q_{\mathrm{CO_2}}$ and oxygen uptake rate $q_{\mathrm{O_2}}$ are linearly dependent on the specific growth rate (Fig. 3). The $Y_{x\mathrm{ATP\_growth\_maintenance}}$ and m_{ATP} were set up to arbitrary values, and the simulations were run for each of the experimentally investigated dilution rates by fixing the specific growth rate and actinorhodin production rate to the experimental values and performing linear optimization for glucose uptake rate minimization (Supplemental Data Set 6). The obtained q_{gluc}, $q_{\mathrm{CO_2}}$ and $q_{\mathrm{O_2}}$ dependence on dilution rate was compared to the experimental rate. The $Y_{x\mathrm{ATP\_growth\_maintenance}}$ and m_{ATP} were changed until a good prediction was obtained. For the further simulations, the maximal P/O ratio was fixed to 1.5 and $Y_{x\mathrm{ATP}}$ and m_{ATP} were set to the corresponding values.

Acknowledgments

We thank D.A. Hodgson for providing his Ph.D. thesis and for the excellent review on the primary metabolism of *Streptomyces*. We are grateful to Jochen Förster for sharing his experience in genome-scale modeling.

References

Avignone, R.C., White, J., Kuiper, A., Postma, P.W., Bibb, M., and Teixeira de Mattos, M.J. 2002. Carbon flux distribution in antibiotic-producing chemostat cultures of *Streptomyces lividans*. *Metab. Eng.* **4:** 138–150.

Barona-Gomez, F. and Hodgson, D.A. 2003. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep.* **4:** 296–300.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., et al. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417:** 141–147.

Blanco, J., Coque, J.J., and Martin, J.F. 1998. The folate branch of the methionine biosynthesis pathway in *Streptomyces lividans*: Disruption of the 5,10-methylenetetrahydrofolate reductase gene leads to methionine auxotrophy. *J. Bacteriol.* **180:** 1586–1591.

Bott, M. and Niebisch, A. 2003. The respiratory chain of *Corynebacterium glutamicum*. *J. Biotechnol.* **104:** 129–153.

Bruheim, P., Butler, M., and Ellingsen, T.E. 2002. A theoretical analysis of the biosynthesis of actinorhodin in a hyper-producing *Streptomyces lividans* strain cultivated on various carbon sources. *Appl. Microbiol. Biotechnol.* **58:** 735–742.

Bucca, G., Brassington, A.M., Hotchkiss, G., Mersinias, V., and Smith, C.P. 2003. Negative feedback regulation of *dnaK, clpB* and *lon* expression by the DnaK chaperone machine in *Streptomyces coelicolor*, identified by transcriptome and in vivo DnaK-depletion analysis. *Mol. Microbiol.* **50:** 153–166.

Burgard, A.P., Pharkya, P., and Maranas, C.D. 2003. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84:** 647–657.

Cayrol, C., Petit, C., Raynaud, B., Capdevielle, J., Guillemot, J.C., and Defais, M. 1995. Recovery of respiration following the SOS response of *Escherichia coli* requires RecA-mediated induction of 2-keto-4-hydroxyglutarate aldolase. *Proc. Natl. Acad. Sci.* **92:** 11806–11809.

Covert, M.W., Knight, E.M., Reed, J.L., Herrgard, M.J., and Palsson, B.Ø. 2004. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429:** 92–96.

Duarte, N.C., Herrgard, M.J., and Palsson, B.Ø. 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **14:** 1298–1309.

Edwards, J.S. and Palsson, B.Ø. 2000a. Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics* **1:** 1.

———. 2000b. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci.* **97:** 5528–5533.

Edwards, J.S., Ibarra, R.U., and Palsson, B.Ø. 2001. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* **19:** 125–130.

Famili, I., Förster, J., Nielsen, J., and Palsson, B.Ø. 2003. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl. Acad. Sci.* **100:** 13134–13139.

Fink, D., Falke, D., Wohlleben, W., and Engels, A. 1999. Nitrogen metabolism in *Streptomyces coelicolor* A3(2): Modification of glutamine synthetase I by an adenylyltransferase. *Microbiology* **145:** 2313–2322.

Flett, F., Platt, J., and Cullum, J. 1987. DNA rearrangements associated with instability of an arginine gene in *Streptomyces coelicolor* A3(2). *J. Basic Microbiol.* **27:** 3–10.

Förster, J., Famili, I., Fu, P., Palsson, B.Ø., and Nielsen, J. 2003a. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13:** 244–253.

Förster, J., Famili, I., Palsson, B.Ø., and Nielsen, J. 2003b. Large-scale evaluation of in silico gene deletions in *Saccharomyces cerevisiae*. *OMICS* **7:** 193–202.

Garrity, G.M. 2002. *Bergey's manual of systematic bacteriology*. Springer-Verlag, New York.

Green, M.L. and Karp, P.D. 2004. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinformatics* **5:** 76.

Gupta, S.C. and Dekker, E.E. 1984. Malyl-CoA formation in the NAD-, CoASH-, and α-ketoglutarate dehydrogenase-dependent oxidation of 2-keto-4-hydroxyglutarate. Possible coupled role of this reaction with 2-keto-4-hydroxyglutarate aldolase activity in a pyruvate-catalyzed cyclic oxidation of glyoxylate. *J. Biol. Chem.* **259:** 10012–10019.

Hesketh, A.R., Chandra, G., Shaw, A.D., Rowland, J.J., Kell, D.B., Bibb, M.J., and Chater, K.F. 2002. Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor*. *Mol. Microbiol.* **46:** 917–932.

Hockenhull, D.J., Fantes, K.H., Herbert, M., and Whitehead, B. 1954. Glucose utilization by *Streptomyces griseus*. *J. Gen. Microbiol.* **10:** 353–370.

Hodgson, D.A. 1980. "Carbohydrate utilization in *Streptomyces coelicolor* A3(2)." Ph.D thesis, University of East Anglia, Norwich, UK.

———. 2000. Primary metabolism and its control in *Streptomycetes*. *Adv. Microbial Physiol.* **42:** 47–238.

Hopwood, D.A. 1999. Forty years of genetics with *Streptomyces*: From in vivo through in vitro to in silico. *Microbiology* **145:** 2183–2202.

Hu, D.S., Hood, D.W., Heidstra, R., and Hodgson, D.A. 1999. The expression of the *trpD*, *trpC* and *trpBA* genes of *Streptomyces coelicolor* A3(2) is regulated by growth rate and growth phase but not by feedback repression. *Mol. Microbiol.* **32:** 869–880.

Huang, G., Lih, C.-J., Pan, K.-H., and Cohen, S.N. 2001. Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. *Genes & Dev.* **15:** 3183–3192.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., and Hood, L. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292:** 929–934.

Ihmels, J., Bergmann, S., and Barkai, N. 2004. Defining transcription modules using large-scale gene expression data. *Bioinformatics* **20:** 1993–2003.

Inbar, L. and Lapidot, A. 1991. ^{13}C nuclear magnetic resonance and gas chromatography–mass spectrometry studies of carbon metabolism in the actinomycin D producer *Streptomyces parvulus* by use of ^{13}C-labeled precursors. *J. Bacteriol.* **173:** 7790–7801.

Ingraham, J.L., Maaløe, O., and Neidhardt, F.C. 1983. *Growth of the bacterial cell*. Sinauer, Sunderland, MA.

Jackson, M., Crick, D.C., and Brennan, P.J. 2000. Phosphatidylinositol is an essential phospholipid of mycobacteria. *J. Biol. Chem.* **275:** 30092–30099.

Karp, P.D., Paley, S., and Romero, P. 2002. The Pathway Tools software. *Bioinformatics* **18:** S225–S232.

Kealey, J.T., Liu, L., Santi, D.V., Betlach, M.C., and Barr, P.J. 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proc. Natl. Acad. Sci.* **95:** 505–509.

Kell, D.B. 2004. Metabolomics and systems biology: Making sense of the soup. *Curr. Opin. Microbiol.* **7:** 296–307.

Kennedy, J., Murli, S., and Kealey, J.T. 2003. 6-Deoxyerythronolide B analogue production in *Escherichia coli* through metabolic pathway engineering. *Biochemistry* **42:** 14342–14348.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. 2000. General introduction to actinomycete biology. In *Practical* Streptomyces *genetics*, pp. 1–42. The John Innes Foundation, Norwich, UK.

Kinoshita, K., Pfeifer, B.A., Khosla, C., and Cane, D.E. 2003. Precursor-directed polyketide biosynthesis in *Escherichia coli*. *Bioorg. Med. Chem. Lett.* **13:** 3701–3704.

Lee, H.S., Ohnishi, Y., and Horinouchi, S. 2001. A σ B-like factor responsible for carotenoid biosynthesis in *Streptomyces griseus*. *J. Mol. Microbiol. Biotechnol.* **3:** 95–101.

Melzoch, K., Teixeira de Mattos, M.J., and Neijssel, O.M. 1997. Production of actinorphodin by *Streptomyces coelicolor* A3(2) grown in chemostat culture. *Biotech. Bioengineer.* **54:** 577–582.

Michal, G. 1999. *Biochemical pathways: An atlas of biochemistry and molecular biology*. Wiley, New York.

Navarrete, R.M., Vara, J.A., and Hutchinson, C.R. 1990. Purification of an inducible L-valine dehydrogenase of *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* **136:** 273–281.

Novotna, J., Vohradsky, J., Berndt, P., Gramajo, H., Langen, H., Li, X.M., Minas, W., Orsaria, L., Roeder, D., and Thompson, C.J. 2003. Proteomic studies of diauxic lag in the differentiating prokaryote *Streptomyces coelicolor* reveal a regulatory network of stress-induced proteins and central metabolic enzymes. *Mol. Microbiol.* **48:** 1289–1303.

Pandya, K.P. and King, H.K. 1966. Ubiquinone and menaquinone in bacteria: A comparative study of some bacterial respiratory systems. *Arch. Biochem. Biophys.* **114:** 154–157.

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

Poralla, K., Muth, G., and Hartner, T. 2000. Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiol. Lett.* **189:** 93–95.

Reed, J.L., Vo, T.D., Schilling, C.H., and Palsson, B.Ø. 2003. An

expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol*. **4:** R54.
Roubos, H. 2002. "Bioprocess modeling and optimization. Fed-batch clavulanic acid production by *Streptomyces clavuligerus*." Ph.D thesis, Delft University of Technology, Ponsen & Looijen, Netherlands.
Schilling, C.H. and Palsson, B.Ø. 2000. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. Theor. Biol*. **203:** 249–283.
Schilling, C.H., Edwards, J.S., and Palsson, B.Ø. 1999. Toward metabolic phenomics: Analysis of genomic data using flux balances. *Biotechnol. Prog*. **15:** 288–295.
Schilling, C.H., Covert, M.W., Famili, I., Church, G.M., Edwards, J.S., and Palsson, B.Ø. 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol*. **184:** 4582–4593.
Schulbach, M.C., Brennan, P.J., and Crick, D.C. 2000. Identification of a short (C15) chain Z-isoprenyl diphosphate synthase and a homologous long (C50) chain isoprenyl diphosphate synthase in *Mycobacterium tuberculosis*. *J. Biol. Chem.* **275:** 22876–22881.
Shahab, N., Flett, F., Oliver, S.G., and Butler, P.R. 1996. Growth rate control of protein and nucleic acid content in *Streptomyces coelicolor* A3(2) and *Escherichia coli* B/r. *Microbiology* **142:** 1927–1935.
Stanier, R.Y., Ingraham, J.L., Wheelis, M.L., and Painter, P.R. 1986. *General microbiology*, 5th ed. Macmillan Education LTD, London.
Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E.D. 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420:** 190–193.
Stouthamer, A.H. and Bettenhaussen, C. 1973. Utilization of energy for growth and maintenance in continuous and batch cultures of microorganisms. A reevaluation of the method for the determination of ATP production by measuring molar growth yields. *Biochim. Biophys. Acta* **301:** 53–70.
Tang, L. and Hutchinson, C.R. 1993. Sequence, transcriptional, and functional analyses of the valine (branched-chain amino acid) dehydrogenase gene of *Streptomyces coelicolor*. *J. Bacteriol*. **175:** 4176–4185.
Thomas, M.G., Burkart, M.D., and Walsh, C.T. 2002. Conversion of L-proline to pyrrolyl-2-carboxyl-S-PCP during undecylprodigiosin and pyoluteorin biosynthesis. *Chem. Biol*. **9:** 171–184.
van Wezel, G.P., White, J., Bibb, M.J., and Postma, P.W. 1997. The *malEFG* gene cluster of *Streptomyces coelicolor* A3(2): Characterization, disruption and transcriptional analysis. *Mol. Gen. Genet*. **254:** 604–608.
Varma, A. and Palsson, B.Ø. 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.
Yassin, A.F., Brzezinka, H., Schaal, K.P., Truper, H.G., and Pulverer, G. 1988. Menaquinone composition in the classification and identification of aerobic actinomycetes. *Zentralbl. Bakteriol. Mikrobiol. Hyg*. **267:** 339–356.
Zhang, Y.X., Tang, L., and Hutchinson, C.R. 1996. Cloning and characterization of a gene (*msdA*) encoding methylmalonic acid semialdehyde dehydrogenase from *Streptomyces coelicolor*. *J. Bacteriol*. **178:** 490–495.

Web site references

ftp://ftp.genome.ad.jp/pub/kegg/pathways/sco/; *Streptomyces coelicolor* KEGG pathway database (on ftp server).
http://dbkweb.ch.umist.ac.uk/StreptoBASE/s_coeli/referencegel/; *Streptomyces coelicolor* 2D gel protein database.
http://www.expasy.org/cgi-bin/search-biochem-index; ExPASy Biochemical Pathways database.
http://www.expasy.org/enzyme; ExPASy Enzymes nomenclature database.
http://www.expasy.org/sprot/sprot-top.html; SWISS-PROT protein knowledgebase.
http://www.genome.ad.jp/dbget-bin/get_htext?S.coelicolor.kegg+-f+T+w+C; *Streptomyces coelicolor* KEGG Pathways database.
http://www.genome.ad.jp/kegg/ligand.html; KEGG Ligand database.
http://www.sanger.ac.uk/Projects/S_coelicolor/scheme.shtml; The Wellcome Trust Sanger Institute database of *Streptomyces coelicolor*.
http://www.sanger.ac.uk/Software/Pfam/; Protein families database of alignments and HMMs.

Received October 15, 2004; accepted in revised form March 1, 2005.