

Natarajan Vijayasankaran · Ross Carlson ·  
Friedrich Srienc

## Metabolic pathway structures for recombinant protein synthesis in *Escherichia coli*

Received: 11 January 2004 / Revised: 10 January 2005 / Accepted: 13 January 2005 / Published online: 1 March 2005  
© Springer-Verlag 2005

**Abstract** *Escherichia coli* is a valuable commercial host for the production of heterologous proteins. We used elementary mode analysis to identify all possible genetically independent pathways for the production of three specific recombinant proteins, green fluorescent protein, savinase and an artificial protein consisting of repeating units of a five-amino-acid cassette. Analysis of these pathways led to the identification of the most efficient pathways for the production of each of these proteins. The results indicate that the amino acid composition of expressed proteins has a profound effect on the number and identity of possible pathways for the production of these proteins. We show that several groups of elementary modes produce the same ratio of biomass and recombinant protein. The pattern of occurrence of these modes is dependent on the amino acid composition of the specific foreign protein produced. These pathways are formed as systemic combinations of other pathways that produce biomass or foreign protein alone after the elimination of fluxes in specific internal reversible reactions or the reversible carbon dioxide exchange reaction. Since these modes represent pathway options that enable the cell to produce biomass and protein without utilizing these reactions, removal of these reactions would constrain the cells to utilize these modes for producing biomass and foreign protein at constant ratios.

N. Vijayasankaran · R. Carlson · F. Srienc  
Department of Chemical Engineering and Materials Science,  
University of Minnesota,  
151 Amundson Hall, 421 Washington Avenue S.E.,  
Minneapolis, MN 55455-0321, USA

N. Vijayasankaran · R. Carlson · F. Srienc (✉)  
BioTechnology Institute, University of Minnesota,  
240 Gortner Lab, 1479 Gortner Avenue,  
St. Paul, MN 55108, USA  
e-mail: srienc@umn.edu  
Tel.: +1-612-6429776  
Fax: +1-612-6241700

R. Carlson  
Center for Biofilm Engineering and Department of Chemical  
and Biological Engineering, Montana State University,  
Bozeman, MT 59717, USA

### Introduction

*Escherichia coli* is used extensively for producing a wide variety of heterologous proteins (Anderson and Krummen 2002). It is one of the best characterized prokaryotic organisms and years of research have produced detailed knowledge of its genetics, molecular biology and biochemistry (Ingraham et al. 1983). Extensive investigations have also resulted in the development of sophisticated techniques for achieving high-level expression of foreign proteins (Khosla et al. 1990; Swartz 2001) and very high cell densities in bioreactors (Nakagawa et al. 1995; Lee 1996).

In addition to these experimental studies, *E. coli* is a standard for the development of theoretical tools for studying metabolism (Stephanopoulos et al. 1998). One recent method for analyzing the metabolic capabilities of biochemical networks is the rigorous pathway analysis technique known as elementary mode analysis (Schuster et al. 1999, 2000). This method identifies the complete set of genetically independent pathways for a biochemical reaction network and has been used, for instance, to find the most efficient *E. coli* pathways for ATP and biomass production (Carlson and Srienc 2004a,b). These studies indicated that cells likely use sets of defined pathways to grow efficiently under varying levels of culture-stress. Since *E. coli* is extensively used for foreign protein production, it is of significant practical interest to identify the most efficient pathway options for the synthesis of these macromolecules. Metabolic networks for the production of amino acids were recently analyzed using linear optimization-based approaches (See et al. 1996; Pharkya et al. 2003). However, little work has been dedicated so far towards the analysis of metabolic flux distributions during the synthesis of heterologous proteins.

This study examines the production, in *E. coli*, of three example proteins, poly(glycine-valine-glycine-isoleucine-proline) (GVGIP), green fluorescent protein (Gfp) and savinase. GVGIP is a protein from the class of elastomeric polypeptides poly(GVG-X-P), where X stands for any amino acid. These proteins are of commercial interest since they are biocompatible, biodegradable polymers that have a

range of other properties which make them suitable for drug-delivery vehicles, surgical scaffolds and other applications (Urry 1988, 1999). Gfp is widely utilized as a reporter protein to test expression systems and to develop protein production strategies because of its convenient auto-fluorescence (Chalfie et al. 1994; Natarajan et al. 1998; Subramanian and Srienc 1996). Finally, savinase is a protease used in detergents (Gupta et al. 2002).

We describe here an analytical technique for identifying and analyzing the metabolic flux distributions for optimum heterologous protein production. Different amino acid compositions of the foreign protein result in different optimal flux patterns, suggesting the potential for tailoring recombinant hosts to efficiently produce specific recombinant proteins. Furthermore, we identify subclasses of elementary flux modes that are formed as a systemic combination of other modes. We show that these modes could be useful in certain situations for identifying genetic modifications for efficient protein production by forcing the metabolite fluxes toward these pathways. Moreover, the presented methodology is a further example for the rational analysis of metabolic networks which should be generally useful for the optimization of production of any foreign protein or metabolite in a recombinant host, if its metabolic pathway structure is known. Information about pathway structure can be obtained, in principle, from genome sequences, which are available for an increasing number of organisms.

## Materials and methods

### Pathway analysis

The publicly available program METATOOL (ver. 352\_double; <http://mudshark.brookes.ac.uk/sware.html>) was used for the elementary mode analysis. The output from the METATOOL program was analyzed using a Microsoft Excel spreadsheet. The modes were sorted based on various criteria, such as biomass yield and protein yields on glucose. The yield of a certain mode was defined as the ratio of carbon atoms in the amount of product produced to the number of carbon atoms in the glucose consumed. A separate simulation was used for each protein analyzed.

The simulations used the network described by Carlson and Srienc (2004a). Glucose and carbon dioxide were considered as the sole carbon sources. The metabolic requirements for biomass and specific recombinant protein synthesis were modeled using the theory developed by Ingraham et al. (1983) by calculating the corresponding metabolic drain from the central metabolic pathway. The results presented in this study involve simulations utilizing a biomass macromolecular composition corresponding to a 200-min doubling time. The results with other growth rates were qualitatively similar to the results presented in this work. Appropriate reaction equations were constructed for each considered protein by taking into account the me-

tabolite drain from the central metabolism for each required amino acid and the energy requirements for the production of each peptide bond. Information about metabolite requirements for the production of each amino acid was obtained from the literature (Ingraham et al. 1983). The production of a protein with  $n$  residues requires the hydrolysis of approximately  $4n$  high-energy phosphate bonds to form the polypeptide bonds (Mathews et al. 2000). Accordingly, the model required four ATP equivalents to form each peptide bond. Table 1 illustrates the method used for calculating the metabolite drain from intermediate metabolism for the synthesis of Gfp, taking into account its amino acid composition. It also lists the final metabolic requirements for the production of savinase and GVGIP. It is to be noted that the equation for biomass production does not account for the production of the specific recombinant protein in consideration. Hence, the term “biomass” as used in this work does not include the recombinant protein. Details about the construction of the biomass and the protein terms and assumptions involved are provided by Carlson and Srienc (2004a). The systemic dependence between modes was analyzed using MATLAB code developed based on the algorithm detailed in the next section.

### Algorithm

This algorithm was used for identifying complementary biomass and protein-producing modes that form constraint line modes.

- (1) Choose a specific constraint line. Choose a mode on this constraint line:  $M^{CL}$ .
- (2) Select the set of all modes with a yield of protein on glucose greater than that of  $M^{CL}$ . From this set, eliminate all modes that have a positive flux in any irreversible reaction that has a zero flux in  $M^{CL}$ . The rest of the modes comprise set I.
- (3) Select the set of all modes with a yield of protein on glucose less than that of  $M^{CL}$ . From this set, eliminate all modes that have a positive flux in any irreversible reaction that has a zero flux in  $M^{CL}$ . These modes comprise set II.
- (4) Select a reversible reaction ( $R_{elim}$ ) in  $M^{CL}$  with a zero flux.
- (5) Choose a single mode from set I:  $M^I$ .
- (6) Choose a single mode from set II:  $M^{II}$ .
- (7) Compute a new mode,  $M^{Sys}$ , by forming a systemic combination of  $M^I$  and  $M^{II}$  to eliminate flux through  $R_{elim}$ .
- (8) Check that every reaction with a zero flux in  $M^{CL}$  also contains a zero flux in  $M^{Sys}$ . If not, go to step 5.
- (9) For each reaction, calculate the ratio of the raw flux through  $M^{CL}$  and  $M^{Sys}$ . Check that this ratio is the same for each of the reactions. If yes, then  $M^{CL}$  is a systemic combination of  $M^I$  and  $M^{II}$ . If not, go to step 5.

**Table 1** Metabolic requirements and overall stoichiometry for the production of Gfp. The overall stoichiometry is also given for GVGIP and savinase. The different amino acid composition of each protein results in a different metabolic drain from the intermediary me-

tabolism. *Rib-5-P* Ribose-5-phosphate, *Ery-4-P* erythrose-4-phosphate, *PEP* phosphoenol pyruvate, *PG* 3-phosphoglycerate, *Ac-CoA* acetyl coenzyme A, *AKG*  $\alpha$ -ketoglutarate, *Oxalo* oxaloacetate

Amino acid	Moles	Metabolite (moles)												
		Rib-5-P	Ery-4-P	PEP	PG	Pyruvate	Ac-CoA	AKG	Oxalo	ATP	NADH	NADPH	NH <sub>3</sub>	CO <sub>2</sub>
Alanine	8	0	0	0	0	8	0	0	0	0	0	8	8	0
Arginine	7	0	0	0	0	0	0	7	0	49	-7	28	28	0
Asparagine	13	0	0	0	0	0	0	0	13	39	0	13	26	0
Aspartate	18	0	0	0	0	0	0	0	18	0	0	18	18	0
Cysteine	2	0	0	0	2	0	0	0	0	8	-2	10	2	0
Glutamate	16	0	0	0	0	0	0	16	0	0	0	16	16	0
Glutamine	7	0	0	0	0	0	0	7	0	7	0	7	14	0
Glycine	21	0	0	0	21	0	0	0	0	0	-21	21	21	-21
Histidine	10	10	0	0	0	0	0	0	0	60	-30	10	30	10
Isoleucine	12	0	0	0	0	12	0	0	12	24	0	60	12	-12
Leucine	19	0	0	0	0	38	19	0	0	0	-19	38	19	0
Lysine	20	0	0	0	0	20	0	0	20	40	0	80	40	0
Methionine	6	0	0	0	0	0	0	0	6	42	0	48	6	6
Phenylalanine	13	0	13	26	0	0	0	0	0	13	0	26	13	0
Proline	10	0	0	0	0	0	0	10	0	10	0	30	10	0
Serine	10	0	0	0	10	0	0	0	0	0	-10	10	10	0
Threonine	15	0	0	0	0	0	0	0	15	30	0	45	15	0
Tryptophan	1	1	1	1	0	0	0	0	0	5	-2	3	2	0
Tyrosine	10	0	10	20	0	0	0	0	0	10	-10	20	10	0
Valine	17	0	0	0	0	34	0	0	0	0	0	34	17	0
Peptide bonds										936				
Gfp		11	24	47	33	112	19	40	84	1,273	-101	525	317	-17
GVGIP					2	3		1	1	23	-2	12	5	-3
Savinase		10	12	21	69	142	19	36	62	1,370	-130	484	347	-34

## Results

### Most efficient elementary modes

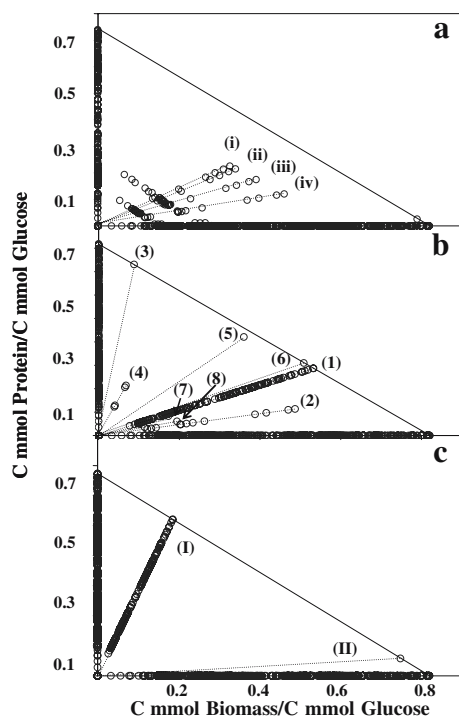
Since elementary mode analysis identifies every possible metabolic pathway for the production of heterologous proteins and/or biomass, it is possible to explicitly list and compare all pathways according to certain criteria. We compared the carbon yield of recombinant protein and biomass on glucose as a measure of the efficiency of different pathways. Elementary mode analysis revealed 370 GVGIP-producing modes for the considered *E. coli* network. In contrast, there are 916 modes which produce Gfp and 826 modes that produce savinase. Thus, the number of possible modes appears to be strongly dependent on the amino acid composition of the expressed protein.

The yield of biomass on glucose versus the yield of protein on glucose was plotted for each of the modes for the different proteins analyzed (Fig. 1). The modes on the abscissa represent modes that make only biomass without any recombinant protein. The modes on the ordinate represent pathways that make the foreign protein without any biomass. Experimentally, it has been shown that it is possible to produce recombinant protein without producing biomass (Flickinger and Rouse 1993). The other modes co-produce biomass and recombinant protein. Examination of these

plots confirms the expected result that the most efficient protein-producing modes do not make biomass while the most efficient biomass-producing mode does not produce any recombinant protein.

Certain common features can be observed for the three proteins analyzed. The line connecting the most efficient protein-producing and biomass-producing modes (optimum line) represents flux possibilities that are based on a linear combination of the two most efficient modes. In all three cases, every elementary mode is located either *under* or *on* the optimum line. Therefore, a linear combination of these two most efficient modes results in the optimum usage of glucose for the efficient and simultaneous production of the considered foreign protein and biomass.

The pathways resulting in most efficient protein production for GVGIP, Gfp and savinase are shown in Fig. 2a, b and c, respectively. For the sake of comparison, the most efficient biomass-producing mode as described by Carlson and Srienc (2004a) is shown in Fig. 2d. The optimum modes for the three proteins analyzed use the complete tricarboxylic acid (TCA) cycle and do not produce any carbon-containing byproducts, apart from carbon dioxide. There are significant differences among the pathways, depending on the specific protein being analyzed. The most efficient GVGIP-producing mode does not utilize the pentose phosphate pathway (PPP), since the protein does not contain any



**Fig. 1** Comparison of available modes for biomass and recombinant protein production. All available modes are illustrated for: **a** GVGIP **b** Gfp, **c** savinase. The *abscissa* represents the yield of biomass on glucose while the *ordinate* represents the yield of the specific protein on glucose. The *points* that lie on the *abscissa* represent modes that produce biomass only, while *points* that lie on the *ordinate* represent modes that make protein only. Some of the modes that produce both biomass and protein align along straight lines called constraint lines. Only constraint lines (i, ii, iii, iv) for GVGIP are indicated

amino acids originating from precursors of the PPP (see Table 2). In contrast, Gfp and savinase require reactions from the PPP to produce precursors for the appropriate amino acids. While the flux through each of the reactions is quantitatively different (Fig. 3), the directions of the flux in certain reactions in the protein modes change compared with the biomass mode. A reversal in the direction of these reactions leads to specific cases where the net fluxes are cancelled, as shown in Fig. 4 and 5.

For Gfp, the direction of flux through the transketolase-catalyzed (Fig. 2, R13r) and the transaldolase-catalyzed (R14r) sugar rearrangement reactions of the PPP and the reactions between PEP and pyruvate (R9, RR9) are reversed in comparison with the corresponding reactions for biomass synthesis. In the case of savinase, the flux through reactions R13r and R14r are reversed but not through R9, RR9, unlike the case for Gfp.

#### Co-production of biomass and recombinant protein—constraint lines

Many modes co-producing biomass and recombinant protein align along straight lines that we call constraint lines

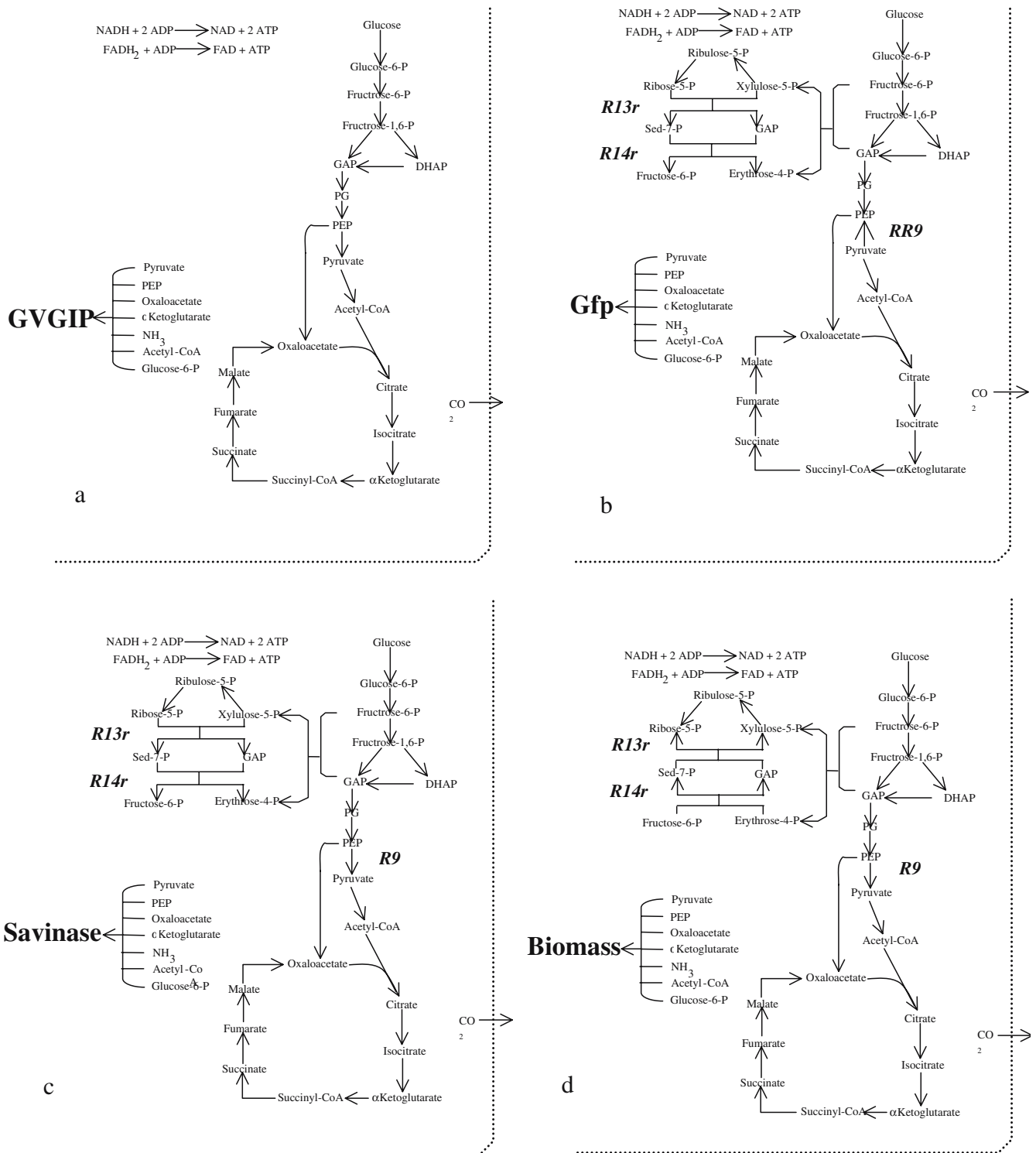
(Fig. 1). The modes that occur on these lines produce foreign protein and biomass in the same ratio. Since constraint lines reflect inherent limitations imposed on the cell by the stoichiometry of the reaction network and the amino acid composition of the expressed proteins, the number and pattern of occurrence of these modes varies depending on the specific protein being produced.

All modes on the same constraint line are characterized by a conservation of the topology of reactions of the PPP. For Gfp, the modes on constraint line 1 do not include the sugar rearrangement reactions of the PPP (R13r, R14r). Flux into the PPP for all modes in constraint line 2 is exclusively through the oxidative branch of the PPP (R10) and these modes do not utilize one of the transketolase-catalyzed reactions (R15r). The most efficient modes on these constraint lines (representing the end-points of the constraint lines) are shown in Fig. 4. Savinase features two constraint lines, I and II. Similar to constraint line 1 for Gfp, the modes on constraint line I for savinase do not contain reactions R13r and R14r. Finally, GVGIP constraint lines ii, iii and iv have no flux through reactions R13r/R14r, R11r and R15r, respectively, while constraint line i has a non-zero flux in all of its PPP reactions (Table 3).

Some of these lines extend to the optimum line, while others do not (see Fig. 1). For instance, the plot for GVGIP shows many constraint lines that do not extend to the optimum line, while the plots for Gfp and savinase contain one constraint line each that extends to the optimum line. Certain properties make them inefficient and prevent specific constraint lines from intersecting with the optimum line. For instance, Gfp constraint line 1 intersects with the optimum line while constraint line 2 does not. Further examination shows that all modes on constraint line 2 utilize the oxidative PPP (R10). This reaction produces two moles of NADPH and one mole of carbon dioxide for every molar flux of glucose-6-phosphate into the PPP. Apart from the loss of carbon as carbon dioxide, partially oxidized by-products are also produced to maintain the redox balance. These factors lead to reduced yields for modes on constraint line 2 and therefore these modes are less efficient.

Constraint line modes are systemic combinations of other modes

There is a common mechanism for the occurrence of a vast majority of these constraint lines: constraint line modes are systemic combinations of modes producing biomass and protein alone, i.e., they can be expressed as a non-trivial, non-negative linear combination of modes producing only biomass or only recombinant protein. The modes occurring on constraint lines result from the cancellation of certain reversible reactions when these specific biomass- and recombinant protein-producing modes are systemically combined. Examination of the modes shows that such cancellation can occur in six different reversible reactions of the network, namely the transketolase-catalyzed reactions, R13r, R15r, the transaldolase-catalyzed reaction, R14r, the phosphopentose epimerase-catalyzed reaction, R11r, the isomerase-



**Fig. 2** **a** Most efficient GVGIP-producing mode. **b** Most efficient Gfp-producing mode. **c** Most efficient savinase-producing mode. **d** Most efficient biomass-producing mode. The most efficient GVGIP mode does not use reactions of the PPP. The transketolase- and transaldolase-catalyzed sugar rearrangement reactions of the PPP are

catalyzed reaction, R2r and the carbon dioxide exchange reaction, R97r.

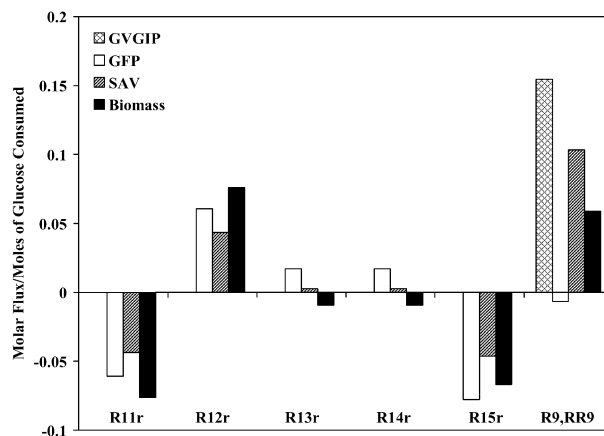
The majority of the modes on constraint line 1 for Gfp result from the elimination of reactions R13r and R14r from

reversed for the Gfp- and savinase-producing modes in reference to the most efficient biomass-producing modes. The flux between PEP and pyruvate for the most efficient Gfp-producing mode is also reversed in reference to the most efficient biomass-producing mode

systemic combinations of modes producing only biomass or recombinant protein. It was pointed out earlier that the optimum Gfp modes differ from the optimum biomass mode in the direction of the flux in the sugar rearrangement

**Table 2** Most efficient GVGIP-, Gfp- and biomass-producing modes. Values in italics represent fluxes that are either zero in contrast to non-zero flux in the biomass mode, or have an opposite sign in comparison with the corresponding biomass flux. The most efficient *Fmt* formate

Modes	Glycolysis															PPP										CO <sub>2</sub>
	R1	R2	R3	R5	R6	R9	RR9	R10	R11	R12	R13	R14	R15	R20	R21	TCA	PEPc	EtOH	Act	Fmt						
$M_{GVGIP}$	123	123	123	123	-123	19	0	0	0	0	0	0	0	0	64	64	38	52	0	0	0	192				
$M_{Gfp}$	20,462	20,462	19,220	19,220	-19,220	0	139	0	-1,242	1,242	351	351	-1,593	0	11,251	9,712	6,472	10,044	0	0	0	18,768				
$M_{Sav}$	19,109	19,109	18,277	18,277	-18,277	1,975	0	0	-832	832	52	52	-884	0	1,008	8,526	5,718	7,566	0	0	0	19,338				
$M_{Bio}$	14,005	13,901	12,835	12,835	-12,835	823	0	0	-1,066	1,066	-130	-130	-936	0	8,666	6,794	4,558	5,850	0	0	0	14,168				



**Fig. 3** Molar flux through different PPP reactions and the CO<sub>2</sub> exchange reaction in the optimal modes for producing GVGIP, Gfp, savinase and biomass. Since GVGIP does not require precursors from the PPP, flux through these reactions is zero. While the value of the flux through each of the reactions is different, the direction of flux in some of the reactions is also reversed

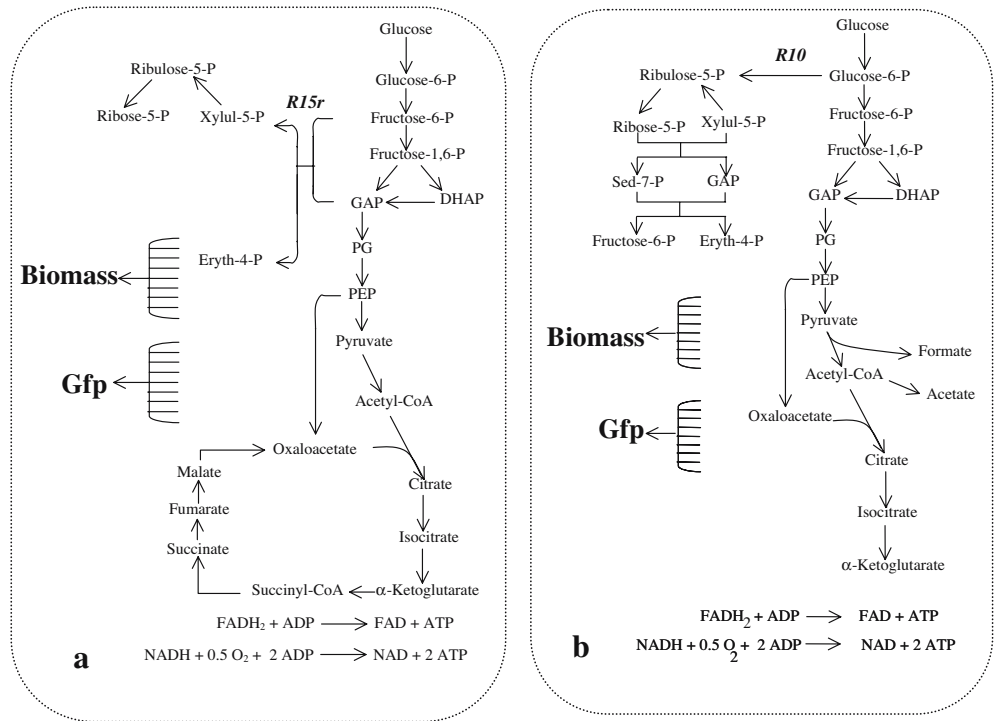
reactions of the PPP (R13r, R14r; see Fig. 3). When the cell is just making biomass without any recombinant protein, the flux is directed towards ribose phosphate in reaction R13r. Alternately, when the cell is making protein only, the flux in this reaction is directed towards glyceraldehyde phosphate. Similarly, the direction of the flux through R14r is also reversed as more and more protein is produced. The modes that occur on constraint line 1 correspond to those flux distributions that occur when the opposing fluxes in R13r and R14r due to biomass and Gfp production cancel each other, resulting in zero net flux. These modes lead to the production of biomass and Gfp at a constant ratio *without* utilizing reactions R13r and R14r.

Similar reasons lead to the creation of constraint line 2 for Gfp. The modes on this constraint line utilize the oxidative branch of the PPP and the sugar rearrangement reactions R13r and R14r. However, these modes do not utilize the transketolase-catalyzed reaction R15r. Thus, these modes utilize the oxidative PPP reaction R10 as the sole entry-point into the PPP. It can be shown that this constraint line arises due to a combination of two sub-optimal biomass- and Gfp-producing modes (Fig. 5). For instance, the most efficient mode on this constraint line occurs at the point where the opposing fluxes in reaction R15r due to sub-optimal biomass and Gfp production cancel each other, resulting in a net flux of zero.

Similarly, it can be shown that the four modes comprising constraint line 4 for Gfp arise due to cancellation of the carbon dioxide exchange reaction, R97r. These modes are formed due to elimination of flux in reaction R97r when a biomass-producing mode consuming carbon dioxide is systematically combined with a protein-producing mode producing carbon dioxide.

Although most of the constraint line modes are systemic combinations of modes producing biomass or recombinant proteins alone, some constraint line modes are formed due to systemic combinations involving other constraint line modes. For example, eight modes on Gfp constraint line 1

**Fig. 4** Most efficient modes for Gfp synthesis on: **a** constraint line 1 and **b** constraint line 2. Most efficient constraint line 1 mode does not produce any partially oxidized by-products, while most efficient constraint line 2 mode produces acetate and formate, making this mode inefficient

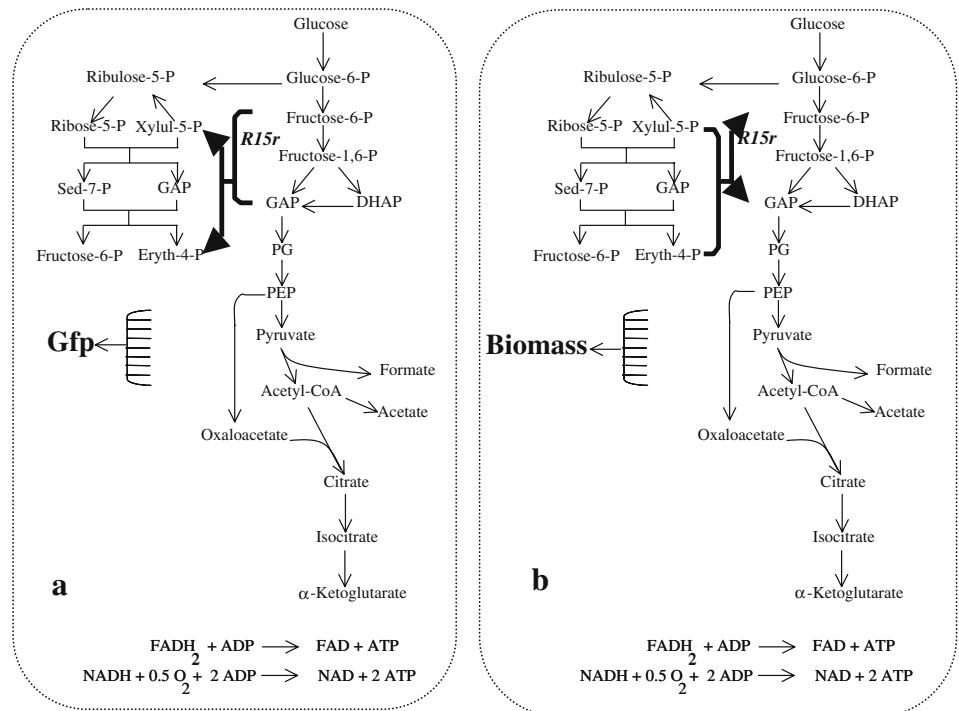


are formed when suboptimal biomass modes are systematically combined with modes on Gfp constraint line 4 to eliminate reactions R13r and R14r. Nine modes on constraint line 1 and the two modes forming constraint line 7 are formed by a similar systemic combination of suboptimal Gfp-producing modes and modes on constraint line 8.

Similar reaction cancellations after systemic combinations of two modes are also responsible for the constraint line modes occurring for the production of savinase and

GVGIP. Constraint line I for savinase is formed due to the elimination of reactions R13r and R14r when biomass- and savinase-producing modes are systematically combined. Constraint lines i, ii, iii and iv for GVGIP are formed by systemic combinations of biomass, GVGIP and other constraint line modes to eliminate reaction R97r. Elimination of reactions R13r, R14r, R15r and R11r are also involved in the formation of several constraint lines for GVGIP that are based on only a few modes.

**Fig. 5** Occurrence of Gfp constraint line 2. The most efficient mode for the constraint line 2 mode is formed due to the elimination of flux in reaction R15r when suboptimal **a** protein-producing and **b** biomass-producing modes are systematically combined



**Table 3** Relative fluxes through the most efficient modes from individual constraint lines. Reactions whose cancellation leads to the creation of the respective constraint lines are in italics (see text for explanation). Constraint lines that are systemically independent of other modes do not contain entries in italics. R11r–R15r are PPP reactions, R97r is the CO<sub>2</sub> exchange reaction and R2r is a glycolysis reaction (for reaction designations, see Carlson and Srienc 2004a)

Reaction constraint line	R2r	R11r	R12r	R13r	R14r	R15r	R97r
<b>GVGIP</b>							
i	5,391	-164	164	-20	-20	-144	0
ii	20,449	-496	736	0	0	-496	0
iii	33,985	0	1,968	496	496	-496	0
iv	846	31	77	31	31	0	0
v	13,163	62	154	62	62	0	-7,993
vi	20,653	0	246	62	62	-62	-12,979
vii	9,315	62	154	62	62	0	756
viii	3,745	-31	46	0	0	-31	-2,493
ix	13,763	0	246	62	62	-62	582
x	6,575	-82	82	-10	-10	-72	-4,501
xi	6,039	62	154	62	62	0	0
xii	11,241	0	246	62	62	-62	0
xiii	5,607	-82	82	-10	-10	-72	0
xiv	8,238	-93	138	0	0	-93	0
xv	26,901	-410	410	-50	-50	-360	0
xvi	5,231	-62	92	0	0	-62	0
xvii	5,121	-82	82	-10	-10	-72	0
xviii	2,224	-31	46	0	0	-31	-87
xix	3,663	-82	82	-10	-10	-72	-324
xx	8,497	589	1,463	589	589	0	10,415
xxi	9,291	-186	276	0	0	-186	0
<b>Gfp</b>							
1	253,558	-17,549	17,549	0	0	-17,549	292,596
2	12,265	688	1,512	688	688	0	0
3	697,866	-43,557	43,557	10,004	10,004	-53,561	651,708
4	52,175	-1,156	1,156	194	194	-1,350	0
5	1,937	182	329	182	182	0	1,590
6	368,331	0	43,557	14,830	14,830	-14,830	423,036
7	89,342	-1,949	2,006	0	0	-1,949	-5,995
8	40,883	-922	922	-38	-38	-884	-2,676
9	0	230,331	145,800	122,905	122,905	107,426	324,242
<b>Savinase</b>							
I	123,347	-6,292	6,292	0	0	-6,292	126,846
II	176,794	12,549	30,569	12,549	12,549	0	239,878

Using the algorithm described in the [Materials and methods](#), we tested whether each of the constraint line modes is a systemic combination of other modes. Table 4 provides an example of complementary biomass- and protein-producing modes that form a constraint line mode. However, a small number of constraint line modes were found to be systemically independent of other modes. All modes on Gfp constraint lines 3, 6 and 8 and GVGIP constraint lines viii, x and xix were found to be systemically independent of other biomass, protein or constraint line modes (Table 3). However, the number of such systemically independent modes is very small. While all constraint line modes for savinase are systemically dependent on other modes, four of the 89 modes on GVGIP constraint lines and nine of the 194 Gfp constraint line modes are systemically independent.

#### Gene-knockout targets for efficient protein production

The identification of all possible non-decomposable pathways using elementary mode analysis gives insight into the metabolic capabilities of a system (Schuster et al. 2000). It provides a means of identifying which enzymatic reactions are required for efficient protein production and which reactions are not. Accordingly, we identified specific mutations that eliminate less efficient pathways for protein production for GVGIP and savinase.

For GVGIP, the identified reactions to be eliminated involve the oxidative branch of the PPP (R10), the enzymatic activity of the malic enzymes (R41), the enzymatic activity associated with the action of NADH dehydrogenase II (R83), the enzymes involved in the production of lactate and succinate (R94, R95) and the transketolase enzyme (R15r). Such a strain would produce GVGIP most efficiently but would be unable to produce biomass. Since this strain



**Table 4** Identification of complementary biomass- and protein-producing modes that form a constraint line mode. R11r–R15r are PPP reactions, R97r is the CO<sub>2</sub> exchange reaction, R2r is a glycolysis reaction, R70 is the recombinant protein synthesis reaction and R71 is the biomass synthesis reaction (for reaction designations, see Carlson and Srienc 2004a)

Mode	Reaction								
	R2r	R11r	R12r	R13r	R14r	R15r	R97r	R70	R71
$M^{CL}$	20,449	-496	736	0	0	-496	0	1,272	16
$M^{Bio}$	6,011	-248	368	0	0	-248	-1,272	0	8
$M^{Pro}$	53	0	0	0	0	0	16	8	0

The algorithm for identifying the complimentary modes that form constraint line modes is described in the [Materials and methods](#).

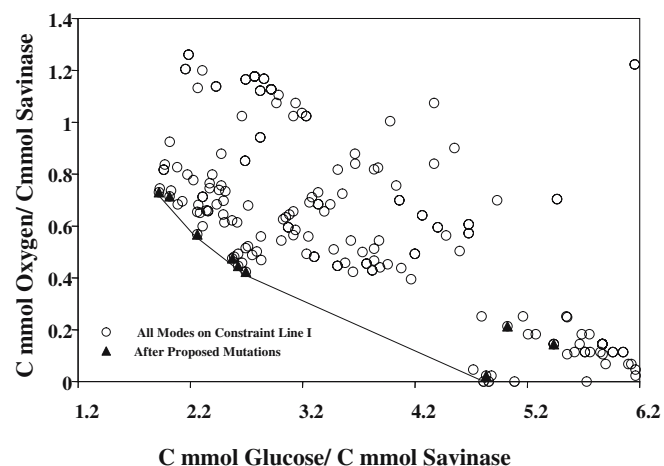
This table illustrates an example for constraint line ii for GVGIP.  $M^{CL}$  is the most efficient mode on this constraint line, while  $M^{Bio}$  and  $M^{Pro}$  are the biomass- and protein-producing modes that form  $M^{CL}$ . It can be verified that:

$$M^{CL} = \frac{M^{Pro} + M^{Bio} \begin{pmatrix} 16 \\ 1, 272 \end{pmatrix}}{0.00629}$$

would be unable to grow, an inducible genetic switch would have to be used to implement this strategy.

In the case of savinase, the removal of the transaldolase enzyme (R14r) along with reactions R10, R41, R83, R94 and R95 leads to the elimination of all modes except a few modes on constraint line 1. These remaining modes force the cell to make savinase at a ratio of 2.8 moles of carbon in savinase per mole of carbon in biomass. This analysis indicates that this strain would still be capable of growth and hence this strain could be used to produce savinase and biomass at a constant ratio in a continuous reactor.

It can be further shown that the proposed mutations have the effect of optimizing the production of savinase at all levels of oxygen availability. Figure 6 is an inverse yield



**Fig. 6** The effect of proposed mutations on the availability of constraint line I modes for the production of savinase. The *abscissa* is the inverse yield of savinase on glucose, while the *ordinate* is the inverse yield of savinase on oxygen. The *points* that lie closest to the origin are the more efficient in the conversion of glucose and oxygen into savinase. The transition from aerobic to anaerobic conditions is characterized by a line segment spanning four modes. The effect of the proposed mutations is to reduce the availability of modes to those that lie *on* or *near* this optimum transition line segment

plot (Carlson and Srienc 2004a) that illustrates the available modes on constraint line I, before and after the proposed mutations. In such a plot, modes that lie closest to the origin are the most efficient in the conversion of the carbon source and oxygen into the desired product. It can be seen that, for savinase, the available modes after the proposed mutations are on or very close to the optimum transition line from aerobic to anaerobic conditions. It can also be similarly shown that the proposed mutations in these cases also optimize the production of maintenance energy at all levels of oxygen availability. It should also be noted that, since all optimal modes utilized under oxygen limitation require the production of acetate, a single mutation eliminating the acetate-producing reaction would eliminate all modes except the most efficient mode. However, the optimal performance of such a strain would be maintained only under aerobic conditions.

## Discussion

We used elementary mode analysis to study the co-production of biomass and recombinant proteins in *E. coli*. We identified special sets of elementary flux modes that co-produce biomass and recombinant protein at constant ratios. Understanding these pathway possibilities is important because it provides unique insight into the nature of the metabolic network, which may suggest specific strain-optimization strategies. These constraint lines are created by the cancellation of opposing fluxes in specific reversible reactions during the co-production of biomass and recombinant protein. Hence, these modes are systemically dependent on other modes. The property of systemic independence has been used in other metabolic network analysis tools, such as extreme pathway analysis (Schilling et al. 2000). We observed in our simulations that many constraint line modes are observed to be formed due to the elimination of the carbon dioxide exchange reaction when other modes are systemically combined. These constraint line modes in turn lead to other constraint line modes when they are combined systemically to eliminate other internal reversible reactions. It is important to note that tools such as extreme pathway analysis do not identify these modes, even though these modes are physiologically relevant (Klamt and Stelling 2003).

We further showed that some of these constraint line modes could be used to optimize the production of recombinant proteins. Since all modes on a constraint line produce foreign protein and biomass at a constant ratio, directing cellular fluxes through these modes could force the cell to produce protein and biomass at these ratios. Since constraint lines are formed by the elimination of fluxes through certain reversible reactions due to systemic combinations of biomass- and protein-producing modes, removal of these reversible reactions through genetic modifications could force the cell to utilize specific constraint line modes. Constraint line I for savinase is formed by the elimination of flux through reaction R14r. Hence, elimination of this reaction forces the cell to utilize modes on this constraint line and

hence produce savinase and biomass at a constant molar ratio of 2.8 moles of carbon in savinase per mole of carbon in biomass. It was shown previously that the elimination of reactions R10, R41, R83, R94 and R95 has the effect of producing biomass for wild-type *E. coli* at the most efficient carbon yield (Carlson and Srienc 2004a). Elimination of these five reactions in addition to reaction R14r ensures that only the most efficient modes on constraint line I are available to the cells at all levels of oxygen stress.

**Acknowledgement** This work was funded by the National Science Foundation (BES-0109383).

## References

- Anderson DC, Krummen L (2002) Recombinant protein expression for therapeutic applications. *Curr Opin Biotechnol* 13:117–123
- Carlson R, Srienc F (2004a) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: identification of reactions. *Biotechnol Bioeng* 85:1–19
- Carlson R, Srienc F (2004b) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: creation of overall flux states. *Biotechnol Bioeng* 86:149–162
- Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. *Science* 263:802–805
- Flickinger MC, Rouse MP (1993) Sustaining protein-synthesis in the absence of rapid cell-division—an investigation of plasmid-encoded protein expression in *Escherichia coli* during very slow growth. *Biotechnol Prog* 9:555–572
- Gupta R, Beg QK, Lorenz P (2002) Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl Microbiol Biotechnol* 59:15–32
- Ingraham JL, Maaloe O, Neidhardt FC (1983) Growth of the bacterial cell. Sinauer Associates, Sunderland, Mass.
- Khosla C, Curtis JE, DeModena J, Rinas U, Bailey JE (1990) Expression of intracellular hemoglobin improves protein synthesis in oxygen limited *Escherichia coli*. *Biotechnology* 8:849–853
- Klamt S, Stelling J (2003) Two approaches for metabolic pathway analysis. *Trends Biotechnol* 21:64–69
- Lee SY (1996) High cell density culture of *Escherichia coli*. *Trends Biotechnol* 14:98–105
- Mathews CK, Holde KE van, Ahern KG (2000) *Biochemistry*. Addison Wesley Longman, San Francisco
- Nakagawa S, Oda H, Anazawa H (1995) High cell density cultivation and high recombinant protein production of *Escherichia coli* strain expressing uricase. *Biosci Biotechnol Biochem* 59:2263–2267
- Natarajan A, Subramanian S, Srienc F (1998) Comparison of mutant forms of the green fluorescent protein as expression markers in Chinese hamster ovary (CHO) and *Saccharomyces cerevisiae* cells. *J Biotechnol* 62:29–45
- Pharkya P, Burgard AP, Maranas CD (2003) Exploring overproduction of amino acids using bilevel optimization framework OptKnock. *Biotechnol Bioeng* 84:887–899
- Schilling CH, Edwards JS, Lestscher D, Palsson BO (2000) Combining pathway analysis with flux balance analysis for comprehensive study of metabolic systems. *Biotechnol Bioeng* 71:286–306
- Schuster S, Dandekar T, Fell D (1999) Detection of elementary flux modes in biochemical networks: a systemic organization and analysis of complex metabolic networks. *Trends Biotechnol* 17:53–60
- Schuster S, Fell D, Dandekar T (2000) A general definition of metabolic pathways useful for systemic organization and analysis of complex metabolic networks. *Nat Biotechnol* 18:326–332
- See SM, Dean JP, Dervakos G (1996) On the topological features of optimal metabolic pathway regimes. *Appl Biochem Biotechnol* 60:251–301
- Stephanopoulos G, Aristidou A, Nielsen J (1998) *Metabolic engineering: principles and methodologies*. Academic, San Diego
- Subramanian S, Srienc F (1996) Quantitative analysis of transient gene expression in mammalian cells using the green fluorescent protein. *J Biotechnol* 49:137–151
- Swartz JR (2001) Advances in *Escherichia coli* production of therapeutic proteins. *Curr Opin Biotechnol* 12:195–201
- Urry DW (1988) Entropic elastic processes in protein mechanisms. I. Elastic structure due to inverse temperature transition and elasticity due to internal chain dynamics. *J Protein Chem* 7:1–34
- Urry DW (1999) Elastic molecular machines in metabolism and soft-tissue restoration. *Trends Biotechnol* 17:249–257