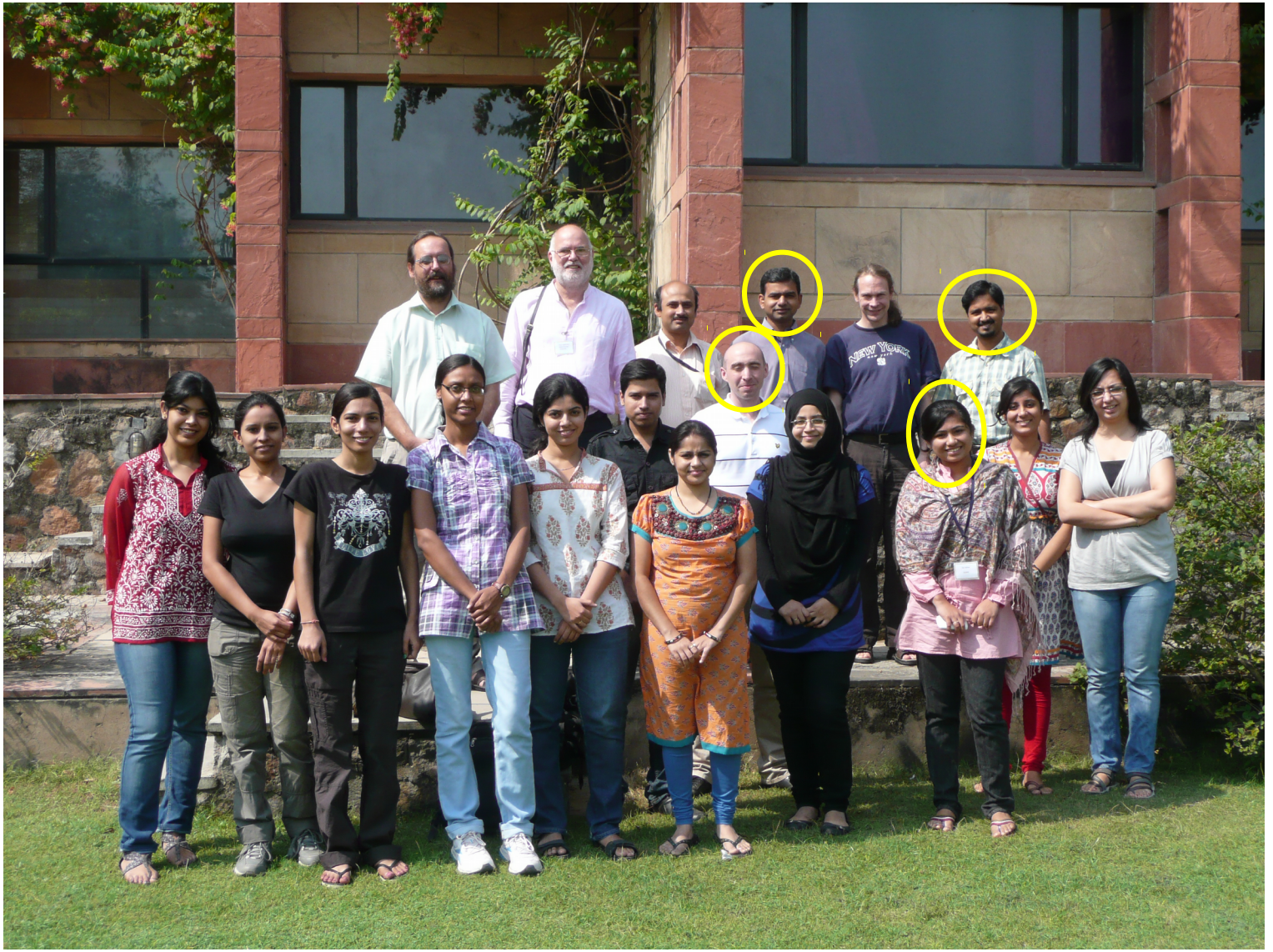
Model-assisted engineering of *Escherischia coli* for biofuel production

David Fell, Mark G Poolman, Hassan Hartman
Oxford Brookes University

Zia Fatma, Shireesh Srivastava, Syed Shams Yazdani
ICGEB, New Delhi

Department of Biotechnology
Govt. of India

ICGEB Delhi 2011



Biofuels

Ethanol	Corn or potato starch, sugar cane	Lignocellulosic waste
Butanol	Corn starch	Lignocellulosic and food waste
Biodiesel (alkanes, esterified fatty acids)	Palm oil, oil seeds	Photosynthesis; waste cooking oil; ? lignocellulosic waste

Pilot plant for production of ethanol from lignocellulosic waste, Kashipur, India. Designed by ICT Mumbai Centre for Energy Biosciences. Photo: courtesy of Prof A Lali

Microbial Routes to Biodiesel

- Cyanobacteria naturally produce alkanes and long-chain fatty alcohols by photosynthesis.
- Some algae and yeasts produce high levels of triglycerides as storage compounds, which could be trans-esterified to fatty methyl esters.
- Engineering of bacteria or yeasts with the cyanobacterial pathway to generate alkanes non-photosynthetically from C-containing substrates

Approach in this Study

- Used *E coli* as the host organism because of:
  - Detailed knowledge of its metabolic network
  - Extensive molecular genetics tools that facilitate experimental modification of metabolism
  - Grows readily
  - Native ability to use pentoses (such as xylose from lignocellulose wastes) in addition to glucose

Heterologous Alkane Synthesis Pathway

- The pathway branches off from fatty acid synthesis by intercepting acyl-ACP intermediates.
- AAR, acyl-ACP reductase, EC 1.2.1.80 from *Synechococcus elongatus* releases a long-chain aldehyde, e.g.:

  palmitoyl-ACP + NADPH + H^+ -> palmitaldehyde + $NADP^+$ + ACP

- ADO. aldehyde oxygenase deformylating, EC 4.1.99.5 from *Nostoc punctiforme* :

  long-chain aldehyde + O_2 + 2NADPH + $2H^+$ -> alkane + formate + H_2O + $2NADP^+$

  The reaction requires ferredoxin and ferredoxin reductase

Alternative to Long-Chain Alcohols

- AAR, acyl-ACP reductase, EC 1.2.1.80 from *Synechococcus elongatus* releases a long-chain aldehyde, e.g.:

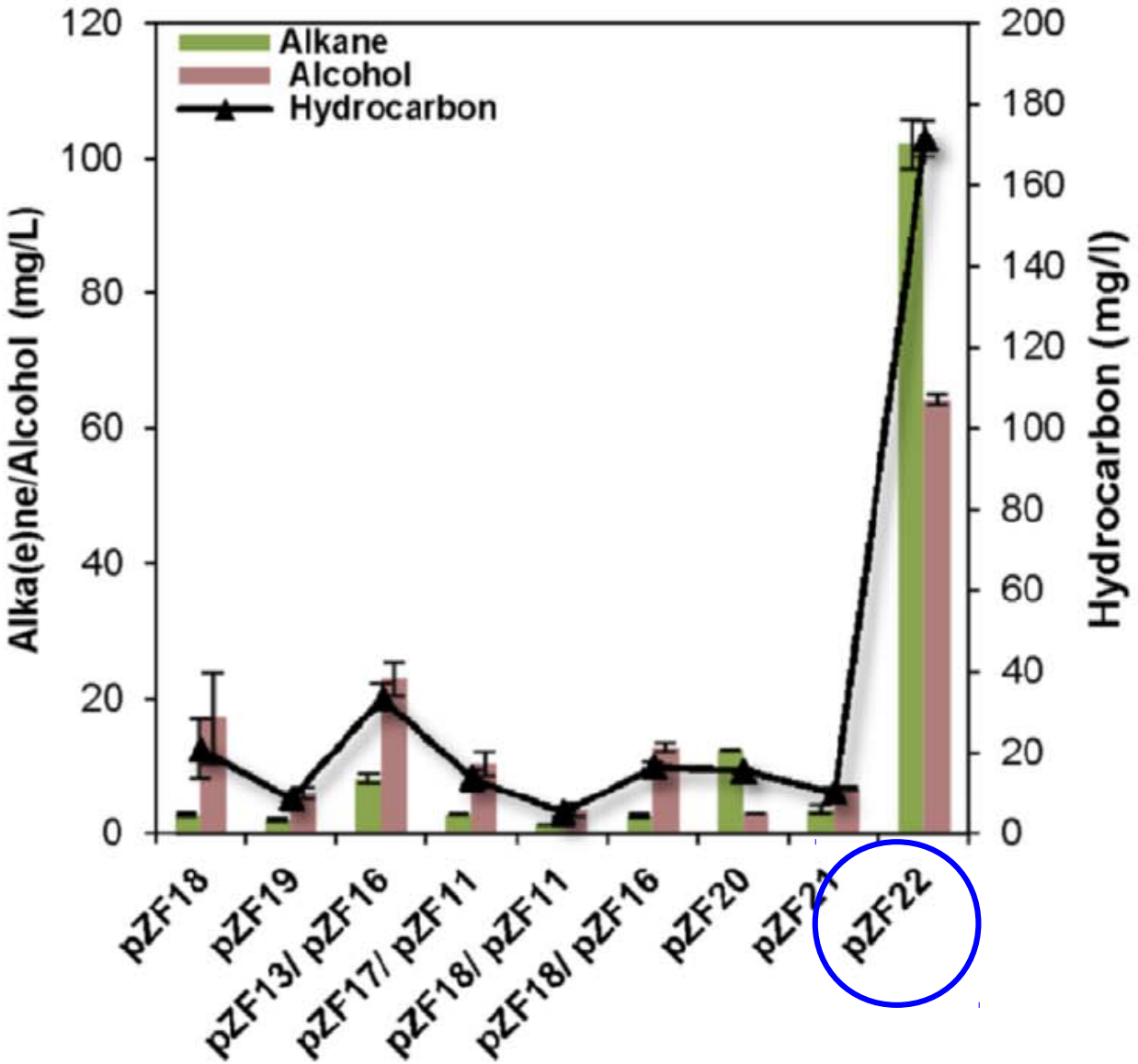
  palmitoyl-ACP + NADPH + H^+ -> palmitaldehyde + $NADP^+$ + ACP

- YbbO, $NADP^+$-dependent aldehyde reductase, EC 1.1.1.2 from E. coli, overexpressed:
  long-chain aldehyde + NADPH + H^+ -> long-chain alcohol + $NADP^+$

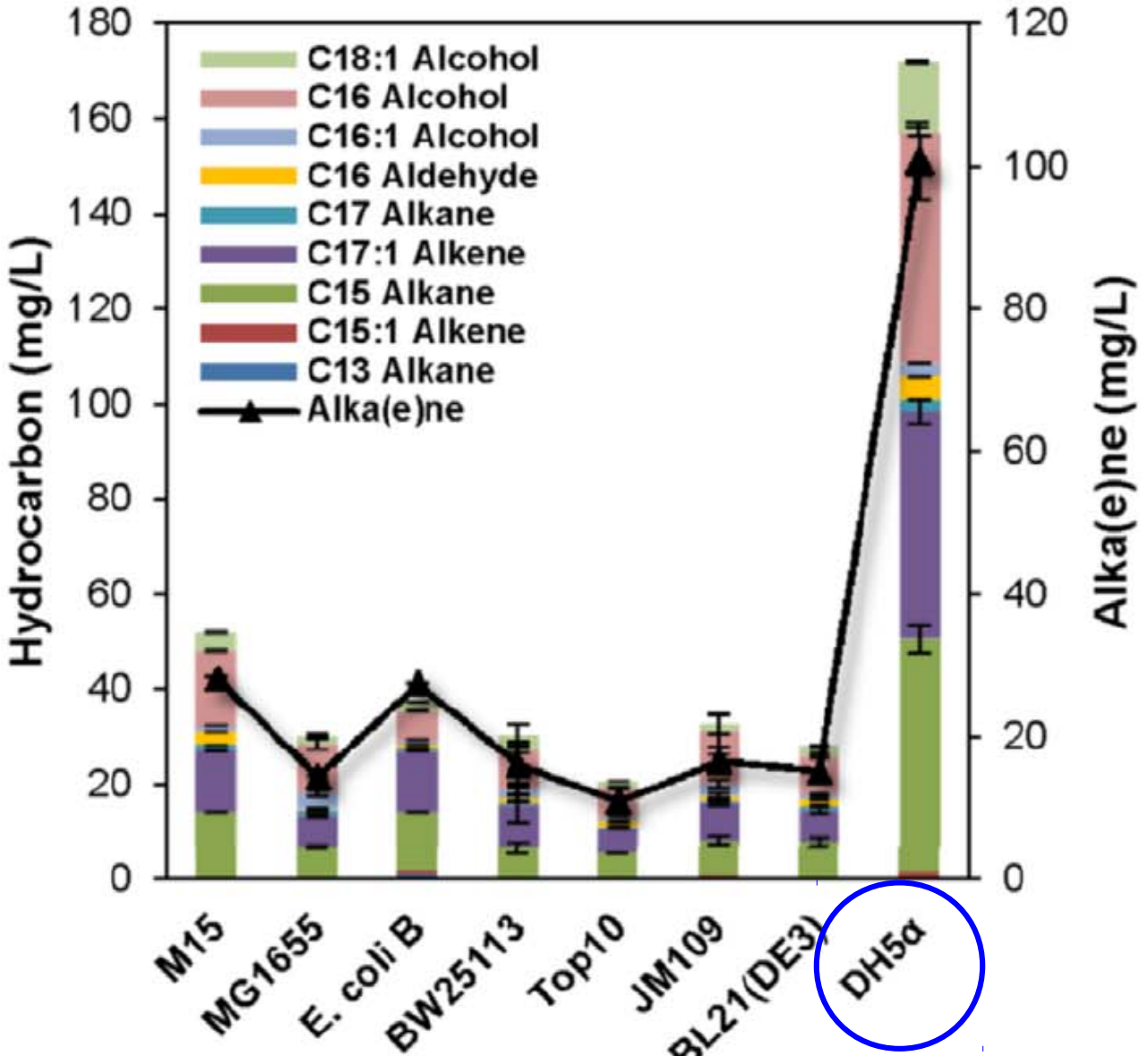
Optimisation at ICGEB

- Different options of enzyme sources (after codon optimisation) and promoters were explored via expression in medium-copy number plasmids.
- A fusion protein of AAR and ADO was tested but not found better than enhanced expression of separate genes with T5 promoters.
- For fatty alcohols, expression of AAR and YbbO from T5 promoters was found better than expression of AAR alone and reliance on native actvity of *E coli* alcohol dehydrogenases.

Outcome for Alkanes



Outcome for Strain Selection



Modelling for Improved Productivity

Metabolic Modelling Methods

Structural modelling techniques

- - need an accurate reaction list from which to generate a stoichiometry matrix; assume metabolic steady state.
- - show existence (and number) of feasible metabolic routes; optimal conversion stoichiometries; network flux values.

Metabolic Modelling Methods

Kinetic modelling techniques

- - need a reaction list **and** full kinetic description of each enzyme/step.
- - predict time-courses, steady-state values of reaction fluxes **and** metabolite concentrations.
- - allows sensitivity analysis (Metabolic Control Analysis) to compute dependence of fluxes and concentrations on enzyme activities.

Structural Modelling Methods

Elementary modes analysis

- - all feasible routes (modes) through a network from nutrients to metabolic products;
- - network flux values and product yields;
- - good for designing knock-out strategies to eliminate metabolism to unwanted products;
- - computationally limited to small to medium sized metabolic networks.

(Schuster, Dandekar & Fell, 1999, 2000)

Structural Modelling Techniques

Linear programming (LP or Flux Balance Analysis)

- - incorporates known metabolic properties, such as nutrient uptake rates, as constraints;
- - computationally feasible even on the largest (genome-scale) metabolic models;
- - can be used to design over-expression strategies for increasing productivity;
- - determines the optimal network route to achieve a specified metabolic objective;
- - several techniques for design of knock-out strategies, though very large models produce less clear results;
- Basic method only produces a single solution; finding multiple optima or near-optimal solutions is more complex.

(Fell & Small, 1985; Varma & Palsson, 1993)

Our Modelling Approach

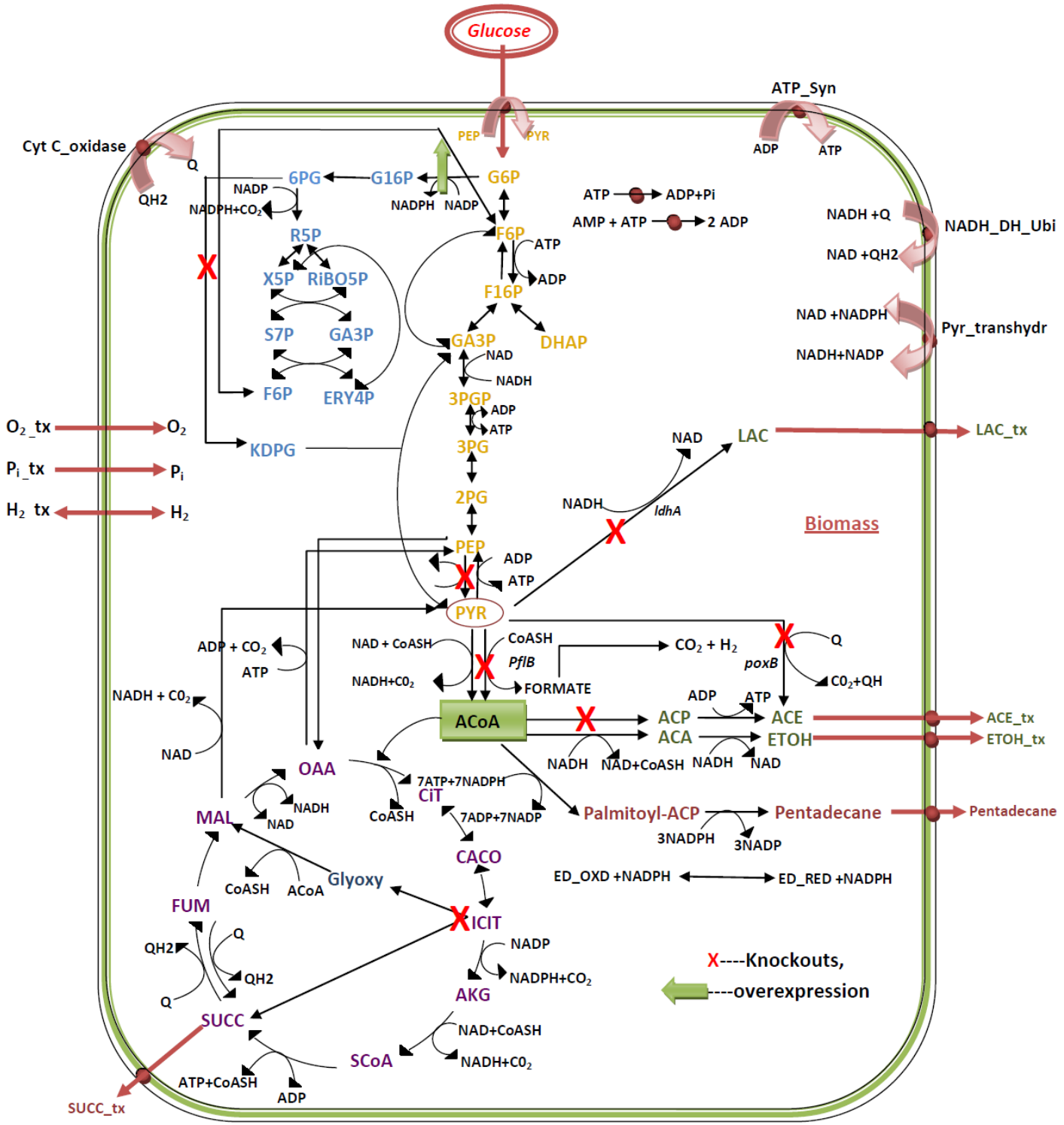
Though an elementary modes analysis would potentially have been feasible:

- As alkanes are not catabolic products, merely cutting our routes to other products would not necessarily induce alkane synthesis, and
- We expected to have to over-express parts of central carbon metabolism to supply enough substrate to allow significantly increased flux through the fatty acid synthesis pathway.

We therefore opted for a linear programming approach.

The Model

- The model was based on a central carbon metabolism (CCM) model developed by Trinh, Unrean & Srienc (2008).
- It was reconstructed using *ScrumPy* from the EcoCyc database, plus the additional heterologous reactions to alkane.
- A single, representative alkane – pentadecane – was modelled as output; the carbon source was glucose in aerobic conditions.
- Growth of the cell was modelled by withdrawal of a small set of CCM intermediates at appropriate rates established from larger-scale *E coli* models.
- The model has 74 reactions and 61 metabolites.


Glucose
PEP
PYR
ATP_Syn
ADP
ATP
Cyt C_oxidase
Q
QH2
6PG
G16P
G6P
NADP
NADPH+CO2
NADPH
NADP
R5P
X5P
RiBO5P
S7P
GA3P
F6P
ERY4P
F6P
ATP
ADP
F16P
GA3P
DHAP
NAD
NADH
3PGP
ADP
ATP
3PG
2PG
PEP
ADP
ATP
PYR
ATP
ADP+Pi
AMP + ATP
2 ADP
NADH +Q
NADH_DH_Ubi
NAD +QH2
NAD +NADPH
Pyr_transhydr
NADH+NADP
O2_tx
O2
Pi_tx
Pi
H2 tx
H2
KDPG
NAD
LAC
LAC_tx
NADH
ldhA
Biomass
ADP + CO2
ATP
NAD + CoASH
CoASH
PflB
CO2 + H2
poxB
Q
CO2+QH
NADH+C02
FORMATE
NADH + C02
NAD
ACoA
ACP
ACA
ADP
ATP
ACE
ETOH
ACE_tx
ETOH_tx
NADH
NAD+CoASH
NADH
NAD
OAA
7ATP+7NADPH
CiT
CoASH
7ADP+7NADP
MAL
NADH
NAD
Palmitoyl-ACP
Pentadecane
Pentadecane
3NADPH
3NADP
CACO
ED_OXD +NADPH
ED_RED +NADPH
CoASH
ACoA
Glyoxy
FUM
ICIT
Q
QH2
QH2
Q
NADP
NADPH+CO2
AKG
SUCC
NAD+CoASH
SCoA
NADH+C02
ATP+CoASH
ADP
SUCC_tx
X----Knockouts,
----overexpression

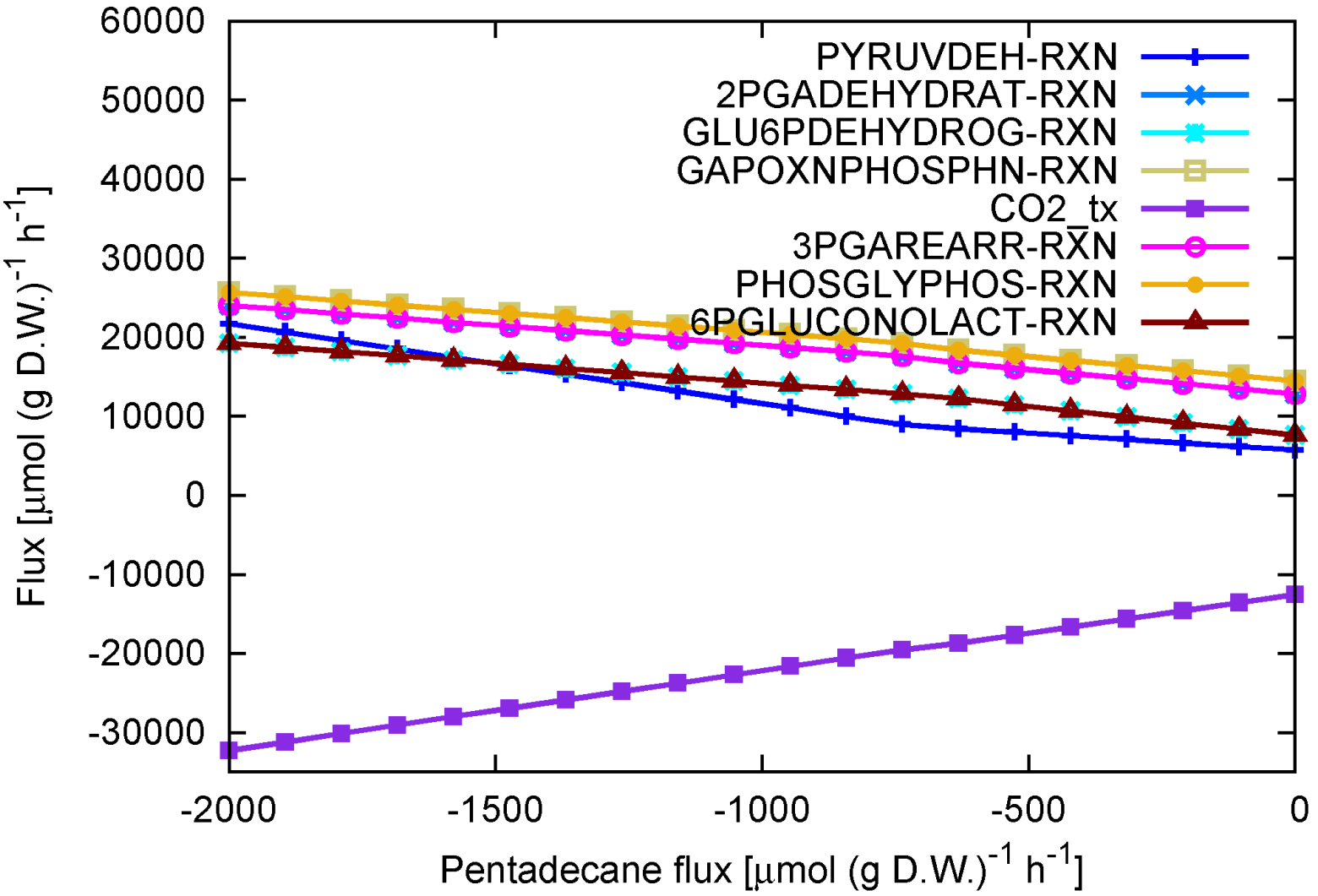
Initial Model Analysis

1. The model was checked for stoichiometric and energetic consistency.
2. LP was used (with *ScrumPy*) to check the ability to supply biomass precursors from glucose at a rate equivalent to a growth rate of 1 g dcw.h^{-1} (as a constraint), with minimisation of total flux in the network as the optimisation criterion.
3. The rate of glucose uptake was then set as a constraint at twice its value in the previous solution to model the fate of excess carbon intake. Only acetate and lactate were formed in addition; no pentadecane.

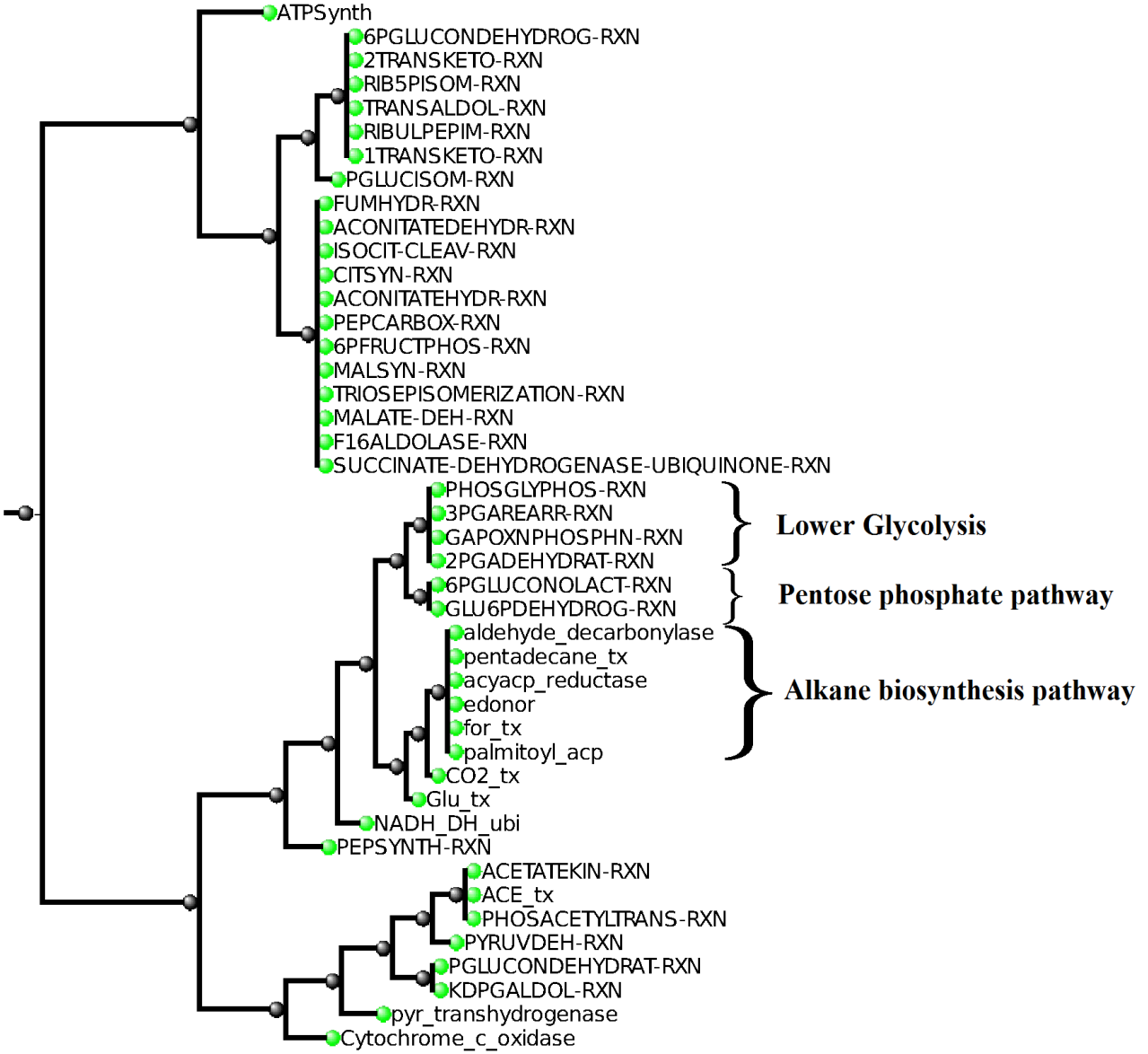
Constraint Scanning

- The model can be solved for simultaneous production of biomass and pentadecane by imposing these as constraints.
- However, if we use ScrumPy to compute a series of LP solutions for fixed biomass but pentadecane from 0 to 2 mmol.$(gDW.h)^{-1}$, we can see how fluxes through the network have to change to support alkane synthesis.

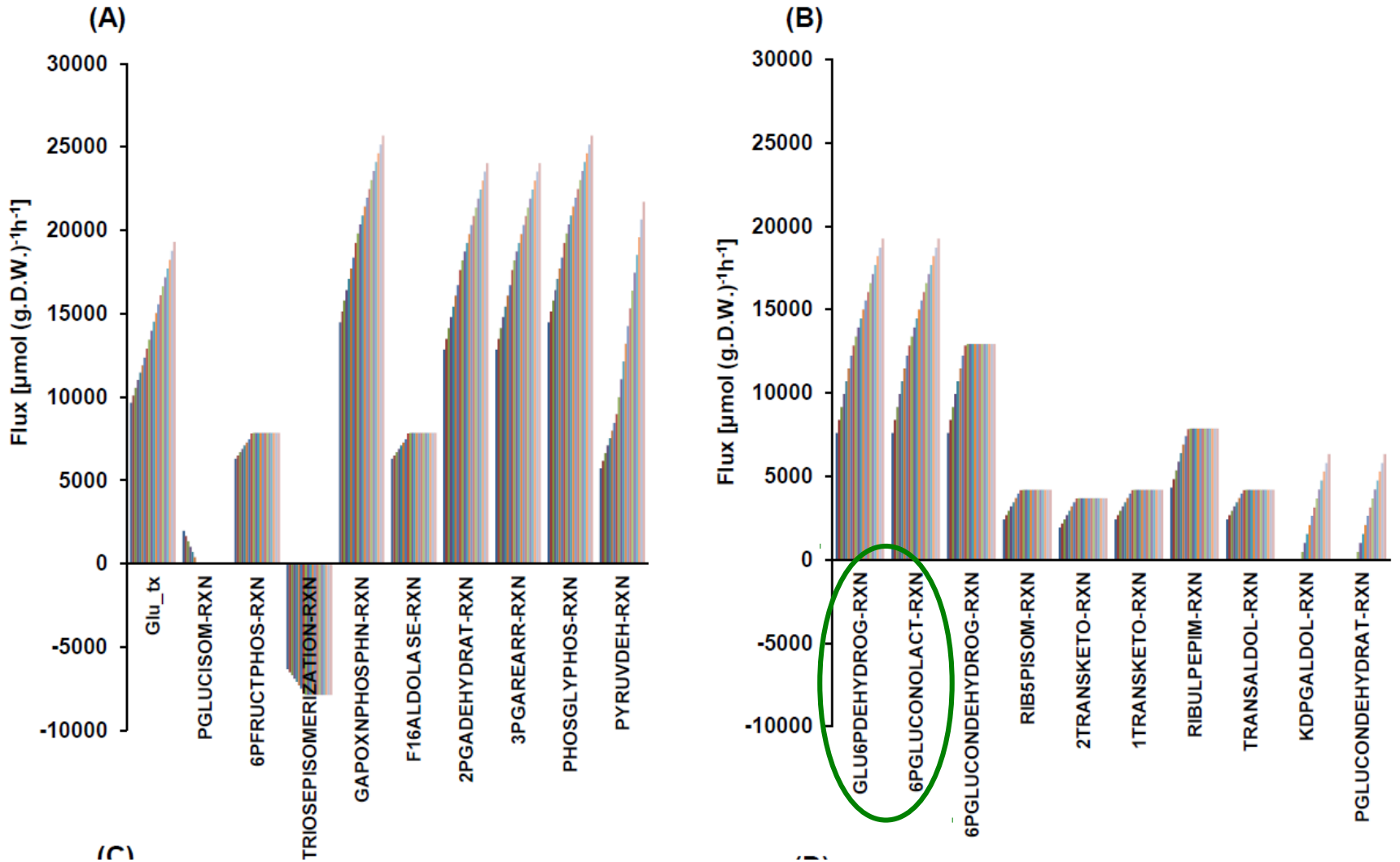
Constraint Scanning Alkane Production



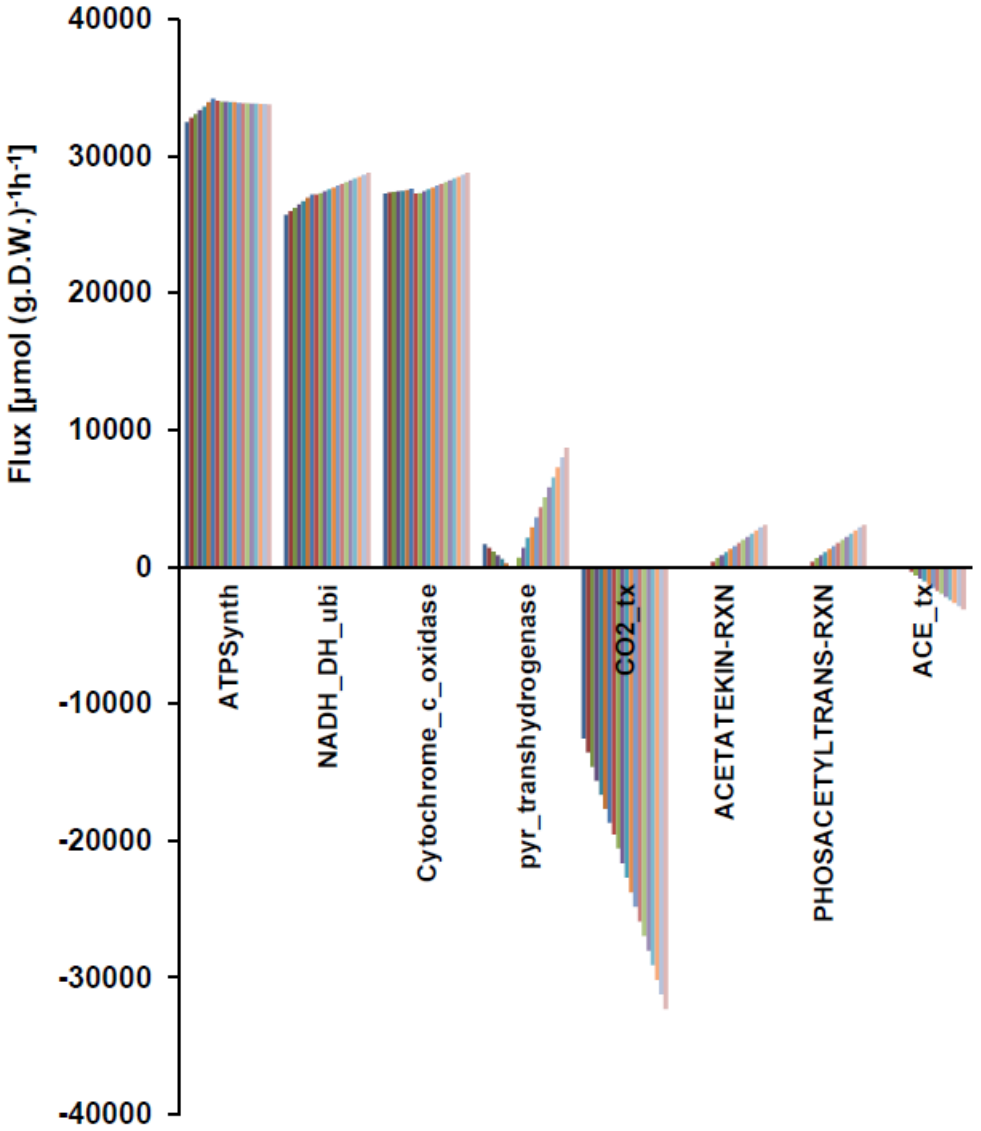
Reaction Correlations

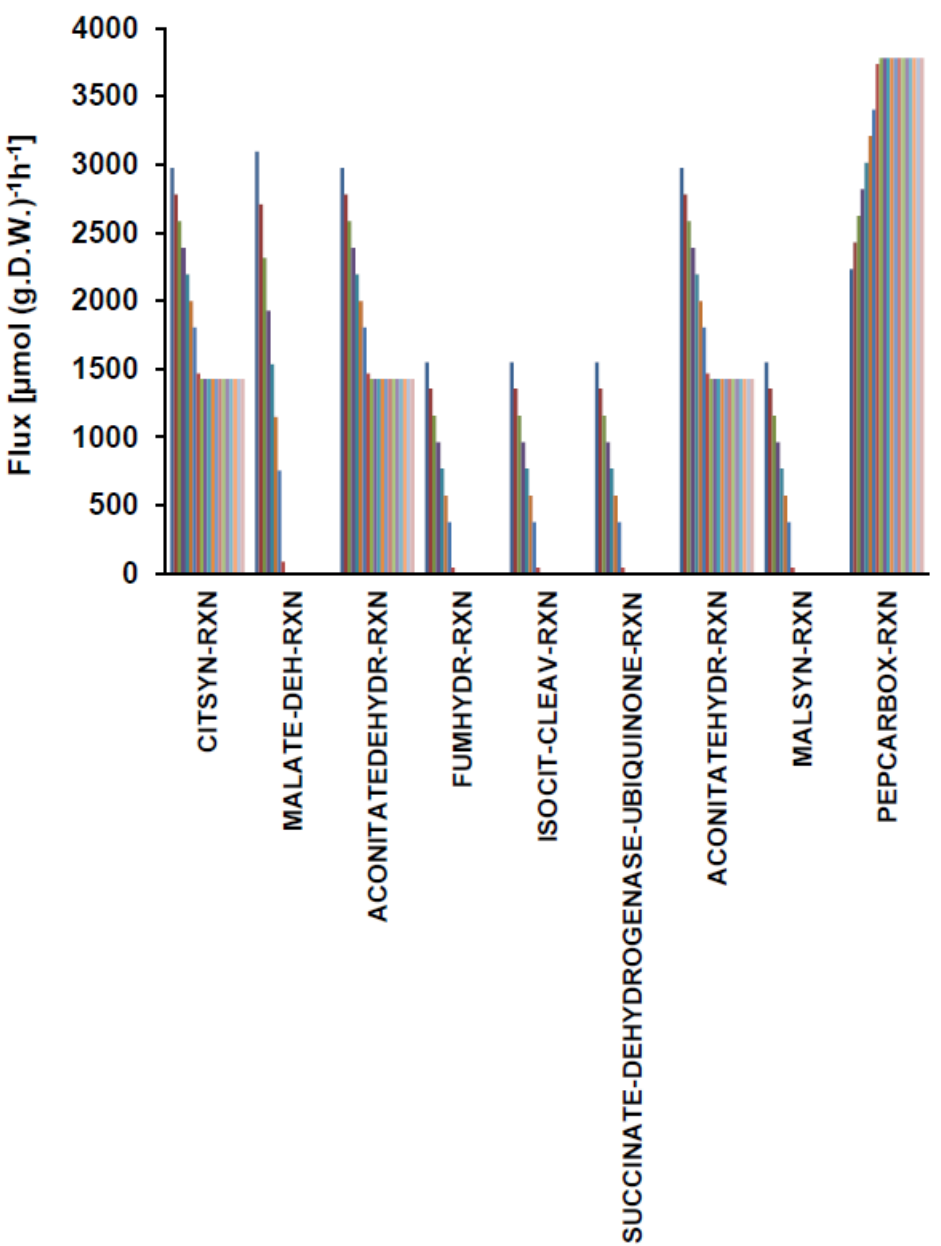


Another View ...



… continued


Flux [µmol (g.D.W.)-1h-1]
40000
30000
20000
10000
0
-10000
-20000
-30000
-40000
ATPSynth
NADH_DH_ubi
Cytochrome_c_oxidase
pyr_transhydrogenase
CO2_tx
ACETATEKIN-RXN
PHOSACETYLTRANS-RXN
ACE_tx


Flux [µmol (g.D.W.)-1h-1]
4000
3500
3000
2500
2000
1500
1000
500
0
CITSYN-RXN
MALATE-DEH-RXN
ACONITATEDEHYDR-RXN
FUMHYDR-RXN
ISOCIT-CLEAV-RXN
SUCCINATE-DEHYDROGENASE-UBIQUINONE-RXN
ACONITATEHYDR-RXN
MALSYN-RXN
PEPCARBOX-RXN

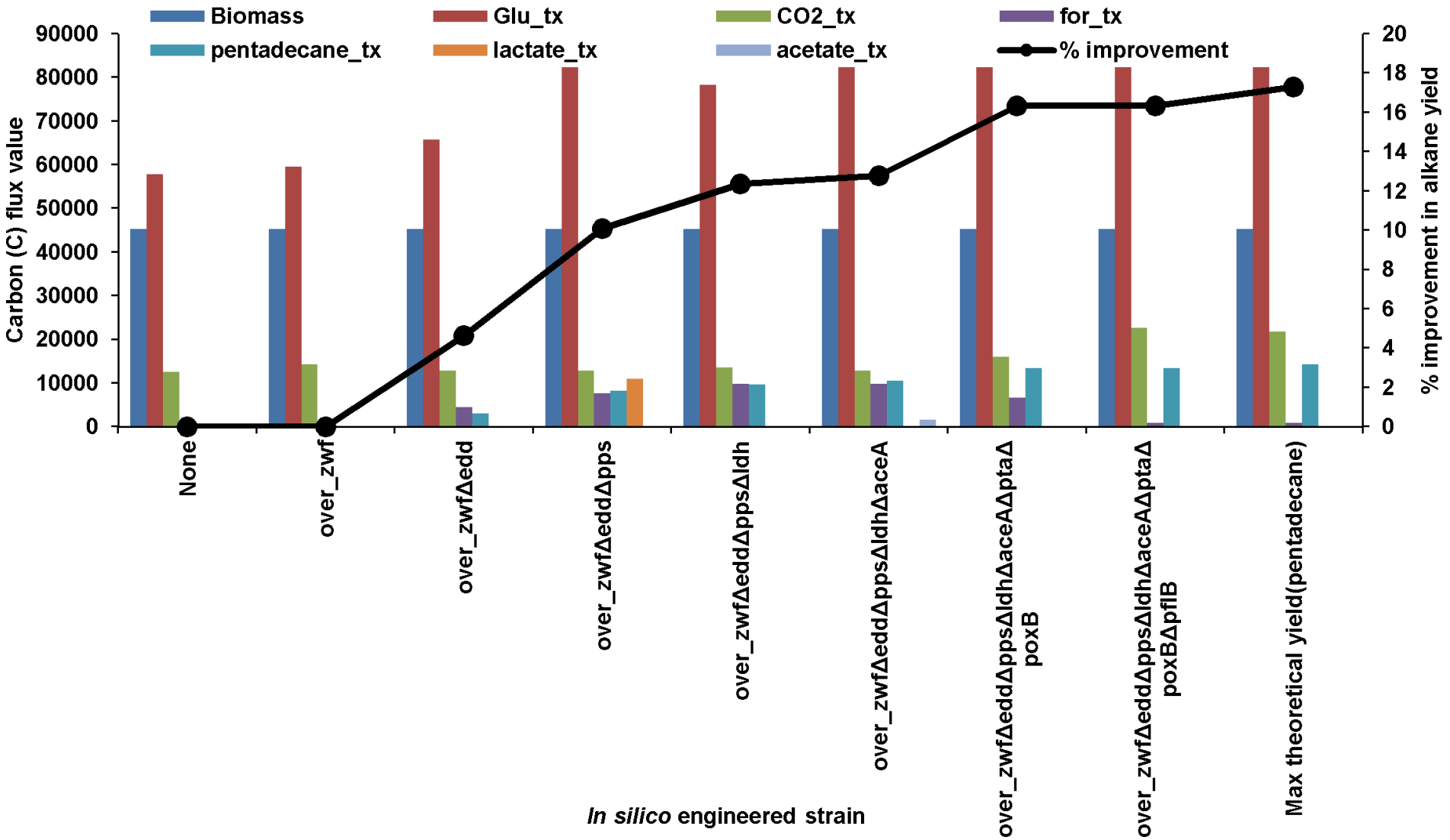
G6PDH Over-Expression

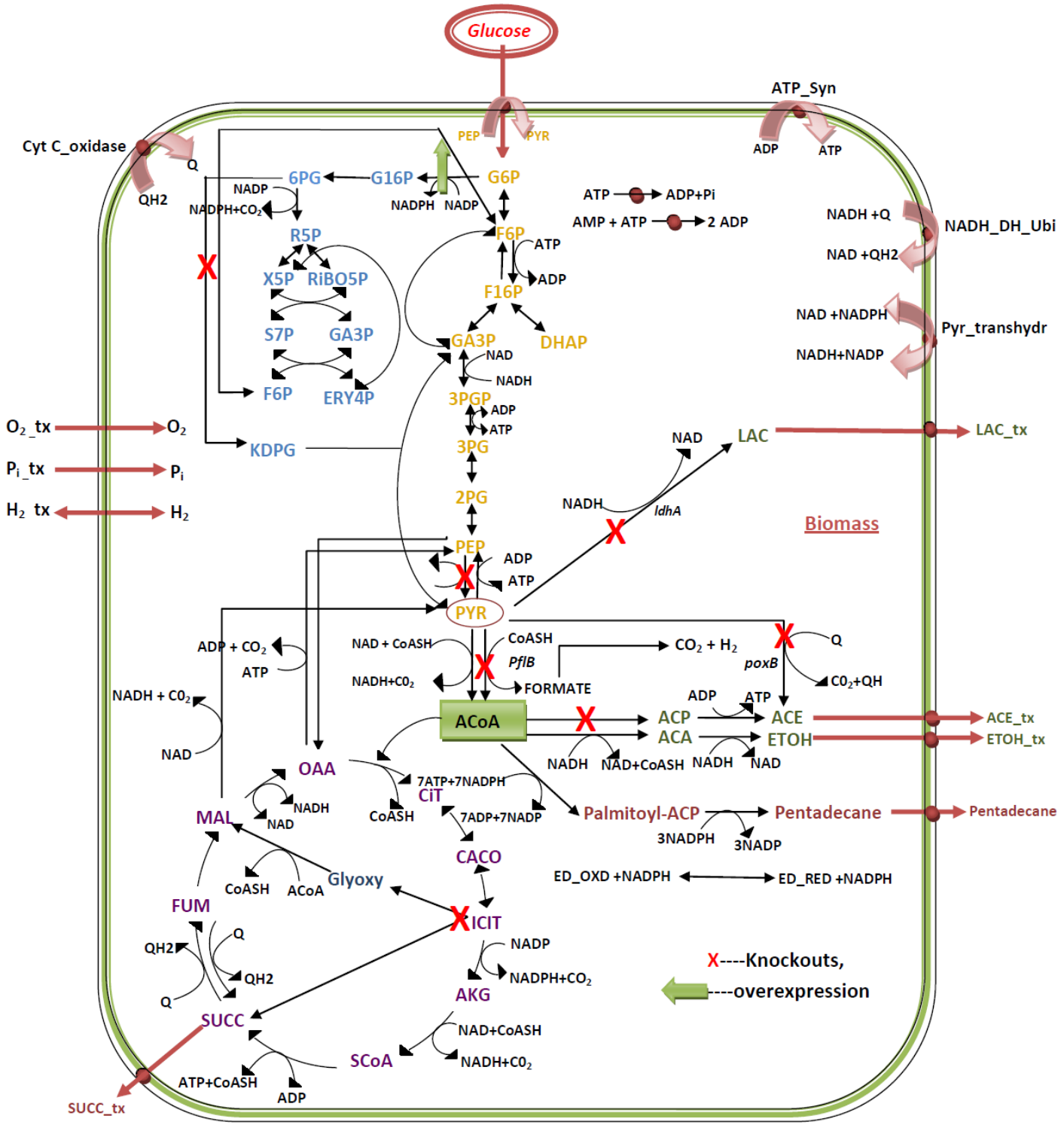
- On the basis that large relative increases in flux with pentadecane production could indicate steps that might become limiting, glucose-6-phosphate dehydrogenase (G6PDH) was selected as a target for over-expression.
- This was represented in the model by setting its flux to twice the value needed for biomass formation. Solving the model with this as a constraint only resulted in extra flux being diverted into the Entner-Duodoroff pathway, and then from pyruvate to PEP via PEP synthase with no change in other outputs apart from CO_2.

Knock-out Selection

- As the increased C-flux was not reaching alkane, we modelled a knockout of the ED pathway by constraining the flux of the reaction catalysed by the *edd* gene product to zero, resulting in small amounts of formate and alkane.
- This was followed by deleting the PEP synthase reaction (Δ*pps*), which resulted in larger amounts of formate, alkane and lactate.
- As additional products appeared, we modelled additional knock-outs in the branches leading to them until the only products were biomass, CO_2 and alkane.

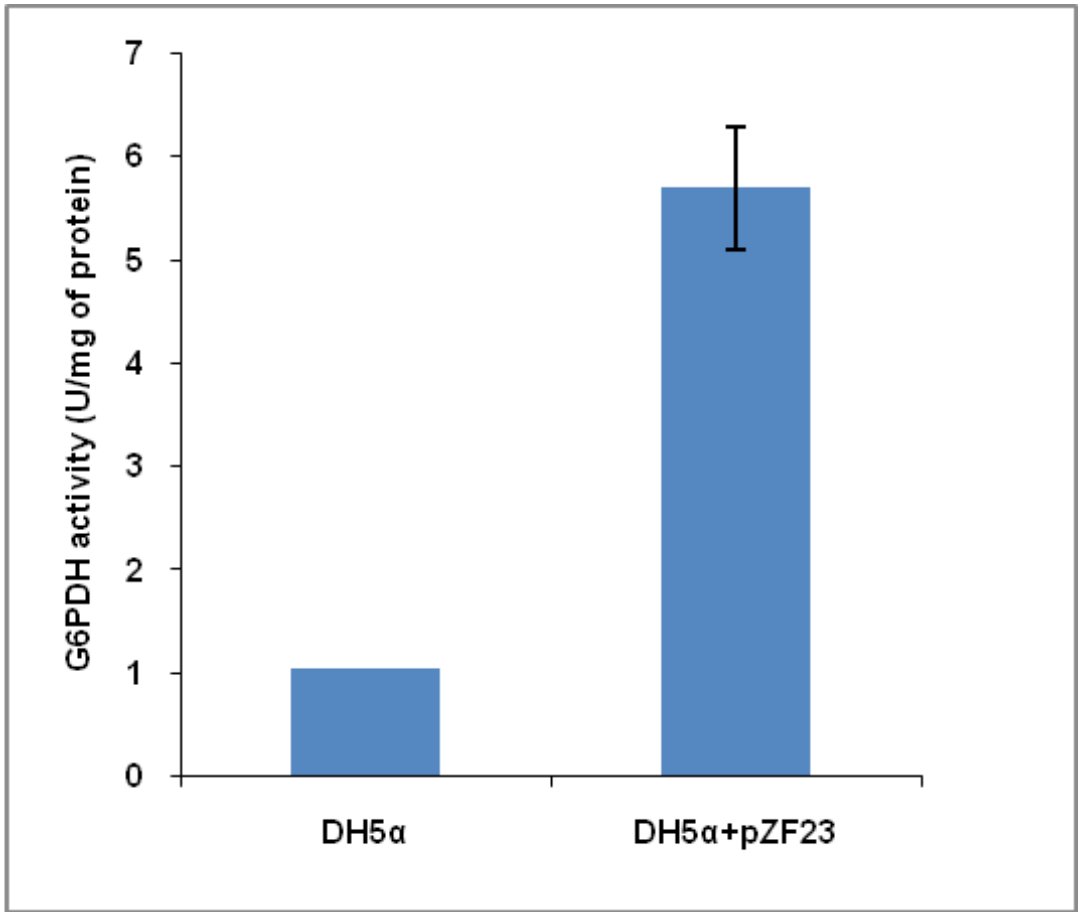
Knock-outs for Alkane production



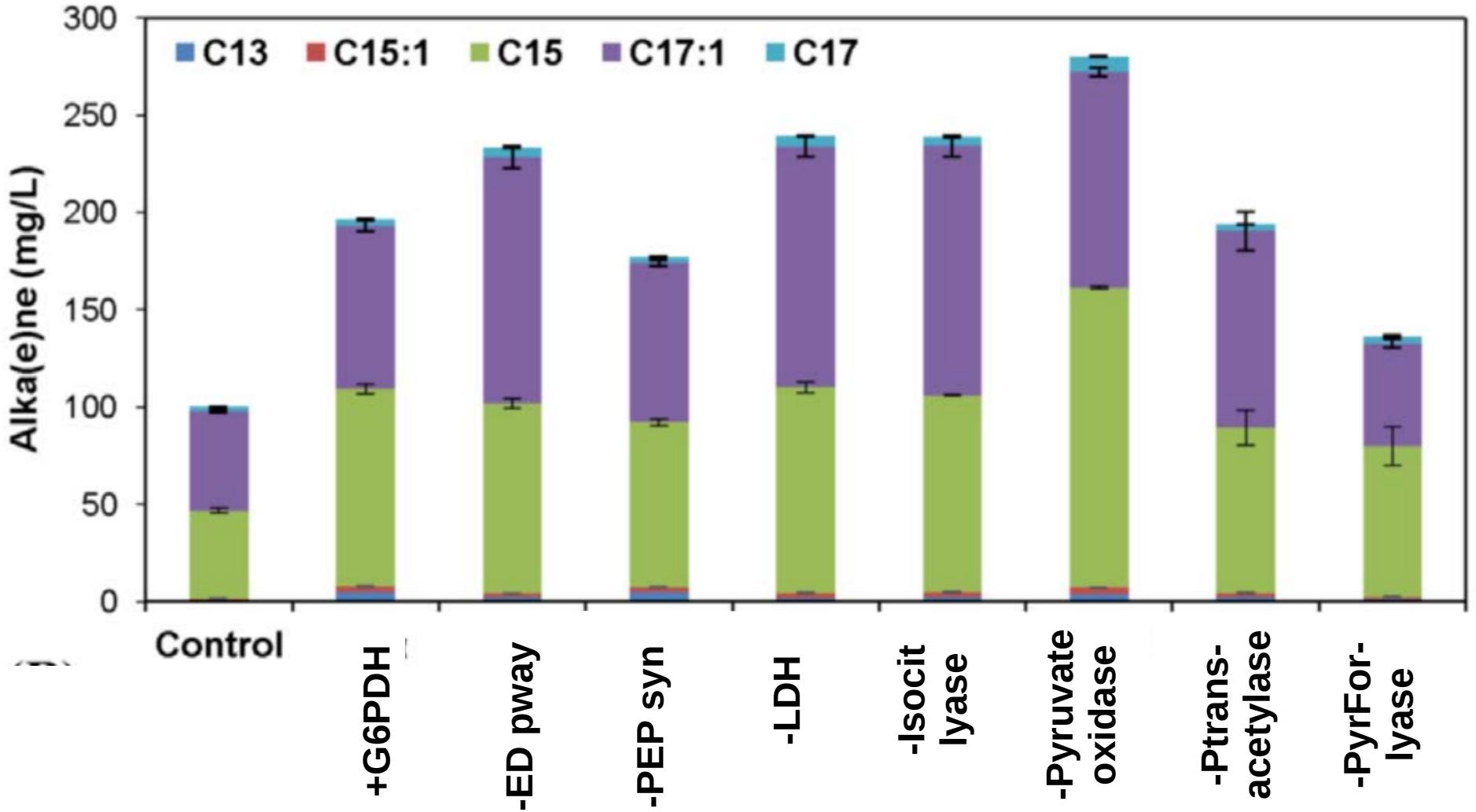

Glucose
PEP
PYR
ATP_Syn
ADP
ATP
Cyt C_oxidase
Q
QH2
6PG
G16P
G6P
NADP
NADPH+CO2
NADPH
NADP
R5P
X5P
RiBO5P
S7P
GA3P
F6P
ERY4P
F6P
ATP
ADP
F16P
GA3P
DHAP
NAD
NADH
3PGP
ADP
ATP
3PG
2PG
PEP
ADP
ATP
PYR
ATP
ADP+Pi
AMP + ATP
2 ADP
NADH +Q
NADH_DH_Ubi
NAD +QH2
NAD +NADPH
Pyr_transhydr
NADH+NADP
O2_tx
O2
Pi_tx
Pi
H2 tx
H2
KDPG
NAD
LAC
LAC_tx
NADH
ldhA
Biomass
ADP + CO2
ATP
NADH + C02
NAD
NAD + CoASH
CoASH
PflB
NADH+C02
CO2 + H2
FORMATE
poxB
Q
C02+QH
ADP
ATP
ACoA
ACP
ACA
ACE
ETOH
ACE_tx
ETOH_tx
NADH
NAD+CoASH
NADH
NAD
OAA
7ATP+7NADPH
CiT
CoASH
7ADP+7NADP
MAL
NADH
NAD
Palmitoyl-ACP
Pentadecane
3NADPH
3NADP
Pentadecane
CACO
ED_OXD +NADPH
ED_RED +NADPH
Glyoxy
CoASH
ACoA
FUM
ICIT
Q
QH2
QH2
Q
NADP
NADPH+CO2
AKG
SUCC
NAD+CoASH
SCoA
NADH+C02
ATP+CoASH
ADP
SUCC_tx
X----Knockouts,
----overexpression

Experimental Implementation

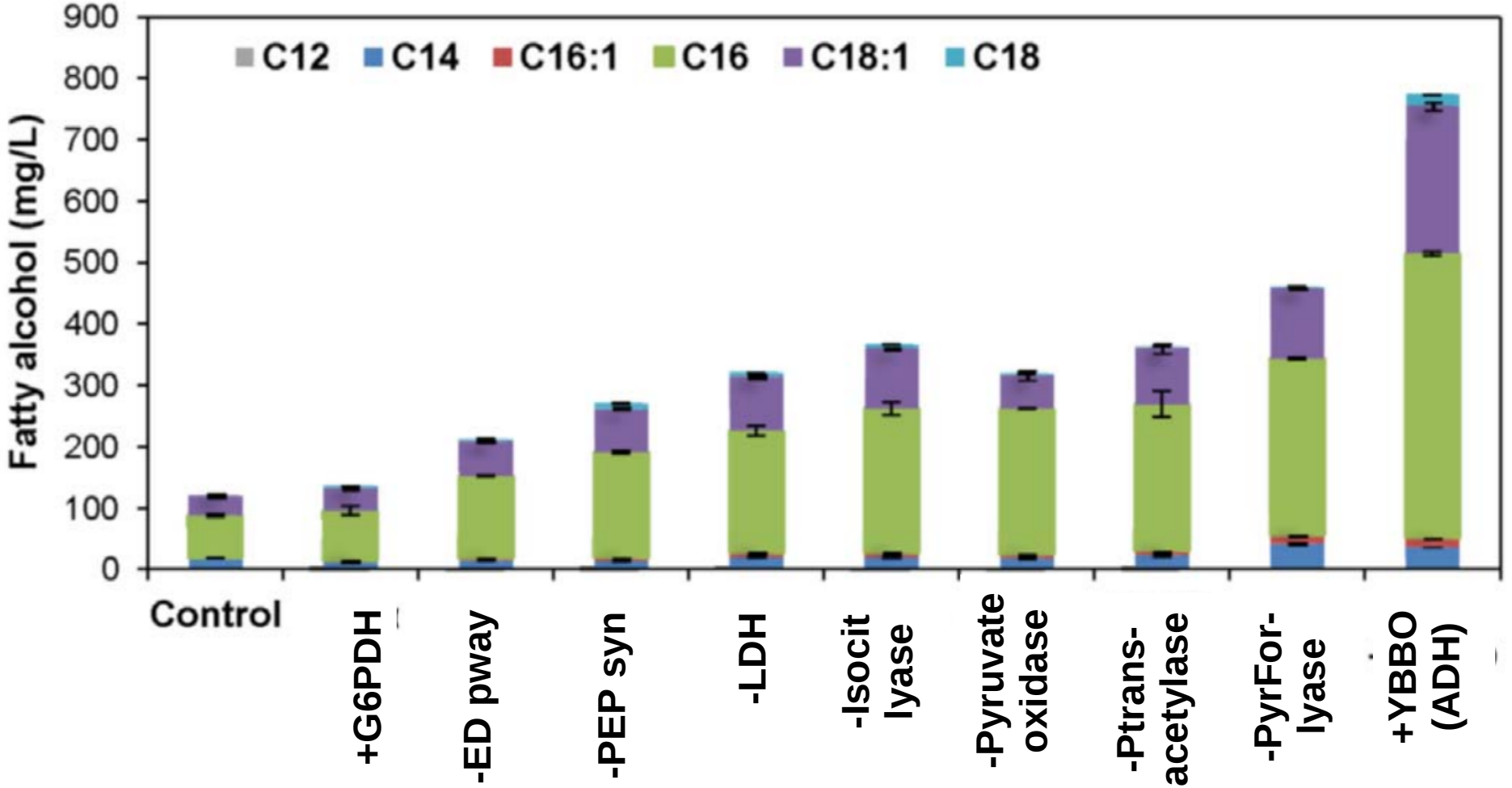
G6PDH Over-expression



Alkane Yields



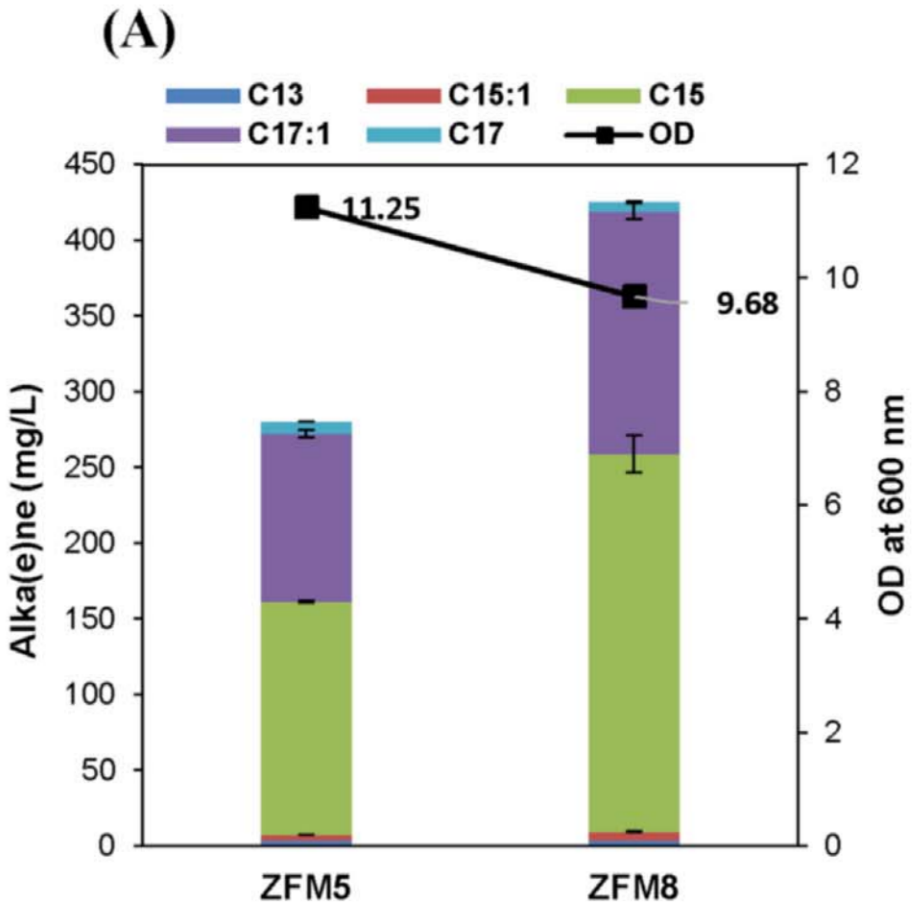
Fatty Alcohol Production

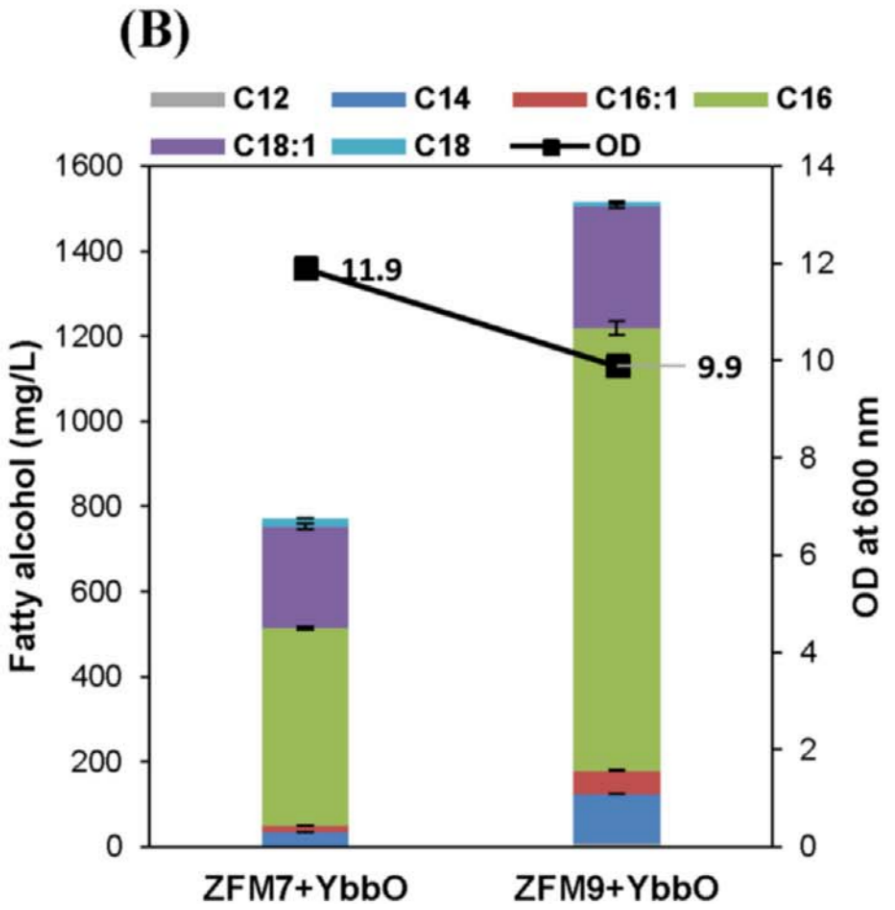


Final Improvement

- In the final strain so far, the carbon yield from glucose was still less than the theoretical maximum attainable with simultaneous biomass synthesis.
- Analysis of the flux values in the model solutions showed that the main competitive flux for alkane was now the phospholipid synthesis for biomass.
- This was therefore attenuated by a non-lethal knockout of a phospholipid pathway gene for the final producing strain.

Increase in Yield by Growth Attenuation





Overall Result

- The initial strain optimisation by ICGEB improved alkane production from an initial 2.8 mg/L to 102 mg/L.
- The further model-designed improvements led to a further increase to 425 mg/L.
- The productivity of fatty alcohols was increased even further to 1500 mg/L.
- These strains, when tested in fed-batch bioreactors produced 2.54 g/L alkane and 12.5 g/L fatty alcohol – the highest alkane yields yet reported for *E coli*.

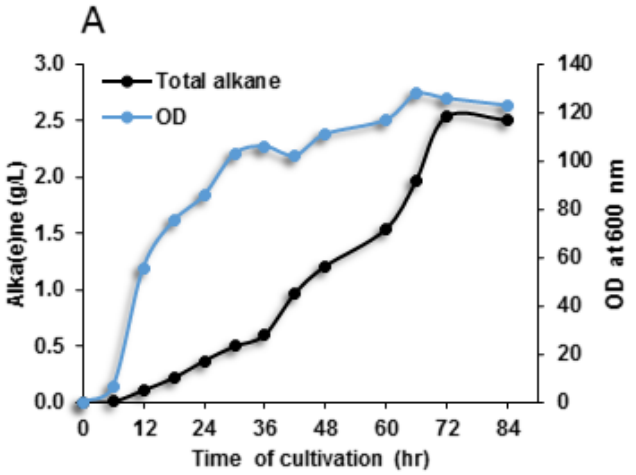
Conclusions

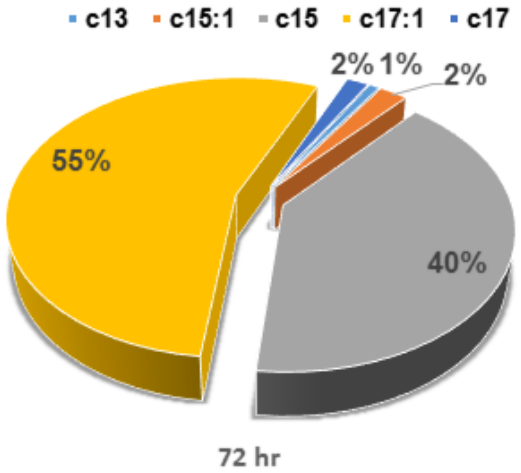
- Theory-led design can narrow the search space for metabolic engineering.

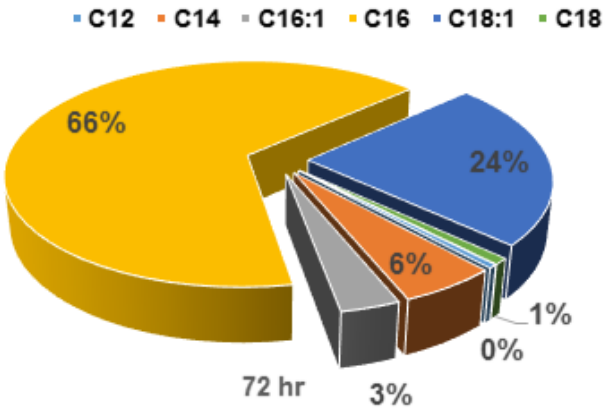
- There is still a long way to go to achieve commercial alkane production by microorganisms from sustainable C-sources.

- Fatma, Z. et al Model-assisted metabolic engineering of Escherichia coli for long chain alkane and alcohol production, Metabolic Engineering, **45**, 134-141 (2018).

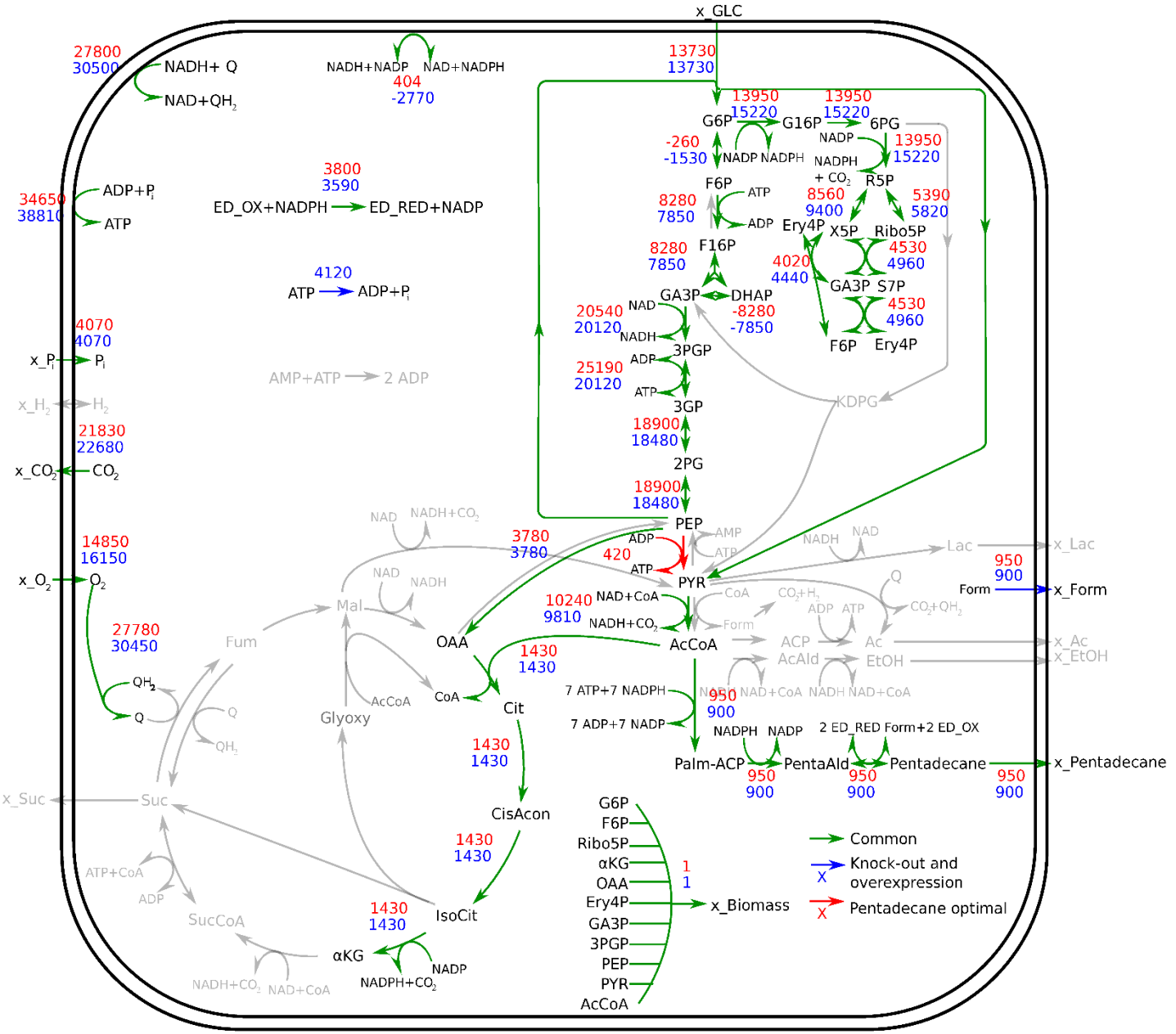
Growth curves

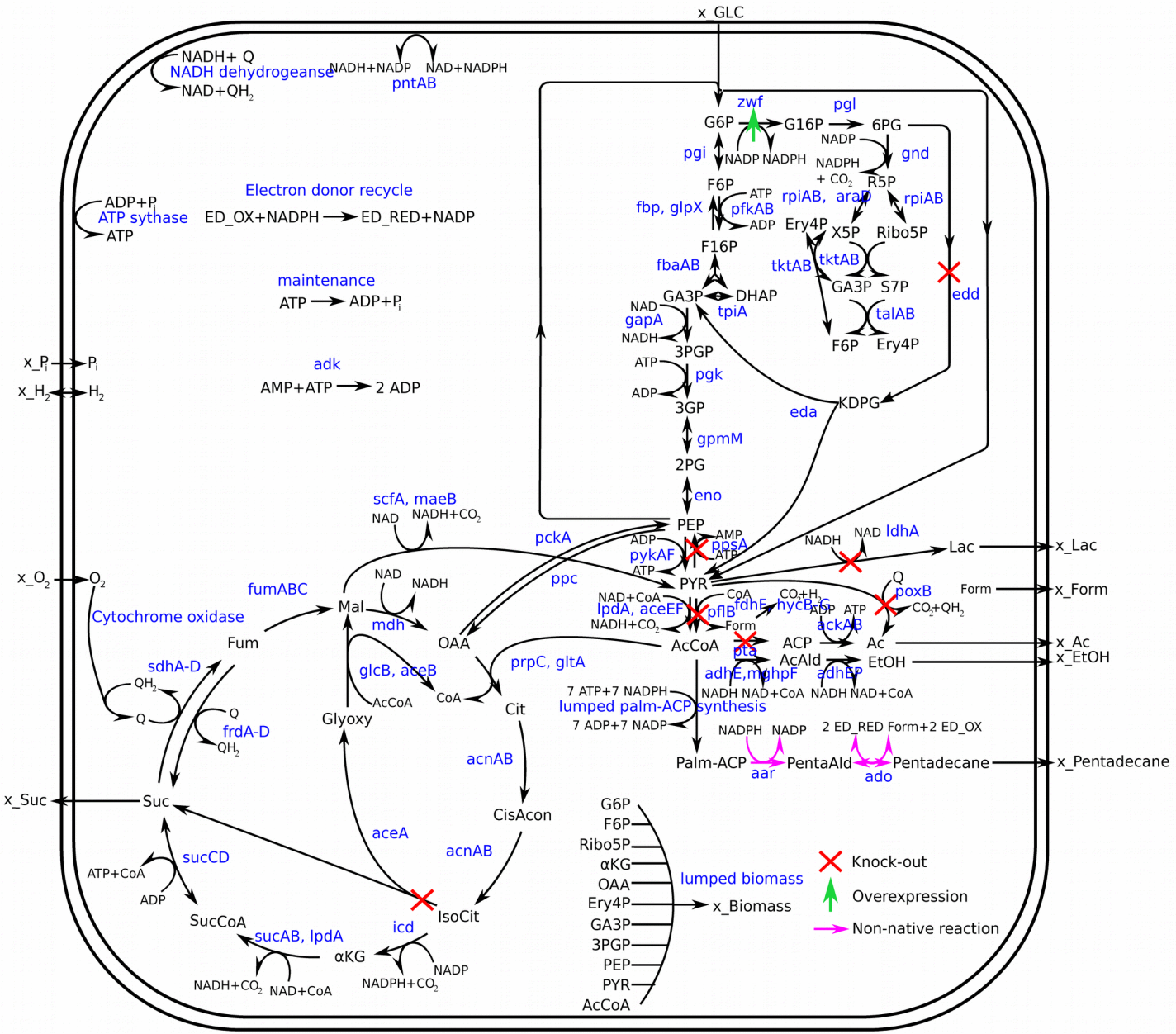






Flux Patterns




x_GLC
NADH+ Q
NADH dehydrogeanse
NAD+QH2
NADH+NADP
NAD+NADPH
pntAB
zwf
pgl
G6P
G16P
6PG
NADP
pgi
NADP NADPH
NADPH
+ CO2
gnd
R5P
F6P
ATP
fbp, glpX
pfkAB
ADP
rpiAB, araD
rpiAB
Ery4P
X5P
Ribo5P
F16P
fbaAB
tktAB
tktAB
GA3P S7P
edd
GA3P
DHAP
tpiA
NAD
gapA
NADH
talAB
F6P
Ery4P
3PGP
ATP
pgk
ADP
3GP
eda
KDPG
gpmM
2PG
eno
PEP
scfA, maeB
NAD
NADH+CO2
pckA
ADP
AMP
ppsA
ATP
pykAF
ATP
NADH
NAD
ldhA
Lac
x_Lac
x_P
Pi
x_H2
H2
adk
AMP+ATP
2 ADP
x_O2
O2
fumABC
Mal
NAD
NADH
ppc
PYR
Q
poxB
Form
x_Form
Cytochrome oxidase
mdh
NAD+CoA
CoA
CO2+H2
fdhF
hycB2C
lpdA, aceEF
pflB
ADP ATP
ackAB
NADH+CO2
Form
CO2+QH2
Fum
OAA
AcCoA
ACP
Ac
x_Ac
pta
AcAld
EtOH
x_EtOH
sdhA-D
QH2
glcB, aceB
prpC, gltA
adhE,mghpF
adhEP
7 ATP+7 NADPH
NADH NAD+CoA
NADH NAD+CoA
Q
AcCoA
CoA
Cit
lumped palm-ACP synthesis
7 ADP+7 NADP
frdA-D
QH2
Glyoxy
NADPH NADP
2 ED_RED Form+2 ED_OX
Palm-ACP
aar
PentaAld
ado
Pentadecane
x_Pentadecane
acnAB
x_Suc
Suc
CisAcon
G6P
F6P
Ribo5P
αKG
OAA
Ery4P
GA3P
3PGP
PEP
PYR
AcCoA
lumped biomass
x_Biomass
aceA
acnAB
sucCD
ATP+CoA
ADP
SucCoA
IsoCit
icd
sucAB, lpdA
αKG
NADH+CO2
NAD+CoA
NADPH+CO2
NADP
Knock-out
Overexpression
Non-native reaction