Model-assisted metabolic engineering of *Escherichia coli* for long chain alkane and alcohol production

Zia Fatma[a,c], Hassan Hartman[d], Mark G. Poolman[d], David A. Fell[d], Shireesh Srivastava[b,c], Tabinda Shakeel[a,c] and Syed Shams Yazdani[a,c]*

[a]*Microbial Engineering Group,* [b]*Systems Biology for Biofuel Group,* [c]*DBT-ICGEB Centre for Advanced Bioenergy Research, International Centre for Genetic Engineering and Biotechnology, New Delhi, India;* [d]*Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK*

*Corresponding Author

Address of correspondence: Microbial Engineering Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi-110067, India. Email – shams@icgeb.res.in; Phone - +91 11 26742357; Fax - +91 11 26742316.

Abstract

Biologically-derived hydrocarbons are considered to have great potential as next-generation biofuels owing to the similarity of their chemical properties to contemporary diesel and jet fuels. However, the low yield of these hydrocarbons in biotechnological production is a major obstacle for commercialization. Several genetic and process engineering approaches have been adopted to increase the yield of hydrocarbon, but a model driven approach has not been implemented so far. Here, we applied a constraint-based metabolic modeling approach in which a variable demand for alkane biosynthesis was imposed, and co-varying reactions were considered as potential targets for further engineering of an *E. coli* strain already expressing cyanobacterial enzymes towards higher chain alkane production. The reactions that co-varied with the imposed alkane production were found to be mainly associated with the pentose phosphate pathway (PPP) and the lower half of glycolysis. An optimal modeling solution was achieved by imposing increased flux through the reaction catalyzed by glucose-6-phosphate dehydrogenase (*zwf)* and iteratively removing 7 reactions from the network, leading to an alkane yield of 94.2 % of the theoretical maximum conversion determined by *in silico* analysis at a given biomass rate. To validate the *in silico* findings, we first performed pathway optimization of the cyanobacterial enzymes in *E. coli* via different dosages of genes, promoting substrate channelling through protein fusion and inducing substantial equivalent protein expression, which led to a 36-fold increase in alka(e)ne production from 2.8 mg/L to 102 mg/L. Further, engineering of *E. coli* based on *in silico* findings, including biomass constraint, led to an increase in the alka(e)ne titer to 425 mg/L (major components being 249 mg/L pentadecane and 160 mg/L heptadecene), a 148.6-fold improvement over the initial strain, respectively; with a yield of 34.2 % of the theoretical maximum. The impact of model-assisted engineering was also tested for the production of long chain fatty alcohol, another commercially important molecule sharing the same pathway while differing only at the terminal reaction, and a titer of 1506 mg/L was achieved with a yield of 86.4 % of the theoretical maximum. Moreover, the model assisted engineered strains had

produced 2.54 g/L and 12.5 g/L of long chain alkane and fatty alcohol, respectively, in the bioreactor under fed-batch cultivation condition. Our study demonstrated successful implementation of a combined *in silico* modeling approach along with the pathway and process optimization in achieving the highest reported titers of long chain hydrocarbons in *E. coli*.

1. Introduction

Biological production of long chain hydrocarbons, including alkanes and alcohols, holds great potential for generating renewable alternatives to petrochemicals. While alkanes can be used as drop-in biofuels (Choi and Lee, 2013; Schirmer et al., 2010), long chain alcohols (>C12) are value added chemicals having significant uses in making detergents, surfactants, lubricants, additives for gasoline, etc. (Cao et al., 2015; Fatma et al., 2016; Liu et al., 2014; Liu et al., 2016; Sheng et al., 2016). The limitation, however, lies at the production level of these molecules. Several different pathways have been explored for microbial production of both long chain alkanes and alcohols, the most common of those being utilizing fatty acid synthesis type I and type II pathways (Fatma et al., 2016; Haushalter et al., 2015). Various well characterized heterologous hosts like *Escherichia coli*, *Saccharomyces cerevisiae* and *Yarrowia lipolytica* have been used for successful production of a diverse range of chemicals including drop-in transportation fuels (Buijs et al., 2015; Cao et al., 2015; Cao et al., 2016; d'Espaux et al., 2017; Fatma et al., 2016; Runguphan and Keasling, 2014; Xu et al., 2016). Amongst these hosts, *E. coli* has been the organism of choice for producing these molecules because of the better understanding of its cellular metabolic network in comparison to eukaryotic hosts. *Yarrowia lipolytica* has been extensively engineered for fatty alcohol and alkane production. It has a very high capacity for lipid synthesis, though diverting this to fuel molecules required extensive rewiring of the acyl-ACP and acyl-CoA metabolism and the decoupling of lipogenesis from the nitrogen starvation response. This modification resulted in yields of 23.3 mg/L alkane and 2.15 g/L fatty alcohol (Xu et al., 2016). Exploiting the fatty acid biosynthesis pathway of *E. coli* led to the production of odd chain alkanes as well as even chain fatty alcohols. However, even chain alkanes and odd chain alcohols have also been produced in *E. coli* by heterologously expressing a 3-oxoacyl-ACP synthase with specificity for CoA thioesters, i.e., *fabH2* from *Bacillus subtilis* (Harger et al., 2012) and an α-dioxygenase gene from *Oryza sativa* (Cao et al., 2015), respectively. The highest reported titer for long chain alkane has been 1.3 g/L (44% and 49.4% being contributed by pentadecane and

heptadecene, respectively) in the bioreactor using AAR/ADO (acyl-ACP reductase/aldehyde deformylating oxygenase) pathway (Cao et al., 2016). Expression of a different pathway involving type I polyketide synthase and a hydrogenation step has been reported to result in a pentadecane titer of 140 mg/L (Liu et al., 2015). Similarly, different titers for long chain alcohol have also been reported in *E. coli* via strain engineering and optimization. For example, 6.33 g/L (Liu et al., 2016), 1.99 g/L (Fatma et al., 2016), 350 mg/L (Akhtar et al., 2013) and 750 mg/L (Liu et al., 2014) of even chain length and 1.95 g/L of odd chain fatty alcohol (Cao et al., 2015) has been reported by exploiting the FASII pathway.

Various transcriptomic and proteomic approaches have also been exploited for improvement in hydrocarbon production (Cao et al., 2016; Liu et al., 2014; Liu et al., 2016), but a metabolic modeling approach has not been implemented thus far. Metabolic engineering efforts can be greatly facilitated by the use of metabolic modeling, since this allows the researcher to verify that a given strain design could result in a viable cell capable of maintaining essential functions (such as producing precursors for biomass and fulfilling ATP maintenance demand) as well as desired products (Fong, 2014; H Chan et al., 2013; Harder et al., 2016; Kim and Reed, 2010). Metabolic modeling is typically divided into kinetic and structural approaches. Kinetic modeling involves describing a metabolic system in terms of its reaction kinetics, stoichiometry and metabolite concentrations. Structural modeling only takes the stoichiometry and reversibility of the reactions into account. Structural modeling is commonly favoured over kinetic modeling, owing to its simplicity regarding model construction and analysis, its scalability to larger metabolic networks, and its insights into the capabilities of the original or engineered networks (Poolman et al., 2003).

Several computational methods based on structural modeling have been suggested for guiding strain engineering: e.g. OptKnock (Burgard et al., 2003; Pharkya et al., 2004); OptReg (Pharkya and Maranas, 2006); OptORF (Kim and Reed, 2010); Robust Knock (Tepper and Shlomi, 2010), and FSEOF (Choi et al., 2010). All of these methods use Flux Balance Analysis (FBA) (Fell and Small,

1986; Varma and Palsson, 1993a; Varma and Palsson, 1993b; Watson, 1986), the application of Linear Programming (LP) to metabolic models, to identify candidate reactions (and ultimately genes) for deletion or overexpression. In addition, there are methods based on elementary modes analysis (Schuster et al., 2000; Trinh et al., 2008), and the related Minimal Cut Set (MCS) analysis (Hädicke and Klamt, 2011; Klamt and Gilles, 2004) and methods derived from it, that have been employed for rational optimization of the network for biomolecule production. The various methods have different properties. Several of them can be applied to large genome-scale models, though this can make the results more complex owing to peripheral routes in the network with very limited capacity for carrying flux being proposed as feasible by-passes of potential knockouts.

We thus, followed Trinh et al. (2008) in our study in concentrating only on central carbon metabolism where reactions have the capacity to carry significant flux, but instead of elementary modes we employed an FBA based flux constraint scanning method, such as those described by Choi et al. (2010), Hartman et al. (2014) and Poolman et al. (2009), to identify targets for overexpression. We first implemented the overexpression as an additional constraint during *in silico* analysis, followed by iterative simulation for identification of further gene targets. The proposed strategy was then implemented experimentally for production of alkane and fatty alcohol. Under the controlled condition acquired through fed-batch cultivation, the engineered strains produced 2.54 g/L of alkane and 12.5 g/L of long chain fatty alcohol, the maximum titers of hydrocarbons reported so far in any microbe.

2. Materials and Methods

2.1. Construction of an E. coli central carbon metabolic model for hydrocarbon production

A model of central carbon metabolism of *E. coli* was constructed to allow heterologous hydrocarbon (alkane, via fatty aldehyde) production based on a previously published *E. coli* model (Trinh et al., 2008). In contrast to the original model, glucose was assumed to be the only carbon

source available during this analysis. Furthermore, the electron transport chain was modified as described by Hartman et al. (2014) and shown in Fig. 1 and Table S2. Reactions for the fatty acid and alkane biosynthesis pathways were incorporated in the model to reflect the production of alkane (pentadecane) in the recombinant *E. coli* strain (Table S2). The incorporated reactions for fatty acid biosynthesis (FAS) were taken from the EcoCyc database, and those for cyanobacterial alkane biosynthesis, which involved reactions catalyzed by acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO), were taken from published literature (Cintolesi et al., 2014; Eser et al., 2011; Li et al., 2012; Schirmer et al., 2010; Zhang et al., 2013).

The model was built according to our standard approach as an assembly of modules as previously described (Hartman et al., 2014; Poolman et al., 2009). The model consisted of a top-level module, which imported the other modules. These are - (A) *Automatically generated reactions*: gene names mentioned in the published model of Trinh et al (2008) were extracted and the corresponding reactions retrieved from the EcoCyc database; (B) *Biomass module*: biomass was written as a lumped reaction where all metabolites required to form 1 gram dry cellular weight per hour (gDW h^{-1}) were written in single reaction; (C) *Transporter module*: transporters were defined for the exchange of metabolites between the cell and environment including glucose, NH_3, PO_4^{2-}, and alkane. All transporters were described with the external component on the left-hand side and the internal on the right-hand side, thus ensuring that positive transport flux is consistently interpreted as import into the system, while negative flux represents export from the system; (D) *Electron Transport Chain*: This module defines a network of coupled reactions which was capable of synthesizing ATP through NADPH/NADH oxidation and proton pumping; (E) *Alkane biosynthesis pathway*: Since hydrocarbon biosynthesis along with fatty acid pathways were not included in the previous model, a lumped reaction corresponding to a fatty acid synthase producing palmitoyl-ACP formed from Acetyl-CoA, NADPH, NADH, and ATP was incorporated in the model. To synthesize alkane from fatty acyl-ACP, reactions catalyzing reduction of palmitoyl-ACP (C16) to corresponding aldehyde (C16) and its

subsequent conversion to alkane (C15) were added; (F) *Extra reactions*: This module contained reactions that were required in the model but could not be extracted from EcoCyc through the gene-reaction association. In this model, we have assumed that water, protons, and oxygen were present in excess. The model in ScrumPy (spy) and Systems Biology Markup Language (sbml) file format is available at https://drive.google.com/open?id=0B4LBWVX8JlHBZDBNblZVbkpuQXc. Common names of all abbreviated reactions and metabolites are given in Table S2.

2.1.1. Model analysis

The metabolic network was analyzed by FBA after ensuring that it was consistent in terms of stoichiometry, mass and energy conservation using the methodology described in (Gevorgyan et al., 2008; Poolman et al., 2013). Its ability to represent biomass formation was checked using the model where the objective function was fixed for minimization of the total flux of reactions present in the model while setting the rate of biomass export and bounding exporters of fermentation by-products to only carry negative flux, as summarized in Equation 1:

$$\begin{array}{lll} \text{minimize} & : & |\mathbf{v}| \qquad\qquad\qquad\quad \text{(objective function)} \\ \text{subject to} & \begin{cases} \mathbf{Nv} = \mathbf{0} & \text{(steady} - \text{state)} \\ v_{biomass} = 1 & \text{(biomass flux)} \\ -\infty \leq v_i \leq 0 & \text{(By} - \text{product export)} \end{cases} \end{array} \qquad (1)$$

In the above Equation, v defines the flux vector of reactions present in the network and N represents the stoichiometric matrix of all reactions.

2.1.2. Identification of gene targets for enhanced hydrocarbon production

To identify candidate genes for overexpression, FBA was used to force export of alkanes at different rates, while maintaining a constant rate of biomass export (1 g/h), as expressed in Equation 2.

$$\begin{aligned} &\text{minimize} \quad : \quad |\mathbf{v}| \\ &\text{subject to} \quad \begin{cases} \mathbf{Nv} = \mathbf{0} \\ v_{biomass} = 1 \\ v_{pentadecane} = J_{pentadecane} \\ -\infty \leq v_i \leq 0 \end{cases} \end{aligned} \tag{2}$$

Equation 2 was then repeatedly solved while imposing a fixed flux in the pentadecane transporter between limits of 0 and 2000 µmol $(gDW)^{-1}hr^{-1}$. The solutions were saved and a dendrogram constructed to visualize correlations between reaction fluxes as described previously (Hartman et al., 2014; Poolman et al., 2009). Genes encoding enzymes that catalyze reactions showing the greatest positive correlation with the pentadecane export flux were considered as candidates for overexpression. This indicated that glucose-6-phosphate dehydrogenase (G6PDH, encoded by *zwf*) as a candidate for overexpression.

Gene candidates for knockout were identified after incorporating the overexpression of *zwf* (encoding G6PDH) as an additional constraint as represented in Equation 3:

$$\begin{aligned} &\text{minimize} \quad : \quad |\mathbf{v}| \\ &\text{subject to} \quad \begin{cases} \mathbf{Nv} = \mathbf{0} \\ v_{biomass} = 1 \\ v_{G6PDH} = x \text{ (G6PDH fixed to twice original value)} \\ -\infty \leq v_i \leq 0 \end{cases} \end{aligned} \tag{3}$$

The flux dataset obtained was then compared with the dataset before the overexpression was implemented. The reactions that were diverting the carbon fluxes away from acetyl-CoA or hydrocarbon without providing any reducing equivalents or just acting as carbon sinks were considered one at a time for knockout by fixing the flux value of the reaction under consideration to zero in Equation 3 and checking the effect on alkane production. This was repeated until the maximum theoretical yield of hydrocarbon was achieved. For determining the optimal fraction of carbon source that can be diverted towards pentadecane formation while fixing a certain value of glucose uptake flux, the objective of Equation 1 was changed from minimization of fluxes to maximization of the pentadecane formation as represented by Equation 4:

$$\text{maximise} \quad : \quad v_{pentadecane}$$
$$\text{subject to} \quad \begin{cases} \mathbf{Nv} = \mathbf{0} \\ v_{biomass} = 1 \\ 0 \leq v_{Glc} \leq \hat{x} \text{ (Upper limit of Glc uptake)} \\ -\infty \leq v_i \leq 0 \end{cases} \tag{4}$$

2.2. Microbial strains, reagents, and media

E. coli and cyanobacterial strains used in the study along with their sources are listed in Table 1. Strains used for preparing phage lysate were procured from the Coli Genetic Stock Centre (CGSC), Yale University, USA. *E. coli* DH5α strain (Invitrogen) was used as host for performing all cloning work as well as genomic manipulation. For genomic DNA isolation from *Synechococcus elongatus* PCC7942 and *E. coli* DH5α, DNeasy plant mini kit (Qiagen) and Axygen genomic DNA isolation kit were used respectively as per the manufacturers' instructions. Chemical reagents were purchased from Sigma-Aldrich and Himedia. Standards for alkane, aldehyde and alcohol were purchased from TCI, India, and Sigma-Aldrich. All restriction endonucleases and other enzymes for molecular biology were procured from New England Biolabs (NEB), while Phusion DNA Polymerase for PCR was procured from Finnzymes. Codon optimized synthetic genes for this study were synthesized commercially at Genscript, USA.

2.3. Strain and plasmid construction

Detailed information about the plasmids used in this study is given in Table 1. Oligonucleotide primers used for making constructs are listed in Table S1.

The codon-optimized aldehyde deformylating oxygenase (*ado*) gene from *N. punctiformes* was cloned in pZS21mcs (a low copy number plasmid, kan^R, tet promoter) and pQE30 (a medium copy number plasmid, Amp^R, T5 lac promoter) to obtain pZF16 and pZF17 plasmids, respectively, with the help of primers as mentioned in Table S1. Different constructs for alkane biosynthesis were made as follows - (i) pZF18 was made by amplifying *ado* and *aar* genes from genomic DNA of *S.*

elongatus and was cloned in pQE30 at *Bam*HI/*Sac*I and *Sac*I/*Sal*I restriction site, respectively, as an operon; (ii) pZF19 was made by cloning codon optimized *ado* gene from *N. punctiformes* and *aar* from *S. elongatus* at *Bam*HI/*Sac*I and *Sac*I/*Sal*I restriction sites, respectively, as an operon; (iii) pZF20 containing fused genes were made by first cloning the codon optimized *aar* gene in pQE30 at *Sac*I/*Sal*I restriction site, followed by cloning of *ado* gene without stop codon having glycine-serine linker repeated twice $(G_4S)_2$ at 3' end at *Bam*HI and *Sac*I site; (iv) pZF21 containing fused genes was made by first cloning the codon optimized *aar* gene in pQE30 at *Sac*I/*Sal*I restriction site, followed by cloning of the *ado* gene without its stop codon and with a glycine-serine linker repeated thrice $(G_4S)_3$ at the 3' end at the *Bam*HI and *Sac*I site; (v) pZF22 having *aar* and *ado* genes with their individual T5 promoter and rrnB terminator was made by amplifying *ado* from pZF17 along with T5 promoter and rrnB terminator sequences and cloning in pZF13 at the *Xba*I site present downstream to the terminator sequence.

For validation of the *in silico* findings, the plasmids and strains were constructed as follows. The *zwf* gene was amplified from *E. coli* DH5α genomic DNA and cloned in the pZS21mcs plasmid at *Kpn*I and *Hind*III restriction sites to generate the pZF23 plasmid. Gene deletions in DH5α were achieved through the P1 transduction method (Fatma et al., 2016; Mattam and Yazdani, 2013; Munjal et al., 2012). The phage lysate was prepared and enriched by using the single gene knockout strain obtained from CGSC, and the targeted gene was replaced by an FRT flanked-kanamycin resistance cassette. For sequential deletion, the kanamycin resistant marker gene was removed by transforming the strain with a temperature sensitive pCP20 plasmid carrying a flippase enzyme. After removal of the kanamycin cassette, the strain was further used for making a consecutive knockout of genes through transduction. All the gene deletions were confirmed by PCR using primers listed in Table S1.

2.4. Growth media and cultivation conditions

Luria-Bertani (LB) broth with appropriate antibiotics (100 µg/mL ampicillin and 30 µg/mL kanamycin) was used for all cloning and propagation of bacterial cells. For hydrocarbon production, M9 modified medium (Schirmer et al., 2010) supplemented with 2% glucose as a sole carbon source was used. For determining the optimal IPTG concentration for induction of proteins from pQE30 based plasmid, the experiment was performed in presence of various concentration of IPTG (0.0 mM, 0.1 mM, 0.01mM and 0.001mM) (Fig. S1), and 0.01 mM IPTG was considered for further experiment. Anhydrotetracycline (100 ng/mL) were used for induction of protein from pZS21mcs based plasmids, respectively.

All test tube and shake flask experiments were conducted in 15 mL test tubes and 100 mL flasks with a working volume of 3 mL and 10 mL, respectively, under aerobic conditions. For preparing bacterial inocula, glycerol stock placed in -80 °C was streaked on LB agar plates with the appropriate antibiotics and placed in a 37 °C incubator overnight. A single isolated colony was inoculated in 5 mL LB and grown overnight at 37 °C. The grown culture was used to inoculate M9 modified medium containing 20 g/L glucose, antibiotics, and inducer to achieve an initial OD_{600} of 0.02 and incubated for 48 hr at 30 °C and 180 rpm. To increase the solubility and reduce the evaporation of hydrocarbon, shake flask culture was overlaid by 10% dodecane and flask were sealed by Parafilm. For growth curve and extracellular metabolites analysis, 1 mL of cell samples were periodically collected and used for measurement of optical density at 600 nm and metabolite analysis by GC-FID/GC-MS.

2.5. Enzyme assay for glucose-6-phosphate dehydrogenase (G6PDH)

A spectrophotometric based assay was used to quantify the change in enzyme activity after

overexpression of the gene (*zwf*) for G6PDH. The assay was conducted in 1 mL volume consisting of 50 mM Tris-HCl (pH 8.0), magnesium chloride (1 mM), glucose-6-phosphate (200 μM), $NADP^+$ (100 μM), cell lysate (50 μL, 100 μL, 150 μL or 200 μL). For the preparation of the cell lysates, overnight frozen cells at -20 °C were resuspended in 50 mM Tris-HCl (pH 8.0) to maintain a cell density of OD_{600} ~ 10. Chloroform (50 μL) was added to 250 μL of cell suspension, quickly vortexed for 10 sec and an appropriate volume was then added to the reaction mixture to get an optimal increase in optical density (Yazdani and Gonzalez, 2008). The rate of increase in OD_{340} indicated the production of NADPH from $NADP^+$. One unit of enzyme activity corresponded to the oxidation of 1 μmol of glucose-6-phosphate per min. Cell density at OD_{600} of 1 corresponded to 0.5 mg/mL of dry cell mass and the total cellular protein corresponded to 50% of the dry cell mass.

2.6. Analytical techniques

Cell growth was measured via optical density at 600 nm in a 1 mL cuvette using an Ultraspec 3100 Pro spectrophotometer (Amersham Biosciences). For hydrocarbon (alkane, alkene, alcohol, and aldehyde) analysis, equal volumes of ethyl acetate having 10 mg/L 1-octadecene as an internal standard and culture were mixed by vortexing for 20 min and then centrifuged for 10 min at 12000 rpm. The volatile upper layer was used for analysis and quantification of hydrocarbon. A gas chromatography system (GC 7890A from Agilent) equipped with an HP-5 column of 30 m length, 0.32 mm internal diameter and 0.25 μm film thickness was used for measurement. The oven program was set as follows: initial 100 °C for 3 min, then the temperature ramped up to 250 °C at a rate of 10 °C/min and held at 250 °C for 10 min. The inlet and detector temperature were maintained at 150 °C and 280 °C, respectively. Further confirmation of products was made through GC-MS analysis using an HP-5 MS column with almost the same program as above except that the final temperature was 300 °C. The MS quadrupole scanned from 50 to 550 *m*/*z*. For confirmation and quantification of

hydrocarbon, standards from TCI, India were used.

For analysis of extracellular metabolites, e.g., glucose, acetate, formate, lactate, and ethanol, 1 mL cell culture was centrifuged at 12000 rpm for 10 min, the supernatant was filtered through a 0.22 μm membrane filter and analyzed using an HPLC system (1260 Infinity Series, Agilent) equipped with an HPX-87H anion exchange column (Bio-Rad). Filtered and degassed 4 mM H_2SO_4 was used as a mobile phase at a constant rate of 0.3 mL/min. The column and RI detector temperatures were maintained at 40 °C and 35 °C, respectively. For quantification of liquid metabolites, standards of 1 g/L from Absolute Standards, USA were used.

2.7. *Fed-batch cultivation for long chain hydrocarbon production*

For fed-batch fermentation, the inoculum was prepared by growing cells taken from glycerol stock in 5 ml LB in 14 ml test-tubes for 12 hr. This culture (1%) was further used to inoculate 120 ml of secondary culture containing modified mineral medium (6g/L NH_4Cl, 8.5g/L KH_2PO_4, 1g/L $MgSO_4.7H_2O$, 0.5 g/L citrate, 0.07g/L $CaCl_2.2H_2O$, 15g/L glucose, 5g/L yeast extract, 4 mL trace metal and 4 mL of 10 mg/mL thiamine) (Cao et al., 2016) with glucose as carbon source and grown until OD_{600} reached 3. The grown secondary culture was then used to inoculate a 5L bioreactor (Applikon) having 3L of modified mineral medium as described earlier (Cao et al., 2016), except that the glucose used was as a carbon source instead of glycerol. Ampicillin (100 μg/mL) and kanamycin (30 μg/mL) were used as a selectable marker to maintain the plasmid stability. The growth of culture in both seed and bioreactor was performed at 30 °C. The pH in the bioreactor was automatically maintained at 7.0 using 5N NaOH and HCl. The dissolved oxygen was maintained between 30 to 40% of saturation throughout the cultivation by a cascade of agitation (500-800 rpm) and oxygen-mixed air supply. The culture was induced with 0.01 mM IPTG and 100 ng/ml of anhydrotetracycline soon after the inoculation in the bioreactor. The culture was grown until the glucose concentration in the bioreactor dropped below 5 g/L. Feeding was then started with a sterile solution containing 200

g/L yeast extract, 500 g/L of glucose, 2.47 g/L $MgSO_4$, 2X antibiotics and 1X inducer to achieve 0.2 h^{-1} of specific growth rate as mentioned earlier (Yazdani et al., 2004) until 18 hr of growth. Beyond 18 hr, the feed rate was maintained in order to keep the glucose concentration lower than 10 g/L in the culture.

For efficient production of long chain fatty alcohol, 5% of dodecane was overlaid upon starting feed as reported previously (Fatma et al., 2016). Samples were collected at regular intervals to monitor growth, glucose consumption and metabolite formation, as mentioned in Section 2.6.

3. Results and Discussion

3.1. Model properties and consistencies

The central carbon metabolism model considered in this study had 74 reactions (45 of which were irreversible) and 61 metabolites and was capable of generating precursor metabolites needed for biomass formation using glucose as the sole carbon source (Fig 1 and Additional data file 1). The model was analyzed for stoichiometric consistency and biomass formation under aerobic conditions. It was found that out of the 74 reactions, only 36 were responsible for biomass formation. Solutions obtained by forcing a 2-fold increase in glucose uptake flux through the reaction of glucose transporter did not lead to the formation of alkane; instead, glucose was channelled towards lactate and acetate (Fig. S2). Interestingly, we noticed an incomplete TCA cycle as no flux was computed through the reactions catalyzed by α-ketoglutarate dehydrogenase and succinyl-CoA synthetase, which is consistent with ^{13}C-metabolic flux analysis reported by Chen et al., (2011).

3.2. Constraint-based analysis of model to identify target genes for improvement in hydrocarbon production

Since simulation of the model of the *E. coli* metabolic network did not lead to the formation of any alkane, we adopted a constraint-based flux balance analysis to ascertain gene candidate(s)

responsible for diverting flux towards hydrocarbon production. We first identified gene candidates for overexpression based on the assumption that the overexpression candidate gene(s) should be those whose flux would co-vary along with the flux through hydrocarbon synthesis pathway. We collected LP (linear programming) datasets of each simulation and plotted bar graphs to visualize the response in flux carried through each pathway (Fig. S3). To visualize the correlation between the reactions, a dendrogram was made as described previously (Fig. 2) (Hartman et al., 2014; Poolman et al., 2009). In the dendrogram, reactions with a common parent node are exactly correlated and carry flux in a constant ratio. Inspection of the dendrogram shows that the alkane biosynthesis pathway was found to be closely correlated with the reactions of the pentose phosphate pathway (PPP) and the reactions of the lower half of glycolysis (Fig. 2). The same correlation can be seen in the analysis of the LP dataset where an increase in pentadecane output flux led to a rise in glucose uptake and subsequently an increase in the proportion of glucose metabolism through the PPP (Fig. 3). This phenomenon reflected the fact that alkane production requires more reducing equivalents in the form of NADPH.

We found that apart from the PPP, where flux increased by 153%, fluxes also increased through reactions catalyzed by pyruvate dehydrogenase by 278% (Fig. S4). These changes could be directly related to the PPP itself because it provides additional reducing equivalents as well as terminal metabolites which finally merge with the lower half of glycolysis (Fig. 2).

From the visualization of the dendrogram, we expected that increase in flux towards PPP would enhance hydrocarbon production; therefore the flux through the first reaction of PPP catalyzed by glucose-6-phosphate dehydrogenase (G6PDH; shown as overexpression of *zwf* in Fig. 4) was increased by 2-fold to 15220 μmol $(gDW)^{-1}hr^{-1}$ in the model. However, an increase in flux through G6PDH did not lead to an increase in flux through PPP and also did not result in hydrocarbon production. Instead, the flux was diverted through the Entner Duodoroff (ED) pathway, with pyruvate converted back to phosphoenolpyruvate (PEP) through PEP synthase (PPS). We therefore introduced another constraint by specifying zero flux through the first reaction of the ED pathway catalyzed by

phosphogluconate dehydratase (corresponding to deletion of *edd*) and the reaction catalyzed by PPS, which led to an increase in pentadecane yield to 10.09% of carbon input (Fig. 4). This change in the network, however, also resulted in the formation of lactate from pyruvate which competed for carbon and thus had an impact on hydrocarbon yield.

To avoid this competition, we blocked flux towards lactate synthesis (corresponding to deletion of *ldhA*), which improved alkane yield to 12.37% (Fig. 4). Deletion of the LDH reaction, however, directed to increased flux through the glyoxylate shunt. We, therefore, blocked the first reaction of the glyoxylate shunt catalyzed by isocitrate lyase (ICL, corresponding to deletion of *aceA* in Fig. 4). The yield of alkane increased to 12.7% of carbon source with the implementation of the *aceA* deletion, but there was then acetate formation following deletion. Acetate formation was blocked by assigning zero flux to the reactions catalyzed by pyruvate oxidase (POX) and phosphate acetyltransferase (PTA) together, as individual deletion did not reduce acetate formation. This led to an improvement in pentadecane production to 16.34% of carbon source (Fig. 4). We also observed significant formate formation, mainly due to the reaction catalyzed by pyruvate formate lyase (PFL, encoded by *pflB*). However, blocking flux through PFL reaction did not lead to any improvement in hydrocarbon production (Fig 4).

We performed further validation of the final solution of the constraint-based analysis by changing the objective function from minimization of the sum of reaction fluxes to maximization of flux for alkane production. This was performed by constraining the flux value of glucose uptake to that obtained in the previous solution, using Equation 4 as mentioned in the Materials and Methods section. We found that a maximum of 17.3% of the glucose could be utilized for alkane production while maintaining synthesis of biomass precursors for growth. Considering the 17.3% utilization of the carbon source as the 100% theoretical maximum yield for alkane formation under the *in silico* conditions, the 16.34% of carbon diverted towards pentadecane production corresponded to 94.2% of the maximum theoretical yield. Moving towards our goal of achieving maximum yield, we

compared the preferred paths through the network in the FBA solutions of the *in silico* engineered strain with the FBA solution for the theoretical maximum, and found similarity in terms of the preferred glucose metabolism pathway (Pentose Phosphate Pathway), with no fluxes towards the fermentative pathways, the ED pathway and the glyoxylate shunt (Fig. S5). This similarity supported the accuracy of our predicted gene targets for engineering *E. coli* for improvement in hydrocarbon production.

3.3. Assembly of the alkane pathway in E. coli

The implementation of the *in silico* findings first required an assembly of the alkane-producing pathway in *E. coli*. We thus assembled a functional cyanobacterial alkane biosynthesis pathway in *E. coli* by cloning genes for AAR and ADO of *S. elongatus* in the form of a single operon in a pQE30 vector under the control of a T5 promoter (pZS18). This resulted in production of ~20 mg/L of total hydrocarbon (i.e., sum of all aldehyde, alcohol and alka(e)ne of C13-C18 chain length), including 17 mg/L of alcohols and 2.86 mg/L of alka(e)ne (Fig. 5A), which was in a similar range to previous reports (Akhtar et al., 2013; Coursolle et al., 2015; Howard et al., 2013; Rahman et al., 2014). Long chain fatty alcohol produced here is due to the activity of endogenous aldehyde reductases, mainly YbbO, that convert long-chain fatty aldehyde, a product of AAR, into the corresponding alcohol (Fatma et al., 2016; Howard et al., 2013; Rodriguez and Atsumi, 2014; Schirmer et al., 2010). Although we did observe alkane production in the recombinant *E. coli,* proving that the assembled hydrocarbon biosynthesis pathway was functional, the amount of hydrocarbon produced was small and needed improvement.

It has been reported (Schirmer et al., 2010) that the ADO enzyme from *Nostoc punctiformes* PCC73102 has a higher activity for long-chain aldehydes compared to the ADO from *S. elongatus* PCC7942. Therefore, codon optimized genes for the ADO of *N. Punctiformes* and the AAR from *S. elongatus* were commercially synthesized and cloned as an operon in pQE30 (pZF19). However, no

significant impact on alkane production was observed. Placing these two genes on separate compatible plasmids also did not improve the alkane titer. To understand which of the two enzymes was limiting in the production of alkane, we supplemented each gene in an *E. coli* expressing both genes in an operon via separate compatible plasmids. Supplementing AAR led to a 3-fold increase in alkane production (8.2 mg/L), while no impact was seen on ADO supplementation (Fig. 5A), suggesting that AAR being the first enzyme in the alkane pathway beyond fatty acid synthesis has much more control over the flux. Although our data did not show significant differences between the ADOs of *S. elongatus* and *N. punctiformes* in terms of alkane production, we proceeded further with ADO of *N. punctiformes* due to previous findings reported in the literature showing its >2-fold higher activity (Schirmer et al., 2010).

Since the product of the AAR enzyme i.e. fatty aldehyde, is a substrate of the ADO enzyme, their co-localization could help in the channeling of the intermediates and higher yield of targeted products. A similar result has also been shown with cellulolytic enzymes through making fusion constructs (Adlakha et al., 2011; Adlakha et al., 2012). However, making fusion proteins of two different catalytic domains may interfere with proper folding and catalytic activities of each other. Therefore, two fusion proteins, ADO$(G_4S)_2$AAR and ADO$(G_4S)_3$AAR, were made where glycine–serine linkers of two different lengths were introduced in between ADO and AAR for independent folding and expressed in *E. coli*. Alkane production was increased to 12 mg/L (4.28-fold higher) and 3.9 mg/L (1.39-fold higher) for strains carrying ADO$(G_4S)_2$AAR (pZF20) and ADO$(G_4S)_3$AAR (pZF21), respectively (Fig 5A). This finding was similar to that reported by Rahman et al., (2014) where the use of a glycine-serine linker led a 4.8-fold higher value than the one without fusion. Apart from channelling, equal molar expression of both enzymes in the fusion polypeptide could be another reason for higher alkane production. We made a further attempt to achieve equivalent high-level expression of these two enzymes by cloning their respective genes in medium copy number plasmids with individual T5 promoters and rrnB terminators (pZS22). We observed a 36-fold increase in alkane

production to 102 mg/L (Figure 5A), with total hydrocarbon reaching to 170 mg/L. Pentadecane and heptadecene were most abundant with 48 mg/L of each amongst the alka(e)nes, and hexadecanol (52 mg/L) was most prevalent amongst fatty alcohols, followed by octadecenol (15 mg/L) (Fig 5B). The results indicated a significant role of vector modulation and individual promoter in guiding gene expression for improvement in metabolite production.

3.4. Implementation of in silico finding under in vivo conditions

Validation of the *in silico* findings required selection of a suitable *E. coli* host strain as the genetic background of the host makes a significant impact on the metabolite profiles (Liu et al., 2014; Mattam and Yazdani, 2013; Song et al., 2016). Therefore we compared hydrocarbon profiles in eight different *E. coli* strains, i.e., *E. coli* B, DH5α, M15, MG1655, BW25113, Top10, BL21, and JM109, by transforming them with plasmid pZS22 and found that DH5α showed the highest alkane production (Fig. 5B). We thus adopted DH5α as the background host to validate the FBA predictions.

3.4.1. Improvement in long chain alkane production

The *in silico* identification approach to select gene(s) had resulted in a combination of one overexpression (*zwf*) and seven knockouts (*edd, pps, ldhA, aceA, ptaA, poxB, pflB*) for the hydrocarbon production in *E. coli* (Table 2). The candidate gene overexpression would likely enhance reducing equivalents (NADPH) for hydrocarbon production, whereas the gene deletions would remove fermentation by-products (lactate, acetate, and formate), prevent diversion of glucose metabolism through the ED pathway and enhance the level of acetyl-CoA.

To validate the prediction, the native *zwf* gene from *E. coli* DH5α was cloned and overexpressed in a low copy number pZS21mcs vector (pZS23); expression and activity were confirmed by Western blotting using an antibody against the His-tag and by measuring glucose-6-phosphate dehydrogenase activity. DH5α cells overexpressing *zwf* gene showed 5.5-fold higher

activity compared to the wild-type strain (Fig. S6). Further, sequential knockouts of *edd, pps, ldhA, aceA, poxB, ptaA* and *pflB* were created in DH5α through phage transduction (Baba et al., 2006; Mattam and Yazdani, 2013). Gene knockouts were confirmed by PCR using external primers unique to the flanking regions of the genes (Table S1, Fig. S7). The deletion up to *aceA* gene did not impair the growth rate, but deletion of *poxB* and *pflB* prolonged the lag phase (Fig. S8).

All the engineered *E. coli* strains, along with DH5α as control, were co-transformed with plasmid pZF23 carrying the gene for glucose-6-phosphate dehydrogenase and the plasmid pZF22 carrying the pathway for alkanes and tested for their ability to produce alkane. We found that over-expression of *zwf* and deletion of *edd*, *ldh, aceA* and *poxB* had a positive impact on alkane production, while deletion of *pps*, *pta* and *pflB* had an adverse impact on alkane production (Fig. 6A). The alkane accumulation had reached to 279 mg/L (Fig. 6A) in the case of ZFM5 (Δ*edd*Δ*pps*Δ*ldhA*Δ*aceA*Δ*poxB*). Layering of dodecane in the culture tube further enhanced the alkane titer to 318 mg/L (Fig. S9A), which was 3.1-fold higher compared to the control strain, and 111-fold higher compared to the un-optimized basal strain at test tube level. When the yield predicted via *in silico* analysis was compared with the experimental value, the strain engineered for alkane production was found to produce alka(e)ne and total hydrocarbon (including alka(e)ne, fatty aldehyde and alcohol) at 19.68% and 27.82% of the theoretical maximum, respectively.

3.4.2. Improvement in long chain fatty alcohol production

It was shown earlier that, when producing alkane in *E. coli*, significant amounts of C16 and C18 alcohols accumulate in the culture medium (Schirmer et al., 2010). This is because the long chain fatty aldehyde produced by the action of cyanobacterial acyl-ACP reductase undergoes a reduction reaction in *E. coli* to produce the corresponding alcohol. We earlier characterized a long-chain specific aldehyde reductase of *E. coli* encoded by *ybbO* carrying out this reaction with the help of NADPH as a co-factor (Fatma et al., 2016). We performed an *in silico* study for fatty alcohol

production identical to that of alkane production and found similar trends in the percentage improvement of fatty alcohol production. Thus all the sequentially mutated strains (ZFM1-ZFM7) were co-transformed with pZF23 and pZF13 (containing the genes for the AAR of *S. elongatus*) to assess the production of long chain alcohols. The final strain (ZFM7) was also co-transformed with pZF23 and pZF15 (carrying genes for AAR and YbbO) to observe the impact of overexpression of *ybb*O on long chain alcohol production. The comparison of results of all the engineered strains demonstrated the stepwise improvement in fatty alcohol production after the successive deletions, except for *pta* deletion, where alcohol titer declined (Fig. 6B). We achieved a maximum fatty alcohol production of 460 mg/L in ZFM7 strain without *ybbO* overexpression, which was 3.83-fold higher than in the control strain and 773 mg/L of fatty alcohol along with ybbO overexpression corresponding to 6.4-fold higher than basal strain (Fig. 6B). Dodecane layering on the ZFM7+YbbO strain during cultivation led to the production of 1022 mg/L of long chain fatty alcohols at 58.6 % of the theoretical yield. The long chain alcohols were mainly composed of hexadecanol (670 mg/L) and octadecenol (203 mg/L) (Fig. S9B).

The yield and titer obtained for fatty alcohol were significantly higher than alkane. Since pathway until second last enzymatic step, i.e., up to AAR, is common among both the metabolites, the relatively low titer of alkane could be attributed to the low activity of ADO as mentioned before (Li et al., 2012; Warui et al., 2015). The turn-over number reported for ADO had been 0.11 min^{-1} (Das et al., 2014) as against 12.39 min^{-1} reported for YbbO (Fatma et al., 2016) using Octadecanal as substrate.

3.5. *Effect of phospholipid pathway modulation on alkane and fatty alcohol production*

The achieved yields of alkane (27.8% of theoretical maximum) and long chain alcohol (58.6% of theoretical maximum) were still significantly lower than the theoretical maximum yield calculated from the model. This could be due to the fact that while we constrained flux towards biomass in the

model, this was not done experimentally. When we measured the carbon source going towards biomass and the fermentative products in the parent and the engineered strains, we found that most of the carbon flux was directed towards biomass formation (Table S3). One way to limit flux towards biomass formation is to limit formation of membrane phospholipids. There are two pathways reported in the literature for phospholipid formation: the first one involves the *plsB* and *plsC* genes, and the second involves the *plsX* and *plsY* genes (Yao and Rock, 2013). The deletion of either *plsB* or *plsC* was found to be lethal. The individual deletion of *plsX* or *plsY* gene was found to be non-lethal, though a combined double knockout of these two genes has been reported to be lethal (Larson et al., 1984; Yoshimura et al., 2007). In our previous study (Fatma et al., 2016), we have reported that deletion of the *plsX* gene (responsible for phospholipid biosynthesis by taking fatty acyl-ACP as a substrate) helped in the improvement of fatty alcohol production. Therefore, we knocked-out the *plsX* gene in the background of ZFM5 and ZFM7 to generate ZFM8 and ZFM9 strains for production of alkane and fatty alcohol, respectively, and checked the overall effect on cell growth and hydrocarbon production (Fig. S10 and Fig. 7). We observed that with *plsX* deletion, cells took longer time to reach to an OD_{600} of 12, and the deletion was ultimately helpful in improving the alkane titer by 1.57 fold (280 mg/L to 425 mg/L, Fig. 7A) and fatty alcohol titer by 1.95 fold (773 mg/L to 1506 mg/L, Fig. 7B), reaching to 34.2% and 86.4% of theoretical yield, respectively, for total hydrocarbon. Therefore, as expected from the *in silico* studies, limiting biomass formation via *plsX* deletion helped in improving flux towards hydrocarbon production.

3.6. *Microbial production of long chain alkane and alcohol by fed-batch cultivation*

Fed-batch cultivation is a useful strategy to improve the titer of the final product by growing the cells to high density with the help of substrate feeding under controlled condition (Cao et al., 2015; Cao et al., 2016). We thus performed fed-batch cultivation for the production of alkane and fatty alcohol in 5L bioreactor using mineral medium with glucose as carbon source.

For alkane production, we used the best alkane producing strain in the shake flask, i.e., *E. coli* ZFM8/pZF22/pZF23, having deletion of *edd, ppsA, ldhA, aceA, poxB* and *plsX* genes and overexpression of *aar*, *ado* and *zwf* genes. Cells were initially grown under batch cultivation until glucose concentration reached ~5 g/L. The feed was then initiated at an OD_{600} of 16 as described in the method section. There was a steep rise in cell density up to OD_{600} of 102 at 30 hr, after which it plateaued and reached to maximum of 128 at 66 hr. On the other hand, there was a gradual increase in alka(e)ne until 72 hr, following which it plateaued. The total accumulation of alka(e)ne (Fig. 8A) and hydrocarbon (sum of long chain alkane, alcohol and aldehyde) (Table S4) in the culture had reached to 2.54 g/L (6.2% of theoretical yield) and 5.68 g/L (13.4% of theoretical yield), respectively, by 84 h of cultivation, where the major contributions were from heptadecene (C17:1, 55%) and pentadecane (C15, 40%). Throughout the cultivation, dissolved oxygen was maintained between 30 to 40% saturation because it has been reported earlier that oxygen is the co-substrate for ADO enzyme (Andre et al., 2013; Pandelia et al., 2013). A previously engineered strain of *E. coli*, where the alkane biosynthesis pathway (AAR/ADO) was overexpressed along with the *S. elongatus* FD/FNR, has produced 1.3 g/L alkane (0.01 g alkane/gDW) (Cao et al., 2016), whereas our model-assisted engineered strain has achieved 2.54 g/L alkane (0.051 g alkane/gDW). Other industrially relevant microbial platforms based on *S. cerevisiae* (22.0 μg/gDW) and *Y. lipolytica* (23.3 mg/L) reported in the literatures have been shown to produce much lesser quantity of alkane (Buijs et al., 2015; Xu et al., 2016).

Similarly, we have evaluated the performance of the best fatty alcohol producing strain, i.e., *E. coli* ZFM9, having deletion of *edd, ppsA, ldhA, aceA, poxB, pta, pflB,* and *plsX* genes and overexpression of *aar*, *ybbO* and *zwf* genes. Here, the rise in cell density as well as fatty alcohol was gradual until 60 hr. The maximum fatty alcohol accumulation reached to 12.5 g/L (0.249 g fatty alcohol/gDW, 22.16% of theoretical yield) after 72 hr of cultivation, of which 90% was contributed by hexadecanol (C16, 66%) and octadecenol (C18:1, 24%) (Fig. 8B). The accumulation of fatty

aldehydes was found to be negligible (Table S4). The maximum titer reported so far in the literature for fatty alcohol has been 6.3 g/L (Liu et al., 2016). Thus, the highest titers were obtained in this study for both alkane and fatty alcohol, which indicated the strength of combining metabolic modelling for rational strain design for value added chemical commodities.

4. Conclusion

In this study, we developed a metabolic model of hydrocarbon production in *E. coli* which was optimized using techniques derived from FBA. To implement the *in silico* findings, we first optimized the genetic components of hydrocarbon synthesis and identified a suitable host platform, which led to a 36-fold overall improvement in alka(e)ne production. Based on the *in silico* predictions, an overexpression of a gene and 7 gene knockouts were introduced sequentially into an *E. coli* DH5α strain, resulting in 273 mg/L alkane production at test tube level representing a 95.4-fold increase over parent strain. Subsequent cultivation with dodecane layering led to the production of 318 mg/L alka(e)ne, an improvement of 111-fold over the basal strain. When the modifications were tested for production of fatty alcohols, 1022 mg/L fatty alcohol was achieved. The hydrocarbon yields obtained in the engineered strains for alkane and fatty alcohol were 27.8% and 58.6% of the theoretical maximum, respectively. We found that the reason for achieving low theoretical yield could be due to the major carbon flux diverting towards biomass formation. We therefor deleted *plsX*, which increased the alkane and fatty alcohol yield to 34.2% and 86.4% of theoretical maximum, respectively. Finally, cultivation of the engineered strains (ZFM8 and ZFM9) under fed-batch cultivation resulted in the higher titer of 2.54 g/L of alka(e)ne and 12.5 g/L of fatty alcohol, respectively. Our results demonstrate the potential of a combined modeling and experimental approach to the rational design and engineering of microbial hosts for increased biofuel production.

Acknowledgements

Authors would like to acknowledge funding support from Department of Biotechnology, Govt of India through Grant Reference No. BT/PB/Center/03/2011 and the Biotechnology and Biological Sciences Research Council, UK, through Grant Reference Nos. BB/J019712/1 and BB/K020358/1.

References

Adlakha, N., Rajagopal, R., Kumar, S., Reddy, V. S., Yazdani, S. S., 2011. Synthesis and characterization of chimeric proteins based on cellulase and xylanase from an insect gut bacterium. Applied and environmental microbiology. 77**,** 4859-4866.

Adlakha, N., Sawant, S., Anil, A., Lali, A., Yazdani, S. S., 2012. Specific fusion of β-1, 4-endoglucanase and β-1, 4-glucosidase enhances cellulolytic activity and helps in channeling of intermediates. Applied and environmental microbiology. 78**,** 7447-7454.

Akhtar, M. K., Turner, N. J., Jones, P. R., 2013. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. Proceedings of the National Academy of Sciences. 110**,** 87-92.

Andre, C., Kim, S. W., Yu, X.-H., Shanklin, J., 2013. Fusing catalase to an alkane-producing enzyme maintains enzymatic activity by converting the inhibitory byproduct H2O2 to the cosubstrate O2. Proceedings of the National Academy of Sciences. 110**,** 3191-3196.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., Mori, H., 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular systems biology. 2.

Buijs, N. A., Zhou, Y. J., Siewers, V., Nielsen, J., 2015. Long-chain alkane production by the yeast *Saccharomyces cerevisiae*. Biotechnology and bioengineering. 112**,** 1275-1279.

Burgard, A. P., Pharkya, P., Maranas, C. D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnology and bioengineering. 84**,** 647-657.

Cao, Y.-X., Xiao, W.-H., Liu, D., Zhang, J.-L., Ding, M.-Z., Yuan, Y.-J., 2015. Biosynthesis of odd-chain fatty alcohols in *Escherichia coli*. Metabolic engineering. 29**,** 113-123.

Cao, Y.-X., Xiao, W.-H., Zhang, J.-L., Xie, Z.-X., Ding, M.-Z., Yuan, Y.-J., 2016. Heterologous biosynthesis and manipulation of alkanes in *Escherichia coli*. Metabolic engineering. 38**,** 19-28.

Chen, B., Ling, H., Chang, M. W., 2013. Transporter engineering for improved tolerance against alkane biofuels in *Saccharomyces cerevisiae*. Biotechnology for biofuels. 6**,** 1.

Chen, X., Alonso, A. P., Allen, D. K., Reed, J. L., Shachar-Hill, Y., 2011. Synergy between 13 C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in *E. coli*. Metabolic engineering. 13**,** 38-48.

Choi, H. S., Lee, S. Y., Kim, T. Y., Woo, H. M., 2010. *In silico* identification of gene amplification

targets for improvement of lycopene production. Applied and environmental microbiology. 76**,** 3097-3105.
Choi, Y. J., Lee, S. Y., 2013. Microbial production of short-chain alkanes. Nature. 502**,** 571-574.
Cintolesi, A., Clomburg, J. M., Gonzalez, R., 2014. *In silico* assessment of the metabolic capabilities of an engineered functional reversal of the β-oxidation cycle for the synthesis of longer-chain (C≥ 4) products. Metabolic engineering. 23**,** 100-115.
Coursolle, D., Lian, J., Shanklin, J., Zhao, H., 2015. Production of long chain alcohols and alkanes upon coexpression of an acyl-ACP reductase and aldehyde-deformylating oxygenase with a bacterial type-I fatty acid synthase in E. coli. Molecular BioSystems. 11**,** 2464-2472.
d'Espaux, L., Ghosh, A., Runguphan, W., Wehrs, M., Xu, F., Konzock, O., Dev, I., Nhan, M., Gin, J., Apel, A. R., 2017. Engineering high-level production of fatty alcohols by *Saccharomyces cerevisiae* from lignocellulosic feedstocks. Metabolic Engineering.
Eser, B. E., Das, D., Han, J., Jones, P. R., Marsh, E. N. G., 2011. Oxygen-independent alkane formation by non-heme iron-dependent cyanobacterial aldehyde decarbonylase: investigation of kinetics and requirement for an external electron donor. Biochemistry. 50**,** 10743-10750.
Fatma, Z., Jawed, K., Mattam, A. J., Yazdani, S. S., 2016. Identification of long chain specific aldehyde reductase and its use in enhanced fatty alcohol production in *E. coli*. Metabolic engineering. 37**,** 35-45.
Fell, D. A., Small, J. R., 1986. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. Biochemical Journal. 238**,** 781-786.
Fong, S. S., 2014. Computational approaches to metabolic engineering utilizing systems biology and synthetic biology. Computational and structural biotechnology journal. 11**,** 28-34.
Gevorgyan, A., Poolman, M. G., Fell, D. A., 2008. Detection of stoichiometric inconsistencies in biomolecular models. Bioinformatics. 24**,** 2245-2251.
H Chan, W., S Mohamad, M., Deris, S., M Illias, R., 2013. A review of computational approaches for in silico metabolic engineering for microbial fuel production. Current Bioinformatics. 8**,** 253-258.
Hädicke, O., Klamt, S., 2011. Computing complex metabolic intervention strategies using constrained minimal cut sets. Metabolic engineering. 13**,** 204-213.
Harder, B.-J., Bettenbrock, K., Klamt, S., 2016. Model-Based Metabolic Engineering Enables High Yield Itaconic Acid Production by *Escherichia coli*. Metabolic engineering.
Harger, M., Zheng, L., Moon, A., Ager, C., An, J. H., Choe, C., Lai, Y.-L., Mo, B., Zong, D., Smith, M. D., 2012. Expanding the product profile of a microbial alkane biosynthetic pathway. ACS synthetic biology. 2**,** 59-62.
Hartman, H. B., Fell, D. A., Rossell, S., Jensen, P. R., Woodward, M. J., Thorndahl, L., Jelsbak, L., Olsen, J. E., Raghunathan, A., Daefler, S., 2014. Identification of potential drug targets in *Salmonella enterica* sv. *Typhimurium* using metabolic modelling and experimental validation. Microbiology. 160**,** 1252-1266.
Haushalter, R. W., Groff, D., Deutsch, S., The, L., Chavkin, T. A., Brunner, S. F., Katz, L., Keasling, J. D., 2015. Development of an orthogonal fatty acid biosynthesis system in E. coli for oleochemical production. Metabolic engineering. 30**,** 1-6.
Howard, T. P., Middelhaufe, S., Moore, K., Edner, C., Kolak, D. M., Taylor, G. N., Parker, D. A., Lee, R., Smirnoff, N., Aves, S. J., 2013. Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli*. Proceedings of the National Academy of Sciences. 110**,** 7636-7641.
Kallio, P., Pásztor, A., Thiel, K., Akhtar, M. K., Jones, P. R., 2014. An engineered pathway for the biosynthesis of renewable propane. Nature communications. 5.
Kim, J., Reed, J. L., 2010. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC systems biology. 4**,** 1.

Klamt, S., Gilles, E. D., 2004. Minimal cut sets in biochemical reaction networks. Bioinformatics. 20**,** 226-234.

Larson, T., Ludtke, D., Bell, R., 1984. sn-Glycerol-3-phosphate auxotrophy of plsB strains of *Escherichia coli*: evidence that a second mutation, plsX, is required. Journal of bacteriology. 160**,** 711-717.

Li, N., Chang, W.-c., Warui, D. M., Booker, S. J., Krebs, C., Bollinger Jr, J. M., 2012. Evidence for only oxygenative cleavage of aldehydes to alk (a/e) nes and formate by cyanobacterial aldehyde decarbonylases. Biochemistry. 51**,** 7908-7916.

Liu, Q., Wu, K., Cheng, Y., Lu, L., Xiao, E., Zhang, Y., Deng, Z., Liu, T., 2015. Engineering an iterative polyketide pathway in *Escherichia coli* results in single-form alkene and alkane overproduction. Metabolic engineering. 28**,** 82-90.

Liu, R., Zhu, F., Lu, L., Fu, A., Lu, J., Deng, Z., Liu, T., 2014. Metabolic engineering of fatty acyl-ACP reductase-dependent pathway to improve fatty alcohol production in *Escherichia coli*. Metabolic engineering. 22**,** 10-21.

Liu, Y., Chen, S., Chen, J., Zhou, J., Wang, Y., Yang, M., Qi, X., Xing, J., Wang, Q., Ma, Y., 2016. High production of fatty alcohols in *Escherichia coli* with fatty acid starvation. Microbial Cell Factories. 15**,** 129.

Mattam, A. J., Yazdani, S. S., 2013. Engineering *E. coli* strain for conversion of short chain fatty acids to bioalcohols. Biotechnology for biofuels. 6**,** 1.

Munjal, N., Mattam, A., Pramanik, D., Srivastava, P., Yazdani, S. S., 2012. Modulation of endogenous pathways enhances bioethanol yield and productivity in *Escherichia coli*. Microbial cell factories. 11**,** 1.

Pandelia, M. E., Li, N., Nørgaard, H., Warui, D. M., Rajakovich, L. J., Chang, W.-c., Booker, S. J., Krebs, C., Bollinger Jr, J. M., 2013. Substrate-triggered addition of dioxygen to the diferrous cofactor of aldehyde-deformylating oxygenase to form a diferric-peroxide intermediate. Journal of the American Chemical Society. 135**,** 15801-15812.

Pharkya, P., Burgard, A. P., Maranas, C. D., 2004. OptStrain: a computational framework for redesign of microbial production systems. Genome research. 14**,** 2367-2376.

Pharkya, P., Maranas, C. D., 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metabolic engineering. 8**,** 1-13.

Poolman, M. G., Fell, D. A., Raines, C. A., 2003. Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. European Journal of Biochemistry. 270**,** 430-439.

Poolman, M. G., Kundu, S., Shaw, R., Fell, D. A., 2013. Responses to light intensity in a genome-scale model of rice metabolism. Plant physiology. 162**,** 1060-1072.

Poolman, M. G., Miguet, L., Sweetlove, L. J., Fell, D. A., 2009. A genome-scale metabolic model of *Arabidopsis* and some of its properties. Plant physiology. 151**,** 1570-1581.

Rahman, Z., Sung, B. H., Yi, J.-Y., Bui, L. M., Lee, J. H., Kim, S. C., 2014. Enhanced production of n-alkanes in *Escherichia coli* by spatial organization of biosynthetic pathway enzymes. Journal of Biotechnology. 192**,** 187-191.

Rodriguez, G. M., Atsumi, S., 2014. Toward aldehyde and alkane production by removing aldehyde reductase activity in *Escherichia coli*. Metabolic engineering. 25**,** 227-237.

Runguphan, W., Keasling, J. D., 2014. Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. Metabolic engineering. 21**,** 103-113.

Schirmer, A., Rude, M. A., Li, X., Popova, E., Del Cardayre, S. B., 2010. Microbial biosynthesis of alkanes. Science. 329**,** 559-562.

Schuster, S., Fell, D. A., Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature biotechnology. 18**,** 326-332.

Sheng, J., Stevens, J., Feng, X., 2016. Pathway Compartmentalization in Peroxisome of *Saccharomyces cerevisiae* to Produce Versatile Medium Chain Fatty Alcohols. Scientific reports. 6.

Song, X., Yu, H., Zhu, K., 2016. Improving alkane synthesis in *Escherichia coli* via metabolic engineering. Applied microbiology and biotechnology. 100**,** 757-767.

Tepper, N., Shlomi, T., 2010. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics. 26**,** 536-543.

Trinh, C. T., Unrean, P., Srienc, F., 2008. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. Applied and environmental microbiology. 74**,** 3634-3643.

Varma, A., Palsson, B. O., 1993a. Metabolic capabilities of *Escherichia coli* II. Optimal growth patterns. Journal of Theoretical Biology. 165**,** 503-522.

Varma, A., Palsson, B. O., 1993b. Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. Journal of theoretical biology. 165**,** 477-502.

Warui, D. M., Pandelia, M.-E., Rajakovich, L. J., Krebs, C., Bollinger Jr, J. M., Booker, S. J., 2015. Efficient Delivery of Long-Chain Fatty Aldehydes from the *Nostoc punctiforme* Acyl–Acyl Carrier Protein Reductase to Its Cognate Aldehyde-Deformylating Oxygenase. Biochemistry. 54**,** 1006-1015.

Watson, M., 1986. A discrete model of bacterial metabolism. Computer applications in the biosciences: CABIOS. 2**,** 23-27.

Xu, P., Qiao, K., Ahn, W. S., Stephanopoulos, G., 2016. Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. Proceedings of the National Academy of Sciences. 113**,** 10848-10853.

Yao, J., Rock, C. O., 2013. Phosphatidic acid synthesis in bacteria. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids. 1831**,** 495-502.

Yazdani, S. S., Gonzalez, R., 2008. Engineering *Escherichia coli* for the efficient conversion of glycerol to ethanol and co-products. Metabolic engineering. 10**,** 340-351

Yazdani, S. S., Shakri, A. R., Chitnis, C. E., 2004. A high cell density fermentation strategy to produce recombinant malarial antigen in *E. coli*. Biotechnology letters. 26**,** 1891-1895.

Yoshimura, M., Oshima, T., Ogasawara, N., 2007. Involvement of the YneS/YgiH and PlsX proteins in phospholipid biosynthesis in both *Bacillus subtilis* and *Escherichia coli*. BMC microbiology. 7**,** 69.

Youngquist, J. T., Schumacher, M. H., Rose, J. P., Raines, T. C., Politz, M. C., Copeland, M. F., Pfleger, B. F., 2013. Production of medium chain length fatty alcohols from glucose in *Escherichia coli*. Metabolic engineering. 20**,** 177-186.

Zhang, J., Lu, X., Li, J.-J., 2013. Conversion of fatty aldehydes into alk (a/e) nes by in vitro reconstituted cyanobacterial aldehyde-deformylating oxygenase with the cognate electron transfer system. Biotechnology for biofuels. 6**,** 1.

Table 1. List of strains and plasmids used in study

Strain	Relevant genotype	Reference
S. elongatus	*Synechococcus elongatus* PCC7942, cyanobacterial strain	PCC
DH5α	*F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoRnupG Φ80dlacZΔM15 Δ (lacZYA-argF)U169, hsdR17(rK- mK+), λ–*	Invitrogen
BW25113	*F-,DE(araD-araB)567, lacZ4787(del)::rrnB-3, LAM-, rph-1, DE(rhaD-rhaB) 568, hsdR514*	CGSC#7636
MG1655	*E. coli* K12 *F- λ-, ilvG-, rfb-50*, *rph-1*Δ*recA*Δ*endA*	CGSC #6300
BL21	*E. coli*B *F- dcmompThsdS$_B$(rB- mB-) gal*	Invitrogen
M15	F-, Φ80ΔlacM15, thi, lac-, mtl-, recA+ ,KmR	Qiagen
TOP10	F- mcrAΔ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-	Invitrogen
JM109	*end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(lac-proAB),	Promega
E. coli B	F-, K12 strain	CGSC #2507
ZFM1	DH5α: Δ*edd*	This study
ZFM2	DH5α: Δ*edd*Δpps	This study
ZFM3	DH5α: Δ*edd*ΔppsΔldhA	This study
ZFM4	DH5α: Δ*edd*ΔppsΔldhAΔaceA	This study
ZFM5	DH5α: Δ*edd*ΔppsΔldhAΔaceAΔpoxB	This study
ZFM6	DH5α: Δ*edd*ΔppsΔldhAΔaceAΔpoxBΔpta	This study
ZFM7	DH5α: Δ*edd*ΔppsΔldhAΔaceAΔpoxBΔptaΔpflB	This study
ZFM8	DH5α: Δ*edd*ΔppsΔldhAΔaceAΔpoxBΔplsX	This study
ZFM9	DH5α: Δ*edd*ΔppsΔldhAΔaceAΔpoxBΔptaΔpflBΔplsX	This study

Plasmid	Origin of replication	Overexpressed Gene	Resistance	Source
pCP20	repA101	Carries gene for flippase enzymes	Amp	CGSC
pZS21mcs	pSC101	$P_{LtetO-1}$ expression vector	Kan	Expressys
pQE30	ColE1	P_{T5} promoter/lac operon	Amp	Invitrogen
pZF11	pSC101	pZS21mcs, $P_{LtetO-1}$: *aar* from *S. elongatus*	Kan	This study
pZF13	ColE1	pQE30, P_{T5}: *aar,* codon optimized	Amp	This study
pZF15	ColE1	pQE30, P_{T5}: aar, P_{T5}: *ybbO*, codon optimized *aar* and *ybbO* under separate P_{T5}	Amp	Fatma et al., 2016
pZF16	pSC101	pZS21mcs, $P_{LtetO-1}$: *ado* from *N. punctiformes*	Kan	This study
pZF17	ColE1	pQE30, P_{T5}: *ado,* codon optimized from *N. punctiformes*	Amp	This study
pZF18	ColE1	pQE30, P_{T5}: *ado* and *aar* from *S. elongatus*	Amp	This study
pZF19	ColE1	pQE30, P_{T5}: *ado* and *aar, ado* from *N. punctiformes* and *aar* from *S. elongatus,* codon optimized	Amp	This study
pZS20	ColE1	pQE30, P_{T5}: *ado* and *aar*, codon optimized, *aar* from *S. elongatus* and *ado* from *N. puctiformes* PCC73102, *ado,* and *aar* linked with GGGGSGGGGS	Amp	This study
pZF21	ColE1	pQE30, P_{T5}: *ado* and *aar*, codon optimized, *aar* from *S. elongatus* and *ado* from *N. puctiformes* PCC73102, *ado* and *aar* linked with GGGGSGGGGSGGGG	Amp	This study
pZF22	ColE1	pQE30, P_{T5}: *aar*, codon optimized from *S. elongatus* P_{T5}: *ado,* codon optimized from *N. punctiformes*	Amp	This study
pZF23	pSC101	pZS21mcs, $P_{LtetO-1}$: *zwf* from *E. coli* DH5α	Kan	This study

Table 2. Details of predicted gene target for enhancing hydrocarbon production

Enzymes	Corresponding gene	Reaction name	Perturbation	Pathway
Glucose-6-phosphate-1-dehydrogenase (G6PDH)	*zwf*	GLU6PDEHYDROG-RXN	Overexpression	Pentose phosphate
Phosphogluconate dehydratase (EDD)	*edd*	6PGLUCONOLACT-RXN	Knockout	Entner-Doudoroff
Phosphoenol pyruvate synthetase (PPS)	*ppsA*	PEPSYNTH-RXN	Knockout	Gluconeogenesis
D-lactate dehydrogenase (LDH)	*ldhA*	DLACTDEHYDROGNAD-RXN	Knockout	Fermentative
Isocitrate lyase (ICL)	*aceA*	ISOCIT-CLEAV-RXN	Knockout	Glyoxylate
Phosphate acetyltransferase (PTA)	*pta*	PHOSACETYLTRANS-RXN	Knockout	Fermentative
Pyruvate oxidase (POX)	*poxB*	RXN-11496	Knockout	Fermentative
Pyruvate formate-lyase (PFL)	*pflB*	PYRUVFORMLY-RXN	Knockout	Fermentative

Figures legends

Fig. 1. Central carbon metabolic network of recombinant *E. coli* expressing the alkane biosynthesis pathway. The model comprises 74 reactions, 45 of which are irreversible, and 61 metabolites. Precursor metabolites involved in biomass formation are glucose-6-phosphate, fructose-6-phosphate, ribose-5-phosphate, 2-ketoglutarate, oxaloacetic acid, erythrose-4-phosphate, phospho-enol-pyruvate, 3-phosphoglycerate, glyceraldehyde-3-phosphate, Acetyl-CoA, ATP, NAD, NADPH, and ammonia. Reactions selected for overexpression and knockout are shown by the symbols ↑ and X, respectively. The common names of all abbreviated metabolites are given in Table S2.

Fig. 2. Dendrogram representing the correlation of reactions present in the model. The dendrogram shows the correlations between the fluxes of the reactions present in the metabolic network after forcefully imposing the production of alkane. The common names of all abbreviated reactions are given in Table S2. The reactions on the same level (such as the alkane biosynthesis pathway) represent perfectly correlated fluxes that change in direct proportion with one another. Reactions on adjacent branches (e.g. lower glycolysis and the start of the pentose phosphate pathway) show higher degrees of correlation than more distant branches. The cluster at the bottom of the diagram is the next nearest correlated (both positively and negatively), containing reactions of pyruvate and acetate metabolism as well as reactions of the Entner-Duodoroff pathway. The cluster at the top is the least correlated and contains the rest of central metabolism such as the remainder of the pentose phosphate pathway and the TCA cycle. Positively and negatively correlated reactions with the alkane flux are distinguished by examination of the flux profiles as in Figs. 3 and S3.

Fig. 3. Co-variation of reactions with imposed alkane production. Fluxes of those reactions showing major changes after imposing pentadecane production are plotted at each level of alkane production. The first two reactions of the PPP, catalyzed by glucose-6-phosphate dehydrogenase and 6-phosphoglucono lactonase along with the pyruvate dehydrogenase complex and the reactions of the lower half of glycolysis show the major changes, though not direct proportionality. A negative value

of flux indicates the export of metabolites from the cells; hence the highest output of pentadecane is on the left of the graph. Common names of all abbreviated reactions are given in Table S2.

Fig. 4. Yield bar graph showing yield of metabolites after imposing overexpression and knockouts of genes under *in silico* conditions. Effect of gene overexpression and knockouts on alkane, acetate, lactate, and formate production are shown. The improvement in alkane yield was determined by calculating the percentage of alkane production per carbon source (glucose). Abbreviations: zwf- denotes the overexpression of only G6PDH; *edd*- phosphogluconate dehydratase; *pps*- phosphoenol pyruvate synthetase; *ldhA*- lactate dehydrogenase; *aceA*- isocitrate lyase; *pta*- phosphate acetyltransferase; *poxB*- pyruvate oxidase; *pflB*-pyruvate formate-lyase. Deletion of genes is denoted by Δ sign. Maximum theoretical yield was determined by using Equation 4.

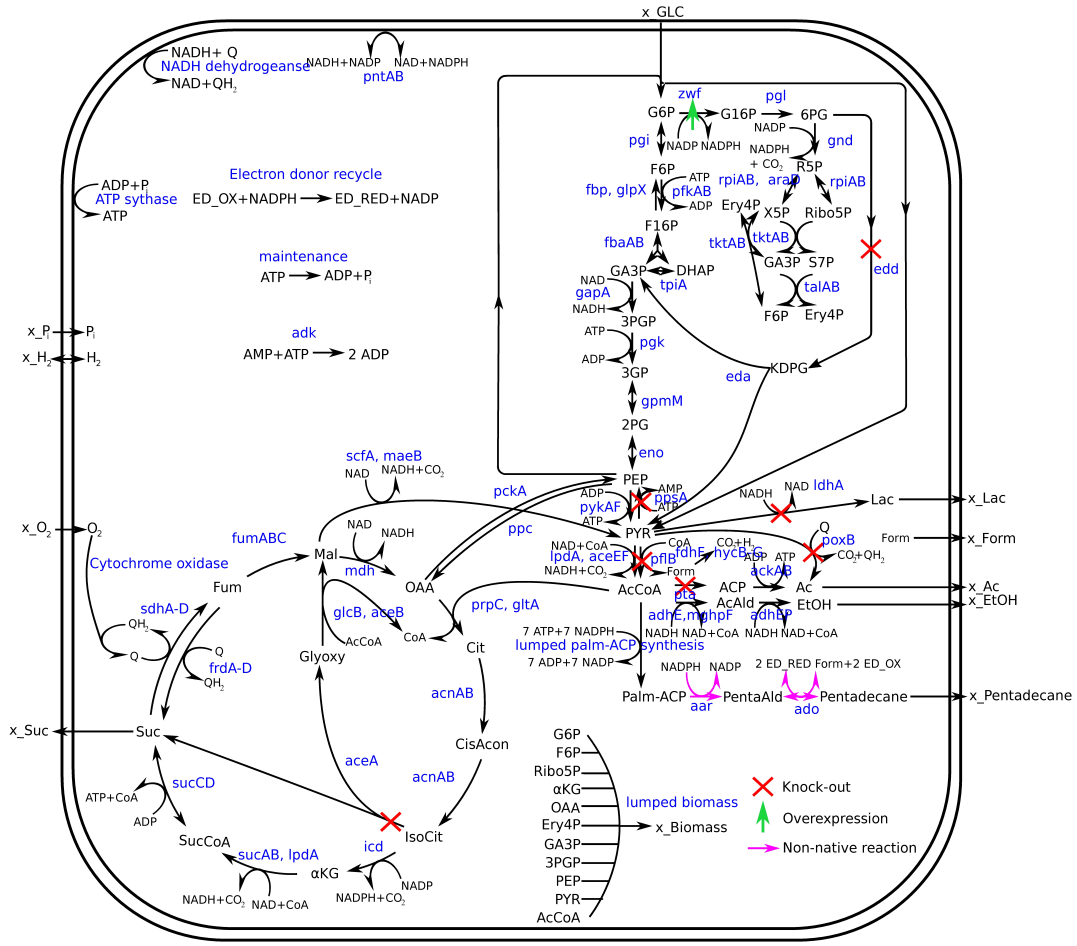
Fig. 5. Comparison of alkane production in different constructs and *E. coli* strains. (A) Optimization of plasmid constructs for alkane production. (B) Eight *E. coli* strains (M15, MG1655, *E. coli* B, BW25113, Top10, JM109, BL21, and DH5α) were transformed with plasmid construct having the gene for *aar* and *ado* with individual T5 promoter and rrnB terminator (pZF22). Single colonies were grown in M9 modified medium supplemented with 2% glucose, and the samples were extracted after 48 hr for hydrocarbon analysis.

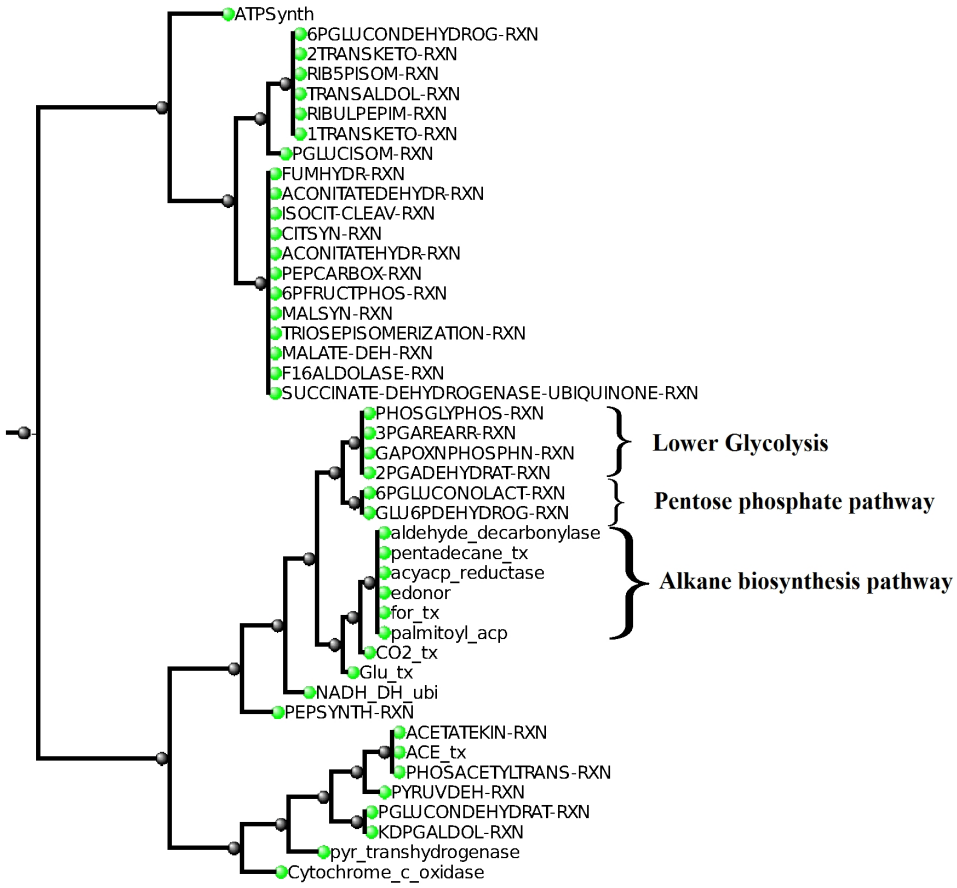
Fig. 6. Alkane and fatty alcohol production analysis in engineered strains. See Table 1 for the gene complements of the strains. (A) The long chain alk(e)ne analysis in engineered strains transformed with a vector construct having a codon optimized *aar* and *ado* gene with individual promoters and terminators (pZF22) and pZF23. (B) Fatty alcohol analysis in strains transformed with a construct having a codon optimized *aar* gene (pZF13) for alcohol production and a *zwf* gene (pZF23) for G6PDH formation. All strains were grown in M9 modified minimal medium supplemented with 2% glucose in a test tube (3ml culture volume), induced with 0.01 mM IPTG and 100 ng/μL of anhydrotetracycline at the time of inoculation. Samples were analyzed after 48 hr.

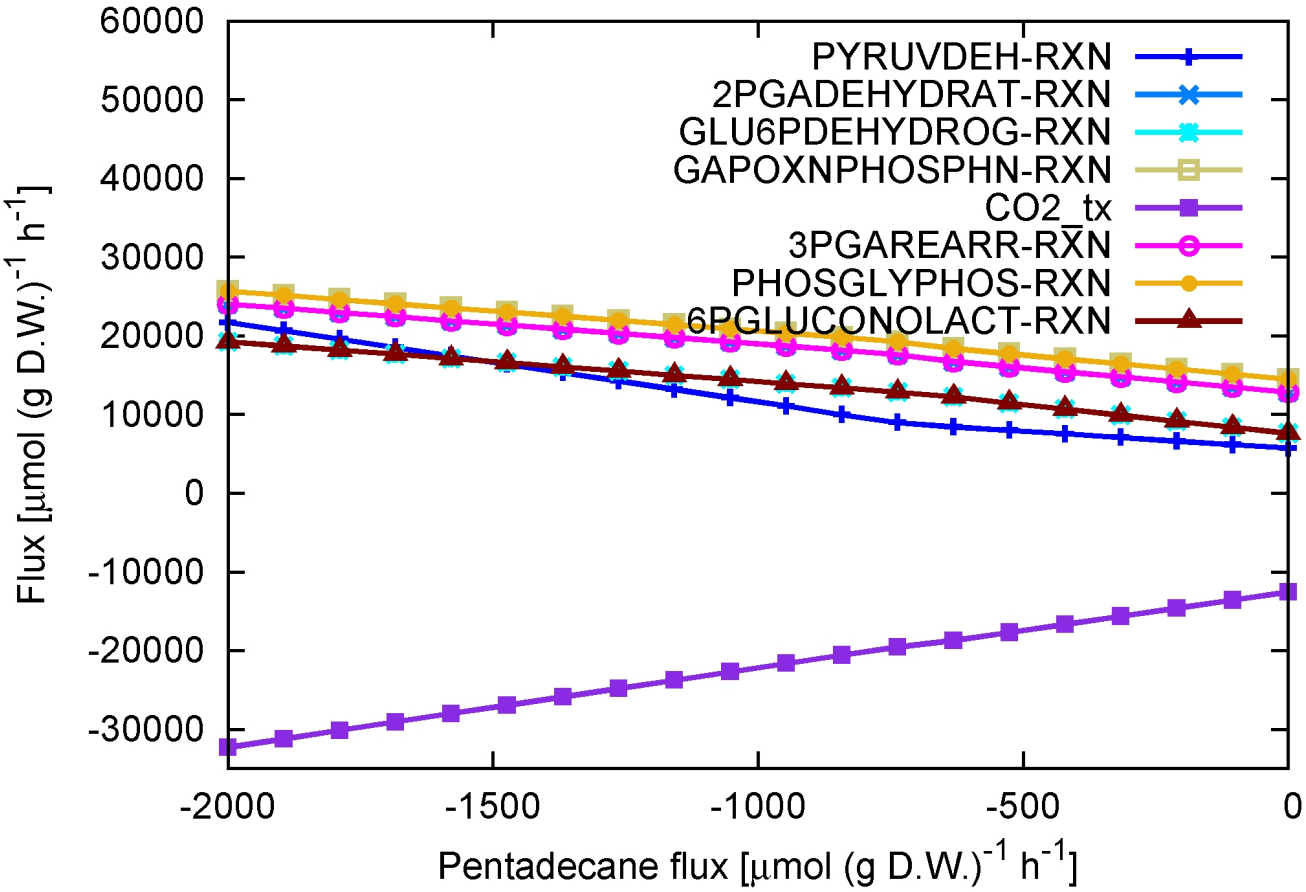
Fig. 7. Effect of phospholipid pathway manipulation on the alkane and fatty alcohol production. (A)

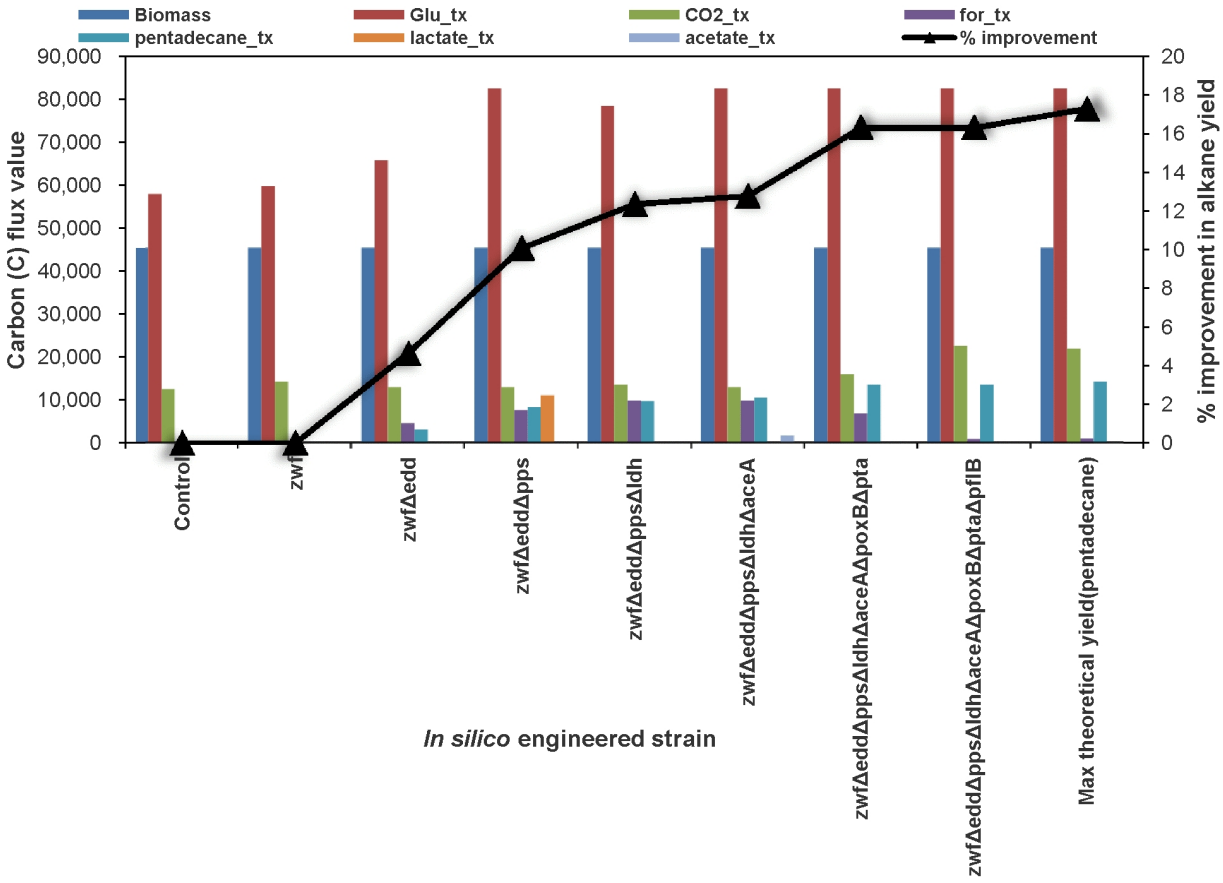
ZFM8 strain was generated by knocking out the *plsX* gene from *E. coli* DH5α strain (ZFM5), and the strain was transformed with pZF22 and pZF23 for the comparative analysis of the alkane production. (B) ZFM9 strain was generated by knocking out the *plsX* gene from *E. coli* DH5α strain (ZFM7), and the strain was transformed with pZF15 and pZF23 for the comparative analysis of the fatty alcohol production.

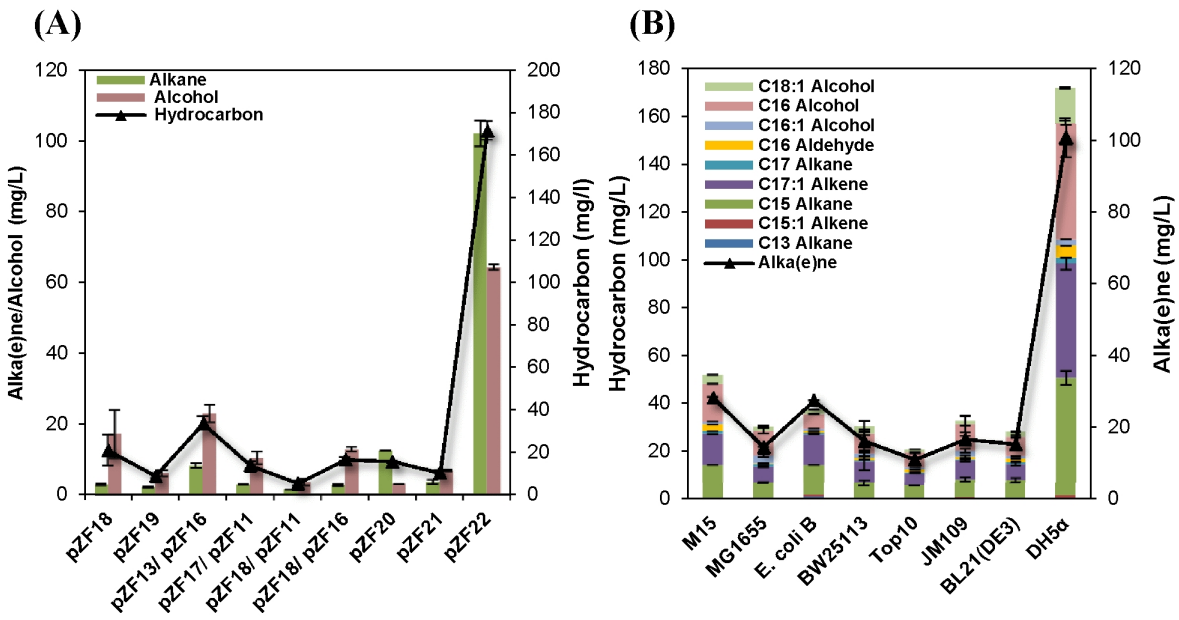
Fig. 8. Fed-batch cultivation for the alkane and fatty alcohol production. (A) Time profile and chain length distribution of alkane producing strain, *E. coli* DH5α ZFM8/pZF22/pZF23 under fed-batch cultivation. (B) Time profile and chain length distribution of fatty alcohol producing strain, *E. coli* DH5α strain ZFM9/pZF15/pZF23 under fed-batch cultivation in mineral medium with glucose as a carbon source.

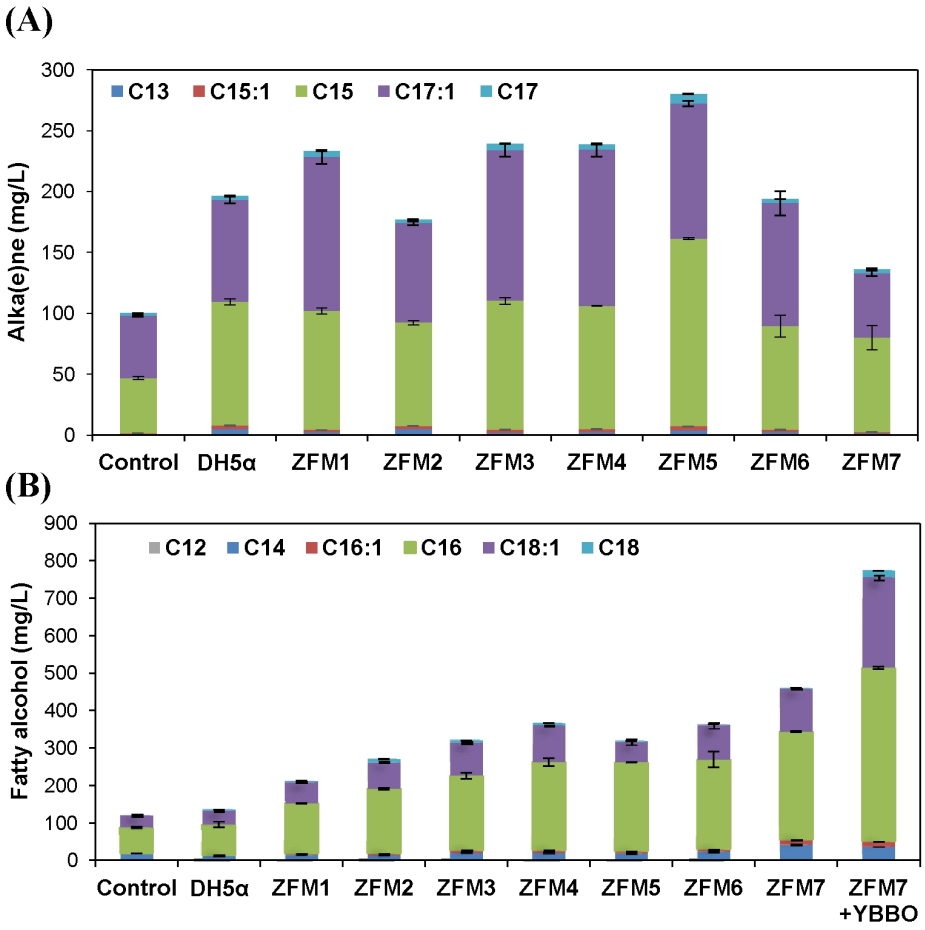

x_GLC
NADH+ Q
NADH dehydrogeanse
NAD+QH2
NADH+NADP
NAD+NADPH
pntAB
zwf
pgl
G6P
G16P
6PG
NADP
pgi
NADPH
NADPH + CO2
gnd
R5P
F6P
ATP
ADP
fbp, glpX
pfkAB
rpiAB, araD
rpiAB
Ery4P
X5P
Ribo5P
F16P
fbaAB
tktAB
tktAB
GA3P
S7P
edd
GA3P
DHAP
tpiA
NAD
gapA
NADH
talAB
F6P
Ery4P
3PGP
ATP
pgk
ADP
3GP
KDPG
eda
gpmM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

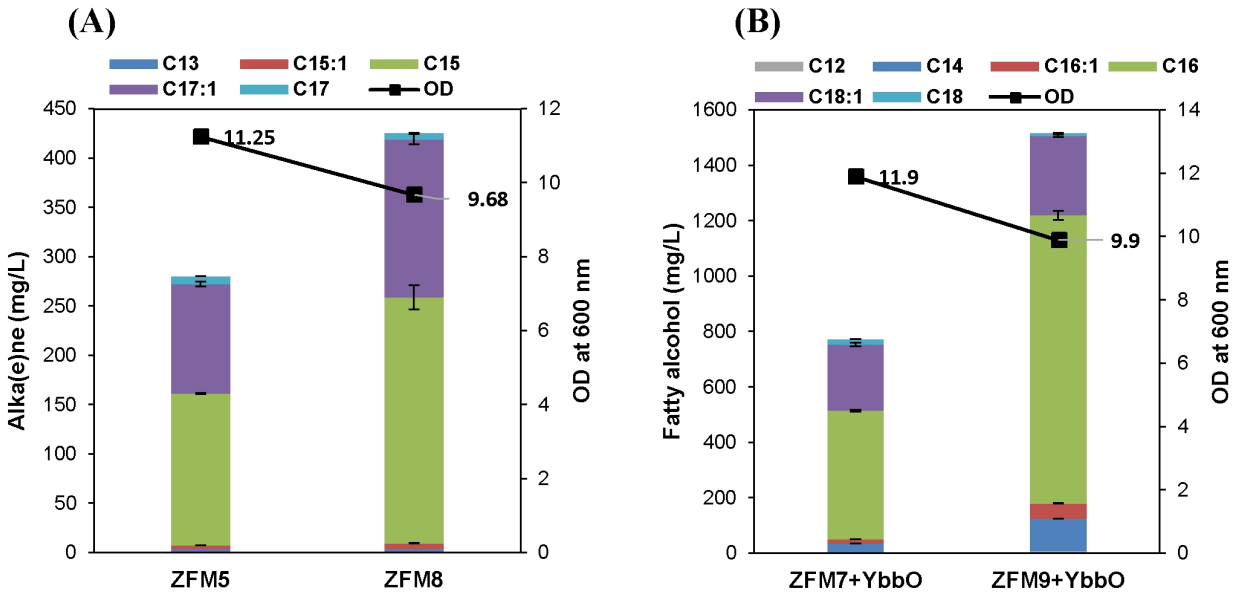

ATPSynth
6PGLUCONDEHYDROG-RXN
2TRANSKETO-RXN
RIB5PISOM-RXN
TRANSALDOL-RXN
RIBULPEPIM-RXN
1TRANSKETO-RXN
PGLUCISOM-RXN
FUMHYDR-RXN
ACONITATEDEHYDR-RXN
ISOCIT-CLEAV-RXN
CITSYN-RXN
ACONITATEHYDR-RXN
PEPCARBOX-RXN
6PFRUCTPHOS-RXN
MALSYN-RXN
TRIOSEPISOMERIZATION-RXN
MALATE-DEH-RXN
F16ALDOLASE-RXN
SUCCINATE-DEHYDROGENASE-UBIQUINONE-RXN
PHOSGLYPHOS-RXN
3PGAREARR-RXN
GAPOXNPHOSPHN-RXN
2PGADEHYDRAT-RXN
6PGLUCONOLACT-RXN
GLU6PDEHYDROG-RXN
aldehyde_decarbonylase
pentadecane_tx
acyacp_reductase
edonor
for_tx
palmitoyl_acp
CO2_tx
Glu_tx
NADH_DH_ubi
PEPSYNTH-RXN
ACETATEKIN-RXN
ACE_tx
PHOSACETYLTRANS-RXN
PYRUVDEH-RXN
PGLUCONDEHYDRAT-RXN
KDPGALDOL-RXN
pyr_transhydrogenase
Cytochrome_c_oxidase
Lower Glycolysis
Pentose phosphate pathway
Alkane biosynthesis pathway

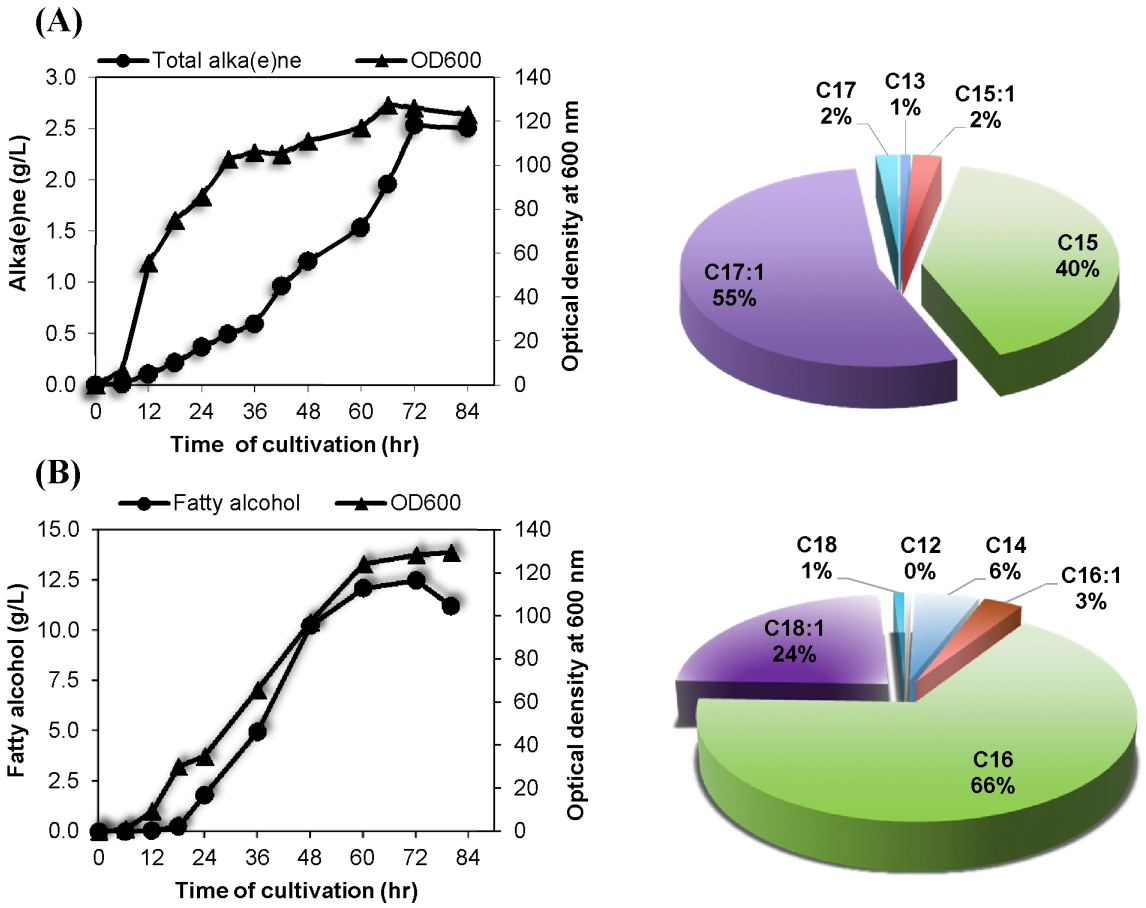

60000
50000
40000
30000
20000
10000
0
-10000
-20000
-30000
-2000
-1500
-1000
-500
0
Flux [µmol (g D.W.)-1 h-1]
Pentadecane flux [µmol (g D.W.)-1 h-1]
PYRUVDEH-RXN
2PGADEHYDRAT-RXN
GLU6PDEHYDROG-RXN
GAPOXNPHOSPHN-RXN
CO2_tx
3PGAREARR-RXN
PHOSGLYPHOS-RXN
6PGLUCONOLACT-RXN


Biomass
Glu_tx
CO2_tx
for_tx
pentadecane_tx
lactate_tx
acetate_tx
% improvement
Carbon (C) flux value
% improvement in alkane yield
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
20
18
16
14
12
10
8
6
4
2
0
Control
zwf
zwfΔedd
zwfΔeddΔpps
zwfΔeddΔppsΔldh
zwfΔeddΔppsΔldhΔaceA
zwfΔeddΔppsΔldhΔaceAΔpoxBΔpta
zwfΔeddΔppsΔldhΔaceAΔpoxBΔptaΔpflB
Max theoretical yield(pentadecane)
In silico engineered strain


(A)
Alka(e)ne/Alcohol (mg/L)
Hydrocarbon (mg/l)
Alkane
Alcohol
Hydrocarbon
pZF18
pZF19
pZF13/ pZF16
pZF17/ pZF11
pZF18/ pZF11
pZF18/ pZF16
pZF20
pZF21
pZF22
(B)
Hydrocarbon (mg/L)
Alka(e)ne (mg/L)
C18:1 Alcohol
C16 Alcohol
C16:1 Alcohol
C16 Aldehyde
C17 Alkane
C17:1 Alkene
C15 Alkane
C15:1 Alkene
C13 Alkane
Alka(e)ne
M15
MG1655
E. coli B
BW25113
Top10
JM109
BL21(DE3)
DH5α


(A)
Alka(e)ne (mg/L)
300
250
200
150
100
50
0
C13
C15:1
C15
C17:1
C17
Control
DH5α
ZFM1
ZFM2
ZFM3
ZFM4
ZFM5
ZFM6
ZFM7
(B)
Fatty alcohol (mg/L)
900
800
700
600
500
400
300
200
100
0
C12
C14
C16:1
C16
C18:1
C18
Control
DH5α
ZFM1
ZFM2
ZFM3
ZFM4
ZFM5
ZFM6
ZFM7
ZFM7 +YBBO


(A)
C13
C15:1
C15
C17:1
C17
OD
Alka(e)ne (mg/L)
OD at 600 nm
11.25
9.68
ZFM5
ZFM8
(B)
C12
C14
C16:1
C16
C18:1
C18
OD
Fatty alcohol (mg/L)
OD at 600 nm
11.9
9.9
ZFM7+YbbO
ZFM9+YbbO


(A)
Total alka(e)ne
OD600
Alka(e)ne (g/L)
Optical density at 600 nm
Time of cultivation (hr)
(B)
Fatty alcohol
OD600
Fatty alcohol (g/L)
Optical density at 600 nm
Time of cultivation (hr)
C17 2%
C13 1%
C15:1 2%
C15 40%
C17:1 55%
C18 1%
C12 0%
C14 6%
C16:1 3%
C18:1 24%
C16 66%

Supplementary data for

Model-assisted metabolic engineering of *Escherichia coli* for long chain alkane and alcohol production

Zia Fatma[a,c], Hassan Hartman[d], Mark G Poolman[d], David A. Fell[d], Shireesh Srivastava[b,c], Tabinda Shakeel[a,c] and Syed Shams Yazdani[a,c]*

[a]*Microbial Engineering Group,* [b]*Systems Biology for Biofuel Group,* [c]*DBT-ICGEB Centre for Advanced Bioenergy Research, International Centre for Genetic Engineering and Biotechnology, New Delhi, India;* [d]*Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK*

*Corresponding Author

Address of correspondence: Microbial Engineering Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi-110067, India. Email – shams@icgeb.res.in; Phone - +91 11 26742357; Fax - +91 11 26742316.

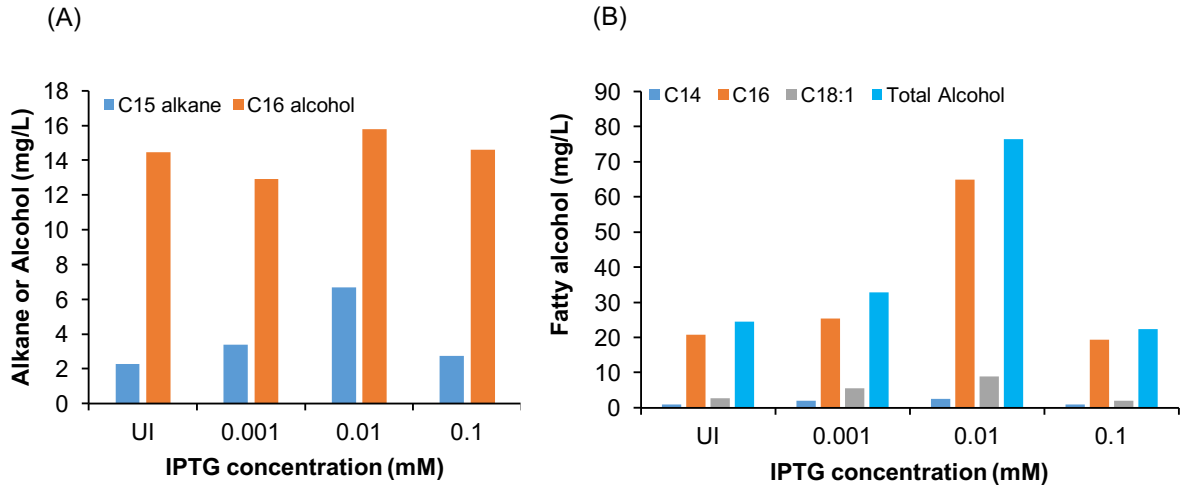


Fig. S1. Effect of IPTG inducer concentration on long chain alkane and fatty alcohol production. (A) The *E. coli* DH5α strain transformed with pZF17 (carrying the *ADO* gene from *N. punctiformes*) plasmid was used for alkane production in presence of exogenous C16 aldehyde (substrate for ADO). (B) The *E. coli* DH5α strain transformed with pZF13 (carrying a codon-optimized *aar* gene from *S. elongatus*) was used for the analysis of various chain length fatty alcohol production.

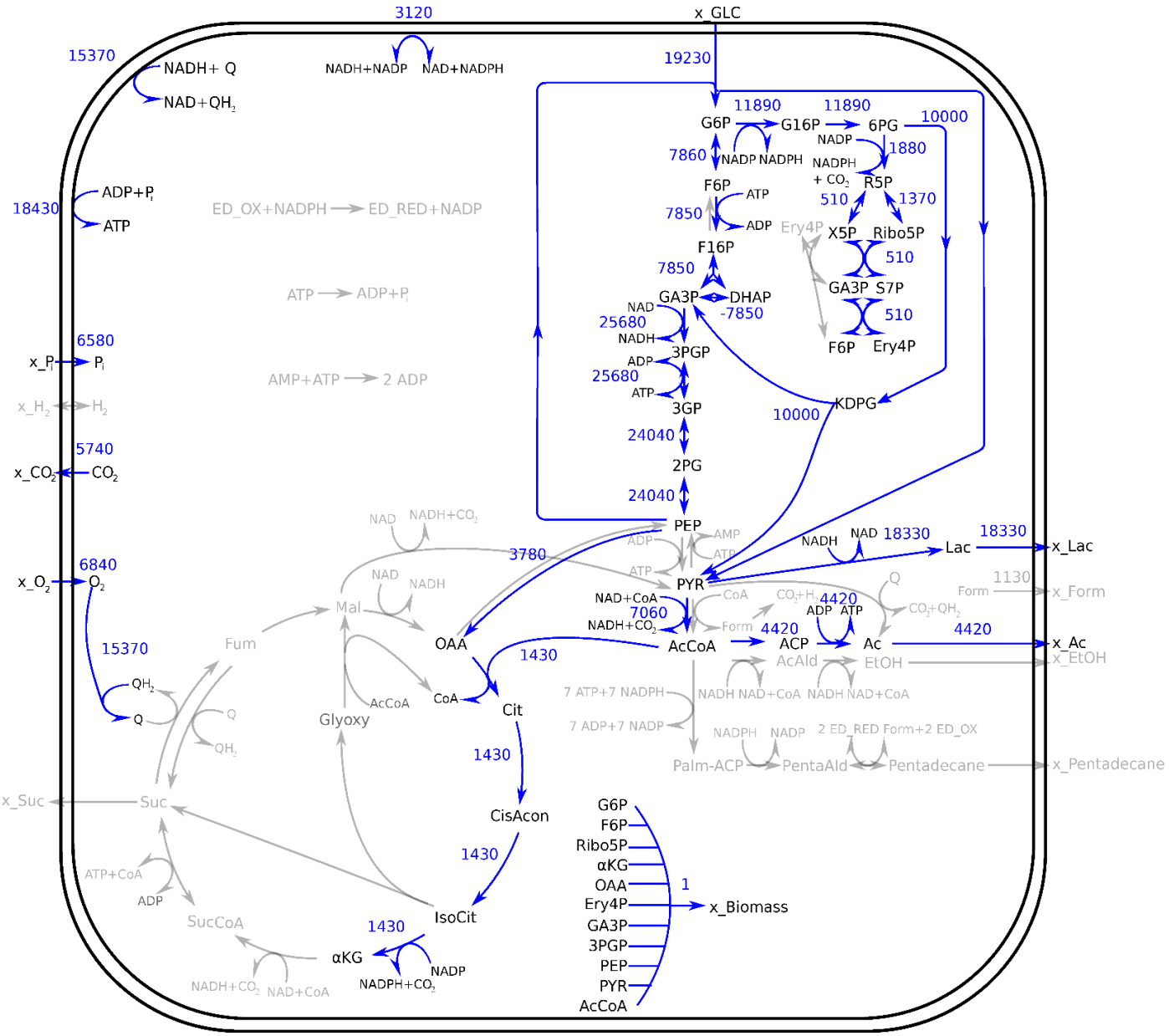
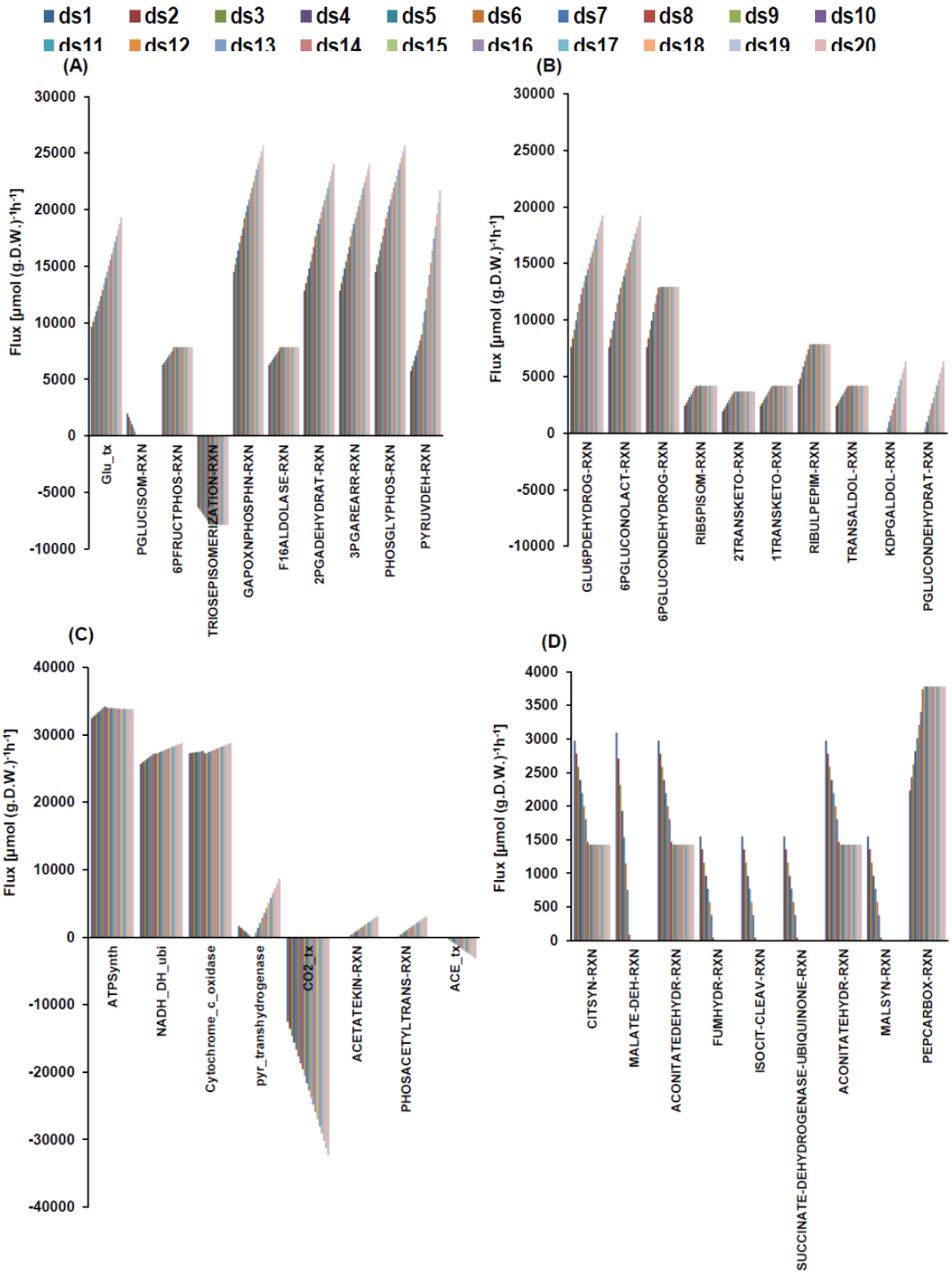


Fig. S2. Representation of flux distribution after increasing glucose uptake by two fold. Unit of represented fluxes is in µmol (gDW^{-1}) h^{-1}


ds1 ds2 ds3 ds4 ds5 ds6 ds7 ds8 ds9 ds10
ds11 ds12 ds13 ds14 ds15 ds16 ds17 ds18 ds19 ds20
(A)
Flux [µmol (g.D.W.)-1h-1]
30000
25000
20000
15000
10000
5000
0
-5000
-10000
Glu_tx
PGLUCISOM-RXN
6PFRUCTPHOS-RXN
TRIOSEPISOMERIZATION-RXN
GAPOXNPHOSPHN-RXN
F16ALDOLASE-RXN
2PGADEHYDRAT-RXN
3PGAREARR-RXN
PHOSGLYPHOS-RXN
PYRUVDEH-RXN
(B)
Flux [µmol (g.D.W.)-1h-1]
30000
25000
20000
15000
10000
5000
0
-5000
-10000
GLU6PDEHYDROG-RXN
6PGLUCONOLACT-RXN
6PGLUCONDEHYDROG-RXN
RIB5PISOM-RXN
2TRANSKETO-RXN
1TRANSKETO-RXN
RIBULPEPIM-RXN
TRANSALDOL-RXN
KDPGALDOL-RXN
PGLUCONDEHYDRAT-RXN
(C)
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
(D)
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

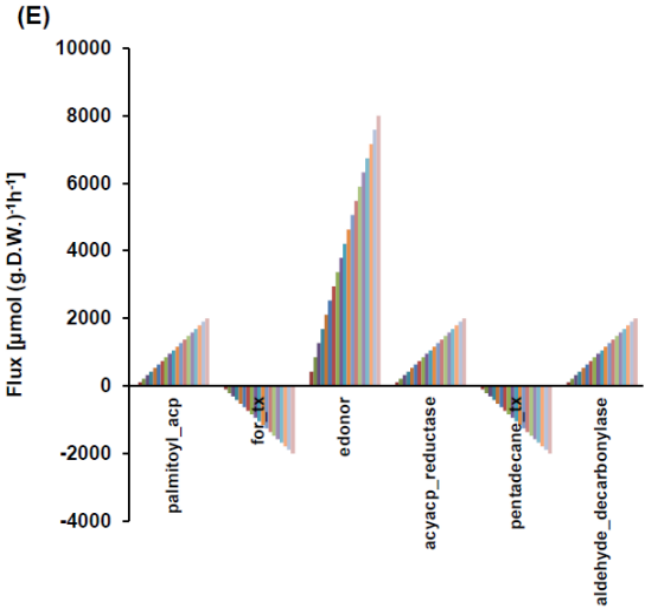


Fig. S3. Effect on the flux of metabolic reactions presents in the model after imposing hydrocarbon production on the system. The graph represents variation in reactions of (A) glycolysis pathway up to acetyl-CoA, (B) pentose phosphate pathway and Entner Doudoroff pathway, (C) electron transport chain along with acetate producing reactions, (D) tricarboxylic acid cycle, glyoxylate shunt, and anaplerotic pathway, and (E) alkane biosynthesis pathway. Graphs were plotted for 20 LP dataset after imposing alkane formation in the range of 0 to 2000 µmol (gDW^{-1}) h^{-1} in 20 steps.

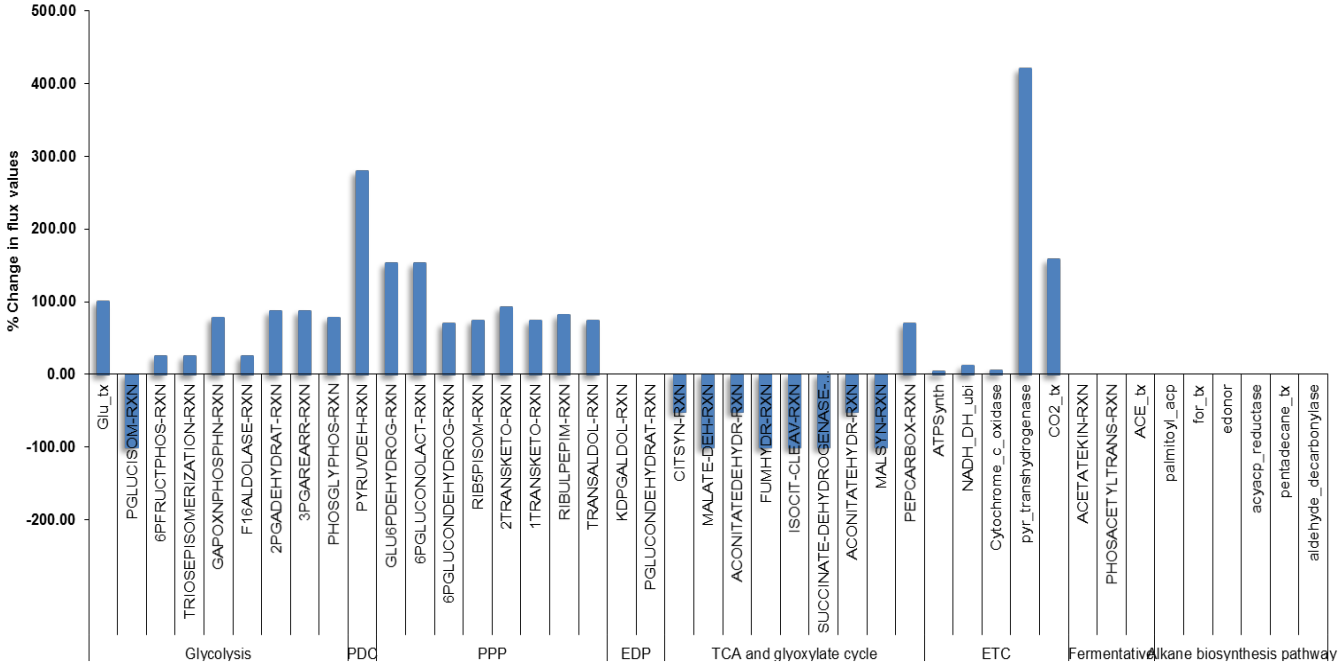


Fig. S4. Percentage change in flux values of reactions after imposing alkane production. A negative value of flux indicates either the reactions are going in opposite direction or metabolites are transported outside the system.

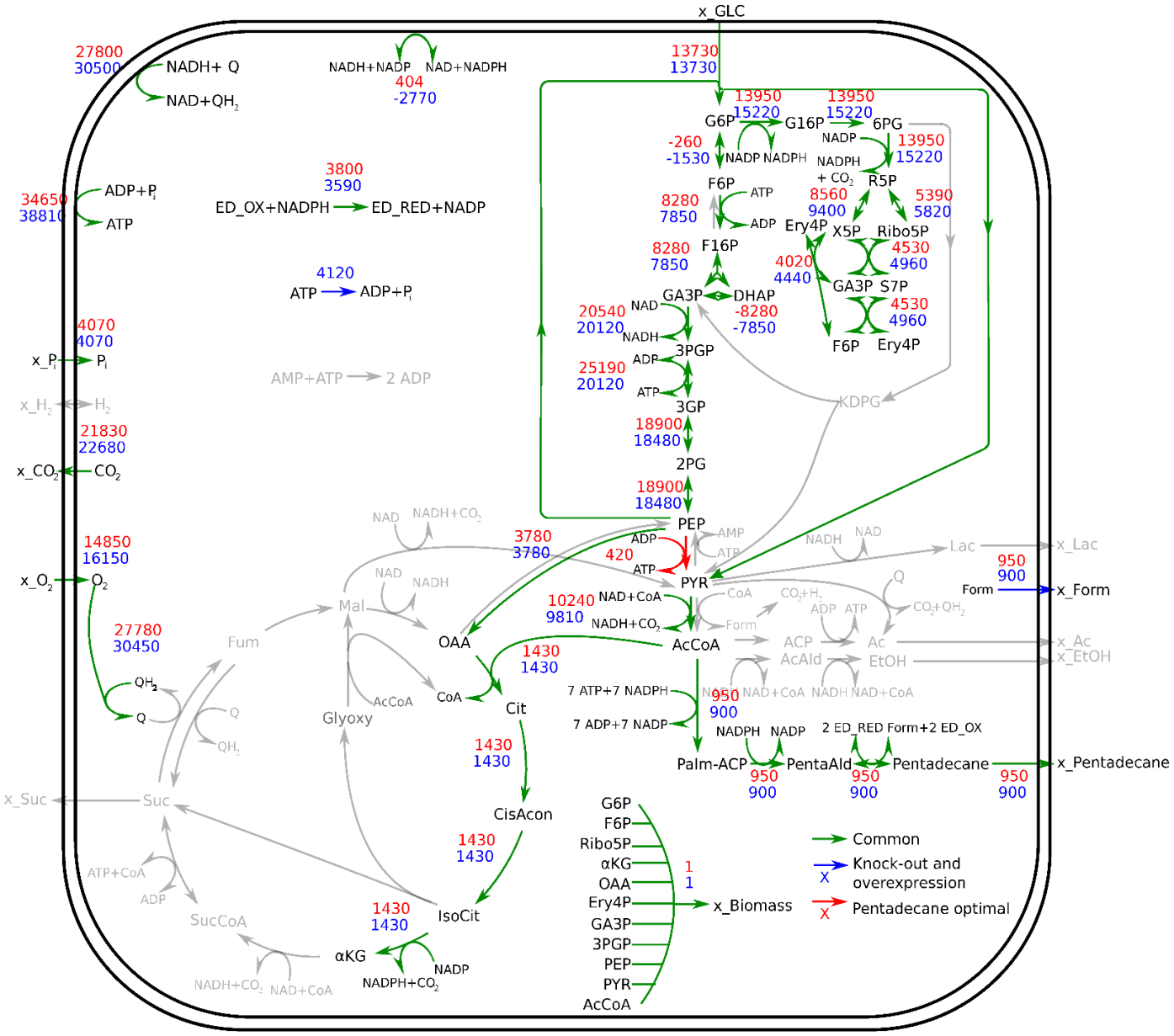


Fig. S5. Metabolic model showing the similarity between the optimal solution (FBA result after simulation of Equation 4) and stepwise predicted solution (FBA result after implementing the overexpression of *zwf* and deletion of *edd*, *pps*, *ldh*A, *ace*A, *pox*B, *pta*, *pfl*B). Unit of represented fluxes is in μmol (gDW^{-1}) h^{-1}

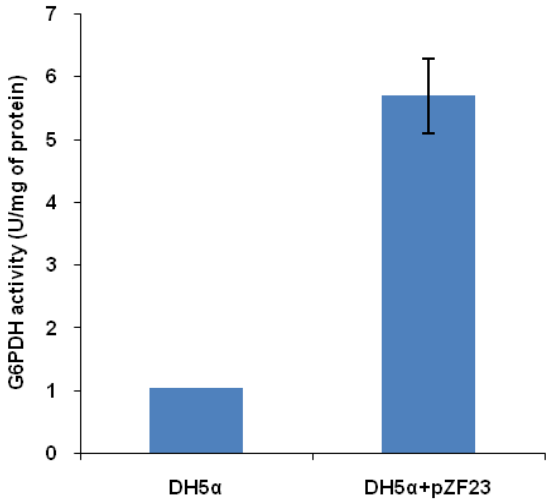


Fig. S6. Enzyme activity of the glucose-6-phosphate dehydrogenase (G6PDH) present in *E. coli* DH5α strain and DH5α overexpressing *zwf* (encodes G6PDH). *zwf* gene was cloned in low copy number plasmid pZS21mcs (pZF23). The cells were grown aerobically in LB broth medium, induced with 100 ng/µL anhydrotetracycline when OD at 600 nm reached to 0.6 and cells were further grown for additional 4 hours, cells were permeabilised through chloroform and then used for determining the G6PDH enzyme activity.

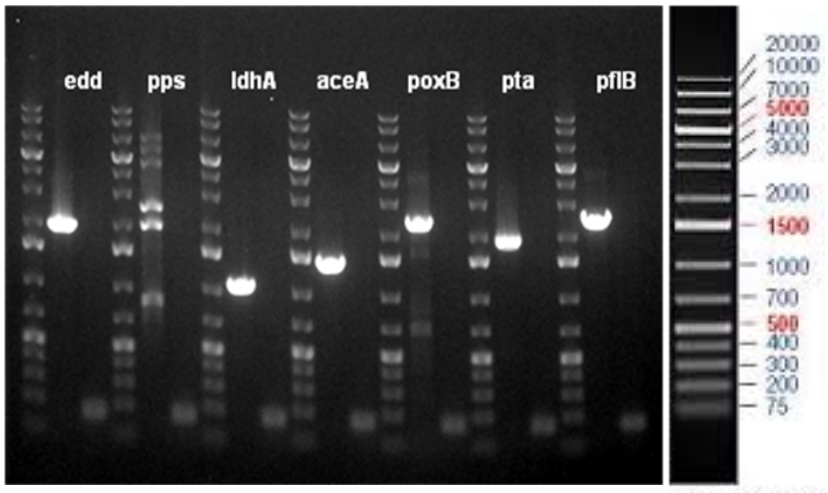


Fig. S7. Confirmation of gene knockout in engineered strain ZFM7 by PCR and gel electrophoresis using primers complementary to a flanking region of the genes. For each gene, the first lane is a 1 Kb plus DNA ladder; the second lane corresponds to the amplified product from the wild-type strain (DH5α), and the third lane is the amplified product from the engineered strain ZFM7. The smaller size (~140 Kb) of DNA bands in the third lane correspond to the FRT-kan-FRT scar in the engineered strain confirm the knockout of the genes.

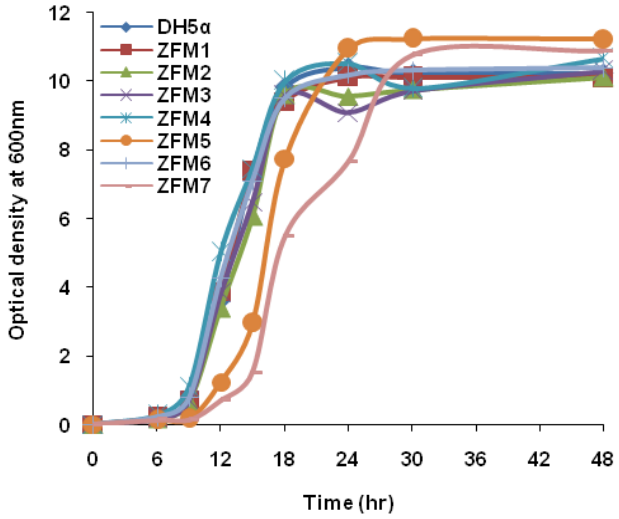


Fig. S8. Growth analysis of consecutive gene knockout strains under aerobic condition. The optical density (cell growth) of sequential knockout strains was compared up to 48 h to observe the effect of gene deletions on cell growth.

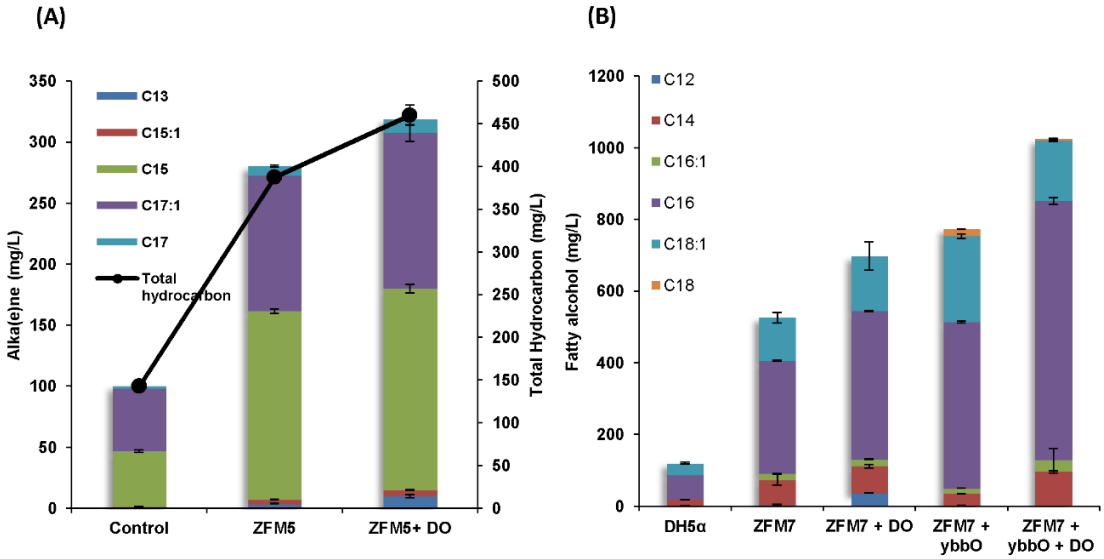


Fig. S9. Effect of dodecane layering on alka(e)ne and fatty alcohol production. (A) The strain used was ZFM5 transformed with pZF23 (carrying the *zwf* gene) and pZF22 (carrying codon optimized *aar* and *ado* genes with individual promoters and terminators) compared for alka(e)ne production in the absence and presence of dodecane. (B) ZFM7 transformed with pZF23 and pZF15 (carrying a codon optimized *aar* and native *ybbO* gene) compared for fatty alcohol production in the absence and presence of dodecane. Dodecane overlaying was represented as DO.

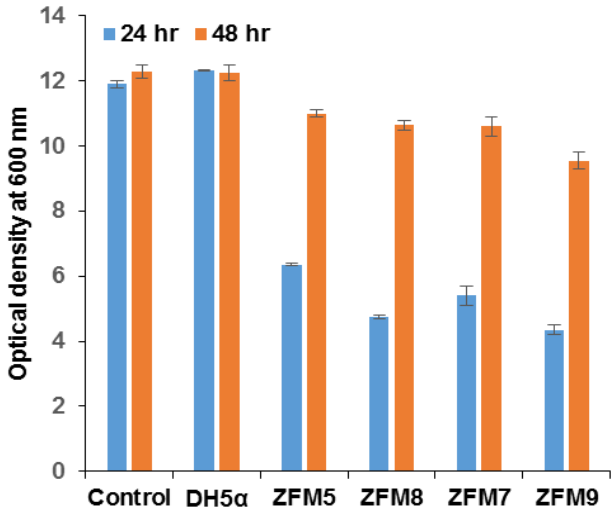


Fig. S10. Effect of *pls*X deletion on cell growth. ZFM8 and ZFM9 strains were generated by deletion of *pls*X gene from strain ZFM5 and ZFM7, respectively. Control is the DH5α cells were transformed with pQE30 and pZS21mcs plasmids; DH5α corresponds to DH5α cells transformed with pZF22 plasmid for alkane production; ZFM5 and ZFM8 are the engineered strains and were transformed with plasmids pZF22 and PZF23 for alkane production; ZFM7 and ZFM9 strains were transformed with plasmids pZF15 and pZF23 for fatty alcohol production. OD of starting culture was maintained at 0.02 in all the culture and OD was measured at 24 and 48 hr. Each data is the average of two biological replicate

Table S1. Oligonucleotide primers used in study

Primers used for gene cloning				
Primer Name	**Sequence (5' to 3')**	**Gene amplified, source**	**Plasmid backbone**	**Name of constructed plasmid**
Synecho_ADO_For	ACTGGATCCATGCCGCAGCTTGAAGCCAG	*ado*, *S. elongates,* native	pQE30	pZF18
Synecho_ADO_Rev	AGTCGAGCTCTCAAACGGCCGCAAGGCCATAGG			
pZS_optADO_For	CACTGTCGACATGCAGCAACTGACGGAT	*ado, N. punctiformes,* codon optimized (CO)	pZS21mcs	pZF16
pZS_optADO_Rev	ACTAAGCTTTTAGTGGTGGTGGTGATG			
pQE_optADO_For	CGCGGATCCATGCAGCAACTGACGGAT	*ado, N. punctiformes*, CO	pQE30	pZF17 and pZF19
pQE_optADO_Rev	AGAGCTCTTAAGCGCCAATCAGACCGTA			
Synecho_AAR_For	AGTAGAGCTCGAAGGAGATATACCATGTTCGGTCTTATCGGTC	*aar, S. elongates,* native	pQE30	pZF19
Synecho_AAR_Rev	ACGCGTCGACTCAATGATGGTGATGATGATGAATTGCCAATGCCAAGGG			
pQE_optAAR_For	AGAGCTCATGTTTGGCCTGATTGG	*aar, S. elongates,* CO	pQE30	pZF13 and pZF19
pQE_optAAR_Rev	CACTGTCGACTTAATGATGGTGATGATGATG GATTGCCAGAGCCAG			
pQE_XbaI_For	GCTCTAGAACGAGGCCCTTTCGTCTTCAC	For cloning with P_{T5} promoter and rrnB terminator	pQE30	pZF22
pQE_XbaI_Rev	GCTCTAGAGGGCGGATTTGT			
ADO_GS2_Rev	AGAGCTCCGACCCACCACCGCCCGAGCCACCGCCACCAGCGCCAATCAGACCGTA	Used for making Fusion constructs, CO *ado* from *N. punctiformes* and *aar* from *S. elongates*	pQE30	pZF21 and pZF22
ADO_GS3_Rev	AGAGCTCCGAGCCACCGCCACCCGACCCACCACCGCCCGAGCCACCGCCACCAGCGCCAATCAGACCGTA			
OptAAR_Rev	CACTGTCGACTTAGATTGCCAGAGCCAG			
E.coli_zwf_For	GGGGTACCGAGCTCGAAGGAGATATACATGGCGGTAACGCAA	*zwf, E. coli,* native	pZS21mcs	pZF23
E.coli_zwf_Rev	CCCAAGCTTTTAATGATGGTGATGATGATGCTCAAACTCATTCCAGGA			
Primers used for confirmation of gene knockout				
Primer Name	**Sequence (5' to 3')**	**Gene name**		
phage_edd_For	TGACAACTCAATTTCAGGAG	*edd*		
phage_edd_Rev	CTCGCCTGATTACAAATTTG			
phage_pps_For	TCATTTATCACAAAAGGATTGT	*pps*		
phage_pps_Rev	GACTAAACGCCGCCGGG			
phage_ldh_For	AACATCACTGGAGAAAGTCTT			

phage_ldh_Rev	TGCAGGGGAGCGGCAAGA	*ldhA*
phage_aceA_For	AACCACCACATAACTATGGA	*aceA*
phage_aceA_Rev	TACAGTCAGCAACGGTTGT	
phage_poxB_For	TATCACATTCAGGAGATGGA	*poxB*
phage_poxB_Rev	ATTATGACGGGAAATGCCAC	
phage_ptaA_For	GTAACGAAAGAGGATAAACC	*ptaA*
phage_ptaA_Rev	AGCTGCGGATGATGACGA	
phage_pflB_For	TCGGGATCCAACCATGCGAGTTAC GGGCCTATAA	*pflB*
phage_pflB_Rev	CCCAAGCTTGTGCCTGTGCCAGTG GTTGCTGTGA	
phage_plsX_For	GGAAAGACCAAACCGGG	*plsX*
phage_plsX_Rev	CGTCACTTGCAAACTGCG	

Note: The restriction sites are underline

Table S2. List of abbreviations and MetaCyc unique identifiers used for reactions and metabolites

Abbreviation/MetaCyc UID	Common name
Reactions	
1TRANSKETO-RXN	Transketolase I
2PGADEHYDRAT-RXN	2-Phosphoglycerate dehydration
2TRANSKETO-RXN	Transketolase II
3PGAREARR-RXN	Phosphoglyceromutase
6PFRUCTPHOS-RXN	6-Phosphofructokinase II
6PGLUCONDEHYDROG-RXN	Phosphogluconate dehydrogenase
6PGLUCONOLACT-RXN	6-Phosphogluconolactonase
ACETATEKIN-RXN	Acetate kinase
ACE_tx	Acetate transporter
ACONITATEDEHYDR-RXN	Aconitase I
ACONITATEHYDR-RXN	Aconitas II
ATPSynth	ATP synthase
CITSYN-RXN	citrate synthase
CO2_tx	Carbon-dioxide transporter
Cytochrome_c_oxidase	Cytochrome oxidase
F16ALDOLASE-RXN	Fructose-bisphosphate aldolase
FUMHYDR-RXN	Fumarase
GAPOXNPHOSPHN-RXN	Triosephosphate dehydrogenase
GLU6PDEHYDROG-RXN	Glucose 6-phosphate-1-dehydrogenase
Glu_tx	Glucose transporter
ISOCIT-CLEAN-RXN	Isocitrate cleavage
KDPGALDOL-RXN	2-Dehydro-3-deoxy-D-gluconate-6-phosphate aldolase
MALATE-DEH-RXN	Malic dehydrogenase
NADH_DH_ubi	Ubiquinone utilizing NADH dehydrogenase
PEPCARBOX-RXN	Phosphoenolpyruvate carboxylase
PEPSYNTH-RXN	Phosphoenolpyruvate synthetase
PGLUCISOM-RXN	Phosphoglucose isomerase
PGLUCONDEHYDRAT-RXN	Phosphogluconate dehydratase
PHOSACETYLTRANS-RXN	Phosphate acetyltransferase
PHOSGLYPHOS-RXN	Phosphoglycerate kinase
PYRUVDEH-RXN	Pyruvate dehydrogenase
RIB5PISOM-RXN	Ribose-5-phosphate isomerase
RIBULPEPIM-RXN	L-Ribulose 5-phosphate 4-epimerase
SUCCINATE-DEHYDROGENASE-UBIQUINONE-RXN	Succinate: quinone oxidoreductase
TRANSALDOL-RXN	Transaldolase
TRIOSEPISOMERIZATION-RXN	Triosephosphate isomerase
acyacp_reductase	Acyl-ACP reductase
aldehyde_decarbonylase	Aldehyde Decarbonylase
edonor	Artificial electron donor reaction
for_tx	Formate transporter
palmitoyl_acp	Lumped palmitoyl-ACP (acetyl carrier protein) forming reaction
pentadecane_tx	Pentadecane transporter
pyr_transhydrogenase	Proton-dependent transhydrogenase
Metabolites	
2PG	2-Phosphoglycerate
3GP	3-Phosphoglycerate
3PGP	1,3-Bisphospho-D-glycerate
6PG	D-Gluconate 6-phosphate
αKG	2-Oxoglutarate
ACP	Acetyl-phosphate

Ac	Acetate
AcAld	Acetaldehyde
CisAcon	Cis-aconitate
Cit	Citrate
DHAP	Glycerone phosphate
ED_OX/ED_RED	Artificial oxidized/reduced electron donor
Ery4P	D-Erythrose 4-phosphate
EtOH	Ethanol
F16P	Fructose 1,6-bisphosphate
F6P	Fructose-6-phosphate
Fum	Fumarate
G16P	Glucose-1,6-bisphosphate
G6P	D-Glucose 6-phosphate
GA3P	D-Glyceraldehyde 3-phosphate
IsoCit	Iso-citrate
KDPG	2-Dehydro-3-deoxy-D-galactonate 6-phosphate
Lac	Lactate
Mal	Malate
OAA	Oxaloacetate
PEP	Phosphoenolpyruvate
PYR	Pyruvate
Palmitoyl-ACP	Palmitoyl-Acetyl carrier protein
Q/QH_2	Quinol/quinone
R5P	D-Ribulose 5-phosphate
Ribo5P	D-Ribose 5-phosphate
S7P	D-Sedoheptulose 7-phosphate
Suc	Succinate
SucCoA	Succinyl-CoA
X5P	D-Xylulose 5-phosphate

Table S3. Analysis of biomass and extracellular metabolites during alkane and fatty alcohol production*

		Strains				
Time (h)	**Metabolites (g/L)**	**DH5α (pQE30, pZS21mcs)**	**ZFM5 (pZF22, pZF23)**	**ZFM8 (pZF22, pZF23)**	**ZFM7+YbbO (pZF15, pZF23)**	**ZFM9+YbbO (pZF15, pZF23)**
0	Glucose	24.44 ±0.02	23.59 ±0.01	24.27 ±0.01	24.47 ±0.23	24.61 ±0.01
24	Glucose	n.d	3.84 ±0.01	8.33 ±0.01	5.69 ± 0.01	8.55 ±0.01
	Biomass	4.048 ±0.1	2.92 ±0.04	2.18 ±0.05	2.48 ±0.3	2.0 ±0.15
	Succinate	0.71 ±0.01	0.66 ±0.01	0.69 ±0.04	0.08 ±0.00	0.05 ±0.00
	Lactate	0.42 ±0.00	n.d	n.d	n.d	n.d
	Formate	0.3 ±0.00	0.9±0.00	1.84 ±0.18	n.d	n.d
	Acetate	3.61 ±0.00	1.91 ±0.00	1.97 ±0.35	0.34 ±0.00	0.32 ±0.00
48	Glucose	n.d	n.d	n.d	n.d	n.d
	Biomass	5.31 ±0.35	5.06 ±0.1	4.9 ±0.15	4.88±0.30	4.4±0.25
	Succinate	n.d	0.042 ±0.00	0.033 ±0.00	0.06 ±0.05	0.015 ±0.00
	Lactate	0.062 ±0.01	n.d	n.d	n.d	n.d
	Formate	0.28 ±0.01	0.76 ±0.01	1.21 ±0.00	n.d	n.d
	Acetate	n.d	n.d	0.09 ±0.00	0.79 ±0.03	0.91 ±0.00
	Alkane	n.d	0.28 ± 0.01	0.43 ±0.001	n.d	n.d
	Fatty alcohol	n.d	0.07 ±0.00	0.29 ±0.002	0.77 ± 0.01	1.50 ±0.023

*DH5α cells transformed with pQE30 and pZS21mcs plasmids were used as control for the analysis. ZFM5 and ZFM8 strain were used for alkane production; ZFM7 and ZFM9 were used for fatty alcohol production. All values given here are in g/L and are average and standard deviations of two independent biological replicates.
n.d. correspond to not detected in extracellular fraction by HPLC or GC

Table S4. Analysis of individual hydrocarbon (in g/L) produced by engineered strains during fed-batch cultivation*.

Metabolites	Carbon chain length	ZFM8 (84 hr)	ZFM9+YbbO (84 hr)
Aldehyde	C12	n.d	n.d
	C14	n.d	0.00
	C16	0.01	0.04
	C18:1	0.03	0.01
	C18	n.d	0.02
Alkane	C13	0.02	n.d
	C15:1	0.05	n.d
	C15	0.96	n.d
	C17:1	1.42	n.d
	C17	0.05	n.d
Alcohol	C12	n.d	0.06
	C14	n.d	0.69
	C16:1	0.09	0.42
	C16	2.56	8.28
	C18:1	0.45	2.95
	C18	0.03	0.11
	Total Aldehyde	0.04	0.08
	Total Alkane	2.50	0.00
	Total Alcohol	3.13	12.51
	Overall Hydrocarbon	5.68	12.59

*n.d. corresponds to not detected in the extracellular fraction by GC-FID. *E. coli* ZFM8 strain was used for alkane production, and the *E. coli* ZFM9 strain was used for fatty alcohol production.