



# *In Silico* Identification of Gene Amplification Targets for Improvement of Lycopene Production

Hyung Seok Choi, Sang Yup Lee, Tae Yong Kim and Han Min Woo

**Applied and environmental Microbiology**

May 2010



# The objective – Improving the yield of bioproducts

## ► Identif

MO  
Opt

## ► Identif difficul

1. G  
cc
2. It  
an

## ► Availa

1. O
2. C

maximize  $v_{\text{biochemical}}$  (OptReg)  
 subject to

$$v_{\text{biomass}} - \varepsilon \cdot \sum_j v_j = (v_{\text{atp\_maint}} \cdot \lambda_{\text{atp}}) + (0.01 \cdot v_{\text{biomass}}^{\text{max}} \cdot \lambda_{\text{bio}})$$

$$+ \sum_j (z_{U,j}^k \cdot v_j^{\text{max}} + z_{L,j}^k \cdot v_j^{\text{min}})$$

$$+ \sum_j [(v_j^{\text{max}} \cdot z_{U,j}^d) + [(v_{j,L}^0 \cdot (1 - C) + v_j^{\text{min}} \cdot (C)) \cdot (q_{U,j}^d - z_{U,j}^d) + (q_{L,j}^d \cdot (v_j^{\text{min}})] + \sum_j [(q_{U,j}^k \cdot v_j^{\text{max}} + (v_j^{\text{min}} \cdot z_{L,j}^u) + [(v_{j,0}^u \cdot (1 - C) + v_j^{\text{max}} \cdot (C)) \cdot (q_{L,j}^u - z_{L,j}^u)],$$

$$\sum_{j=1}^M S_{ij} v_j = 0, \quad \forall i \in \mathbb{N},$$

$$v_{\text{atp}} \geq v_{\text{atp\_maint}},$$

$$v_{\text{biomass}} \geq (0.01) \cdot v_{\text{biomass}}^{\text{max}},$$

$$v_{\text{glc}} = 10 \text{ mmol/gDW} \cdot \text{h},$$

$$v_j \leq v_j^{\text{max}} \cdot y_j^k, \quad \forall j \in \mathbb{M},$$

$$v_j \geq v_j^{\text{min}} \cdot y_j^k, \quad \forall j \in \mathbb{M},$$

$$v_j^{\text{min}} \leq v_j \leq [(v_{j,L}^0 \cdot (1 - C) + (v_j^{\text{min}} \cdot (C)) \cdot (1 - y_j^d) + v_j^{\text{max}} \cdot y_j^d], \quad \forall j \in \mathbb{M},$$

$$[(v_{j,U}^0 \cdot (1 - C) + (v_j^{\text{max}} \cdot (C)) \cdot (1 - y_j^u) + v_j^{\text{min}} \cdot y_j^u \leq v_j \leq v_j^{\text{max}}, \quad \forall j \in \mathbb{M},$$

$$(1 - y_j^k) + (1 - y_j^d) + (1 - y_j^u) \leq 1, \quad \forall j \in \mathbb{M},$$

$$y_j^k \in \{0, 1\}; \quad y_j^d \in \{0, 1\}; \quad y_j^u \in \{0, 1\}, \quad \forall j \in \mathbb{M},$$

$$\sum_j [(1 - y_j^k) + (1 - y_j^d) + (1 - y_j^u)] \leq L$$

$$y_j^k = y_{j+1}^k, \quad y_j^d + y_{j+1}^d \geq 1, \quad y_j^u + y_{j+1}^u \geq 1, \quad \forall j \in \mathbb{M}_{\text{rev}},$$

$$\sum_{i=1}^N \lambda_i S_{i,j} + q_{U,j}^k + q_{L,j}^k + q_{U,j}^d + q_{L,j}^d + q_{U,j}^u + q_{L,j}^u \geq -\varepsilon, \quad \forall j \in \mathbb{M}, j \neq \text{atp, biomass},$$

$$\sum_{i=1}^N \lambda_i S_{i,\text{biomass}} + q_{U,\text{biomass}}^k + q_{L,\text{biomass}}^k + q_{U,\text{biomass}}^d + q_{L,\text{biomass}}^d + q_{U,\text{biomass}}^u + q_{L,\text{biomass}}^u + \lambda_{\text{bio}} \geq 1 - \varepsilon,$$

$$\sum_{i=1}^N \lambda_i S_{i,\text{atp}} + q_{U,\text{atp}}^k + q_{L,\text{atp}}^k + q_{U,\text{atp}}^d + q_{L,\text{atp}}^d + q_{U,\text{atp}}^u + q_{L,\text{atp}}^u + \lambda_{\text{atp}} \geq -\varepsilon,$$

$$0 \leq z_{U,j}^k \leq (q_{U,j}^k)_{UB} \cdot y_j^k, \quad \forall j \in \mathbb{M},$$

$$q_{U,j}^k - (q_{U,j}^k)_{UB} \cdot (1 - y_j^k) \leq z_{U,j}^k \leq q_{U,j}^k, \quad \forall j \in \mathbb{M},$$

$$(q_{L,j}^k)_{LB} \cdot y_j^k \leq z_{L,j}^k \leq 0, \quad \forall j \in \mathbb{M},$$

$$q_{L,j}^k \leq z_{L,j}^k \leq q_{L,j}^k - (q_{L,j}^k)_{LB} \cdot (1 - y_j^k), \quad \forall j \in \mathbb{M},$$

$$0 \leq z_{U,j}^d \leq (q_{U,j}^d)_{UB} \cdot y_j^d, \quad \forall j \in \mathbb{M},$$

$$q_{U,j}^d - (q_{U,j}^d)_{UB} \cdot (1 - y_j^d) \leq z_{U,j}^d \leq q_{U,j}^d, \quad \forall j \in \mathbb{M},$$

$$(q_{L,j}^d)_{LB} \cdot y_j^d \leq z_{L,j}^d \leq 0, \quad \forall j \in \mathbb{M},$$

$$q_{L,j}^d \leq z_{L,j}^d \leq q_{L,j}^d - (q_{L,j}^d)_{LB} \cdot (1 - y_j^d), \quad \forall j \in \mathbb{M},$$

$$0 \leq z_{U,j}^u \leq (q_{U,j}^u)_{UB} \cdot y_j^u, \quad \forall j \in \mathbb{M},$$

$$q_{U,j}^u - (q_{U,j}^u)_{UB} \cdot (1 - y_j^u) \leq z_{U,j}^u \leq q_{U,j}^u, \quad \forall j \in \mathbb{M},$$

$$(q_{L,j}^u)_{LB} \cdot y_j^u \leq z_{L,j}^u \leq 0, \quad \forall j \in \mathbb{M},$$

$$q_{L,j}^u \leq z_{L,j}^u \leq q_{L,j}^u - (q_{L,j}^u)_{LB} \cdot (1 - y_j^u), \quad \forall j \in \mathbb{M},$$

$$q_{U,j}^k, q_{U,j}^d, q_{U,j}^u \geq 0; \quad q_{L,j}^k, q_{L,j}^d, q_{L,j}^u \leq 0; \quad \forall j \in \mathbb{M},$$

$$\lambda_i \in \mathbb{R}, \quad \forall i \in \mathbb{N}; \quad \lambda_{\text{atp}} \leq 0; \quad \lambda_{\text{bio}} \leq 0.$$

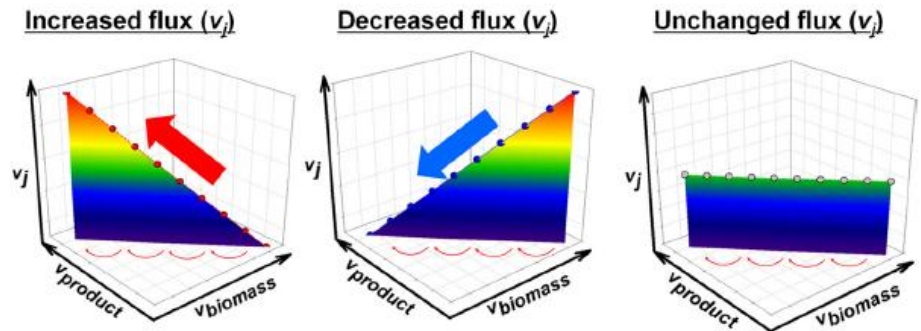
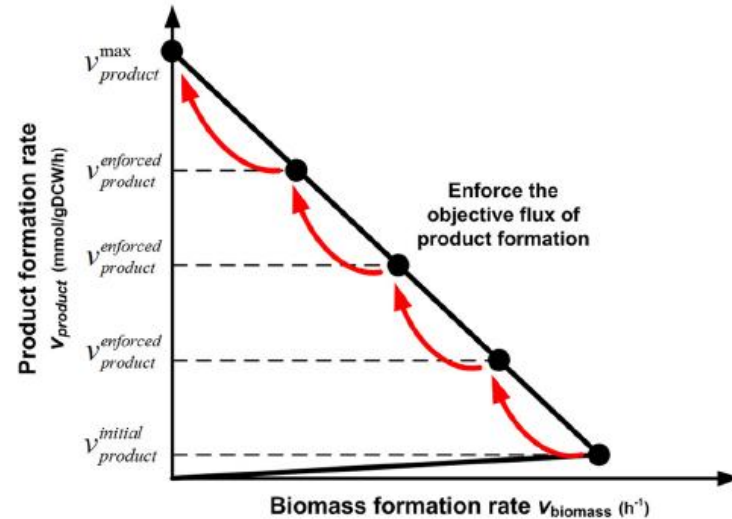




# FSEOF – Flux Scanning based on Enforced Objective Flux

## The computational approach:

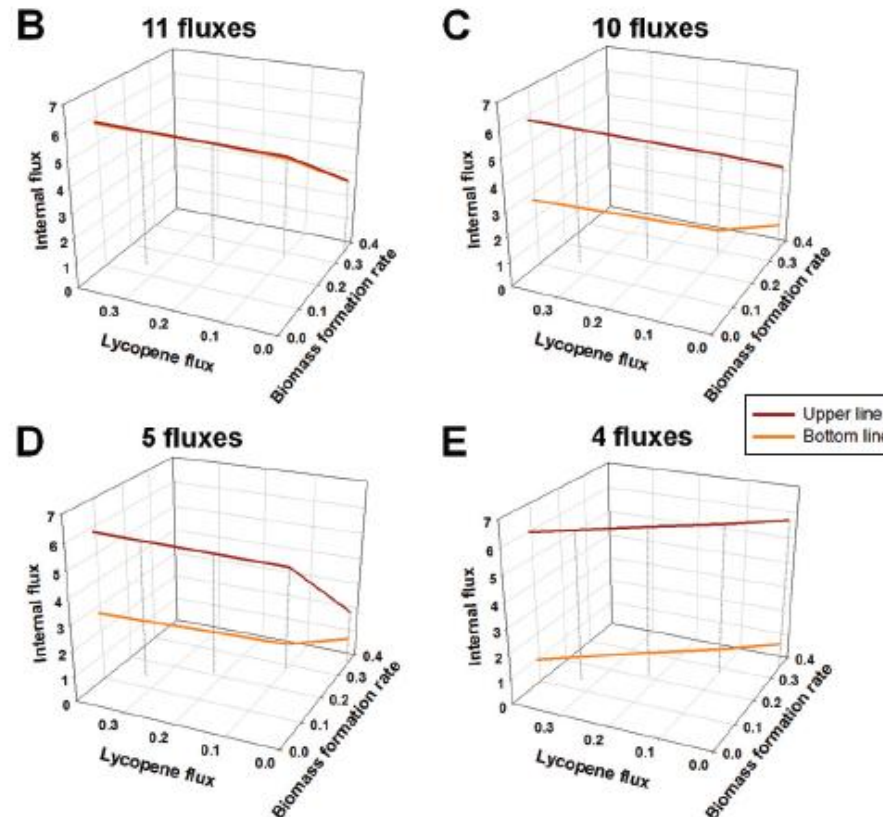
1. Start from an initial flux distribution obtained under maximum biomass production and an initial (measured) lycopene production
2. Calculate the theoretical maximum yield
3. Apply a n-steps procedure maximizing the biomass production while producing the initial lycopene production +  $n$ -th of the difference between the initial and maximal lycopene flux
4. Search for reactions that their final flux is higher than their initial flux





# Identifying gene amplification targets

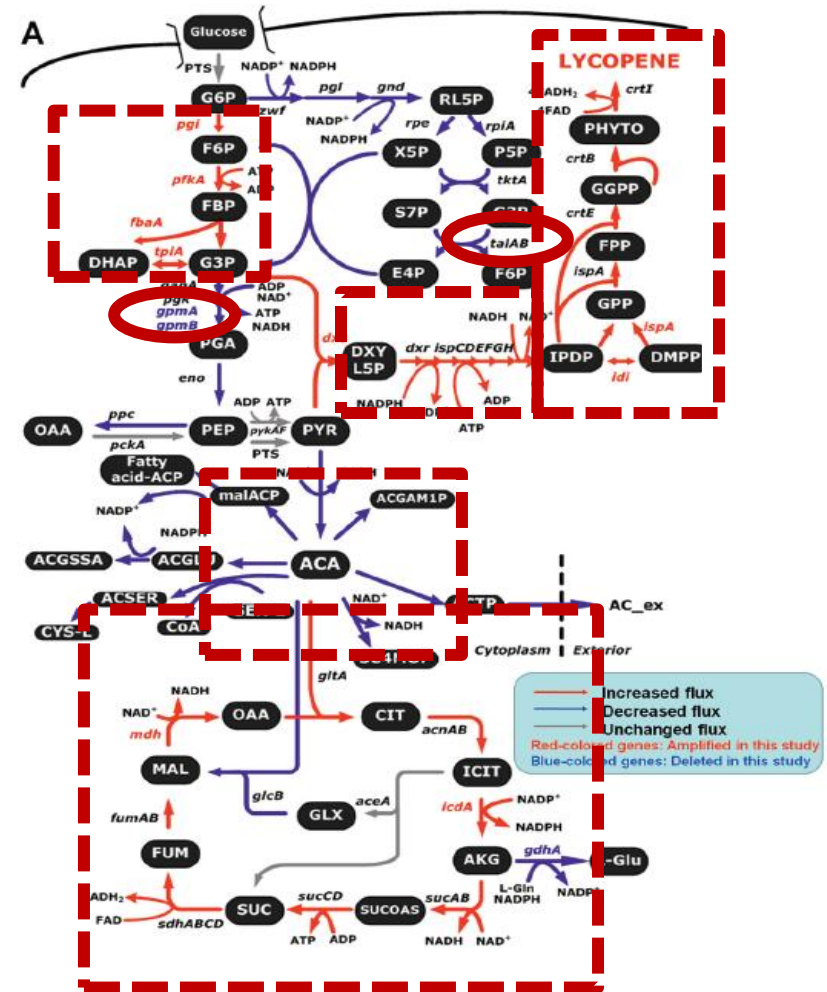
- ▶ 35 reactions were identified as initial gene amplification targets
- ▶ Constraint-based flux analysis does not give a unique flux distribution and therefore FVA was applied





# Identifying gene amplification targets (2)

- The fluxes to G3P and the lycopene biosynthetic pathway should be increased
- The flux to acetyl-CoA decreased and was redirected to DXYL5P

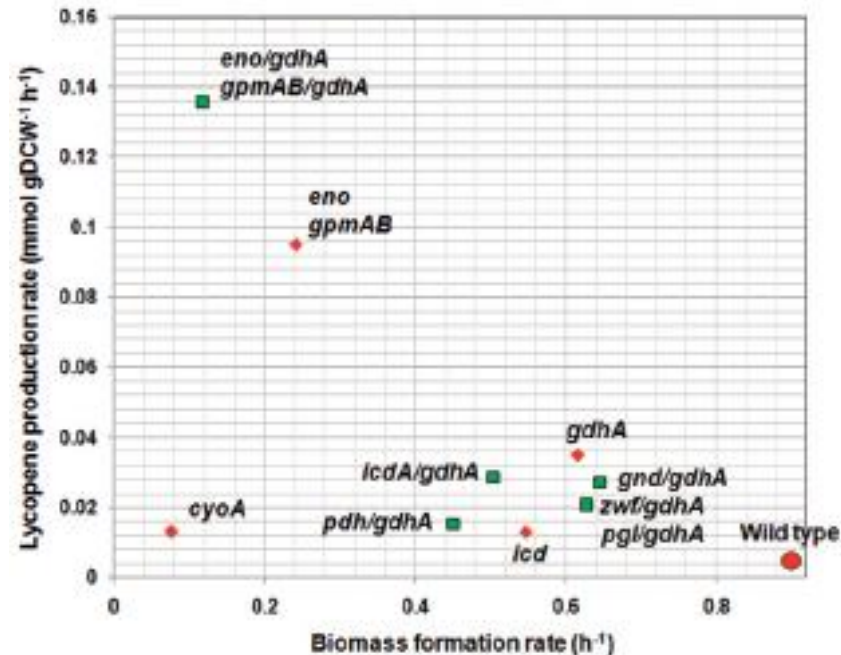
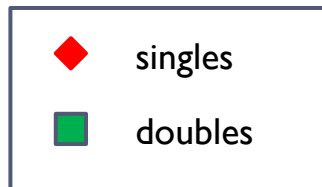


Gene	Enzyme
<i>acnAB</i>	Aconitase
<i>gltA</i>	Citrate (CIT) synthase
<i>fumAB</i>	Fumarase
<i>icdA<sup>a</sup></i>	Isocitrate (ICIT) dehydrogenase (NADP)
<i>mdh<sup>a,c</sup></i>	Malate (MAL) dehydrogenase
<i>sdhABCD</i>	Succinate (SUC) dehydrogenase
<i>sucCD</i>	Succinyl-CoA (SUCOAS) synthetase (ADP-forming)
<i>sucAB</i>	2-Oxoglutarate (AKG) dehydrogenase
<i>sdhABCD</i>	Succinate dehydrogenase
<i>dxr</i>	1-Deoxy-D-xylulose 5-phosphate reductoisomerase
<i>dxs<sup>c</sup></i>	1-Deoxyxylulose-5-phosphate synthase
<i>idi<sup>a</sup></i>	Isopentenyl diphosphate (IPDP) isomerase
<i>ispA<sup>a</sup></i>	Geranyltranstransferase/dimethylallyltranstransferase
<i>ispD</i>	4-Diphosphocytidyl-2C-methyl-D-erythritol synthase
<i>ispE</i>	4-Diphosphocytidyl-2-C-methylerythritol kinase
<i>ispF</i>	2C-Methyl-D-erythritol 2,4-cyclodiphosphate synthase
<i>ispG</i>	1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase
<i>ispH</i>	1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase
<i>crtE</i>	Geranylgeranyl pyrophosphate (GGPP) synthase
<i>crtB</i>	Phytoene (PHYTO) synthetase
<i>crtI</i>	Phytoene dehydrogenase
<i>fbaA<sup>a</sup></i>	Fructose-bisphosphate (FBP) aldolase
<i>pfkAB<sup>a</sup></i>	Phosphofructokinase
<i>pgi<sup>a</sup></i>	Glucose-6-phosphate (G6P) isomerase
<i>tpiA<sup>a,c</sup></i>	Triose-phosphate isomerase



# Identifying gene knockout targets

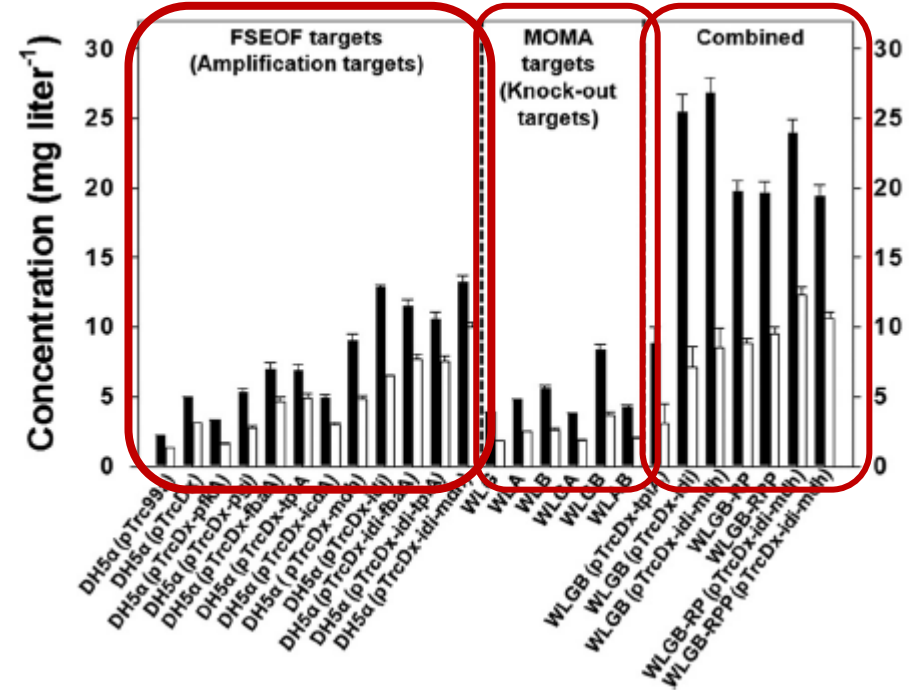
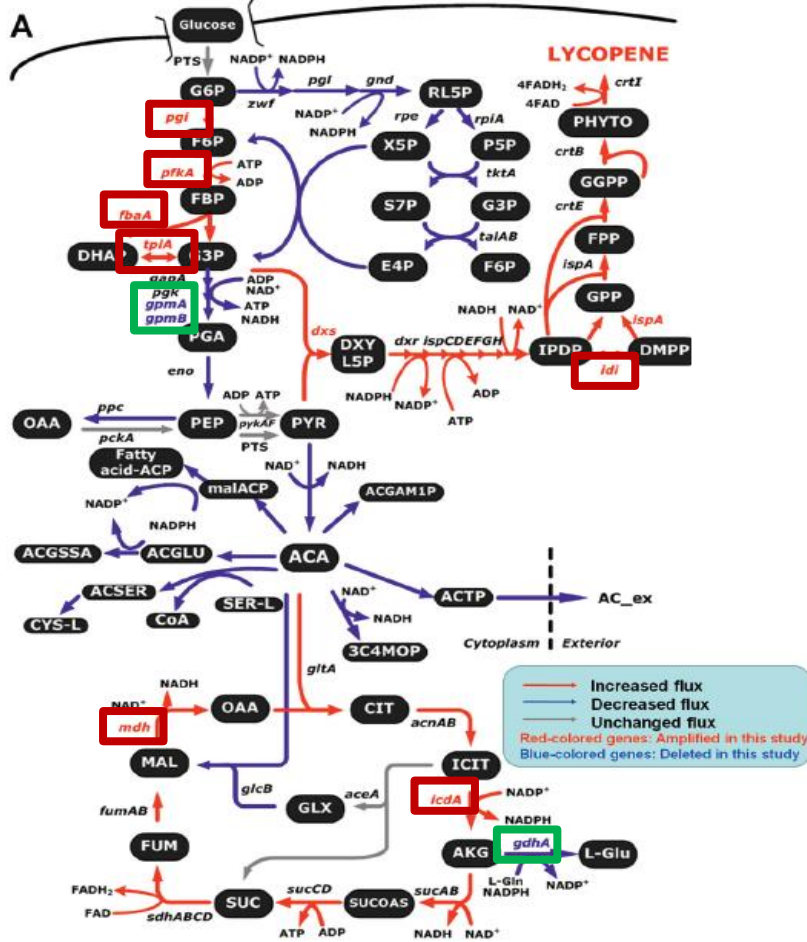
Applying MOMA to identify single and double gene KOs







# Experimental validation



an gene KOs  
KO results in higher production of lycopene



# Summary and Limitations

---

- ▶ FSEOF allows the *in silico* identification of fluxes to be amplified for the enhanced production of desired bioproduct.
  - ▶ Not all targets predicted by FESOF resulted in enhanced production due to model limitations
  - ▶ Perhaps applying a sampling technique would help to improve the model's predictions
  - ▶ Additional validation of non-intuitive gene amplification targets is required
- 







---

# Questions?

