Constraints-based models: Regulation of Gene Expression Reduces the Steady-state Solution Space

Markus W. Covert† and Bernhard O. Palsson*†

†*Department of Bioengineering, University of California, San Diego*, 9500 *Gilman Drive, La Jolla, CA* 92093-0412, *U.S.A.*

Constraints-based models have been effectively used to analyse, interpret, and predict the function of reconstructed genome-scale metabolic models. The first generation of these models used "hard" non-adjustable constraints associated with network connectivity, irreversibility of metabolic reactions, and maximal flux capacities. These constraints restrict the allowable behaviors of a network to a convex mathematical solution space whose edges are extreme pathways that can be used to characterize the optimal performance of a network under a stated performance criterion. The development of a second generation of constraints-based models by incorporating constraints associated with regulation of gene expression was described in a companion paper published in this journal, using flux-balance analysis to generate time courses of growth and by-product secretion using a skeleton representation of core metabolism. The imposition of these additional restrictions prevents the use of a subset of the extreme pathways that are derived from the "hard" constraints, thus reducing the solution space and restricting allowable network functions. Here, we examine the reduction of the solution space due to regulatory constraints using extreme pathway analysis. The imposition of environmental conditions and regulatory mechanisms sharply reduces the number of active extreme pathways. This approach is demonstrated for the skeleton system mentioned above, which has 80 extreme pathways. As regulatory constraints are applied to the system, the number of feasible extreme pathways is reduced to between 26 and 2 extreme pathways, a reduction of between 67.5 and 97.5%. The method developed here provides a way to interpret how regulatory mechanisms are used to constrain network functions and produce a small range of physiologically meaningful behaviors from all allowable network functions.

Introduction

The recent torrent of genome annotations and subsequent metabolic reconstructions has led to the development of mathematical modeling approaches to analyse the integrated behavior of microbial cells on a genome scale (Tomita *et al.*, 1999; Gombert & Nielsen, 2000; Varner, 2000; Covert *et al.*, 2001a). A constraints-based approach to studying metabolic models *in silico* has proven effective in the analysis of these genome-scale models (Palsson, 2000). The statement of governing constraints defines a solution space within which optimal solutions can be found using linear optimization (Varma & Palsson, 1994a; Bonarius *et al.*, 1997; Edwards *et al.*, 1999). This constraints-based approach

*Corresponding author. Tel.: +1-858-534-5668; fax: +1-858-822-3120.
E-mail address: palsson@ucsd.edu (B.O. Palsson).

has been useful in generating hypotheses *in silico* which may be tested experimentally (Varma & Palsson, 1994b, c). The result thus far has been a surprising degree of correlation between the predictions of genome-scale models and independently obtained experimental data (Edwards & Palsson, 2000; Edwards *et al.*, 2001). The study of the characteristics of the solution space uses principles of convex analysis (Schilling *et al.*, 1999; Schilling & Palsson, 2000). "Extreme pathways" are calculated as edges of the solution space, where the optimal solution must lie (Schilling *et al.*, 2000a). Taken together, the formulation of "hard" constraints associated with network connectivity, reaction irreversibility, and maximal flux restrictions, has led to the formulation and testing of what might be called the first generation of constraints-based models of biochemical reaction networks.

A logical next step in adding to the functionality of these genome-scale models is to incorporate regulation of gene expression. Several approaches have been used to model metabolic and regulatory events, beginning with Boolean representation of genetic circuits (Thomas, 1973, 1991; Kaufman *et al.*, 1985; Kauffman, 1993; Thieffry & Thomas, 1995; Somogyi & Sniegoski, 1996), which is advantageously simple but lacks the ability to make quantitative predictions. On the other hand, deterministic approaches (Reich & Sel'kov, 1981; Fell, 1996; Heinrich & Schuster, 1996; Wong *et al.*, 1997; Stephanopoulos *et al.*, 1998) may be combined with Boolean logic (McAdams & Shapiro, 1995), fuzzy logic (Lee *et al.*, 1999), or cybernetic principles (Kompala *et al.*, 1986; Varner, 2000) to model regulatory and metabolic events. Stochastic modeling approaches have also been developed to account for the low concentrations, spatial isolation, and slow reaction rates which characterize metabolic reactions in a single cell (McAdams & Arkin, 1997, 1998, 1999). Both deterministic and stochastic modeling approaches have been limited by the lack of experimental methods to determine kinetic and other parameters, as well as the computational difficulties of developing these models on a genome-scale (Bailey, 2001).

The constraints-based approach to modeling microbial metabolism has recently been expanded to incorporate regulatory constraints

(Covert *et al.*, 2001b). Regulatory constraints differ from the rigid physico-chemical constraints in two important ways; they are (1) *self-imposed*, meaning that over time, evolution has selected the development of complex mechanisms to restrict the allowable behaviors of these organisms under various conditions; and (2) *time dependent*, in that the state of the external and internal environment at a given time point determines transcriptional activity. As a result of these two features, the effects of transcriptional regulation can be treated as temporary constraints on the metabolic system. These constraints reduce the size of the solution space and change its shape from one environmental condition to another (Fig. 1).

In this paper we continue the work described in an earlier manuscript published in this journal (Covert *et al.*, 2001b), where regulatory events and gene expression were described using Boolean logic equations. For a given environment, we determine the corresponding regulatory constraints (e.g. repression of gene transcription) and eliminate extreme pathways that are inconsistent with the imposed regulatory constraints. This procedure reduces the solution space and customizes it for the given environmental conditions. The inclusion of known regulatory mechanisms effectively moves us toward the formulation of second-generation constraints-based models of complex biochemical reaction networks: models that combine metabolic flux-balance formalism and regulation of gene expression.

Methods

SAMPLE METABOLIC NETWORK

A skeleton network of core metabolism was formulated earlier (Covert *et al.*, 2001b). It includes 20 reactions, seven of which are governed by regulatory logic. This network is a highly simplified representation of core metabolic processes (e.g. glycolysis, the pentose phosphate pathway, TCA cycle, fermentation pathways, amino acid biosynthesis and cell growth), along with corresponding regulation (e.g. catabolite repression, aerobic/anaerobic regulation, amino acid biosynthesis regulation and carbon storage regulation). A schematic of
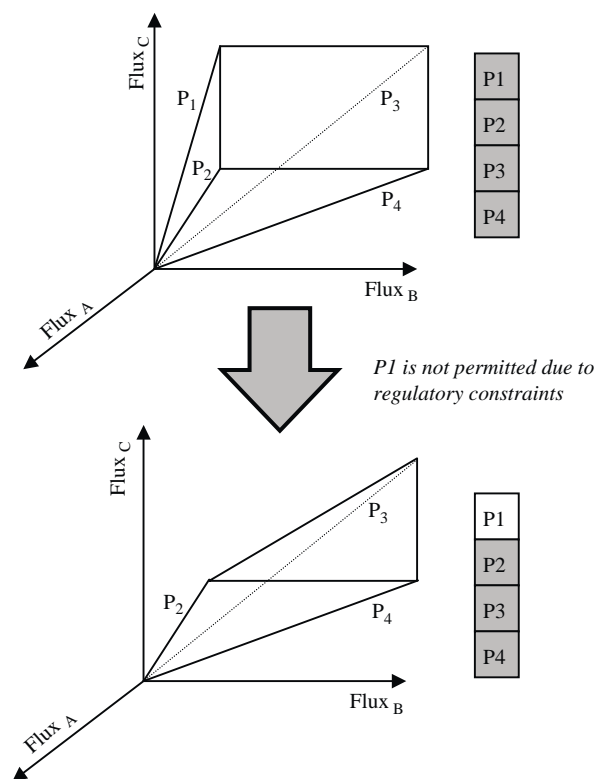
FIG. 1. Regulatory constraints reduce the steady-state solution space of a metabolic network. A solution space bounded by invariant constraints on the network is shown. Extreme pathways may be calculated as the unique, systemically independent generating vectors for the space. In the space on top, all of the pathways are considered available to the system (denoted by the highlighted gray boxes at right). Under certain environments however, regulatory constraints may cause one or more of the extreme pathways to be temporarily unavailable to the system, P1 in the case shown here. This results in a more restricted space with a reduced volume and/or dimension (bottom), corresponding to a metabolic network with fewer available behaviors.

this skeleton network is shown in Fig. 2, together with a table containing all of the relevant chemical reactions and regulatory rules which govern the transcriptional regulation.

EXTREME PATHWAYS

The steady-state behavior of a metabolic system may be characterized as

$$\boldsymbol{Sv} = \boldsymbol{0}, \qquad (1)$$

where $\mathbf{S}_{m \times n}$ (m metabolites $\times$ n fluxes) is the stoichiometric matrix for the network and $\boldsymbol{v}_{n \times 1}$ is a vector of the flux levels through each reaction in the system. Certain constraints on the system,

such as the thermodynamics of reactions or the constraints associated with enzyme capacity (e.g. maximum metabolite uptake and/or secretion rates) may be represented as upper and lower bounds on reaction flux levels. If reaction fluxes are also constrained to positive values by decomposing reversible reactions in **S** into forward and reverse components, a solution space may be geometrically defined for the system as a convex polyhedral cone in n-dimensional space (Schilling et al., 2000a). Such a space contains every possible steady-state flux distribution available to the system, subject to the given constraints.

A minimal set of generating vectors may be determined for the convex polyhedral cone using principles of convex analysis. This set is approximately analogous to a basis in linear algebra. The generating vectors span the null space of **S** and are the edges of the cone (Schilling & Palsson, 1998). Given that the cone represents the metabolic flux solution space at steady state, each generating vector or extreme ray corresponds to a particular pathway or active set of fluxes in the metabolic network and is termed an extreme pathway. Every possible steady-state flux distribution of a metabolic network may therefore be represented as a positive combination of extreme pathways:

$$C = \left\{ v : v = \sum_{i=1}^{k} \alpha_i \boldsymbol{p}_i, \alpha_i \geqslant 0, \forall i \right\}, \qquad (2)$$

where C is the polyhedral cone representing the metabolic network at steady state and $\mathbf{p}_i$ represents the extreme pathway vectors (Schilling et al., 1999). The algorithm used to generate extreme pathways has been described in detail (Schilling et al., 2000b).

BOOLEAN LOGIC DESCRIPTION OF ENVIRONMENTAL AND REGULATORY CONSTRAINTS

Boolean logic formalism can be used to represent regulatory constraints (Covert et al., 2001b) where a gene is considered in one of two states: active (ON) or inactive (OFF). Here, this formalism is also used to describe environmental constraints where a substrate is considered either present in (ON) or absent from (OFF) the external medium. The regulatory rules are listed

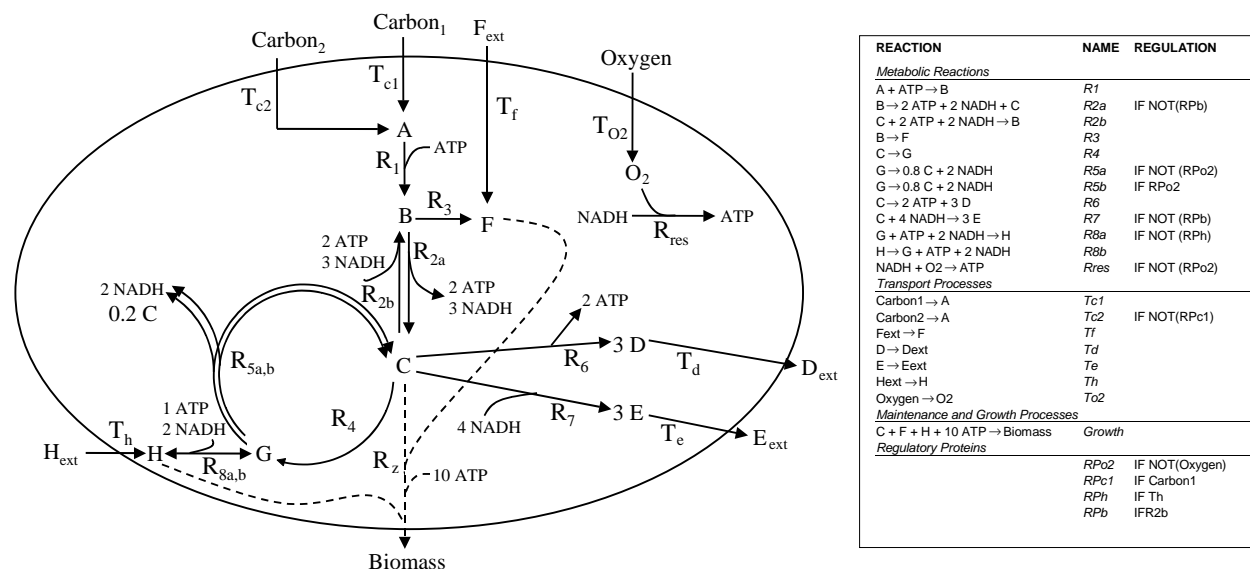| REACTION	NAME	REGULATION
Metabolic Reactions		
A + ATP → B	R1	
B → 2 ATP + 2 NADH + C	R2a	IF NOT(RPb)
C + 2 ATP + 2 NADH → B	R2b	
B → F	R3	
C → G	R4	
G → 0.8 C + 2 NADH	R5a	IF NOT (RPo2)
G → 0.8 C + 2 NADH	R5b	IF RPo2
C → 2 ATP + 3 D	R6	
C + 4 NADH → 3 E	R7	IF NOT (RPb)
G + ATP + 2 NADH → H	R8a	IF NOT (RPh)
H → G + ATP + 2 NADH	R8b	
NADH + O2 → ATP	Rres	IF NOT (RPo2)
Transport Processes		
Carbon1 → A	Tc1	
Carbon2 → A	Tc2	IF NOT(RPc1)
Fext → F	Tf	
D → Dext	Td	
E → Eext	Te	
Hext → H	Th	
Oxygen → O2	To2	
Maintenance and Growth Processes		
C + F + H + 10 ATP → Biomass	Growth	
Regulatory Proteins		
	RPo2	IF NOT(Oxygen)
	RPc1	IF Carbon1
	RPh	IF Th
	RPb	IFR2b

FIG. 2. A schematic of the simplified core metabolic network, together with a table containing the stoichiometry of the 20 metabolic reactions, seven of which are regulated by four regulatory proteins. This network is a highly simplified representation of core metabolic processes, including a glycolytic pathway with primary substrates carbon1 (C1) and carbon2 (C2), as well as a pentose phosphate pathway and a TCA cycle, through which "amino acid" H enters the system. Fermentation pathways as well as amino acid biosynthesis are also represented. The regulation modeled in this study includes simplified versions of catabolite repression (e.g. preferential uptake of C1 over C2), aerobic/anaerobic regulation, amino acid (H) biosynthesis regulation and carbon storage regulation, and is also listed. The *Growth* reaction is indicated by a dashed line.

in Fig. 2. Using these restrictions, an extreme pathway may be determined infeasible (1) if the external nutrient is absent from the external medium (e.g. pathways that have oxygen as an input would be inoperative in a simulation in an anaerobic environment); or (2) when expression of the gene responsible for producing a metabolic flux has been repressed (e.g. if extracellular carbon1 is present, then pathways which utilize T_{c2} are considered infeasible due to the repression of T_{c2} transcription by carbon1). Conversely, a pathway is feasible for a certain condition if it is consistent with all the applicable environmental and regulatory constraints.

PHENOTYPIC PHASE PLANE ANALYSIS

Phenotypic phase planes (PhPPs) are two-dimensional representations of the solution space (Edwards & Palsson, 1999; Schilling *et al.*, 2000a). Once the axes of the PhPP are set (generally uptake rates of two nutrients such as oxygen and a carbon source), flux maps which are optimal with respect to growth may be calculated for each point in the plane. The lines in a PhPP demarcate a change in the flux map;

these lines are determined using the shadow prices (the dual solution in linear programming). PhPP analysis has recently been used to demonstrate the optimal growth of *E. coli* on succinate and acetate minimal media (Edwards *et al.*, 2001).

Results

A total of 80 extreme pathways were calculated for the simplified metabolic system shown in Fig. 2. They are all shown graphically in the Appendix. The numbering of the extreme pathways is arbitrary and corresponds to the order in which the algorithm generates them. Given the five inputs to the metabolic network and representing these inputs using Boolean logic, considering each as ON if present or OFF if absent, there are a total of $2^5 = 32$ possible environments which may be recognized by the cell. These environments are listed in Table 1. For each environment, the transcription of several of the enzymes in the network may be restricted due to regulation. The constraints imposed on the metabolic system by both the substrates available to (i.e. the external environment) and the enzymes expressed in the cell (i.e. the internal

TABLE 1

A list of all the possible environments which can be recognized by the system shown in Fig. 1, of which six environments do not enable the cell to produce biomass

Environments					Repressed enzymes							Pathways	Pathway list
C1	C2	F	H	O2			R_{5b}		R_{8a}		T_{c2}	26	P2, P4, P5, P6, P8, P9, P10, P12, P29, P30, P31, P32, P33, P34, P35, P36, P37, P38, P45, P46, P47, P48, P49, P50, P51, P52 Detail in Fig. 4.
C1	C2	F	H			R_{5a}			R_{8a}	R_{res}	T_{c2}	10	P39, P40, P41, P42, P43, P44, P49, P50, P51, P52
C1	C2	F		O2			R_{5b}				T_{c2}	8	P29, P30, P33, P34, P45, P46, P49, P50
C1	C2	F				R_{5a}				R_{res}	T_{c2}	4	P41, P42, P49, P50
C1	C2		H	O2			R_{5b}		R_{8a}		T_{c2}	14	P2, P5, P6, P9, P10, P30, P31, P34, P35, P37, P46, P47, P50, P51
C1	C2		H			R_{5a}			R_{8a}	R_{res}	T_{c2}	5	P39, P42, P43, P50, P51
C1	C2			O2			R_{5b}				T_{c2}	4	P30, P34, P46, P50 Detail in Fig. 3
C1	C2					R_{5a}				R_{res}	T_{c2}	2	P42, P50
C1		F	H	O2			R_{5b}		R_{8a}		T_{c2}	26	P2, P4, P5, P6, P8, P9, P10, P12, P29, P30, P31, P32, P33, P34, P35, P36, P37, P38, P45, P46, P47, P48, P49, P50, P51, P52
C1		F	H			R_{5a}			R_{8a}	R_{res}	T_{c2}	10	P39, P40, P41, P42, P43, P44, P49, P50, P51, P52
C1		F		O2			R_{5b}				T_{c2}	8	P29, P30, P33, P34, P45, P46, P49, P50
C1		F				R_{5a}				R_{res}	T_{c2}	4	P41, P42, P49, P50
C1			H	O2			R_{5b}		R_{8a}		T_{c2}	14	P2, P5, P6, P9, P10, P30, P31, P34, P35, P37, P46, P47, P50, P51
C1			H			R_{5a}			R_{8a}	R_{res}	T_{c2}	5	P39, P42, P43, P50, P51
C1				O2			R_{5b}				T_{c2}	4	P30, P34, P46, P50
C1						R_{5a}				R_{res}	T_{c2}	2	P42, P50
	C2	F	H	O2			R_{5b}		R_{8a}			26	P3, P4, P5, P7, P8, P9, P11, P12, P57, P58, P59, P60, P61, P62, P63, P64, P65, P66, P73, P74, P75, P76, P77, P78, P79, P80
	C2	F	H			R_{5a}			R_{8a}	R_{res}		10	P67, P68, P69, P70, P71, P72, P77, P78, P79, P80
	C2	F		O2			R_{5b}					8	P57, P58, P61, P62, P73, P74, P77, P78
	C2	F				R_{5a}				R_{res}		4	P69, P70, P77, P78
	C2		H	O2			R_{5b}		R_{8a}			14	P3, P5, P7, P9, P11, P58, P59, P62, P63, P65, P74, P75, P78, P79
	C2		H			R_{5a}			R_{8a}	R_{res}		5	P67, P70, P71, P78, P79
	C2			O2			R_{5b}					4	P58, P62, P74, P78
	C2					R_{5a}				R_{res}		2	P70, P78
		F	H	O2			R_{5b}		R_{8a}			5	P4, P5, P8, P9, P12
		F	H		R_{2a}	R_{5a}		R_7	R_{8a}	R_{res}		0	
		F		O2			R_{5b}					0	
		F				R_{5a}				R_{res}		0	
			H	O2	R_{2a}		R_{5b}	R_7	R_{8a}			2	P5, P9
			H		R_{2a}	R_{5a}		R_7	R_{8a}	R_{res}		0	
				O2			R_{5b}					0	
												0	

Note: For each environment, there is a set of enzymes which are repressed under the given environmental conditions. The extreme pathways which remain feasible even under the combination of environmental and regulatory constraints are listed. For a schematic of each of the pathways, see the Appendix. C1 = carbon1, C2 = carbon2, O2 = oxygen. Two of the environments are shown in more detail in Figs 3 and 4 and are labeled correspondingly.

environment) reduce the number of extreme pathways accessible to the cell at a given time.

Several interesting observations may be made from Table 1: first, 21 extreme pathways (P1, P13–P28 and P53–56, enclosed by a box in the Appendix), although stoichiometrically feasible, are always impossible due to regulatory constraints. Pathways P13–28 and P53–56 are infeasible due to the fact that R_{res} is only expressed aerobically while R_{5b} is only expressed

anaerobically. Therefore, any pathway that includes a flux through both R_{res} and R_{5b} is eliminated. Similarly, pathways P1 and P13 are eliminated because a flux through R_{2b} activates a regulatory protein that represses transcription of R_7. Therefore, R_{2b} and R_7 cannot both be expressed together. Note that P13 is infeasible in either case.

Another interesting observation from Table 1 is that several environments have identical or near-identical sets of available extreme pathways. For example, the environment containing carbon1 (C1), carbon2 (C2), F, H, and oxygen (O2) has an identical extreme pathway list to that for the environment containing C1, F, H, and O2. The reason is that T_{c2}, the transport flux for C2, is repressed in the presence of C1. Furthermore, the extreme pathway list for the environment containing C2, F, H, and O2 is similar to the pathway lists for the previously mentioned environments, different only in that the pathways which utilize the T_{c1} flux in the former pathway list are replaced by pathways which utilize the T_{c2} flux in the latter.

Finally, Table 1 shows that the highest number of extreme pathways available to the cell is 26; the lowest is 2, corresponding to a reduction in the number of available extreme pathways between 67.5 and 97.5%. A relatively simple dual-substrate environment and the most complex environment were examined in more detail to more closely investigate the effect of regulation on available pathways.

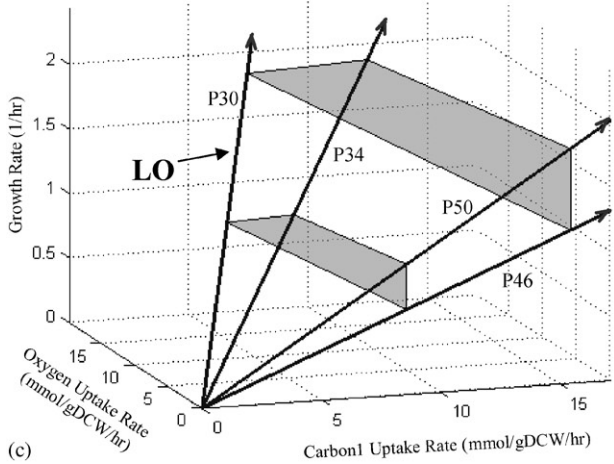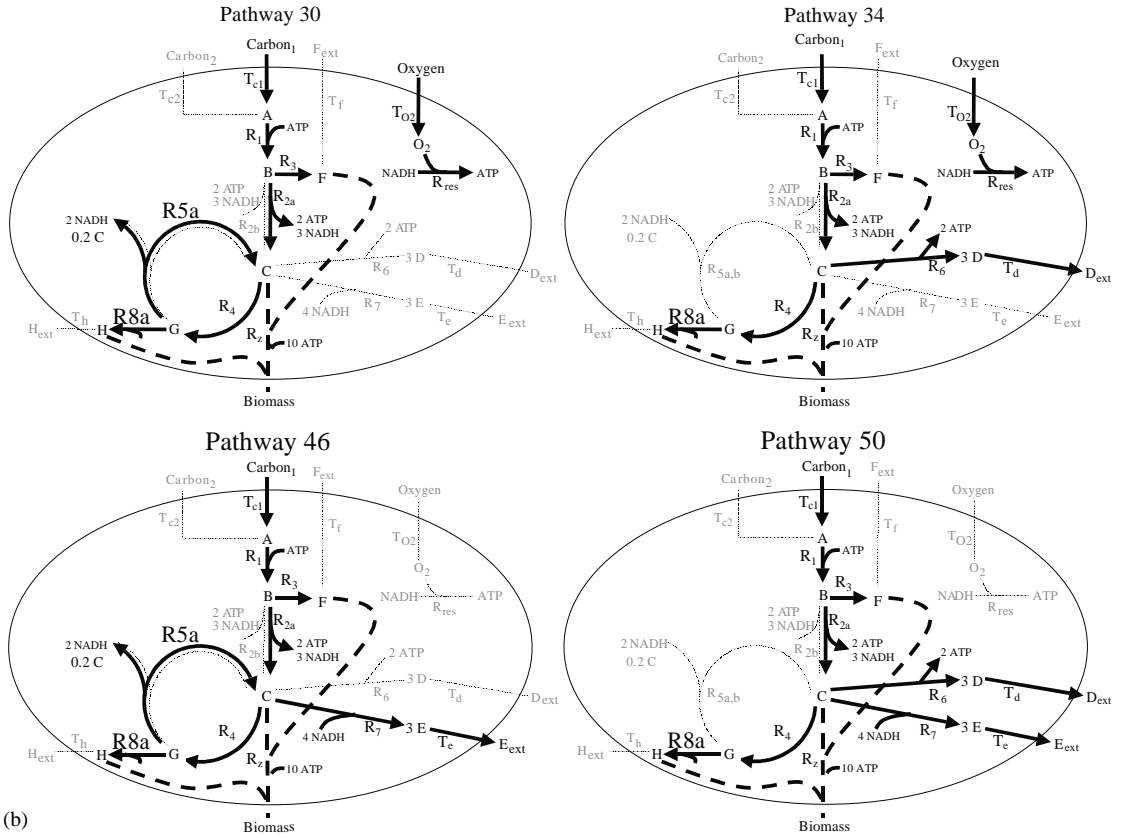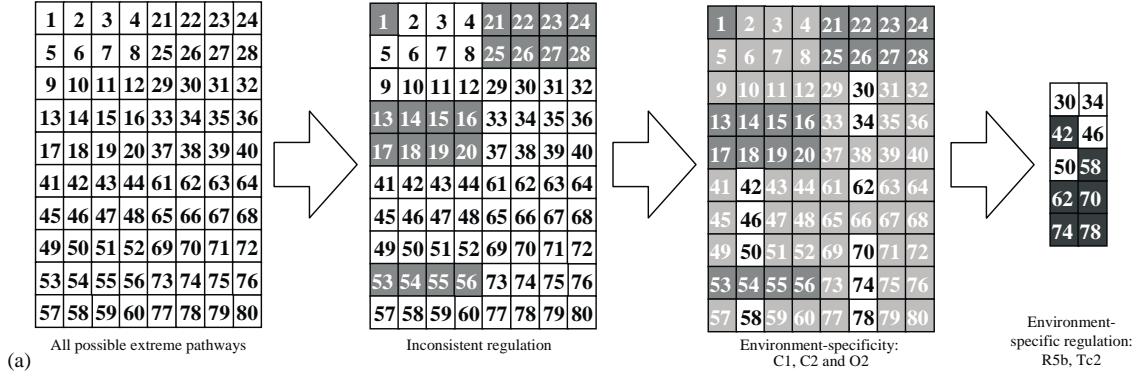### EXAMPLE 1: GROWTH ON C1, C2, AND O2

The metabolic network was given C1, C2, and O2 as inputs and allowed biomass, D_{ext} and E_{ext} as outputs. These conditions reduce the number of extreme pathways available to the cell (Fig. 3). Initially, all 80 pathways are considered, and are represented schematically in Fig. 3(a). Twenty-one of the extreme pathways are always restricted by regulation, as discussed earlier; the boxes representing these pathways are darkened in gray. By considering only the pathways with appropriate inputs and outputs based on the cell environment, 49 more pathways are eliminated (shaded in light gray). Of the ten remaining pathways, six are inconsistent with the given regulation (C1 catabolite repression of the C2 transport protein T_{c2} or regulation due to the aerobic environment) (shown in black) and the flux maps for the remaining four extreme pathways are shown in Fig. 3(b).

The resulting solution space is projected in three dimensions (C1 uptake rate, oxygen uptake rate and growth rate), as shown in Fig. 3(c), with the four feasible extreme pathways labeled. All the volume defined by these edges is accessible to the cell. The corresponding range of growth and uptake rates can be attained by the cell under these conditions.

The two-dimensional PhPP for growth on C1 and O2 is shown in Fig. 3(d). This PhPP has two feasible phases between the lines shown, which represent the four extreme pathways available to the system. For this case, the two-dimensional projection of the extreme pathways lies on the region boundaries of the PhPP; pathways 46 and 50 are both fermentative and therefore overlap in the PhPP (oxygen uptake rate = 0). Pathway 30 is the line of optimality (Edwards et al., 2001, 2002) as none of the carbon is lost in secretion of by-products; pathway 34 includes secretion of D_{ext} and therefore gives a lower biomass yield (inset) than pathway 30.

FIG. 3. Extreme pathway reduction by constraints, using growth of the sample system in a C1 and C2 aerobic medium as an example. (a) The 80 total extreme pathways calculated for the system are represented by a grid, arranged to correspond with the pathway figures in the Appendix. The number of pathways is reduced by 21 when pathways are removed which are always inconsistent with the regulatory rules (dark gray), then by 49 due to the specific environmental constraints (light gray), then by 6 as the regulation corresponding to the specific environment is considered (black). The four remaining pathways which are consistent with all the regulatory and environmental constraints are shown schematically in (b), where the thick dark arrows represent active fluxes. (c) The solution space of the system, projected on a three-dimensional space, with the pathways and the line of optimality (the pathway with the greatest growth yield) noted. (d) A two-dimensional projection of the space, superimposed on a two-dimensional phenotypic phase plane for the C1 and oxygen uptake. The region at left lies outside of the space and is therefore infeasible.

(a)

(b)

(c)

(d)
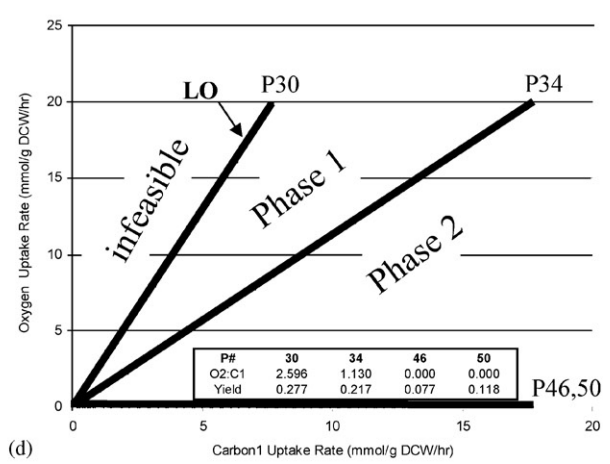
The allowable extreme pathways for growth on a medium containing C1, C2, F, and H in an aerobic environment were determined. For this case, the environment offers no restrictions—all possible inputs are available—and therefore the restriction of the solution space by elimination of extreme pathways is entirely due to regulatory effects. From Table 1 it is shown that R_{2a}, R_{5b}, R_7, R_{8a}, and T_{c2} are constrained to zero by the regulatory rules. Consequently, the 33 corresponding extreme pathways were removed from the solution space, resulting in a list of 26 available pathways which may be used by the cell under these conditions [Fig. 4(b)]. These remaining pathways, normalized by the total input of C1, C2, F, and H for comparison purposes, are shown in Fig. 4(a).

The 26 allowable pathways shown in Fig. 4(a) may be grouped by biomass yield. In the top two sets, all pathways are optimal or very near-optimal in terms of biomass yield, with no by-product secretion. The middle pathway sets involve secretion of either D or E while biomass is generated, and the bottom set of pathways represents purely fermentative use of the network. Again, it is seen that even with a higher number of allowable extreme pathways, the actual degree of variation in possible network behavior is surprisingly small once regulatory constraints are taken into account. The multiplicity of extreme pathways with near-optimal biomass yield gives the metabolic network robustness characteristics as the cell has many alternatives with nearly the same outcome.

The reduced solution space is projected in three and two dimensions, as shown in Fig. 4(c) and (d), respectively. The dimensions are the same as in Fig. 3, with the exception of the C1 uptake rate axis, which has been replaced by a normalized axis of all possible routes for substrate uptake (C1 + C2 + F + H). The three-dimensional projection is bounded by the solid black vectors, many of which overlap in the projection (note that these extreme pathways do not actually overlap; however, in a projection of the high-dimensional solution space onto a lower dimension, they share certain characteristics such as the relationship between growth, substrate uptake rate, and oxygen uptake rate). Five pathways (P4, P9, P10, P34, P35, shown with dashed black lines) lie inside the three-dimensional projection. The structure of the solution space is more complex as compared to the simple growth condition analysed in Example 1.

The solution spaces are compared on the same axis to illustrate the concept of solution space reduction further [Fig. 5(a)]. The solution space discussed in Example 1 is a subset of the space defined by the complex medium. Figure 5(b) and (c) shows cross sections of the space where combined (C1 + C2 + F + H) uptake and oxygen uptake are set at a constant rate of $5 \, \text{mmol}^{-1} \, \text{g} \, \text{DCW} \, \text{hr}^{-1}$, respectively. The line of optimality, P30 in Example 1, is shifted to any of three pathways, P8, P29, and P32, in Example 2. These three pathways obtain similar growth yields but exhibit different behaviors in terms of substrate uptake. With slightly smaller growth yields, extreme pathways P5, P6, P30, and P31 also bound the space and are in close proximity to the optimal pathways. The cross section where the oxygen uptake rate is constant (bottom) is also unbounded, as shown by the dotted lines.
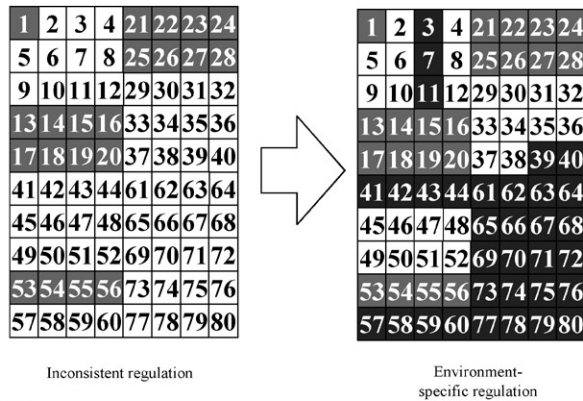
Discussion

Constraints-based models have been formulated to analyse, interpret, and predict cell behavior from reconstructed metabolic networks under given environmental conditions. They have given useful and surprisingly predictive results (Edwards & Palsson, 2000; Edwards *et al.*,
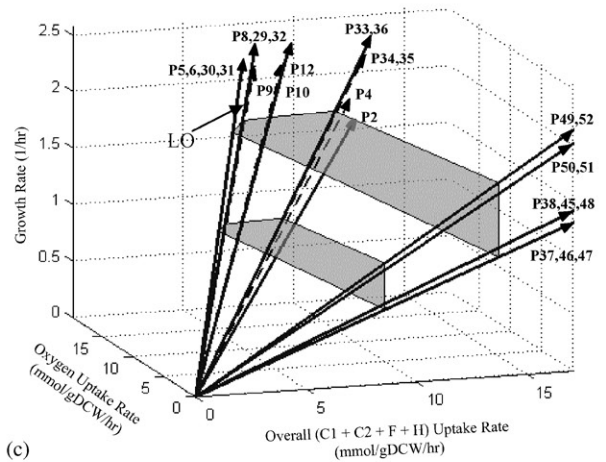
FIG. 4. Aerobic growth of the sample system in a complex medium. (a) A table summarizing the 26 extreme pathways which are feasible for this enviroment. The four boxes indicate significant pathway groupings based on by-product secretion and yield. Groups 1 and 2 are the sets of pathways with the highest growth yield. (b) Extreme pathway reduction in the complex medium, similar in format to Fig. 3(a). Note that no environmental constraints are imposed upon the space in this example, and reduction of the solution space is therefore only due to regulatory constraints. Here the pathways have been normalized by the combined uptake of C1, C2, F, and H in order to be represented together in the plots shown in (c) and (d), which are also similar in format to their counterparts in Fig. 3(c) and (d), respectively. In (c), five extreme pathways lie inside the three-dimensional projection (dashed lines).

P#											Pathway									
8				0.283	Fext	+	0.717	Hext	+	2.391	Oxygen								0.283	Biomass
29	0.717	Carbon1	+	0.283	Fext				+	2.391	Oxygen								0.283	Biomass
32	0.435	Carbon1	+	0.283	Fext	+	0.283	Hext	+	2.391	Oxygen								0.283	Biomass
5							1.000	Hext	+	2.596	Oxygen								0.277	Biomass
6	0.277	Carbon1				+	0.723	Hext	+	2.596	Oxygen								0.277	Biomass
30	1.000	Carbon1							+	2.596	Oxygen								0.277	Biomass
31	0.723	Carbon1				+	0.277	Hext	+	2.596	Oxygen								0.277	Biomass
9							1.000	Hext	+	2.390	Oxygen	0.146	Dext				+		0.268	Biomass
12				0.262	Fext	+	0.738	Hext	+	1.905	Oxygen	0.357	Dext				+		0.262	Biomass
10	0.252	Carbon1				+	0.748	Hext	+	1.985	Oxygen	0.435	Dext				+		0.252	Biomass
33	0.773	Carbon1	+	0.227	Fext				+	1.091	Oxygen	0.955	Dext				+		0.227	Biomass
36	0.545	Carbon1	+	0.227	Fext	+	0.227	Hext	+	1.091	Oxygen	0.955	Dext				+		0.227	Biomass
34	1.000	Carbon1							+	1.130	Oxygen	1.043	Dext				+		0.217	Biomass
35	0.783	Carbon1				+	0.217	Hext	+	1.130	Oxygen	1.043	Dext				+		0.217	Biomass
4				0.188	Fext	+	0.813	Hext	+	1.250	Oxygen				0.938	Eext	+		0.188	Biomass
2	0.170	Carbon1				+	0.830	Hext	+	1.208	Oxygen				1.075	Eext	+		0.170	Biomass
49	0.875	Carbon1	+	0.125	Fext							0.750	Dext	+	1.125	Eext	+		0.125	Biomass
52	0.750	Carbon1	+	0.125	Fext	+	0.125	Hext				0.750	Dext	+	1.125	Eext	+		0.125	Biomass
50	1.000	Carbon1										0.794	Dext	+	1.147	Eext	+		0.118	Biomass
51	0.882	Carbon1				+	0.118	Hext				0.794	Dext	+	1.147	Eext	+		0.118	Biomass
38	0.357	Carbon1	+	0.083	Fext	+	0.560	Hext									+	1.964	Eext	+	0.083	Biomass
45	0.917	Carbon1	+	0.083	Fext												+	1.964	Eext	+	0.083	Biomass
48	0.833	Carbon1	+	0.083	Fext	+	0.083	Hext									+	1.964	Eext	+	0.083	Biomass
37	0.429	Carbon1				+	0.571	Hext									+	2.011	Eext	+	0.077	Biomass
46	1.000	Carbon1															+	2.011	Eext	+	0.077	Biomass
47	0.923	Carbon1				+	0.077	Hext									+	2.011	Eext	+	0.077	Biomass

Group 1: rows 8, 29, 32.
Group 2: rows 5, 6, 30, 31.

(a)

Inconsistent regulation → Environment-specific regulation

(b)

(c)

(d)

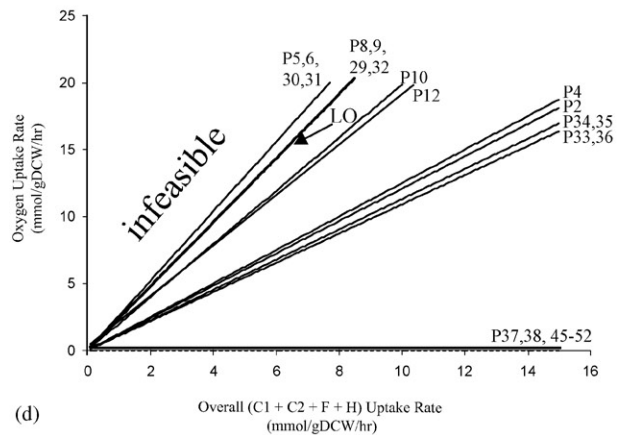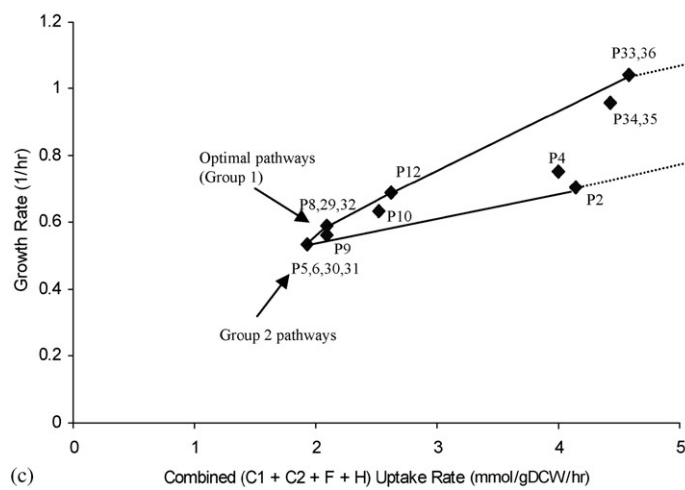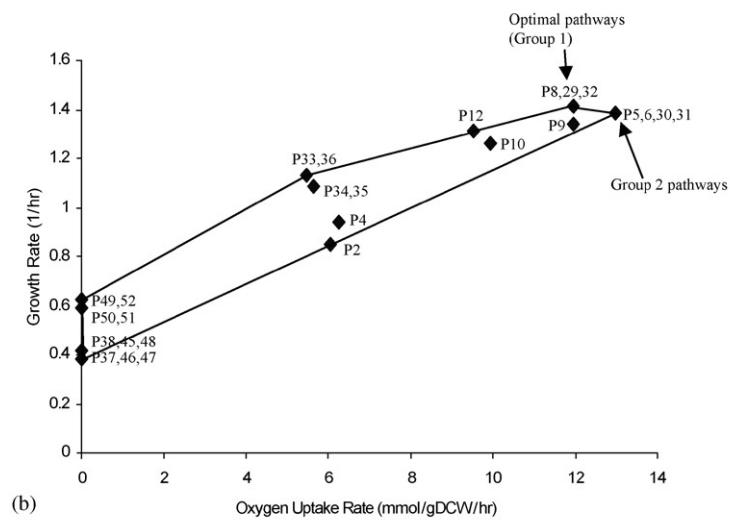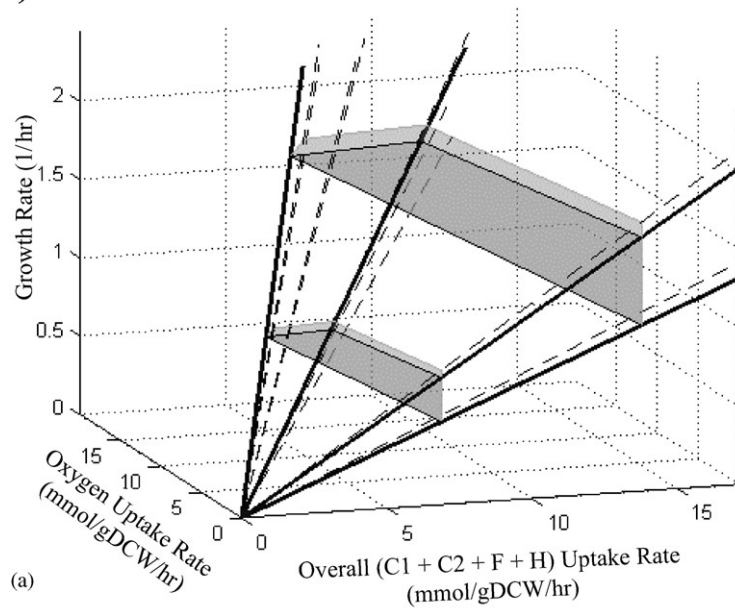(a)

(b)

(c)

2001). The constraints imposed on the reconstructed genome-scale metabolic networks thus far have been the invariant constraints associated with stoichiometry, thermodynamics, and enzyme capacity (Palsson, 2000). These may be considered as first-generation constraints-based models of metabolism.

Here we continue to develop second-generation constraints-based models by including the temporary and self-imposed constraints associated with regulation of gene expression and environmental conditions that further constrain allowable functions of the network. A Boolean logic formalism was used to describe regulation and derived logic regulatory rules may be used in conjunction with extreme pathway analysis to examine the effects of regulation on the allowable range of network function. The most important finding of this study is that the imposition of regulatory constraints significantly reduces the size of the solution space. For the skeleton core metabolic network used, the number of extreme pathways was reduced from 80 to as few as two, in some cases, as a result of the imposition of relatively simple environmental and regulatory constraints. In a simulated rich medium, the skeleton network was unconstrained with respect to the environment and yet 67.5% of the pathways were eliminated by regulation. This large reduction in the solution space seems to indicate that despite the complex interaction of many genes to produce an integrated cellular function, simple behaviors can relatively easily be selected by the cell.

Another noteworthy observation that can be made from these results regards the extreme pathways that remain after regulatory constraints are applied. In both examples given here, many fermentative pathways were left available to the network despite the presence of oxygen. Although these extreme pathways may be unused, the ability to implement them without the delays associated with transcription and translation would give the cell the ability to rapidly adapt to oxygen deprivation.

Finally, the close proximity and even overlap of optimal and near-optimal extreme pathways to one another in the three- and two-dimensional projections of the solution space suggest that the system has numerous means by which to obtain its growth objectives. Extreme pathway analysis of metabolic networks at the genome scale has indicated a high degree of pathway redundancy (Papin *et al.*, 2002). Although regulatory constraints greatly reduce such redundancy as has been shown, it seems that a certain amount of flexibility is beneficial to the cell. Such flexibility may be useful to an organism, for example, in colonizing diverse and changing environments.

In summary, the application of transcriptional regulatory constraints to metabolic networks results in a large reduction of behaviors available to the network under a given environment. The present study leads to the formulation of a second generation of constraints-based models that can be used to interpret how regulation is used to keep a restricted portion of the total solution space accessible, and thereby, by the process of elimination, force a particular set of phenotypic behaviors to be expressed.

REFERENCES

BAILEY, J. E. (2001). Complex biology with no parameters. *Nature Biotechnol.* **19,** 503–504.

BONARIUS, H. P. J., SCHMID, G. & TRAMPER, J. (1997). Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotechnol.* **15,** 308–314.

COVERT, M. W., SCHILLING, C. H., FAMILI, I., EDWARDS, J. S., GORYANIN, I. I., SELKOV, E. & PALSSON, B. O. (2001a). Metabolic Modeling of Microbial Strains in silico. *Trends Biochem. Sci.* **26,** 179–186.

FIG. 5. Diagrammatic representation of solution space reduction. (a) Schematic of the two solution spaces superimposed on one another. The solution space of Example 1 is a subspace of the space of the complex medium. (b) Cross section of the solution space on the oxygen uptake rate–growth rate plane where the combined (C1 + C2 + F + H) uptake is set at $5 \, \text{mmol}^{-1} \, \text{g DCW hr}^{-1}$. (c) Cross section of the solution space on the combined (C1 + C2 + F + H) uptake rate–growth rate plane where the oxygen uptake rate is set at $5 \, \text{mmol}^{-1} \, \text{g DCW hr}^{-1}$.

COVERT, M. W., SCHILLING, C. H. & PALSSON, B. O. (2001b). Regulation of gene expression in flux balance models of metabolism. *J. theor. Biol.* **213,** 73–88, doi:10.1006/jtbi.2001.2405.

EDWARDS, J. & PALSSON, B. (1999). Properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.* **274,** 17 410–17 416.

EDWARDS, J. S. & PALSSON, B. O. (2000). The *E. coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl Acad. Sci.* **97,** 5528–5533.

EDWARDS, J. S., RAMAKRISHNA, R., SCHILLING, C. H. & PALSSON, B. O. (1999). Metabolic flux balance analysis. In: *Metabolic Engineering* (Lee, S. Y. & Papoutsakis, E. T., eds), pp. 13–57. New York: Marcel Dekker.

EDWARDS, J. S., IBARRA, R. U. & PALSSON, B. O. (2001). In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol.* **19,** 125–130.

EDWARDS, J. S., RAMAKRISHNA, R. & PALSSON, B. O. (2002). Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol. Bioeng.* **77,** 27–36.

FELL, D. (1996). *Understanding the Control of Metabolism.* London: Portland Press.

GOMBERT, A. K. & NIELSEN, J. (2000). Mathematical modelling of metabolism. *Curr. Opin. Biotechnol.* **11,** 180–186.

HEINRICH, R. & SCHUSTER, S. (1996). *The Regulation of Cellular Systems.* New York: Chapman & Hall.

KAUFFMAN, S. A. (1993). *The Origins of Order.* New York: Oxford University Press.

KAUFMAN, M., URBAIN, J. & THOMAS, R. (1985). Towards a logical analysis of the immune response. *J. Theor. Biol.* **114,** 527–561.

KOMPALA, D. S., RAMKRISHNA, D., JANSEN, N. B. & TSAO, G. T. (1986). Investigation of Bacterial Growth on Mixed Substrates. Experimental Evaluation of Cybernetic Models. *Biotechnol. Bioeng.* **28,** 1044–1056.

LEE, B., YEN, J., YANG, L. & LIAO, J. C. (1999). Incorporating qualitative knowledge in enzyme kinetic models using fuzzy logic. *Biotechnol. Bioeng.* **62,** 722–729.

MCADAMS, H. H. & ARKIN, A. (1997). Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. U.S.A.* **94,** 814–819.

MCADAMS, H. H. & ARKIN, A. (1998). Simulation of prokaryotic genetic circuits. *Ann. Rev. Biophys. Biomol. Struct.* **27,** 199–224.

MCADAMS, H. H. & ARKIN, A. (1999). It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15,** 65–69.

MCADAMS, H. H. & SHAPIRO, L. (1995). Circuit simulation of genetic networks. *Science* **269,** 651–656.

PALSSON, B. O. (2000). The challenges of in silico biology. *Nature Biotechnol.* **18,** 1147–1150.

PAPIN, J. A., Price, N. D., EDWARDS, J. S. & PALSSON, B. O. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J. theor. Biol.* **215,** 67–82, doi:10.1006/jtbi.2001.2496.

REICH, J. G. & SEL'KOV, E. E. (1981). *Energy Metabolism of the Cell.* New York: Academic Press.

SCHILLING, C. H. & PALSSON, B. O. (1998). The underlying pathway structure of biochemical reaction networks. *Proc. Natl Acad. Sci. U.S.A.* **95,** 4193–4198.

SCHILLING, C. H. & PALSSON, B. O. (2000). Assesment of the Metabolic Capabilities of Haemophilus influenzae Rd through a Genome-scale Pathway Analysis. *J. theor. Biol.* **203,** 249–283, doi:10.1006/jtbi.2000.1088.

SCHILLING, C. H., SCHUSTER, S., PALSSON, B. O. & HEINRICH, R. (1999). Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* **15,** 296–303.

SCHILLING, C. H., EDWARDS, J. S., LETSCHER, D. & PALSSON, B. O. (2000a). Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* **71,** 286–306.

SCHILLING, C. H. LETSCHER, D. & PALSSON, B. O. (2000b). Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective. *J. theor. Biol.* **203,** 229–248, doi:10.1006/jtbi.2000.1073.

SOMOGYI, R. & SNIEGOSKI, C. A. (1996). Modeling the complexity of genetic networks: understanding multigenic and pleitropic regulation. *Complexity* **1,** 45–63.

STEPHANOPOULOS, G., ARISTODOU, A. & NIELSEN, J. (1998). *Metabolic Engineering.* New York: Academic Press.

THIEFFRY, D. & THOMAS, R. (1995). Dynamical behaviour of biological regulatory networks—II. Immunity control in bacteriophage lambda. *Bull. Math. Biol.* **57,** 277–297.

THOMAS, R. (1973). Boolean formalization of genetic control circuits. *J. theor. Biol.* **42,** 563–585.

THOMAS, R. (1991). Regulatory networks seen as asynchronous automata: a logical description. *J. theor. Biol.* **153,** 1–23.

TOMITA, M., HASHIMOTO, K., TAKAMASHI, K., SHIMIZU, T., MATSUZAKI, Y., MIYOSHI, F., SAITO, K., TANIDA, S., YUGI, K., VENTER, J. C. & HUTCHISON, C. A. (1999). E-CELL: software environment for whole-cell simulation. *Bioinformatics* **15,** 72–84.

VARMA, A. & PALSSON, B. O. (1994a). Metabolic flux balancing: basic concepts, scientific and practical use. *Bio. Technol.* **12,** 994–998.

VARMA, A. & PALSSON, B. O. (1994b). Predictions for oxygen supply control to enhance population stability of engineered production strains. *Biotechnol. Bioeng.* **43,** 275–285.

VARMA, A. & PALSSON, B. O. (1994c). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *E. coli* W3110. *Appl. Environ. Microbiol.* **60,** 3724–3731.

VARNER, J. D. (2000). Large-scale prediction of phenotype: concept. *Biotechnol. Bioeng.* **69,** 664–678.

WONG, P. GLADNEY, S. & KEASLING, J. D. (1997). Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol. Prog.* **13,** 132–143.

APPENDIX

The flux distribution maps for all 80 of the extreme pathways calculated for the sample network are shown. The active fluxes are shown with thick dark lines, inactive fluxes are denoted by thin light dotted lines and the biomass flux is shown with a dashed line. Note that reactions R_{5a} and R_{5b} are isozymes. The 21 extreme pathways which are infeasible under every environmental condition are enclosed in boxes.
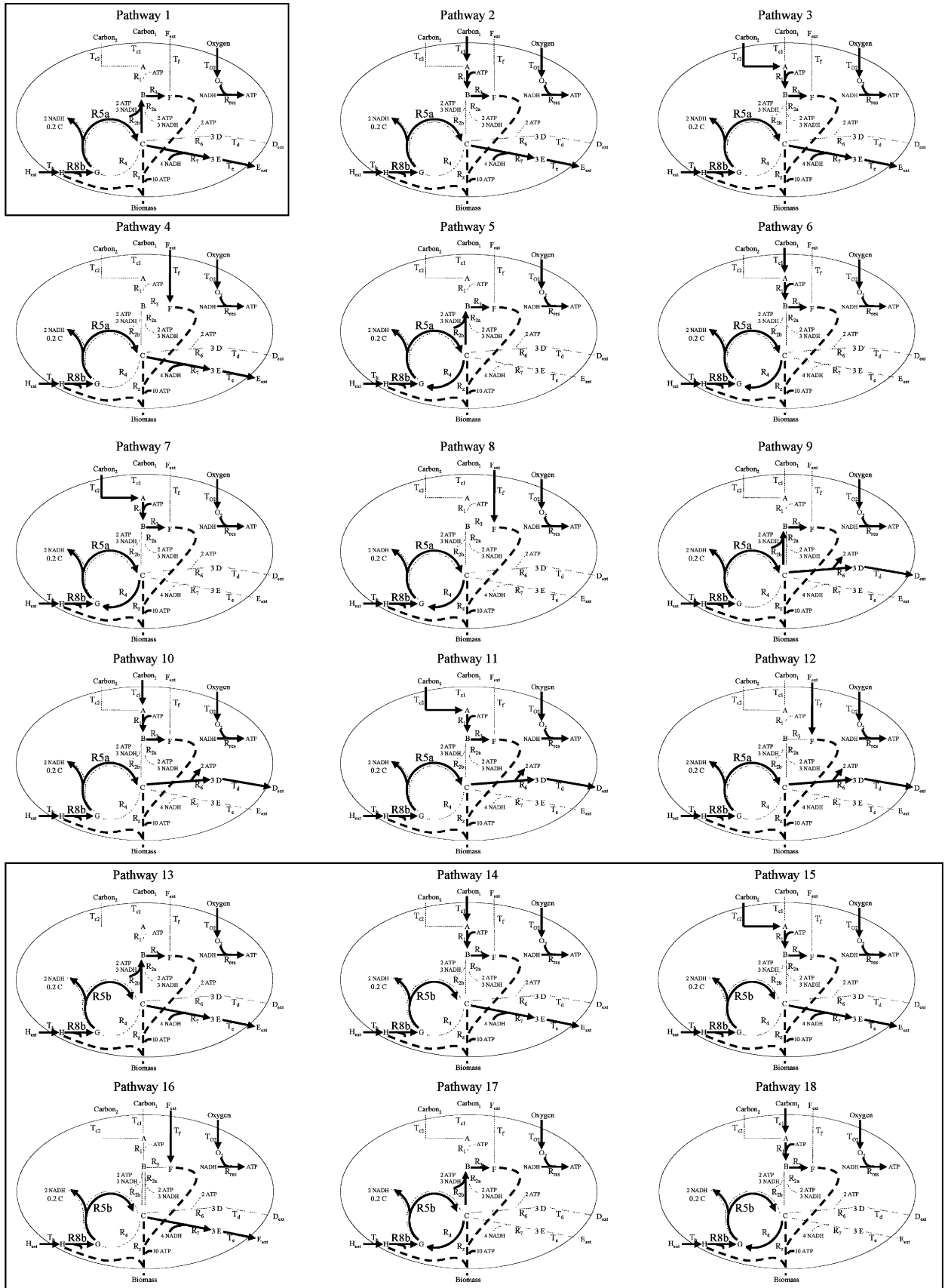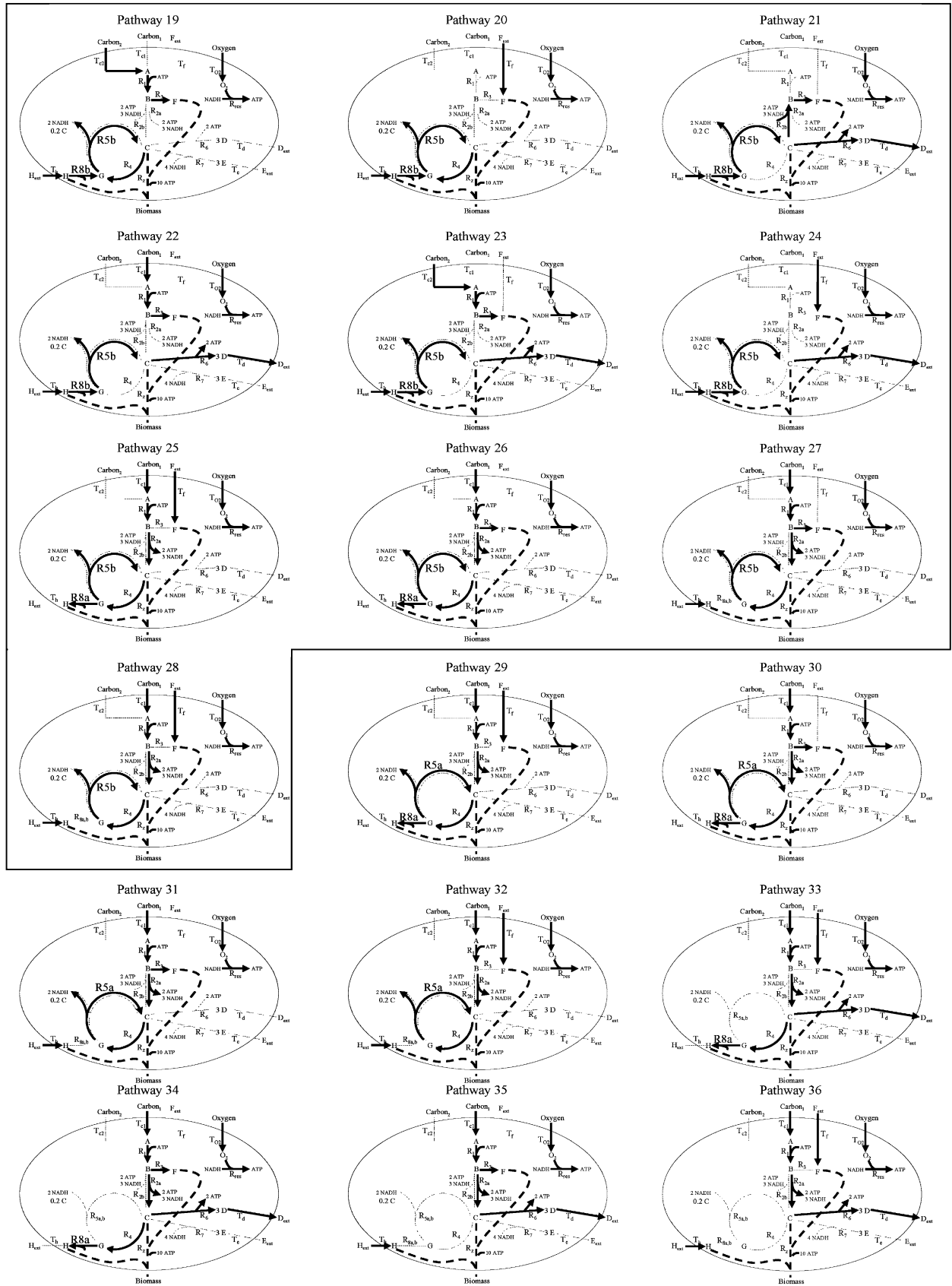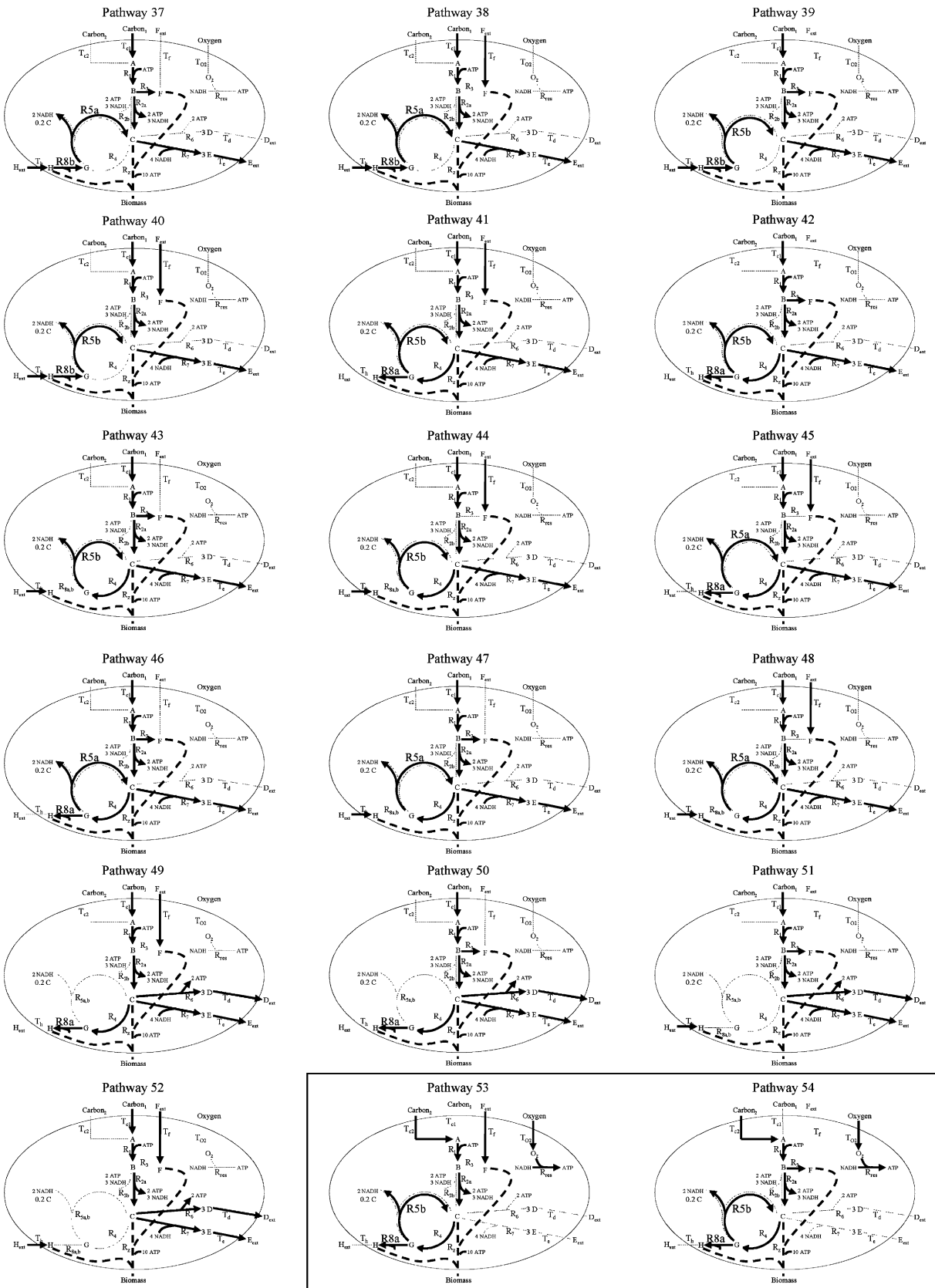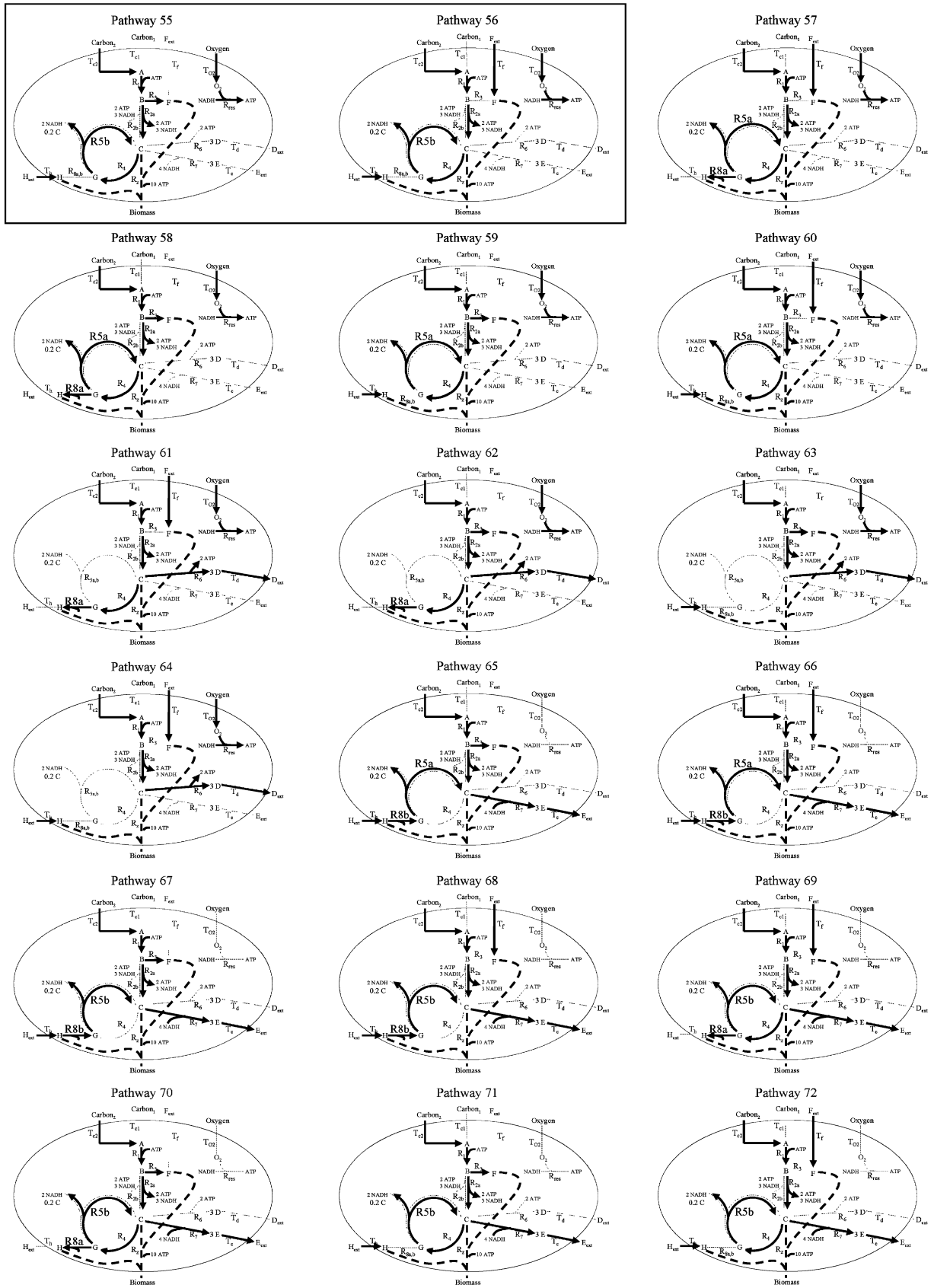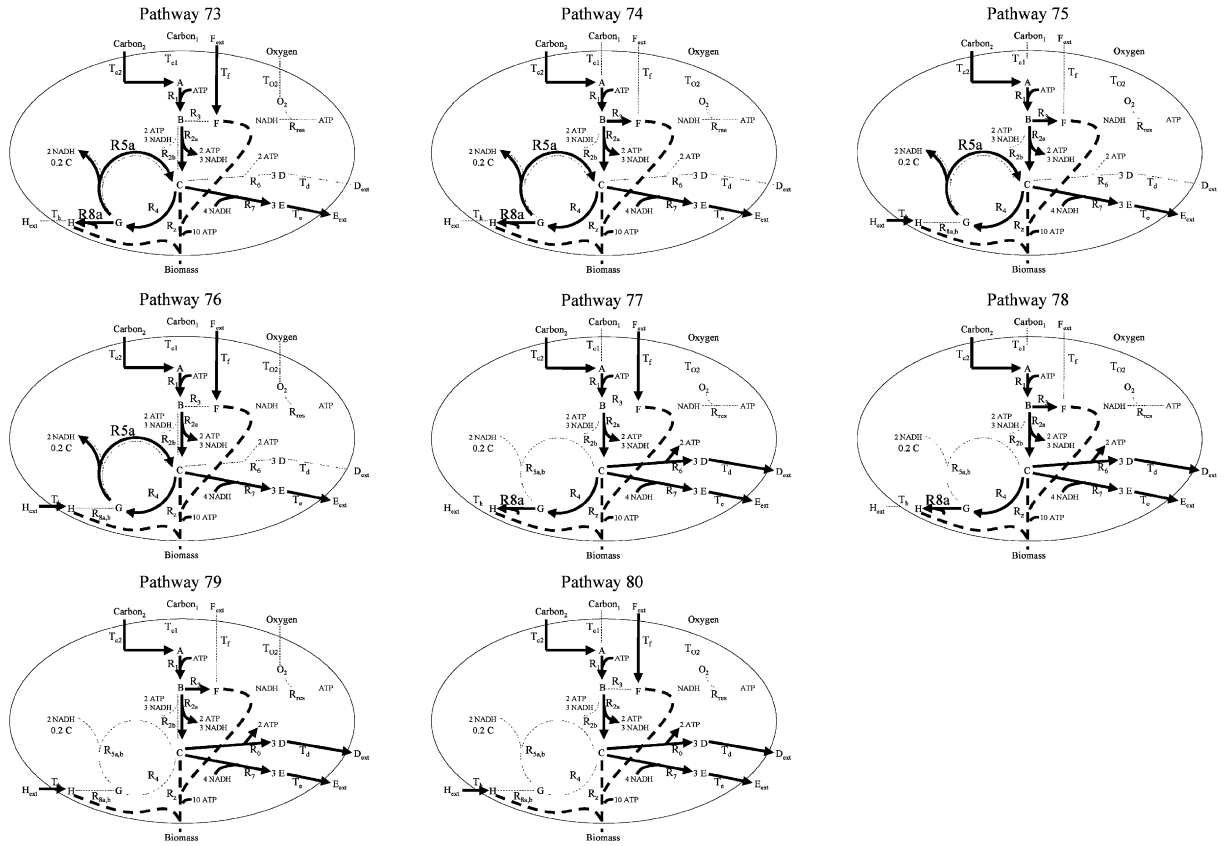
FIG. A1.

FIG. A2.

FIG. A3.

Fig. A4.

FIG. A5.