Chapter 12
Genome-Scale Models and the Genetic Basis for *E. coli* Adaptation

M. Kenyon Applebee and Bernhard Ø. Palsson

Contents

Abstract In this chapter we describe work stemming from the development of a stoichiometrically-constrained model of *Escherichia coli* metabolism, to experimental evolution of strains, to the analysis of the function of acquired adaptive mutations with the goal of understanding their system-wide effect on phenotype.

12.1 Introduction

E. coli is among the most extensively characterized microorganisms (Feist et al. 2007, Karp et al. 2007), making it a good candidate organism to initiate studies of how systems of biological molecules function and interact to produce an organism's physiological behavior – an endeavor that falls under the umbrella of systems

B.Ø. Palsson (✉)
Department of Bioengineering, University of California – San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412 USA
e-mail: palsson@ucsd.edu

S.Y. Lee (ed.), *Systems Biology and Biotechnology of Escherichia coli*,
DOI 10.1007/978-1-4020-9394-4_12,

biology. The study of systems with numerous interacting elements is often aided by the development of computational models of the systems, both to facilitate calculations and to discover emergent properties (Cohen and Harel 2007). The models can then be used to predict the cell's simulated response to a perturbation, which can be compared to the system's actual response; failed predictions can be analyzed to determine what may be missing or incorrect in the model and thus direct research towards critical elements or interactions, creating an iterative process of model testing and discovery.

12.2 Metabolic Models

Genome-scale *in silico* metabolic models simulate how nutrients are processed by an organism's metabolic network to harness energy and biomass, based on the set of reactions the organism's complement of enzymes can catalyze. The dynamics of enzyme-catalyzed reactions are traditionally described using kinetic constants and the concentrations of reactants and products. However, the catalytic efficiency of many enzymes is sensitive to differences that exist between most assay conditions and physiological cellular conditions, such as pH and ionic strength, making the values reported in the literature unrepresentative of *in vivo* conditions. Additionally, the concentration of many metabolites and enzymes within the cell are not accurately known, and they are condition-dependent. These factors have hampered efforts to construct robust metabolic models based on kinetic parameters (Pramanik and Keasling 1997, Varma and Palsson 1994).

For this reason, many recent metabolic modeling efforts have focused on simulating flux states of metabolic maps using only stoichiometric constraints. In this approach, metabolism is represented by a set of stoichiometrically-balanced equations of each metabolic reaction. This can be used to calculate all possible steady-state flux distributions that can result from passing a defined supply of simulated nutrients through the metabolic network. Additional algorithms can then be applied to identify solutions that represent physiological behavior. For example, since the exponential growth of bacteria has been hypothesized to optimize biomass (Lenski et al. 1991), the model can be used to make predictions about the exponential growth rate of the *E. coli* under specified growth conditions by searching the solution space for flux distributions that use the simulated nutrients to produce the most biomass (Pramanik and Keasling 1997), using a method known as flux balance analysis (FBA). FBA uses linear programming to find flux states that maximize a determined objective (Reed and Palsson 2003, Varma and Palsson 1994), such as biomass. Many of these methods are discussed in more depth in Chapter 11 of this book, and elsewhere (Feist and Palsson 2008, Reed and Palsson 2003).

FBA with the biomass objective function predicts the maximum growth rates that can be produced given the specified availability of oxygen and other limiting nutrients (usually the carbon source), represented as the oxygen and substrate uptake rate (OUR and SUR, respectively) (Varma and Palsson 1994). The results of this calculation can be visualized as a three-dimensional phenotypic phase plane (PhPP)

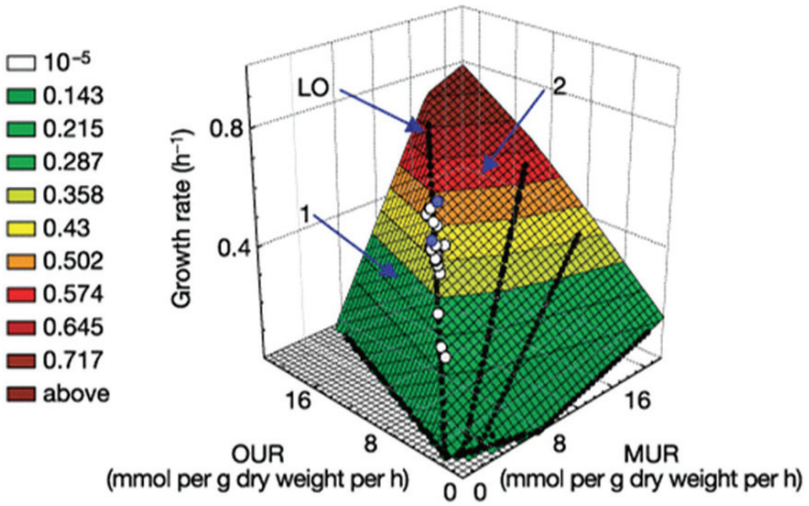


Fig. 12.1 The Phenotype Phase Plane. This figure depicts the maximum growth rate that is predicted to be produced across the range of feasible oxygen and substrate uptake rates. This example shows predicted optimized growth rates of *E. coli* grown aerobically on malate. The line of optimality (LO) is highlighted. White and dark grey circles represent the experimentally-observed growth phenotypes of *E. coli* cultures, where open circles represent wild type cultures grown over a range of malate concentrations (0.25–3g/L) and temperatures (29–37C). The filled circles represent wild type before and after adaptive evolution on malate (2 g/L). Note that all experimentally-derived measurements cluster on the line of optimality. OUR, oxygen uptake rate; MUR, malate uptake rate. Figure originally published in (Ibarra et al. 2002)

(Fig. 12.1) that show the maximum growth rate across the range of allowable uptake rates. These phase planes often have several faces, each of which represents a mode of growth such as aerobic growth or fermentation of a byproduct (Ibarra et al. 2002). The intersection separating oxygen-limited and carbon source-limited growth is known as the line of optimality (LO), and represents the growth mode in which the uptake rate of oxygen and the carbon source are stoichiometrically balanced, producing completely aerobic growth without any energy loss to futile cycles (for example, due to concurrent activity of glycolytic and gluconeogenic pathways). Interestingly, more than one set of uptake rates or even intracellular fluxes can produce the same predicted maximum growth rate (Reed and Palsson 2004).

The growth rate predictions made by genome-scale metabolic models of *E. coli* K12 MG1655 have generally been accurate when compared to growth rates of cultures grown under the relevant conditions during exponential growth (Ibarra et al. 2002). Figure 12.1 shows the results of plotting the measured growth rate, SUR and OUR of growing cultures onto the generated phenotypic phase plane. Cultures were grown under various conditions, including variable temperatures, substrates (glucose, malate, succinate, and acetate), and substrate concentrations in minimal media. When the growth phenotypes of the cultures are plotted on the phenotypic phase plane, most of cultures cluster around the line of optimality as predicted (Ibarra et al. 2002).

12.3 Adaptive Evolution

12.3.1 Adaptation to a Substrate Challenge

Experimentally-measured growth phenotypes do not always fall on line of optimality predicted by biomass-optimized flux balance analysis. For example, *E. coli* K12 MG1655 was found to grow significantly slower than the model-predicted optimum on glycerol and lactate. These prediction failures were not believed to be due to model errors, since the metabolic pathways for utilizing these substrates have been well-characterized. Rather, the discrepancy was hypothesized to be due the organism not being suitably adapted to maximally utilize these substrates.

Given that the metabolic pathways for these substrates exist, the growth capacity could be constrained by factors not included in the metabolic model that regulate flux through the metabolic network, such as enzyme expression, kinetics, or feedback regulation. Such constraints, especially transcriptional regulation, are genetically malleable and are refined by evolution. The short generation time of *E. coli* makes it possible to observe the evolution of adaptive traits in long-term cultures (Lenski et al. 1991). If the prediction failures are due to the wild type strains not being well adapted, subjecting the strains to natural selection by long term exponential growth on the challenging substrate should allow the strains to approach or achieve the FBA-predicted optimal growth rates.

Such experiments, known as laboratory adaptive evolution experiments, were conducted by culturing *E. coli* K12 MG1655 on each of the challenging substrates for a prolonged ($\geq$ 500 generations) period of time (Fong et al. 2003, Ibarra et al. 2002). The resulting strains all had increased growth rates on the targeted substrate (Table 12.1) (Ibarra et al. 2002). Additionally, the growth profiles of each strain migrated towards the predicted line of optimality, and once the growth phenotype aligned with the predicted line of optimality generally it only migrated along it (Table 12.1, Fig. 12.2). Further growth rate increases were accomplished by increasing both uptake rates proportionally so that the phenotype moved up the line of optimality (Edwards et al. 2001, Ibarra et al. 2002). However, it should be noted that several strains eventually increased their sugar uptake rates beyond what could be fully aerobically metabolized, due to physical limits on the oxygen uptake rate – these strains were driven by growth-rate dependent selection to ferment the excess sugar, and consequently migrated off of the predicted line of optimality.

Overall, the outcomes indicate that actual growth limits are captured by biomass-optimized flux-balance analysis of the stoichiometric reconstruction of the *E. coli* metabolic network. Additionally, they suggest that the cellular architecture is efficient at relieving constraints aside from those imposed by chemical or catalytic limitations. Unfortunately, FBA alone cannot be used to determine how the evolved strains metabolize nutrients, since each growth rate can be associated with multiple flux distributions (Reed and Palsson 2004). Quantitative metabolomic profiling is necessary to both identify and validate the flux distribution predictions, which is

Table 12.1 Summary of substrate-challenged evolution experiments

	Growth phenotype near predicted LO					
	Temp. (°C)	Generations evolved	Wild type	Evolved strains	Percent of GR increase	Ref.
Glucose	37	500	Yes	Yes	17	(Fong and Palsson 2004)
Malate	37	500	Yes	Yes	19	(Fong et al. 2005a)
Acetate	37	700	Yes	Yes	20	(Fong et al. 2005b)
Lactate	30	950	No (acetate sec.)	Yes	147	
Lactate	30	950	No (acetate sec.)	Yes	132	
Glycerol	30	1000	No (futile cycles)	Yes	140	
Lactate	37	870	No (futile cycles)	No (O_2 maxed)	80	
Pyruvate	37	1200	No (acetate sec.)	No (O_2 maxed)	69	
Pyruvate	30	1000	Yes	No (O_2 maxed)	115	
α-Ketoglutarate	37	625	Yes	No	41	
	30	440	Yes	No	48	

This table compares outcomes of adaptive evolution experiments that challenged MG1655 wild type *E. coli* to increase growth on a variety of substrates with *in-silico* flux-balance analysis (FBA) growth rate predictions. The growth phenotype of wild type on some substrates did not meet the FBA prediction because of carbon or oxygen uptake imbalance, that resulted in acetate secretion (too much carbon uptake) or futile cycles (too much oxygen uptake). The growth phenotype of some evolved strains did not fall on the LO because they increased their carbon uptake beyond the limits of physiologically-available O_2. LO – Line of Optimality. GR – Growth rate.

still a very challenging experimental endeavor. Additionally, these results do not begin to identify the mechanics of the adaptation, which will be addressed later in this chapter.

12.3.2 Adaptation to Deletion of a Metabolic Gene

In addition to growth on rarely-encountered carbon sources, strains can also be perturbed from the optimal growth state by deletion of a metabolic gene. Deleting a metabolic gene produces a strain with a different potential optimum growth rate that can be predicted by removing the catalyzed reaction from the *in silico* metabolic model, changing the solution space calculated by flux balance analysis (Reed and Palsson 2003). The new line of optimality is composed of solutions that most effectively redistribute the metabolic flux around the lost reaction to produce the most biomass (deletion of essential genes are not considered since they have no

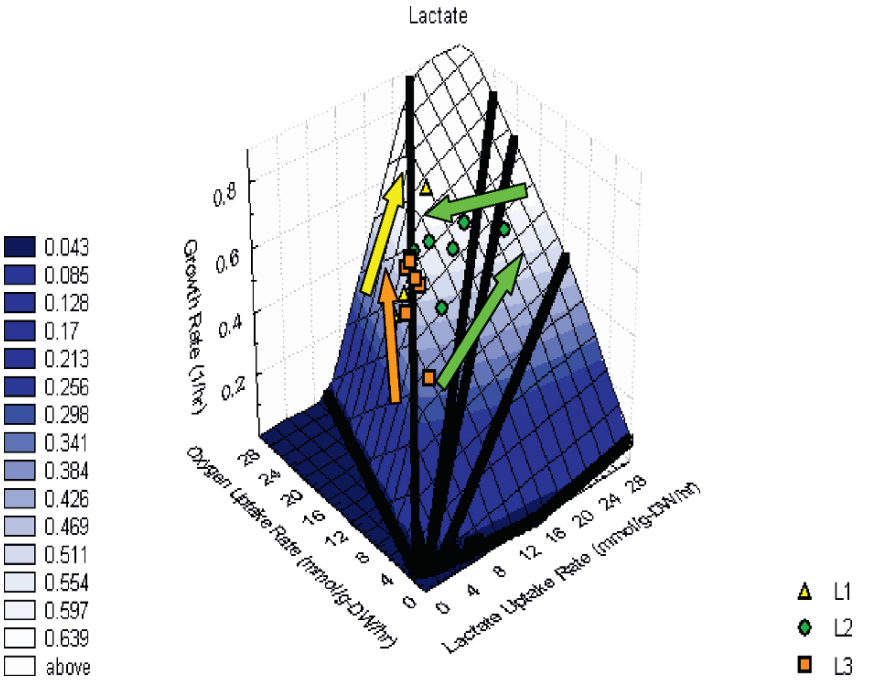


Fig. 12.2 Migration of the growth phenotype of evolving *E. coli* strains towards the line of optimality. This example shows the growth phenotype three replicate experimental evolutions of wild type on lactate (L1-L3), which all eventually migrate towards the line of optimality. Figure originally published in Fong et al. 2003

solution space that supports growth). Strains of *E. coli* with single metabolic gene deletions were adaptively evolved to test whether the predicted growth phenotype reflects the actual potential of *E. coli* to adapt to gene loss (Fong and Palsson 2004). These experiments test the plasticity of cell regulation to allow the redistribution of metabolic fluxes.

One of six genes was deleted from strains of *E. coli*, (acetate kinase A (*ackA*), fumarate reductase (*frd*), phosphoenolpyruvate carboxykinase (*ppc*), phosphoenol pyruvate carboxylase (*ppc*), triosephosphate isomerase (*tpi*), or glucose 6-phospate-1-dehydrogenase (*zwf*)). These genes encode enzymes required for gluconeogenesis, fermentation, or the pentose phosphate pathway (Fig. 12.3). Adaptive evolution experiments were performed with these strains across a set of substrates that enter central metabolism at a variety of points (Fong and Palsson 2004). Approximately 80% of the adaptively evolved gene-deletion strains made gains in growth rate within 10% of their respective FBA prediction. This success rate further validates biomass-optimized flux balance analysis of the stoichiometrically-constrained metabolic model, and indicates that the stoichiometric model captures many of the physiologically-relevant constraints on growth. The results of both these and the substrate-challenged adaptive evolution studies suggests that *E. coli* is fairly adept

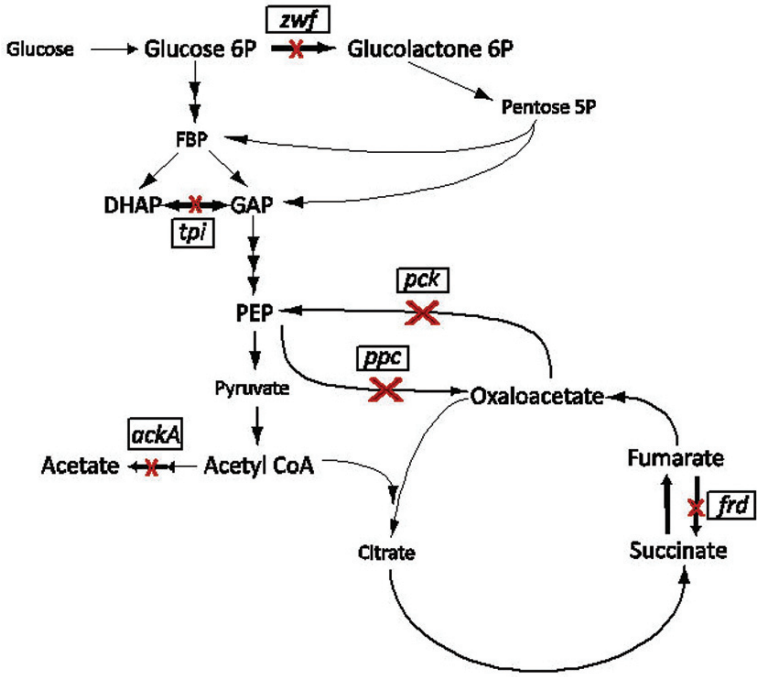


Fig. 12.3 Genes deleted in evolved gene-deletion strains described in Fong and Palsson, 2004. Deleting *zwf* (glucose 6-phosphate dehydrogenase) prevents forward flux through the pentose phosphate pathway. Deleting *tpi* (triosphosphate isomerase) interferes with glycolysis and gluconeogenesis. Deleting *pck* (phosphoenolpyruvate carboxykinase) can interfere with gluconeogenesis from citric acid cycle intermediates. Deletion of *ppc* (phosphoenolpyruvate carboxylase) impedes the ability to replenish oxaloacetate to the citric acid cycle. Deletion of *frd* (fumarate reductase) obstructs utilization of the reductive pathway of the citric acid cycle. Deletion of *ackA* (acetate kinase A) blocks a pathway needed to secrete acetate. These strains were experimentally evolved on a variety of carbon sources to select for increased growth rate

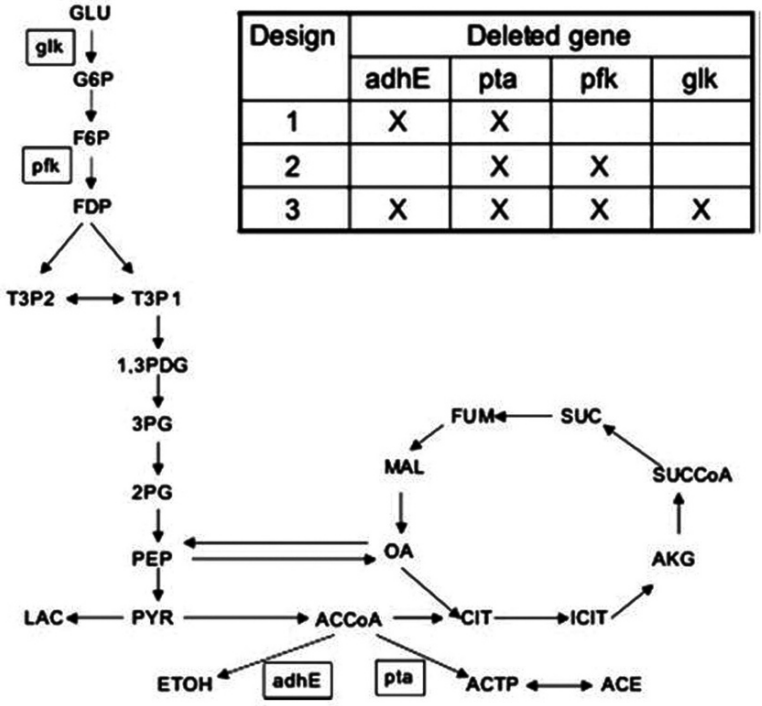
at rerouting its metabolic flux to achieve the optimal growth phenotype available within the limits of its metabolic chemical capacity.

12.3.3 Application of Experimental Evolution for Rational Design of Production Strains (OptKnock)

A practical application for metabolic modeling is to assist the design of strains that secrete desired product(s). A number of algorithms, which can be used to find growth-coupled strains (including OptKnock (Burgard et al. 2003, Pharkya et al. 2003), OptStrain (Pharkya et al. 2004), and OptGene (Patil et al. 2005)), calculate the predicted growth phenotype across all possible gene deletion strains and identify permutations of the metabolic network that maximize both biomass formation

and production of the secreted compound. These two objectives are simultaneously met when metabolic reactions that allow growth by secretion of alternative, more-energetically favorable fermentation products are removed, making growth dependent on secretion of the desired compound. Since these designs involve deleting metabolic genes, adaptive evolution can be employed to drive the generated strains to recover and to optimize their growth rate and secretion rate. Manual examination of the metabolic network can also be used to predict gene deletions that will couple growth to product secretion; however, the computational algorithms can identify designs that are not intuitively obvious.

The accuracy of both intuitive designs and non-intuitive designs predicted by the algorithm OptKnock to optimize the production of lactic acid were tested experimentally by generating the strains with the indicated gene deletions, and the growth rates of the constructs were optimized by adaptive evolution (Fong et al. 2005a). Three strain designs were tested, (1) *pta-adhE* double deletion strain, (2) *pta-pfk* double deletion strain, and (3) *pta-adhE-pfk-glk* quadruple deletion strain, summarized in Fig. 12.4. The first strain design, *pta-adhE*, is an intuitive design as it deletes reactions in the ethanol and acetate fermentation pathways. In the second design, *pta-pfk*, the reason for deleting phosphofructokinase in not intuitively



Design	Deleted gene			
	adhE	pta	pfk	glk
1	X	X		
2		X	X	
3	X	X	X	X

Fig. 12.4 OptKnock strain designs. Strain designs 1–3 were generated by deleting the genes indicated on the table. The reactions lost by each gene deletion are indicated on the diagram of central metabolism. Figure originally published in Fong et al. 2005a

obvious, but it promotes lactic acid production by increasing NADH and pyruvate by forcing flux through the Entner-Doudoroff and pentose phosphate pathway. In the third strain, the additional deletion of glucose kinase also contributes to lactate secretion by increasing pyruvate production, via forcing all metabolized glucose to be phosphorylated by the phosphotransferase system.

Adaptive evolution increased the growth rates of all four strain designs interestingly to approximately the same rate (0.24–0.26 hr^{-1}). Lactic acid production increased with growth rate in each case, proving that all of the designs successfully coupled lactate secretion to growth. The two-gene deletion strains agreed well with the predictions made by OptKnock in terms of migration of both the growth rate and lactate secretion rate; the quadruple gene-deletion strain, Δ*pta-adhE-pfk-glk*, actually acquired slightly faster growth rates than predicted. However, even though the lactic acid secretion rates in all three designs increased over the adaptation, the final lactate titer in the media recovered after culturing did not, possibly suggesting the existence of metabolic feedback mechanisms not included in the current model that can arrest secretion beyond some threshold. Additionally, the intuitively-designed strain, *pta-adhE*, produced the most lactic acid in terms of both secretion rate and titer among the three designs, suggesting that the non-intuitive upstream gene deletions (*pfk* and *glk*) may have more complex effects on metabolism than those currently captured by the model – again, this could easily be attributable to metabolic feedback mechanisms.

This study demonstrated that the OptKnock algorithm can successfully be used identify gene deletions that couple a secondary objective, such as fermentation product secretion, to growth. The useable set of solutions is of course limited to those that do not require the loss of an essential gene or set of genes. And even though in this case the intuitive design were the most effective, this may not always be so – indeed, this tool allows researchers to search for designs that produce compounds when there are none that are intuitively obvious. Additionally, the range of compounds that can be produced by modeled organisms like *E. coli* can conceivably be increased by including reactions and pathways carried out by enzymes that can be transferred from other microorganisms, as performed by the OptStrain algorithm (Pharkya et al. 2004). Growth-coupling algorithms have the potential to allow systems-scale models to be used to drive the development of strains for practical applications.

12.4 Characterizing Intracellular Mechanisms of Adaptation

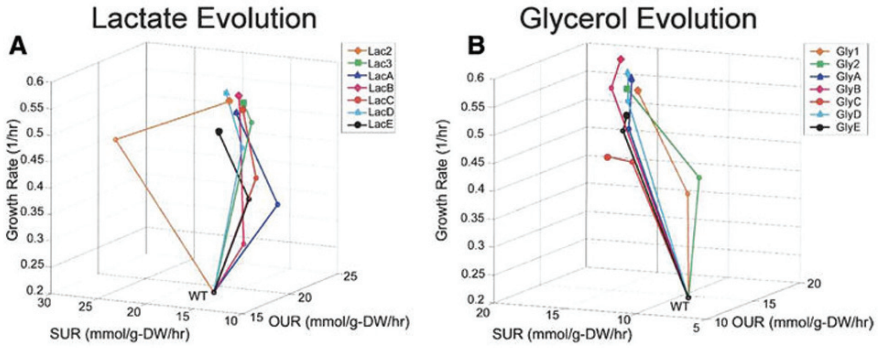
Beyond illustrating that selection during exponential growth produces optimized phenotypes that converge with biomass-optimized FBA predictions, the intracellular changes that facilitate the phenotype shifts are also of inherent interest. Adaptations can act through many intracellular activities, including but not limited to metabolism, transcriptional regulation, protein turnover, and intracellular feedback mechanisms. Thus identifying the mechanism through which adaptive mutations act can require multiple high-throughput experimental methods, including

mRNA transcription profiling, global metabolomic and flux profiling, and chromatin immunoprecipitation-on-chip – and requires an integrated analysis of such data sets. Additionally, the timescale required to change the growth phenotype implies the adaptations involve genetic mutations, rather than an adjustment to an existing mode of regulation. Therefore it is necessary to identify the acquired mutations in order to pursue the fundamental goal of understanding the mechanism that underlies the phenotypic change.

12.4.1 Phenotypic Characterization of Replicate Endpoints

As previously touched upon, flux balance analysis of an *in-silico* model cannot directly be used to assess the intracellular state. This is because the line of optimality actually represents a set of flux states that produce the same, maximum biomass (although, the differences between the solutions in the set are often restricted to variation in flux among a small set of reactions (Reed and Palsson 2004)). The existence of multiple metabolic states that can achieve the same optimal biomass objective parallels the frequent observation that replicated adaptive evolution experiments often achieve nearly identical growth rates and nutrient uptake rates in the evolved environment, but have distinct phenotypes such as growth rate on alternative substrates and byproduct secretion rates (Fong et al. 2005b, Fong et al. 2003, Fong and Palsson 2004). Such observations suggest these replicate lineages acquire different adaptive changes that produce the same or similar adjustment to growth in the environment of the evolution, but which have different (pleiotropic) effects on growth under other conditions. To sample the range of genetically-distinct endpoint strains that can be produced by replicate evolutions to the same metabolic challenge, we have extensively characterized replicate endpoint strains adaptively evolved on glycerol, lactate, or after deletion of various single metabolic genes (Fong et al. 2005b, 2006, Fong and Palsson 2004).

The glycerol- and lactate-adapted endpoint strains all achieved growth rates, SURs, and OURs within 12% of each of the other replicates. Additionally, while the replicate evolved strains generally had similar growth rates on other substrates, there was sufficient variability to suggest each endpoint strain had acquired different adaptations. The variation between replicate strains is even more pronounced between the mid-point evolution cultures (day 20) than between the replicate endpoint strains, in terms of growth rate, and the oxygen and substrate uptake rate during growth on the alternative substrates (Fig. 12.5). This variation may indicate that there were multiple adaptive strategies available during the initial stages of adaptation that lead to divergence between replicate cultures, followed by a period of more discriminate selection during the later period that caused the phenotypes of the endpoints to converged towards a single optimal phenotype. Interestingly, the endpoint strains generally grow faster than the wild type strain on many other carbon sources in minimal media, suggesting that some of the acquired changes were generally beneficial.



Strain	SUR (mmol/g-DW/hr)	OUR (mmol/g-DW/hr)	GR (1/hr)
Wild-type	13.8 ± 0.4	15.9 ± 2.1	0.23
Lac2 D60	14.9 ± 0.2	19.9 ± 1.0	0.56
Lac3 D60	14.5 ± 0.04	20.2 ± 1.1	0.53
LacA D60	14.6 ± 0.03	18.9 ± 0.5	0.54
LacB D60	16.4 ± 1.4	21.2 ± 0.6	0.55
LacC D60	13.0 ± 0.2	17.7 ± 0.2	0.56
LacD D60	14.4 ± 0.4	17.8 ± 0.5	0.59
LacE D60	16.4 ± 0.1	18.9 ± 0.7	0.50

Strain	SUR (mmol/g-DW/hr)	OUR (mmol/g-DW/hr)	GR (1/hr)
Wild-type	7.6 ± 0.4	12.0 ± 0.7	0.22
Gly1 D44	14.7 ± 0.2	17.3 ± 0.7	0.54
Gly2 D44	14.9 ± 0.2	16.6 ± 0.4	0.55
GlyA D44	15.1 ± 0.2	17.3 ± 0.3	0.56
GlyB D44	17.2 ± 0.1	19.3 ± 0.2	0.58
GlyC D44	13.4 ± 0.1	12.2 ± 0.1	0.46
GlyD D44	14.5 ± 0.2	16.0 ± 0.2	0.58
GlyE D44	15.0 ± 0.2	16.6 ± 0.1	0.50

Fig. 12.5 Phenotypic variability between replicate lactate- and glycerol-evolved strains. (**A**) Lactate Evolution. The top figure shows the trajectory of each replicate as it evolved from wild type, in terms of changing growth, substrate uptake, and oxygen uptake rate. Note that most strains begin to converge to the same growth phenotype. The bottom table lists the values of the measured parameters for the wild type and the final (day 60) endpoint strains. (**B**) Similar to (**A**), except applies to strains evolved to glycerol. Figure originally published in Fong et al. 2005b

12.4.2 Identifying Changes in Metabolic Pathway Utilization

Since increasing the growth rate involves increasing the efficiency with which available nutrients are metabolized, it is informative to determine how flux through various metabolic pathways has changed over the course of adaptation. One method for comparing functional flux states in microorganisms tracks the catabolism of ^{13}C-labeled substrates, and has been successfully used to access *in-vivo* reaction rates (Fischer et al. 2004, Sauer 2004). This technique has been used to track flux changes following adaptation to lactate and deletion of various single metabolic genes.

Results from the ^{13}C-labeling experiments on the lactate evolutions (Hua et al. 2007) showed that the first major flux change was a dramatic (up to 80%) increase in the uptake of lactate, and increased flux capacity through most metabolic reactions, over the first 20 days of evolution. Additionally, though the replicate strains showed significant phenotypic diversity at day 10 that later converged, their flux profiles were relatively similar at day 10 relative to later in the evolution. During this period, flux profiling showed that all of the strains shift more metabolites into the TCA cycle rather than acetate fermentation, allowing the cells to generate more energy and anabolic precursors, and thus achieve a faster growth rate. After day

10 the flux changes among the replicate evolutions diverged, though an important consistency was discovered across the course of all of the replicate evolutions – approximately two-thirds of the metabolized lactate was consistently partitioned to the TCA cycle through pyruvate dehydrogenase, indicating that partitioning between gluconeogenic and catabolic fluxes is tightly regulated – likely by feedback mechanisms that recognize cellular concentrations of phosphoenolpyruvate (Chulavatnatol and Atkinson 1973).

An additional ^{13}C-labeling experiment was performed with adaptively evolved gene deletion strains, though with deletion of different genes than those described in Section 12.3.2. The genes deleted in this study encode metabolic enzymes that catalyze key metabolic branch points in central metabolism, and were chosen because their loss is expected to most dramatically change pathway utilization (Fong et al. 2006) (phosphoglucose isomerase (*pgi*), phosphoenolpyruvate carboxylase (*ppc*), triose-phosphate isomerase (*tpi*), or phosphate transacetylase (*pta*)) (Fig. 12.6). As

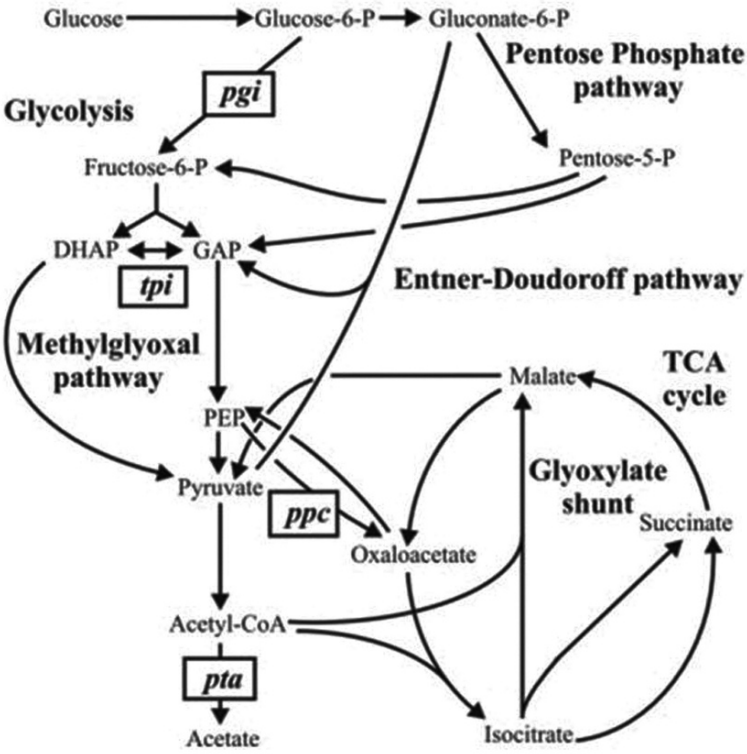


Fig. 12.6 Map of metabolism highlighting the gene deletion strains that were adaptively evolved on glucose in minimal medium. Deleting *pgi* (phosphoglucose isomerase) impedes glycolysis, forcing flux to be rerouted through the pentose phosphate or Entner-Doudoroff pathway. Deleting *pta* (phosphate acetyltransferase) impedes acetate fermentation. The *tpi* and *ppc* deletions are described in the Fig. 12.3 caption. Figure originally published in Fong et al. 2006

in the previous studies, the generated endpoint strains achieved near-wild type growth rates, and replicate endpoints differed in terms of phenotypic traits like byproduct secretion, suggesting they had acquired different specific adaptations.

The outcome of the ^{13}C-labeling experiments revealed that most the flux changes involved activating alternative pathways to locally reroute metabolites around the lost gene function. This rerouting was accomplished by activating pathways typically used for growth under other conditions, such as the pentose-phosphate pathway (*pgi* strains), glyoxylate shunt (*ppc* strains), and the methyl-glyoxyl bypass (*tpi* strains) (Fig. 12.6). No evidence of new enzyme activities or metabolic pathways was observed. The evolved gene deletion strains often utilized the same pathways to circumvent the lost gene activity as those used in pre-evolved deletion strain, and evolution involved increasing the flux capacity through those pathways – similar to increasing lactate uptake in the previously described study. However, the outcome of this study suggests that *E. coli* immediately responds to a breakdown in metabolic capacity by activating repressed pathways to search for a way to redistribute flux through the metabolic network (a process that may favor short routes that cause the least disruption of the wild type flux configuration (Segre et al. 2002)). Thus the adaptation process may often consist of refining the most easily established solutions, rather than searching for.

However, this should not be taken to suggest that every replication of adaptive evolution under the same conditions will eventually converge to the same flux configuration. Although all of the evolved replicate strains circumvented their gene deletions with the same latent pathway, there were several significant differences in how some replicate strains utilized other pathways. For example, one evolved *pgi* strain primarily utilizes the TCA cycle and secretes acetate while another has more flux through the glyoxylate shunt and secretes no acetate. These outcomes suggest that much of the variability between replicate evolutions may stem from variable means of making downstream metabolic adjustments that are necessary to refine usage of the major adaptive flux shift.

12.4.3 Gene Expression Changes

While growth rate increases are dependent on improving flux through the metabolic network, those improvements are a result of refining the activity of metabolic elements. Metabolic activity is regulated at multiple levels, including allosteric control of enzymes, turnover rates, and transcriptional regulation. We used mRNA transcription profiling to identify the adaptive mechanism utilized in each strain, since high throughput methods are not available to screen the other types of regulation, such as feedback regulation.

We measured changes in genome-wide mRNA transcription levels of the replicate glycerol- and lactate-evolved strains over the course of their evolutions, as well as the cultures at day 1 and day 20 of each evolution, and wild type grown on glucose (Fong et al. 2005b). Interestingly, the transcription state of each evolved endpoint

Table 12.2 Number of expression changes across evolutions on glycerol and lactate

		Glycerol evolved		Lactate evolved	
relative to wild type (glucose)	Day 1	39%	1687 genes	18%	756 genes
	Day 20	18%	770 genes	4%	194 genes
	Day 44	11%	498 genes	7%	323 genes

This table shows the average number of significant gene expression changes in adaptively evolved strains relative to wild type grown on glucose at different time points in the evolution experiments. Data from Fong et al. 2005b

was distinct, despite similarity in endpoint growth phenotypes, providing further evidence that endpoint cultures result from acquiring different sets of adaptations.

Analysis of the transcription profiles revealed that the largest number of changes in gene expression were between wild-type grown on glucose and wild-type grown on the challenging substrate (day 1 cultures), and that the course of adaptive evolution returned the expression of many of these genes to pre-evolution levels (Table 12.2). This may be attributable to a large scale carbon-scavenging response intended to allow metabolism of any "low-quality" carbon sources available in the absence of preferred substrates (Liu et al. 2005), in which case the adaptation process may involve refining the regulatory response to only activate genes specific to glycerol or lactate metabolism.

Expression changes that developed over the course of evolution common across replicate strains were also identified. Approximately 70 gene expression changes were identified among the glycerol-evolved strains, and only two among the lactate-evolved strains – a striking difference. The small number of genes identified across multiple lactate strains may be due to their adaptive pathways being more divergent compared to glycerol strains, which is also discussed later in this chapter (Section 12.5). It appears to be very difficult to identify adaptive mechanisms through examination of expression changes alone. Adaptation generally results in changed expression of a large number of genes, and it is difficult without knowledge of changes to other intracellular systems, such as metabolism or of specific mutations, to identify critical expression shifts that mediate the adaptive mechanism rather than result as a down-stream response.

Additionally, the expression profiles of the evolved deletion strains have also been measured, and have been compared to the flux changes discussed in the previous section (Section 12.5.2) to try to identify gene expression changes responsible for the metabolic and phenotypic adaptations. However, as previously stated, not all of the flux changes are caused by expression changes since there are other

mechanisms that regulate metabolism, such as post-translational regulation, enzyme kinetics, and allosteric control. Additionally, not all gene expression changes directly alter the phenotype or metabolic flux, or their effects may not be adaptive but rather are a secondary response to other changes. Wide-spread expression levels can also be caused by altered mRNA stability or RNase activities. But both data sets can be combined to identify adaptive flux changes that are caused by changes to gene expression.

Many of the observed flux changes among the evolved deletion strains can be linked to expression changes (Fong et al. 2006). Expression changes correlated well to changes in flux through the glyoxylate shunt, methylglyoxylate shunt, and TCA cycle, suggesting that these pathways are at least partially controlled at the transcriptional level. On the other hand, no expression changes were identified that correlate to observed flux changes through glycolysis or the pentose phosphate pathway, suggesting that other mechanisms may predominantly regulate flux through those pathways. Interestingly, no correlation was found between flux and gene expression changes among the pre-evolved deletion strains. This may indicate that the major, shared flux shifts were initially mostly dependent on other mechanisms of metabolic regulation, and that evolution was necessary to acquire changes involving transcriptional regulation.

Additionally, the flux and expression data collected from the evolved gene-deletion strains has been used to inform the metabolic model to try to identify why some strains failed to reach model-predicted growth rates (Herrgard et al. 2006). The developed method, called optimal metabolic network identification (OMNI), searches for metabolic reactions that, when eliminated, can produce flux distributions that are the closest fit to those actually measured in the relevant strains. The enzymes of these reactions may act as bottlenecks to optimal growth, that for some reason are disadvantageously regulated in a manner that had not been overcome by the point at which the experimental evolution was terminated. The validity of this approach is supported by the fact that expression of many of the genes identified as possibly causing flux bottlenecks can be seen to have reduced expression in the evolved strain compared to wild type.

The previously described studies successfully correlated flux change to shifts in expression of several relevant genes, partially illustrating the adaptive mechanism. It is likely that additional conclusions can be made from the expression profiling data if it were analyzed with a more sophisticated view of cell regulation, possibly facilitated by a comprehensive model of transcription regulation.

12.5 Genome Resequencing

Adaptive changes in the phenotype of evolved strains are ultimately caused by mutations (and possibly epigenetic changes to the DNA), and a comprehensive understanding of the adaptation process requires their identification. Further, the effect of each mutation on the DNA-encoded function must be determined, whether it

effects regulation of nearby genes, causes an amino acid change in a protein that affects its function or activity, or some other effect. Ultimately, we would like to know how acquired mutations cause all of the observed changes in metabolic flux, gene expression, and the overall systems dynamics that translate into the observed phenotype. Such an accomplishment would significantly contribute to building an understanding of how genetic structure produces phenotype.

Conclusively identifying all of the mutations acquired during adaptive evolution requires a search of the entire genome that can find changes as small as of a single nucleotide (or a single nucleotide polymorphism – SNP). Six of the glycerol-evolved strains have undergone whole-genome resequencing (Herring et al. 2006). With one exception, the strains acquired two to three SNPs within the coding region of annotated genes, as described in Table 12.3. All five resequenced strains acquired a different mutation within *glpK*, which encodes glycerol kinase which catalyzes the rate-limiting step in glycerol metabolism. Three of the strains also acquired mutations in RNA polymerase subunits β and β', encoded by *rpoB* and *rpoC*, which was surprising given the extensive influence these genes may have on global transcription regulation. Additionally, mutations were acquired by two strains that effect peptidoglycan biosynthesis (within genes *dapF* and *murE*).

Table 12.3 Mutations identified in glycerol-evolved strains

Clone	Gene	Product/Function	Mutation	Gene position nt	Region	Genome Position
GB-1	*glpK*	Glycerol kinase	a-> t	218	Coding	4115028
	rpoC	RNA polymerase	27 bp deletion	3132–3158	Coding	4186504–4186530
GC-1	*glpK*	Glycerol kinase	g-> t	184	Coding	4115062
	n/a	All genes between insC-5 & insD6	1313 kb duplication[1]	n/a	n/a	~3189209–4497523
GD-1	*glpK*	Glycerol kinase	g-> a	816	Coding	4114430
	rpoB	RNA polymerase	a-> t	1685	Coding	4180952
	murE	peptidoglycan biosynthase	a-> c	8	Coding	93173
GE-1	*glpK*	Glycerol kinase	a-> c	113	Coding	4115133
	rpoC	RNA polymerase	c-> t	2249	Coding	4185621
	dapF	Lysine/peptidoglycan biosynthase	c-> a	512	Coding	3993293
G2-1	*glpK*	Glycerol kinase	9 bp duplication	705	Coding	4114541
	rph-pyrE	RNAse PH/pyrimidine synthesis	82 bp deletion	rph: 610-end	Coding + Intergenic	3813882–3813963
	pdxK-crr	Pyridoxal kinase/enzyme IIa glucose	28 bp deletion	pdxK: 833-end	Coding + Intergenic	2534400–2534427

Mutations identified by whole-genome resequencing of *E. coli* strains adaptively-evolved to increase growth on glycerol minimal media. Previously published. 1. Evident in CGS mapping data; not independently validated.

The impact of individual mutations on the fitness phenotype was determined by assessing strains that had been altered to carry one or more of the discovered mutations using site-directed mutagenesis (Herring et al. 2006). Importantly, the set of mutations identified in each strain have been proven to be responsible for the phenotype change, since the endpoint phenotype has been shown to be reproduced by inducing the mutation sets into wild type by site-directed mutagenesis. The phenotype of mutant strains carrying only single mutations indicate that the *rpoB/C* mutations have the greatest impact on growth rate, followed by the *glpK* mutations. Additionally, results of competitions between strains with different induced mutations indicated that some of the *glpK* and *rpoB/C* mutations may have cooperative (epistatic) effects (Applebee et al. 2008); possible mechanisms are still being investigated. Mutations to the peptidoglycan synthesis genes (*murE* & *dapF*) were also only shown to have a significant effect on growth rate in strains that also carried the co-acquired *glpK* and *rpoB/C* mutations.

The *rpoB/C* mutations clearly perform a critical function in optimizing the growth of *E. coli* on glycerol, and this function most likely involves adjusting transcriptional regulation. These mutations have the greatest impact on fitness among those acquired by glycerol-evolved stains, and they may additionally be responsible for the increased growth capacity in minimal media on a wide range of non-glycerol substrates (unpublished results). One hypothesis suggests that these mutations improve growth by reducing the sensitivity of RNA polymerase to stress response signals, particularly ppGpp, that may be induced by the transition from rich to minimal media. This may prevent the expression of unnecessary or detrimental proteins that are associated with the stress response (Liu et al. 2005). The effect of these mutations on RNA polymerase activity and global gene expression is currently being investigated.

Strains evolved on lactate or in response to deletion of the *pgi* gene have also undergone whole-genome resequencing (unpublished data). Interestingly, there was more variation in the number of mutations these strains acquired (0–7 mutations per strain, versus 2–3 in glycerol), and they appeared across a more diverse set of genes than glycerol-evolved strains. It is not yet clear why these evolutions produced more genetically divergent replicates compared to the glycerol-evolved strains, though it suggests there may simply be more adaptive routes that these replicate evolutions can sample.

A developing pattern seen across the different experimental evolutions is that strains frequently acquire mutations to both a metabolic gene and a global transcription factor. The function of mutations to metabolic genes (PEP synthase in lactate-evolved strains, NADPH/NADH transhydrogenase genes in evolved Δ*pgi* strains) is assumed to increase the activity of rate-limiting enzymes under the growth conditions, as has been shown for the *glpK* mutations in the glycerol evolved strains (Herring et al. 2006). The function of mutated global regulatory or transcription factors is likely more complicated. Among the mutations discovered in lactate and Δ*pgi* strains are mutations to *crp, cyaA*, and *rpoS* (unpublished data). This trend suggests that the *E. coli* regulatory network is both robust to mutations that alter

the function of major regulatory elements, and that these are more accessible or advantageous than mutations that effect the transcriptional regulation of a smaller, more specific set of genes.

12.6 Summary Remarks

This chapter has covered how stoichiometrically-constrained models of *E. coli* metabolism have been used to predict optimal growth rates on a variety of carbon sources. The physiological relevance of these predictions has been validated by demonstrating they approximate the actual growth phenotype observed on these substrates in exponential growth phase. Not all substrates produce growth in wild type that approximates the model-derived optimum, but adaptive evolution studies have demonstrated that prolonged exposure to these substrates generates adaptations that cause the growth phenotype to converge towards the predicted optimum. Additionally, the models have been successful at predicting the optimum growth phenotype that *E. coli* will adapt to following loss of a metabolic gene function. These outcomes suggest that the model captures the critical constraints on growth capacity before considering metabolic regulation, without requiring kinetic constraints. Additionally, it indicates that metabolic regulation is readily malleable to selection pressure, to allow the optimum growth phenotype to be found in response to a wide range of environmental or metabolic challenges.

Each evolution experiment was performed multiple times, and although each of the replicate strains generally acquired the same growth phenotype (converging towards the predicted optimum), they generally differed in terms of other phenotypic traits such as byproduct secretion or growth capacity on alternative substrates. Significant differences between replicate strains were also validated by differences between their flux and transcription profiles. The basis for those differences have now been genetically identified – no two replicate strains have acquired identical mutations. Thus different genetic changes can produce similar phenotypic changes. The existence of multiple metabolic flux shifts that produce the optimal phenotype is actually predicted from the results of flux balance analysis (Reed and Palsson 2004). The degree of variation among replicate evolved strains may indicate how many adaptive routes exist to produce this phenotype.

Attempts to understand the intracellular changes causing the growth phenotype changes have involved measuring changes in pathway utilization and transcription expression over the course of evolution, and identifying acquired mutations. At this point the challenge involves deducing the mechanism by which the discovered mutations alter metabolic flux through regulatory mechanisms to produce increased growth rate. Among the most intriguing discoveries are the mutations to so-called "global regulators" like RNA polymerase subunits and elements of catabolite repression, that appear to play a significant role in inducing the adapted phenotype. It remains to be proven whether this truly is a general mechanism of adaptation, and if so how it functions.

These studies begin to highlight the potential of using adaptive evolution to discover important cellular dynamics that exist on the systems-biology level. The eventual goal is to be able to predict what regulatory and metabolic changes are necessary to produce a desired phenotype. We will have accomplished a true understanding of the relationship between phenotype and genotype when we understand the adaptive function of acquired mutations – and are able to predict them.

Acknowledgments Thank you to Karsten Zengler, Jan Schellenberger, and Adam Feist for their insightful feedback, to Addiel de Alba for lending his graphics expertise and to Jessica Na for her encouragement during the writing of this chapter.

References

Applebee MK, Herrgard MJ, Palsson BO (2008) Impact of individual mutations on increased fitness in adaptively evolved strains of *Escherichia coli*. J Bacteriol 190(14):5087–94

Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84(6):647–57

Chulavatnatol M, Atkinson DE (1973) Kinetic competition *in vitro* between phosphoenolpyruvate synthetase and the pyruvate dehydrogenase complex from *Escherichia coli*. J Biol Chem 248(8):2716–21

Cohen IR, Harel D (2007) Explaining a complex living system: dynamics, multi-scaling and emergence. J R Soc Interface 4(13):175–82

Edwards JS, Ibarra RU, Palsson BO (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol 19(2):125–30

Feist AM, Henry CS, Reed JL et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3:121

Feist AM, Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat Biotechnol 26(6):659–67

Fischer E, Zamboni N, Sauer U (2004) High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. Anal Biochem 325(2):308–16

Fong SS, Burgard AP, Herring CD et al. (2005a) *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol Bioeng 91(5):643–8

Fong SS, Joyce AR, Palsson BO (2005b) Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. Genome Res 15(10):1365–72

Fong SS, Marciniak JY, Palsson BO (2003) Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale *in silico* metabolic model. J Bacteriol 185(21):6400–8

Fong SS, Nanchen A, Palsson BO et al. (2006) Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. J Biol Chem 281(12):8024–33

Fong SS, Palsson BO (2004) Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nat Genet 36(10):1056–8

Herrgard MJ, Fong SS, Palsson BO (2006) Identification of genome-scale metabolic network models using experimentally measured flux profiles. PLoS Comput Biol 2(7):e72

Herring CD, Raghunathan A, Honisch C et al. (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat Genet 38(12):1406–12

Hua Q, Joyce AR, Palsson BO et al. (2007) Metabolic characterization of *Escherichia coli* strains adapted to growth on lactate. Appl Environ Microbiol 73(14):4639–47

Ibarra RU, Edwards JS, Palsson BO (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. Nature 420(6912):186–9

Karp PD, Keseler IM, Shearer A et al. (2007) Multidimensional annotation of the *Escherichia coli* K-12 genome. Nucleic Acids Res 35(22):7577–90

Lenski RE, Rose MR, Simpson SC et al. (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am Nat 138:1315–1341

Liu M, Durfee T, Cabrera JE et al. (2005) Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. J Biol Chem 280(16):15921–7

Patil KR, Rocha I, Forster J et al. (2005) Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinformatics 6:308

Pharkya P, Burgard AP, Maranas CD (2003) Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. Biotechnol Bioeng 84(7):887–99

Pharkya P, Burgard AP, Maranas CD (2004) OptStrain: a computational framework for redesign of microbial production systems. Genome Res 14(11):2367–76

Pramanik J, Keasling JD (1997) Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. Biotechnol Bioeng 56(4):398–421

Reed JL, Palsson BO (2003) Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. J Bacteriol 185(9):2692–9

Reed JL, Palsson BO (2004) Genome-scale *in silico* models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res 14(9):1797–805

Sauer U (2004) High-throughput phenomics: experimental methods for mapping fluxomes. Curr Opin Biotechnol 15(1):58–63

Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 99(23):15112–7

Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol 60(10):3724–31