



OPEN

Improved production of the non-native cofactor F₄₂₀ in *Escherichia coli*

Mihir V. Shah^{1,2,4}, Hadi Nazem-Bokaei^{1,2,4}, James Antony^{1,2,3}, Suk Woo Kang^{2,3}, Colin J. Jackson^{2,3} & Colin Scott^{1,2}✉

The deazaflavin cofactor F₄₂₀ is a low-potential, two-electron redox cofactor produced by some Archaea and Eubacteria that is involved in methanogenesis and methanotrophy, antibiotic biosynthesis, and xenobiotic metabolism. However, it is not produced by bacterial strains commonly used for industrial biocatalysis or recombinant protein production, such as *Escherichia coli*, limiting our ability to exploit it as an enzymatic cofactor and produce it in high yield. Here we have utilized a genome-scale metabolic model of *E. coli* and constraint-based metabolic modelling of cofactor F₄₂₀ biosynthesis to optimize F₄₂₀ production in *E. coli*. This analysis identified phospho-enol pyruvate (PEP) as a limiting precursor for F₄₂₀ biosynthesis, explaining carbon source-dependent differences in productivity. PEP availability was improved by using gluconeogenic carbon sources and overexpression of PEP synthase. By improving PEP availability, we were able to achieve a ~40-fold increase in the space–time yield of F₄₂₀ compared with the widely used recombinant *Mycobacterium smegmatis* expression system. This study establishes *E. coli* as an industrial F₄₂₀-production system and will allow the recombinant in vivo use of F₄₂₀-dependent enzymes for biocatalysis and protein engineering applications.

Cofactor F₄₂₀ is required for methanogenesis in Archaea^{1–5}, anaerobic oxidation of methane by anaerobic methanotrophs^{6–9}, and is involved in secondary metabolism in some Eubacteria^{10,11}. As a deazaflavin, it is structurally similar to flavin, but given its lower redox mid-point potential (–360 mV for F₄₂₀ cf. ~–230 mV for flavins) and obligate 2-electron transfer it functions analogously to NAD/NADP^{12,13}. It has been suggested that cofactor F₄₂₀-dependent enzymes have significant potential as biocatalysts for the reduction of enoates, imines and ketones, and potentially for other unexplored reactions and processes^{14–17}. However, the lack of a cost-effective production system for cofactor F₄₂₀ is a major deterrent in exploring F₄₂₀-dependent reactions for biocatalytic applications. Development of a low-cost production system for F₄₂₀ will be essential for advancing the application of F₄₂₀-dependent enzymes. An ultimate goal of this study is to devise the development of a system for F₄₂₀ production as an end-product and at scale.

For research applications, cofactor F₄₂₀ is produced via fermentation of organisms that naturally produce it, in particular several species of *Mycobacteria*^{15,18}. However, *Mycobacteria* are not well suited for large scale-production of the cofactor because they are not generally recognised as safe (GRAS) organisms, tend to form dense aggregated “clumps”, and are slow growing¹⁹. Two 13-step chemical syntheses of F₄₂₀ isomers have been reported, differing only in the peptide linkage between the two glutamate residues. Both reported extensive use of protecting groups and low overall yield^{20,21}. While improved syntheses of the deazaflavin moiety F_O have been reported (Fig. 1)²², it is unlikely that a full chemical-synthesis route to F₄₂₀ will be economical due to low yield, poor atom economy and the instability of several intermediates. Although some F₄₂₀-dependent enzymes have limited activity with F_O, it is unlikely to be a suitable substitute due to poor kinetics²². More recently, an efficient chemoenzymatic approach to producing the F₄₂₀ analogue F_OP (phosphate group attached to F_O) was reported and F₄₂₀-dependent enzymes had substantially greater activity with F_OP than F_O, albeit still lower than with authentic F₄₂₀²³.

The bacterial F₄₂₀ biosynthesis pathway proposed by Bashiri et al.¹⁸ was further updated in this study. F_O synthase (CofGH), one of the first steps in the F₄₂₀ biosynthesis pathway, was found to be a radical SAM enzyme

¹Biocatalysis and Synthetic Biology Team, CSIRO Land and Water, Black Mountain Science and Innovation Precinct, Canberra, ACT, Australia. ²Synthetic Biology Future Science Platform, Black Mountain Science and Innovation Precinct, Canberra, ACT, Australia. ³Research School of Chemistry, Australian National University, Canberra, ACT, Australia. ⁴These authors contributed equally: Mihir V. Shah and Hadi Nazem-Bokaei. ✉email: colin.scott@csiro.au

Figure 1. F_{420} biosynthesis pathway: F_{420} pathway has two branches, one branch begins with the formation of (A) F_0 from 5-amino-6-(D-ribitylamino) uracil which is the intermediate of the riboflavin pathway, tyrosine and S-adenosyl methionine. The Reaction in this branch is catalysed by F_0 synthase (FbiC/CofGH). Another branch of F_{420} is known to utilize different metabolites; PEP (FbiD/CofC)¹⁸ (B1); 2-PL (FbiD/CofC)²⁶ (B2) and 3-PG (FbiD)²⁶ (B3) producing EPPG, LPPG and 3-GPPG respectively. PEP and 3-PG are intermediates of the glycolytic pathway. In the next step LPPG (C2; FbiA/CofD) and 3-GPPG (C3; FbiA) together with F_0 produce F_{420-0} and 3PG F_{420-0} respectively. In case of EPPG (C1) an intermediate Dehydro- F_{420-0} ¹⁸ is produced which is further converted to F_{420-0} (D). In the final step glutamylation of F_{420} is catalysed by CofE/FbiB (E1 and E2) to produce either F_{420-n} or 3PG- F_{420-n} ; number n depends on the F_{420} producing species. In this figure Cof genes are derived from archaea and Fbi genes are derived from bacteria. 3-PG derived F_{420} was observed recently in *P. rhizoxinica* bacteria²⁶ and not yet been discovered in archaea.

capable of generating two molecules of 5'-deoxyadenosine²⁴. Therefore, F_0 synthase accepts two molecules of S-adenosyl-L-methionine as substrate in addition to tyrosine, producing two molecules of L-methionine, one molecule of ammonia, and two molecules of 5'-deoxyadenosine (Fig. 1A).

Ideally, cofactor F_{420} would be produced via fermentation in a well-characterised microorganism, such as *Escherichia coli*, for which genetic and metabolic engineering tools are well developed and can be easily cultivated at large scale. Recently, we engineered the model laboratory bacterium *E. coli* to produce F_{420} via heterologous expression of biosynthetic genes sourced from *Mycobacterium smegmatis* (FbiD, FbiC, FbiB) and *Methanosarcina mazei* (CofD; equivalent to FbiA)¹⁸ (Fig. 1). The yield of cofactor F_{420} achieved in *E. coli* was 0.38 $\mu\text{mol/g DCW}$ (grams of dry cell weight)¹⁸, which is comparable to yield obtained in wild type *Mycobacterium smegmatis*²⁵. Thermodynamic analysis of cofactor F_{420} biosynthesis revealed that the overall pathway is energetically favourable, with the final steps being effectively irreversible. This suggests that yields attained could be improved upon through metabolic engineering¹⁵.

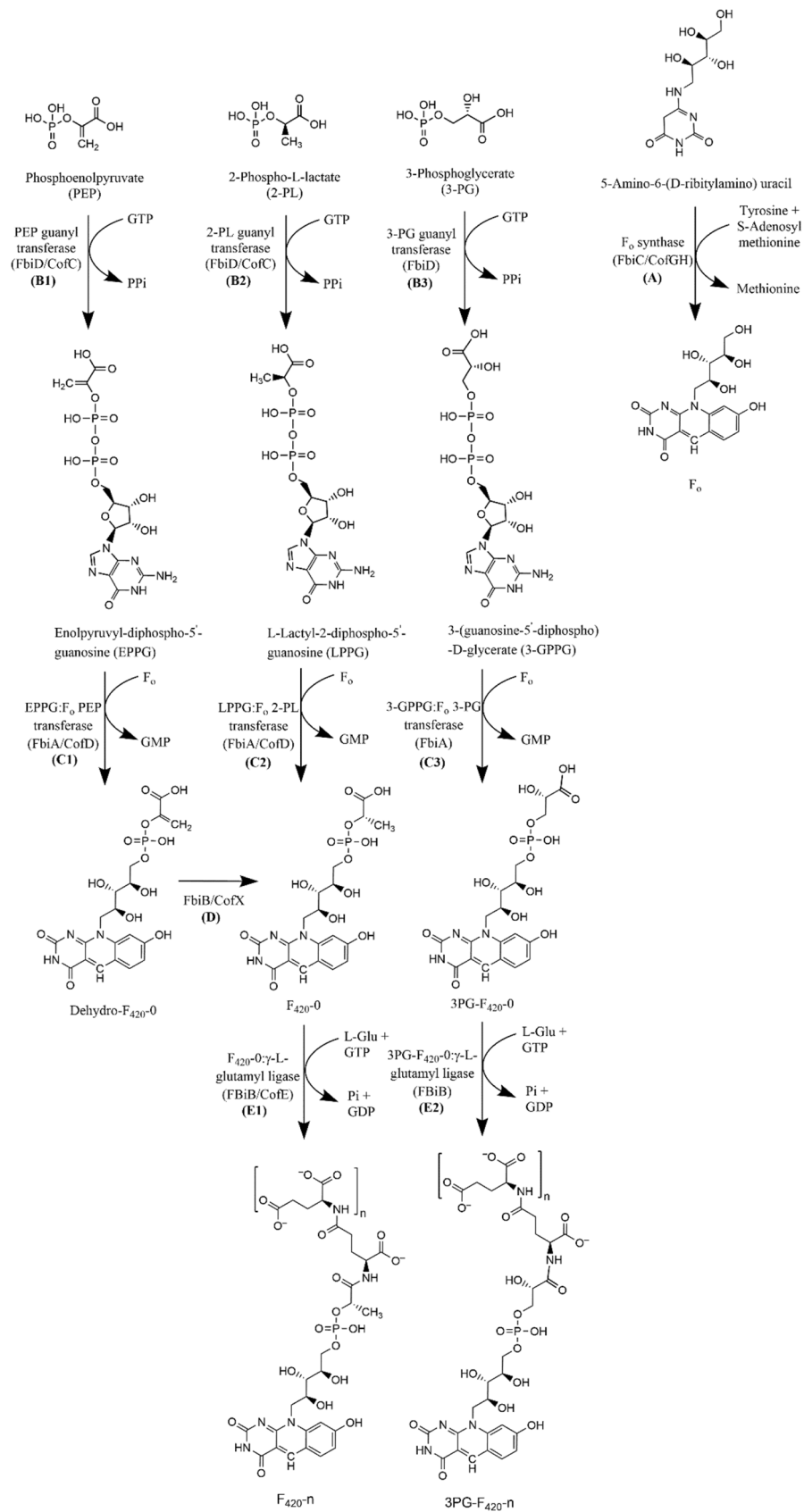
It has recently been shown that FbiD/CofC has species-specific substrate preferences (Fig. 1). FbiD from *Paraburkholderia rhizoxinica* prefers 3-phosphoglycerate (3PG)²⁷ (Fig. 1, Step B3), *M. smegmatis* and *M. mazei* prefer phospho-enol pyruvate (PEP)^{18,28} (Fig. 1, Step B1), and *Methanococcus jannaschii* prefers 2-phospho-L-lactate (2PL)²⁹ (Fig. 1, Step B2). It has been suggested that this step in the biosynthetic pathway may be particularly sensitive to the intramolecular concentration of its substrates, as thermodynamic analysis showed that this step is only just favourable in the forward direction¹⁵. The diversity of substrates used in various organisms may reflect the adaptation of this step to use highly abundant metabolites. Preliminary findings suggest that variation of phospho-carbohydrate moiety has less effect on F_{420} -dependent enzymes than variations of tail length³⁰.

Herein, we report that cofactor F_{420} biosynthesis in *E. coli* is heavily influenced by the carbon source. We used metabolic modelling to understand the underlying causes of carbon-source dependent differences in yield, which identified several potential bottlenecks. As a suitable genome-scale metabolic model was unavailable for production of F_{420} in *E. coli*, nor for any natural F_{420} -producing organisms, we incorporated both the phospho-enol pyruvate-dependent and 3-phosphoglycerate-dependent F_{420} biosynthesis pathways into the iHK1487 genome-scale metabolic model for *E. coli* BL21³¹. The updated model (iEco-F420) was used to identify potential flux bottlenecks and to explore the theoretical limits of F_{420} production in this organism. Although the overall thermodynamics of the pathway are favourable, calculations revealed unfavourable energetics for the reaction catalysed by FbiD/CofC, which converts PEP and guanosine triphosphate into enolpyruvyl-diphospho-5'-guanosine¹⁵. Several strategies were explored to improve availability of PEP for this reaction. Through a combination of metabolic engineering and rational carbon source selection, we were able to improve the yield of cofactor F_{420} from 0.28 to 1.60 $\mu\text{mol/g DCW}$. The highest productivity observed with *E. coli* was a yield of 1.60 $\mu\text{mol/g DCW}$ and culture time of 13 h (equivalent to 123 nmol/h/g DCW); this space-time yield is fourfold higher than is in recombinant *M. smegmatis*, for which the highest published yield of cofactor F_{420} achieved was 3.0 $\mu\text{mol/g DCW}$ with culture time of 96 h (equivalent to 31 nmol/h/gDCW)³².

Results and discussion

The effect of different carbon sources on cofactor F_{420} yield and growth of *E. coli*. To investigate the effects of different carbon sources on the production of F_{420} , we tested acetate, fumarate, glucose, glycerol, pyruvate, and succinate as carbon sources, as these carbon sources enter central metabolism at different points, have varied uptake mechanisms and therefore distinct bioenergetic consequences for the cell³³. Pyruvate and fumarate, (followed closely by acetate and succinate) supported the greatest F_{420} production per gram of dry cell weight (DCW; Fig. 2A, B; Table 1). However, it should be noted that these carbon sources did not support high levels of biomass formation (Fig. 2C). Indeed, the cell density (measured as OD_{600}) varied significantly by carbon source (Fig. 2C). With respect to overall productivity of F_{420} production (expressed as $\mu\text{mol } F_{420}/\text{L/h}$), glycerol was the most productive carbon source (Fig. 2B); F_{420} yield with pyruvate was 0.90 $\mu\text{mol/g DCW}$ which is close to the yield of the cofactor NADPH in *E. coli* of 1.3 $\mu\text{mol/g DCW}$ ³⁴ (Fig. 2A). High F_{420} yield and productivity with pyruvate implicitly indicated the impact of this intracellular metabolite as well as its precursor, PEP, on F_{420} biosynthesis.

Phospho-enol pyruvate (PEP) is a key metabolite for F_{420} biosynthesis. To systematically understand the effect of F_{420} biosynthesis on the distribution of flux through the entire metabolic network of the engineered *E. coli* grown with different carbon sources, we created and utilized the iEco-F420 metabolic model (see Methods) to compare flux profiles. Figure 3 summarises Flux Balance Analysis (FBA) results for two main pathways; glycolysis and the TCA cycle, for in silico growth with glucose, glycerol, and succinate as sole carbon sources, which were selected because of their different F_{420} productivity profiles. FBA predicted assimilation



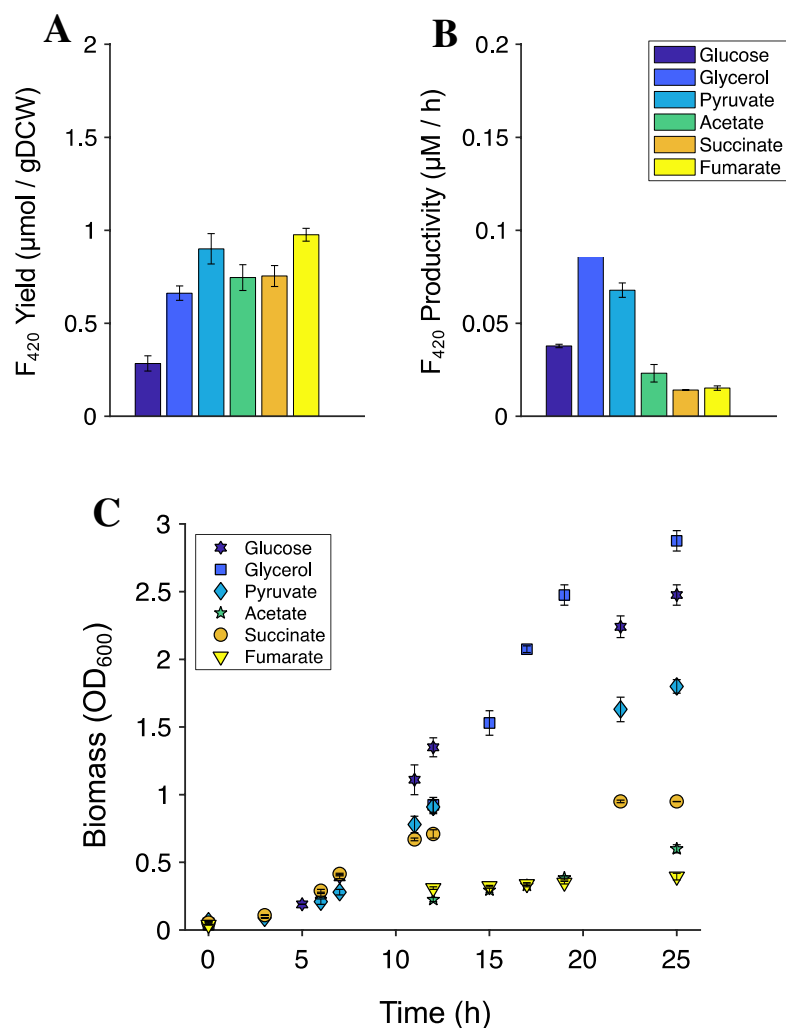


Figure 2. Cofactor F_{420} production and growth of *E. coli* expressing phospho-enol pyruvate-dependent F_{420} biosynthesis pathway using different carbon sources. **(A)** F_{420} yield (in μmol per grams of dry cell weight) at the end of exponential phase of *E. coli* growth. **(B)** Productivity of F_{420} (in μmol per liter per hour). **(C)** Growth of F_{420} -producing *E. coli* (measured as optical density at 600 nm) over time. *E. coli* was cultivated in minimal media at 30 °C in shake flasks with different sole carbon sources. Error bars show standard errors of the mean of at least two replicates.

of 72% of glucose, as the sole carbon source, via the phosphoenolpyruvate (PEP): phosphotransferase system (PTS); all enzymes involved in glycolysis were active. Given the defined criteria, FBA predicted no flux through PEP synthase (PPS) or PEP carboxykinase (PPCK) indicating tight control over the pool of PEP during in silico growth with glucose (Fig. 3). These simulation results suggest a key role for PEP during F_{420} biosynthesis.

The metabolic model indicated that growth on succinate results in activation of the gluconeogenesis pathway and PPCK. With glycerol as the sole carbon source the upper glycolytic pathway was turned off (Fig. 3B), resulting in up to 27% higher overall ATP generation. On the other hand, fumarate was predominantly metabolised through aspartase since the glyoxylate shunt was highly active when glycerol was the carbon source, which leads to a reduction in total flux through TCA cycle. These modelling results explain the higher growth (Fig. 2C) and higher capacity for F_{420} production when engineered *E. coli* cells expressing the F_{420} biosynthetic pathway are grown with glycerol compared with glucose or succinate. Interestingly, with succinate as the carbon source, the iEco-F420 model predicted that pyruvate was produced mainly through malate dehydrogenase (decarboxylating) (Fig. 3A), leaving the PEP pool more accessible for incorporation into F_{420} production, consistent with the experimental yields. These results are consistent with the empirical growth experiments and also indicate a key role for PEP in controlling flux through the F_{420} biosynthesis pathway.

The iEco-F420 model contains 35 reactions consuming PEP: 19 are PEP-dependent phosphotransferases, 10 reactions participate in central carbon metabolism, two occur in cell envelope metabolism, two in tyrosine metabolism, and one in F_{420} biosynthesis (Supplementary File 1; Table S1). In an effort to increase the PEP pool, we used the model to test whether any of these competing reactions were dispensable in silico. However, single

Plasmid used	Carbon source	F ₄₂₀ yield (μmol/g DCW)	F ₄₂₀ concentration (μmol/L)	F ₄₂₀ productivity (μM/h)	Comments
pF420	Glucose	0.28 ± 0.01	0.39 ± 0.06	0.037 ± 0.001	PEP-dependent F ₄₂₀ biosynthesis pathway
pF420	Glycerol	0.66 ± 0.04	1.07 ± 0.06	0.096 ± 0.004	
pF420	Pyruvate	0.90 ± 0.08	0.91 ± 0.08	0.067 ± 0.004	
pF420	Acetate	0.74 ± 0.07	0.25 ± 0.04	0.023 ± 0.004	
pF420	Succinate	0.75 ± 0.06	0.40 ± 0.03	0.014 ± 0.000	
pF420	Fumarate	0.97 ± 0.03	0.22 ± 0.01	0.015 ± 0.001	
pF420 + PPS-pRSF	Glucose	0.26 ± 0.00	0.49 ± 0.00	0.050 ± 0.004	PEP-dependent F ₄₂₀ biosynthesis pathway along with over-expressed PPS (uninduced)
pF420 + PPS-pRSF	Glycerol	0.53 ± 0.10	0.76 ± 0.15	0.078 ± 0.001	
pF420 + PPS-pRSF	Pyruvate	1.61 ± 0.11	2.33 ± 0.16	0.175 ± 0.023	
pF420 + PPS-pRSF	Acetate	0.73 ± 0.01		0.021 ± 0.000	
pF420 + PPS-pRSF	Succinate	0.30 ± 0.02	0.19 ± 0.01	0.019 ± 0.002	
pF420 + PPS-pRSF	Fumarate	0.33 ± 0.00	0.08 ± 0.00	0.007 ± 0.000	PEP-dependent F ₄₂₀ biosynthesis pathway along with over-expressed PPS (induced)
pF420 + PPS-pRSF	Glucose	0.54 ± 0.03	0.76 ± 0.05	0.076 ± 0.001	
pF420 + PPS-pRSF	Glycerol	0.80 ± 0.03	0.94 ± 0.03	0.086 ± 0.001	
pF420 + PPS-pRSF	Pyruvate	0.97 ± 0.09	1.03 ± 0.10	0.070 ± 0.002	
pF420 + PPS-pRSF	Acetate	0.81 ± 0.06	0.43 ± 0.03	0.024 ± 0.002	
pF420 + PPS-pRSF	Succinate	0.32 ± 0.00	0.17 ± 0.00	0.016 ± 0.004	
pF420 + PPS-pRSF	Fumarate	0.27 ± 0.00	0.07 ± 0.00	0.005 ± 0.000	PEP-dependent F ₄₂₀ biosynthesis pathway along with over-expressed PPCK (uninduced)
pF420 + PPCK-pRSF	Glucose	0.19 ± 0.01	0.40 ± 0.01	0.049 ± 0.003	
pF420 + PPCK-pRSF	Glycerol	0.42 ± 0.04	1.09 ± 0.10	0.098 ± 0.039	
pF420 + PPCK-pRSF	Pyruvate	0.87 ± 0.04	1.56 ± 0.07	0.094 ± 0.008	
pF420 + PPCK-pRSF	Acetate	0.77 ± 0.05	1.47 ± 0.10	0.047 ± 0.005	
pF420 + PPCK-pRSF	Succinate	0.25 ± 0.02	0.27 ± 0.02	0.005 ± 0.000	
pF420 + PPCK-pRSF	Fumarate	0.22 ± 0.01	0.04 ± 0.00	0.002 ± 0.000	PEP-dependent F ₄₂₀ biosynthesis pathway along with over-expressed PPCK (induced)
pF420 + PPCK-pRSF	Glucose	0.27 ± 0.01	0.52 ± 0.02	0.035 ± 0.005	
pF420 + PPCK-pRSF	Glycerol	0.36 ± 0.06	0.70 ± 0.11	0.047 ± 0.005	
pF420 + PPCK-pRSF	Pyruvate	0.87 ± 0.07	1.19 ± 0.10	0.067 ± 0.002	
pF420 + PPCK-pRSF	Acetate	0.66 ± 0.04	0.95 ± 0.05	0.029 ± 0.005	
pF420 + PPCK-pRSF	Succinate	0.27 ± 0.04	0.27 ± 0.04	0.003 ± 0.000	
pF420 + PPCK-pRSF	Fumarate	0.25 ± 0.03	0.04 ± 0.01	0.002 ± 0.000	3PG-dependent F ₄₂₀ biosynthesis pathway
pF420-3PG	Glucose	0.21 ± 0.01	0.53 ± 0.00	0.034 ± 0.003	
pF420-3PG	Glycerol	0.38 ± 0.01	0.82 ± 0.02	0.048 ± 0.002	
pF420-3PG	Pyruvate	0.54 ± 0.01	0.88 ± 0.01	0.059 ± 0.000	
pF420-3PG	Acetate	0.32 ± 0.03	0.53 ± 0.02	0.014 ± 0.000	

Table 1. Summary of the results obtained for engineered *E. coli* producing PEP-dependent or 3PG-dependent F₄₂₀ using different carbon sources with and without overexpression of either PPS or PPCK.

gene deletion in silico predictions suggested that removing the reactions involved in cell envelope and tyrosine metabolism would result in cell death.

We next performed flux variability analysis (FVA) for all reactions in the metabolic network, including the PEP-consuming reactions (Supplementary File 1; Tables S2–S13) to specifically explore flux variations in PEP-consuming/producing reactions as a result of maximization of flux through biosynthesis of F₄₂₀. PEP hydratase (enolase) was chosen to interpret flux variations with respect to PEP availability for cellular growth versus F₄₂₀ production. Figure 4 shows the flux profile of PEP hydratase using all six carbon sources. When glucose is the sole carbon source, PEP must be produced through glycolysis to meet cellular objective (i.e., maximizing growth). At maximum biomass (where the blue and red lines showing minimum and maximum fluxes meet in Fig. 4), PEP hydratase flux is positive, meaning that 2-phospho glycerate (2-PG) is fully metabolized to PEP. One engineering objective for increasing the heterologous production of F₄₂₀ requires more carbon to be diverted into the target product rather than biomass, up to the point where the growth of the host is so negatively affected that it becomes uneconomical. When biomass yield drops to 80% of its maximum, for example, the minimum and maximum fluxes through PEP hydratase are still both positive, meaning that essential cellular processes take priority. As a result, 2-PG needs to be metabolized to provide stoichiometric requirements of PEP. However, at 50% of maximum biomass yield, the minimum flux (Fig. 4) through PEP hydratase becomes negative, meaning that the system is more relaxed to divert a portion of PEP for other processes including F₄₂₀ production.

Unlike the flux predictions for PEP hydratase using glucose, PEP is significantly more available for processes other than cellular growth when the carbon source is succinate, fumarate or pyruvate, even at maximum biomass yields (Fig. 4). This is consistent with reports that glucose uptake in *E. coli* occurs primarily via the PTS, consuming up to 50% of the available PEP in cell^{35,36}, thereby reducing its availability for F₄₂₀ biosynthesis. Gluconeogenic

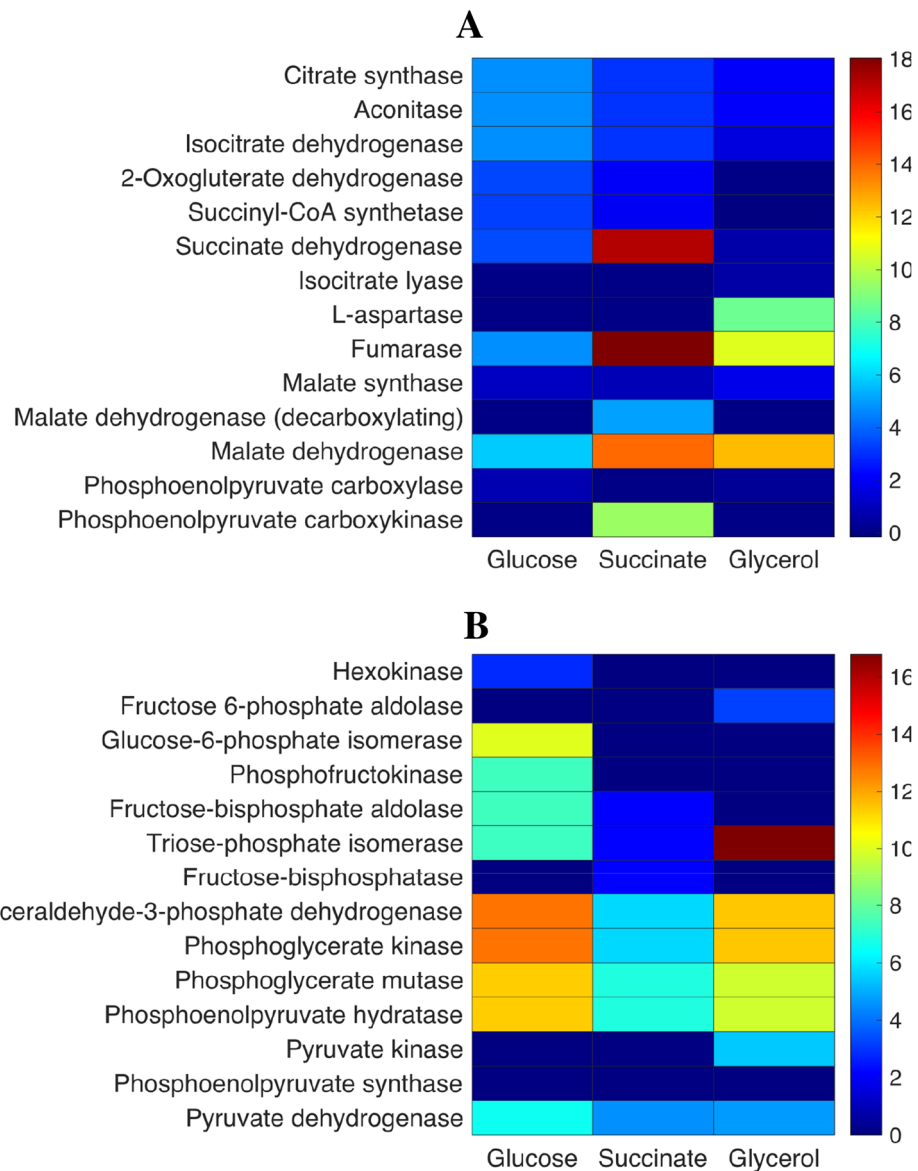


Figure 3. Flux balance analysis of TCA cycle and anaplerotic reactions of the TCA cycle (A) along with glycolysis/gluconeogenesis (B) pathways predicted by iEco-F420 metabolic model of *E. coli* for independent simulations using glucose, glycerol, or succinate as sole carbon sources (60 C-mol of carbon source). Objective is maximizing F_{420} production while maintaining growth at 30% of its max. Maintenance ATP requirements is fixed at 5.17 mmol/g DCW. Colormap shows absolute flux values in mmol/gDCW/hr. Fructose-bisphosphate aldolase, Triose-phosphate isomerase, Glyceraldehyde-3-phosphate dehydrogenase, and phosphoenolpyruvate hydratase are active in favor of gluconeogenesis pathway with succinate as the carbon source.

carbon sources such as pyruvate, succinate, and fumarate increase intracellular PEP levels compared to glucose³³ as their uptake is PEP-independent³⁶. PEP hydratase flux variation with glycerol is the highest among other carbon sources, meaning that glycerol assimilation could potentially lead to greater flexibility in utilising PEP for F_{420} biosynthesis. However, glycerol uptake occurs through the glycolysis pathway and although its uptake requires half the energy (in form of ATP) of glucose, most of the PEP is still required for cellular activities rather than biosynthesis of F_{420} . Nonetheless, glycerol remains a candidate carbon source for large-scale F_{420} production compared with glucose when maintaining high cell masses is essential because it allows for higher cellular mass yields while bypassing PTS-dependent PEP depletion. In the case of acetate, ATP-dependent acetate assimilation is the only route for producing acetyl-CoA, which is an essential precursor for the biosynthesis of most amino acids and fatty acids and therefore biomass yield drops significantly (Fig. 4). However, FVA for PEP hydratase indicates the feasibility of utilising PEP for non-cellular activities.

We measured intracellular PEP for engineered *E. coli* grown with glucose and glycerol to validate the model predictions. When glycerol was used as the sole carbon source, PEP and F_{420} levels were 1.43 and 1.82-fold higher, respectively, compared with when glucose was used as the carbon source. This difference was also borne out in

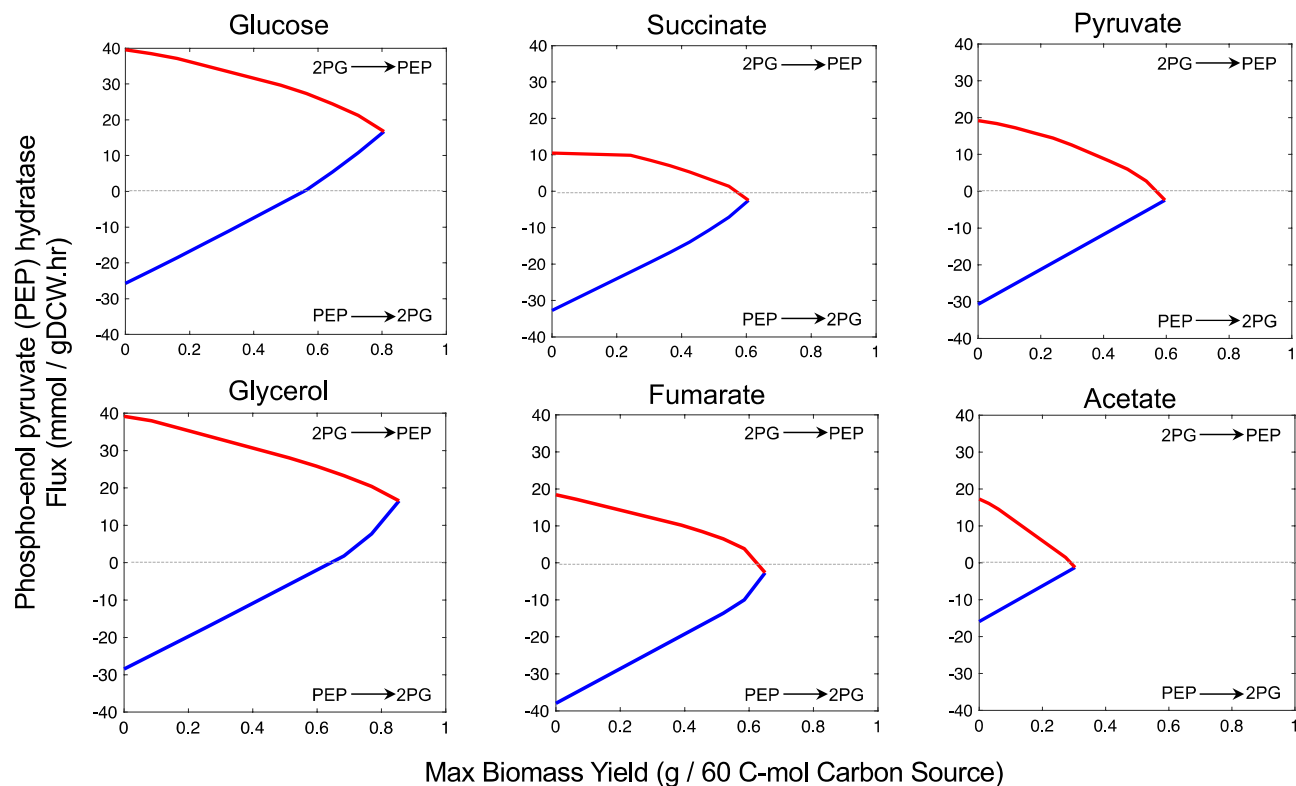


Figure 4. Flux variability analysis showing PEP utilization capacity represented by the flux through PEP hydratase as a function of cellular biomass yield for different carbon sources predicted by the iEco-F420 metabolic model of *E. coli*. Minimum and maximum fluxes are shown as blue and red, respectively. PEP: phospho-enol pyruvate; 2PG: 2-phospho glycerate. Horizontal dashed lines do not correspond to any actual flux values; instead, they serve as hypothetical lines above which glycolysis drives carbon flow (2PG is metabolised to PEP).

the simulation data (Supplementary Table S1). These results, collectively, demonstrate that the choice of carbon source directly affects intracellular availability of PEP, which, in turn, influences F_{420} levels.

Using 3PG as an alternative to PEP. As PEP is likely to be a flux-limiting metabolite, we explored the possibility of using an alternative metabolite in its place. Three different metabolites have been proposed to be incorporated in the sidechain of F_{420} : PEP, 2-phospho-L-lactate and 3-phospho-D-glycerate^{18,27,37}. While 2-phospho-L-lactate has not been observed in *E. coli*¹⁸, 3-phosphoglycerate (3PG) is a glycolytic pathway intermediate present in *E. coli* at 10 times the concentration of PEP³⁸. Moreover, in the context of the iEco-F420 model, PEP-dependent F_{420} biosynthesis requires an additional FMN-dependent reduction step (the FbiB-dependent conversion of dehydro- F_{420} -0 into F_{420} -0) indicating that additional carbon would need to be diverted into FMN biosynthesis^{28,39}. Preliminary evidence suggests that 3PG- F_{420} , unlike F_{420} and F_{420} P, is accepted as a cofactor by F_{420} -dependent enzymes with similar kinetics to standard F_{420} ³⁰. Given the relative abundance of 3PG, we investigated it as an alternative to PEP by substitution of *M. smegmatis* FbiD with that of *P. rhizoxinica*.

Although 3PG is present at a higher intracellular concentration than PEP (1.5 mM *cf.* 0.18 mM)³⁸ and is predicted to provide relatively similar maximum theoretical F_{420} yields (Supplementary Fig. S1), the experimentally determined yield of F_{420} -3PG was found to be lower than for F_{420} -PEP (Fig. 5). Moreover, no F_{420} -3PG formation was observed with either succinate or fumarate as carbon source. This contrasts with the model predictions of feasible theoretical yields for F_{420} -3PG with all carbon sources tested (Supplementary Fig. S1). It is possible that the *P. rhizoxinica* FbiD product, glyceryl-2-diphospho-5'-guanosine, 3PG- F_{420} -0 and/or its polyglutamated derivatives are poor substrates for the enzymes catalysing subsequent steps in F_{420} biosynthesis, which had been sourced from *Mycobacteria* and may have low specificity for 3PG containing F_{420} metabolites (Fig. 5).

Over-expression of PEP synthase increases the yield of F_{420} . Given that PEP is a limiting metabolite in F_{420} biosynthesis, we investigated whether production of PEP could be increased. Growth on fumarate and succinate is known to increase the expression of PEP-producing enzymes PPS and PPCK⁴⁰ (Fig. 3A). Indeed, overexpression of PPS has been used to increase PEP concentrations *in vivo*^{41,42} to improve the yield of shikimic acid⁴³, aromatic amino acids^{42,44} and lycopene⁴⁵ biosynthesis. However, overexpressing PPS has been reported to negatively affect cell growth due to the excretion of pyruvate and acetate⁴².

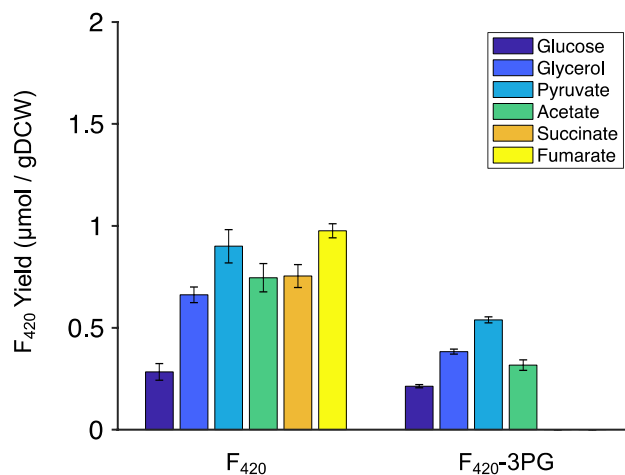


Figure 5. F_{420} yield (in μmol per grams of dry cell weight) at the end of exponential phase of *E. coli* growth with different carbon sources. 3PG- F_{420} and F_{420} refer to the biosynthesis of the cofactor F_{420} derived from 3-phospho glycerate and PEP, respectively, as the substrates of FbiD. Error bars show standard errors of the mean of at least two replicates.

In an attempt to increase intracellular PEP concentrations, we overexpressed PPS and PPCK from an IPTG-inducible expression plasmid and studied the effect on F_{420} yield. Consistent with previous reports⁴¹, overexpression of PPS resulted in growth inhibition. Therefore, to improve final biomass concentration, PPS was only induced once cell density (OD_{600}) was greater than 1.0, which resulted in significant improvement in F_{420} yield. We tested overexpression of PPS when grown on different carbon sources, as shown in Fig. 6. Overexpression of PPS improved the yield of F_{420} from 0.27 to 0.54 $\mu\text{mol/g}$ DCW using glucose and from 0.53 to 0.80 $\mu\text{mol/g}$ DCW using glycerol. When grown on pyruvate, an F_{420} yield of 1.60 $\mu\text{mol/g}$ DCW was observed without the addition of IPTG. With the addition of IPTG, the yield of F_{420} yield decreased to 0.90 $\mu\text{mol/g}$ DCW. The yield of F_{420} also decreased after PPS induction when grown on succinate or fumarate (Fig. 6A). The pyruvate:PEP node of *E. coli* metabolism is highly regulated at both the transcriptional and metabolic levels⁴⁶, it is possible that PPS is metabolically regulated during gluconeogenesis or that the reversible flux through PPS is being driven thermodynamically towards pyruvate formation when grown on gluconeogenic carbon sources. With glucose and glycerol, induction of PPS with IPTG resulted in significant improvement in the yield of F_{420} as compared to non-induced PPS. On the contrary, with pyruvate, non-induced PPS resulted in significantly higher yield and productivity of F_{420} compared to IPTG induced PPS. It may be that optimal PPS expression levels differ with different carbon sources. The highest yield of F_{420} obtained was 1.60 $\mu\text{mol/g}$ DCW, with a productivity of 0.17 $\mu\text{mol/h}$, using pyruvate as carbon source with leaky expression of PPS.

The impact of PPCK overexpression on F_{420} yield was also studied (Fig. 7). Unlike the expression of PPS, no improvement in F_{420} production was observed during PPCK overexpression. We confirmed the protein was expressed in soluble form (Supplementary Fig. S6). It is quite likely that we saw no difference in F_{420} concentration when PPCK was over-expressed because *E. coli* PPCK activity is metabolite controlled, either by the cellular PEP concentration or PEP:pyruvate ratio⁴⁶. We therefore investigated the potential of uncontrolled PPCK overexpression using the iEco- F_{420} model. We explored the overall capability of the metabolic network to improve flux through FbiB (i.e., production of mature F_{420}) by simulating over-expression of PPS or PPCK. The results, shown in Supplementary Fig. S2, indicate that by forcing a higher flux through PPS or PPCK, the maximum FbiB flux (shown by black arrows) drops unless it occurs at a non-zero flux through PPS or PPCK. These results indicate the maximum stoichiometric capacity for F_{420} biosynthesis as a result of over-expressing PPS or PPCK; however, the overall kinetics of the system and regulatory mechanisms for growth with different carbon sources would significantly influence F_{420} yields, in vivo. The experimental results (Fig. 6) confirmed improved F_{420} biosynthesis when using glucose and pyruvate as a result of PPS overexpression, in agreement with the simulation results shown in Supplementary Fig. S2 for these carbon sources. It should be noted that the simulation results of Supplementary Fig. S2 also demonstrate the potential impact of the type of transporter on the flux through CofE when over-expressing PPS or PPCK.

The effect of time and carbon source on polyglutamate chain length. The final step in F_{420} biosynthesis is the addition of between one and nine glutamate residues to the F_{420} -0 intermediate to yield F_{420} -n (n: number of glutamate residues)^{47,48}. What influences the tail length of F_{420} is still not clear, although in vitro analysis of F_{420} -0:g-glutamate ligases from different organisms has revealed that they typically produce F_{420} species with polyglutamate chain lengths consistent with F_{420} obtained from the native organisms^{48,49}. The number of glutamate residues influences the cofactor affinity of some F_{420} -dependent enzymes; for example, the F_{420} -dependent oxidoreductases MSMEG_2027, MSMEG_0777 and MSMEG_3380 from *M. smegmatis* reportedly having a high affinity for long chain F_{420} rather than shorter-chain F_{420} ⁵⁰. Similar effects are seen with polyglutamylated folates and folate mimics^{51–53}. Interestingly, F_{420} -n composition changes with different growth

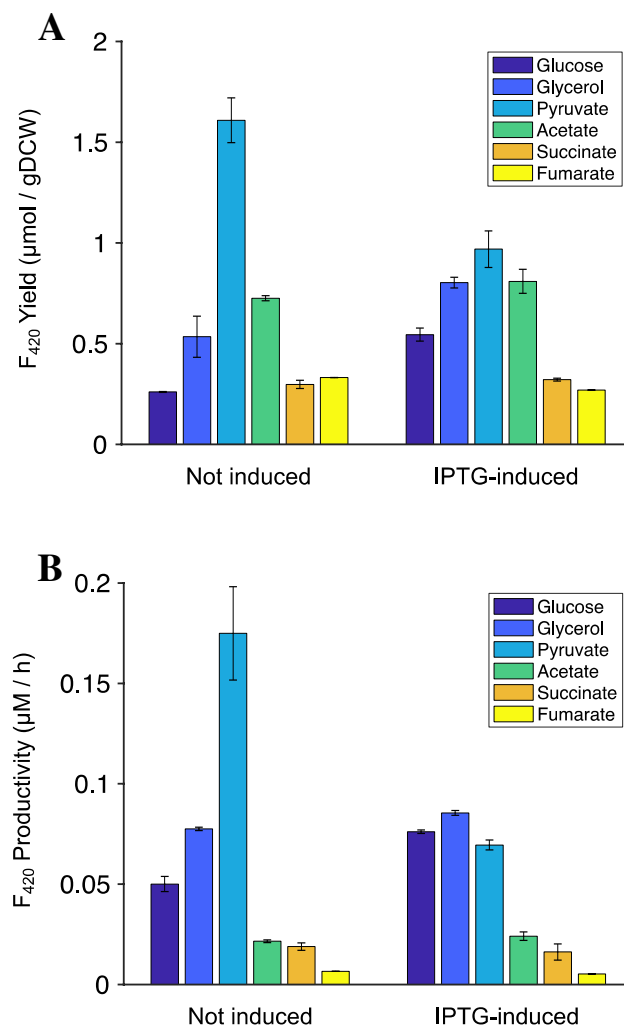


Figure 6. Effect of PPS over-expression on (A) F₄₂₀ yield with and without IPTG (0.1 mM) induction; (B) F₄₂₀ productivity (µM/h) with and without IPTG (0.1 mM). Leaky expression of PPS was observed without the addition of IPTG (Supplementary Fig. S3). *E. coli* was cultivated in a minimal media with different sole carbon sources as shown. Error bars show standard errors of the mean of at least two replicates.

phases of *Methanosarcina barkeri* and *M. mazei*⁵⁴. We therefore investigated the composition of F₄₂₀ during different growth phases of *E. coli*. The composition of F_{420-n} at various time points is shown in Fig. 8. When grown with glucose or glycerol as the carbon source, *E. coli* initially produced short chain F₄₂₀-(1–4) in higher proportions, which shifted over time to predominantly longer chain F₄₂₀-(5–8) (Fig. 8A, B). CofE from *M. smegmatis* (the enzyme used in this system) has been shown to produce predominantly longer F₄₂₀ species (5–8) in stationary phase⁵⁰.

Interestingly, we found that the tail length distribution at the end of the exponential phase was influenced by the carbon source used (Fig. 8C, D). Growth on succinate yielded the highest proportion of long chain F₄₂₀, with F₄₂₀-(5–8) comprising >90%. Glycerol has the next highest proportion of F₄₂₀-(5–8) at >80%, with glucose and acetate with the lowest levels of F₄₂₀-(5–8) products (<30% and <25%, respectively) (Fig. 8C). The iEco-F420 model was used to guide interpretation of carbon source-dependent tail length distribution. According to the cofactor biosynthesis pathway shown in Fig. 1, two molecules of GTP per molecule of glutamate are required to metabolise an F₄₂₀ molecule with only one glutamate residue. Likewise, in an ideal case where all incoming carbon to the *E. coli* has to end up in F₄₂₀ with only one glutamate residue, the iEco-F420 model predicted that the ratio of sum of fluxes through all glutamate-producing reactions (v_{glu}^t) to sum of fluxes through all GTP-producing reactions (v_{gtp}^t) has to be equal to two regardless of the type of carbon source. However, when the model was used to simulate F₄₂₀ biosynthesis with chain length compositions observed experimentally, flux predictions suggested deviations in the ratio of v_{glu}^t to v_{gtp}^t , which depends on the type of carbon source. Interestingly, the ratio of v_{glu}^t to v_{gtp}^t was predicted to be 1.731 and 1.772 using succinate and glycerol, respectively, showing the largest deviation for a ratio of two. On the other hand, the ratio of v_{glu}^t to v_{gtp}^t was predicted by the model to be 1.994, 1.960, and 1.873 using glucose, acetate, and pyruvate, respectively, explaining why the lowest proportion of long chain F₄₂₀ was observed with these carbon sources.

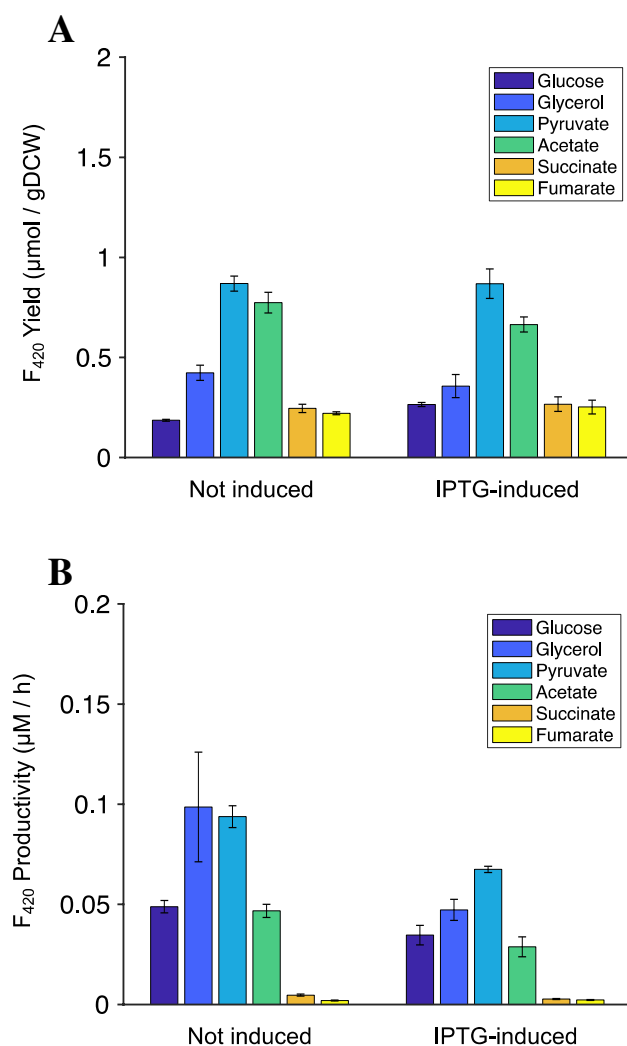


Figure 7. Effect of PPCK over-expression on (A) F₄₂₀ yield with and without IPTG (0.1 mM) induction; (B) F₄₂₀ productivity (µM/h) with and without IPTG (0.1 mM). Leaky expression of PPS was observed without the addition of IPTG (Supplementary Fig. S3). *E. coli* was cultivated in a minimal media with different sole carbon sources as shown. Error bars show standard errors of the mean of at least two replicates.

3PG-F₄₂₀ yielded significantly higher fraction of short chain F₄₂₀-(1–4) > 70% (Fig. 8D) compared to PEP derived F₄₂₀ irrespective of the carbon source used. This could be due to the difference in the kinetics of the enzymes for 3PG-F₄₂₀ and PEP-F₄₂₀.

The iEco-F420 metabolic model additionally provided some insights into the energetic differences in F₄₂₀ biosynthesis with only one glutamate tail as well as with varying number of glutamate tails. For all carbon sources examined, yields were higher for F₄₂₀ with only one glutamate than those for a mixture of F₄₂₀ molecules with different chain-lengths. This is because at a fixed growth rate (i.e., constant cell mass yield), total energy production (in the form of ATP) is higher for biosynthesis of F₄₂₀ with one glutamate than that for biosynthesis of a mixture of F₄₂₀ molecules (Fig. 9). Based on the results illustrated in Fig. 9, glucose maintains the highest cell mass per mole of ATP produced by cells, which explains the low F₄₂₀ yield from glucose compared to other carbon sources as shown in Fig. 2A. Assimilation of acetate as the sole carbon source requires the activation of ATP-dependent acetate kinase. Therefore, cells have to produce ATP in order to uptake carbon source for survival, which results in low growth rates (Fig. 2C) and maintaining the lowest cell mass yield per mole ATP produced among other carbon sources (Fig. 9) but, relatively high F₄₂₀ yields (Fig. 2A; Supplementary Fig. S1). According to the modeling predictions, acetate might provide benefits from industrial perspective because, cells would be forced to produce ATP for fueling F₄₂₀ production rather than for their growth.

Conclusion

This work establishes that intracellular PEP concentration is the key limiting metabolic bottleneck for heterologous F₄₂₀ biosynthesis in *E. coli*, at least when using biosynthetic enzymes, such as those from *M. smegmatis*, that natively use PEP. An updated metabolic model of *E. coli* incorporating the recombinant F₄₂₀ biosynthetic pathway was developed and used to identify differences in metabolic flux distribution through the entire metabolic

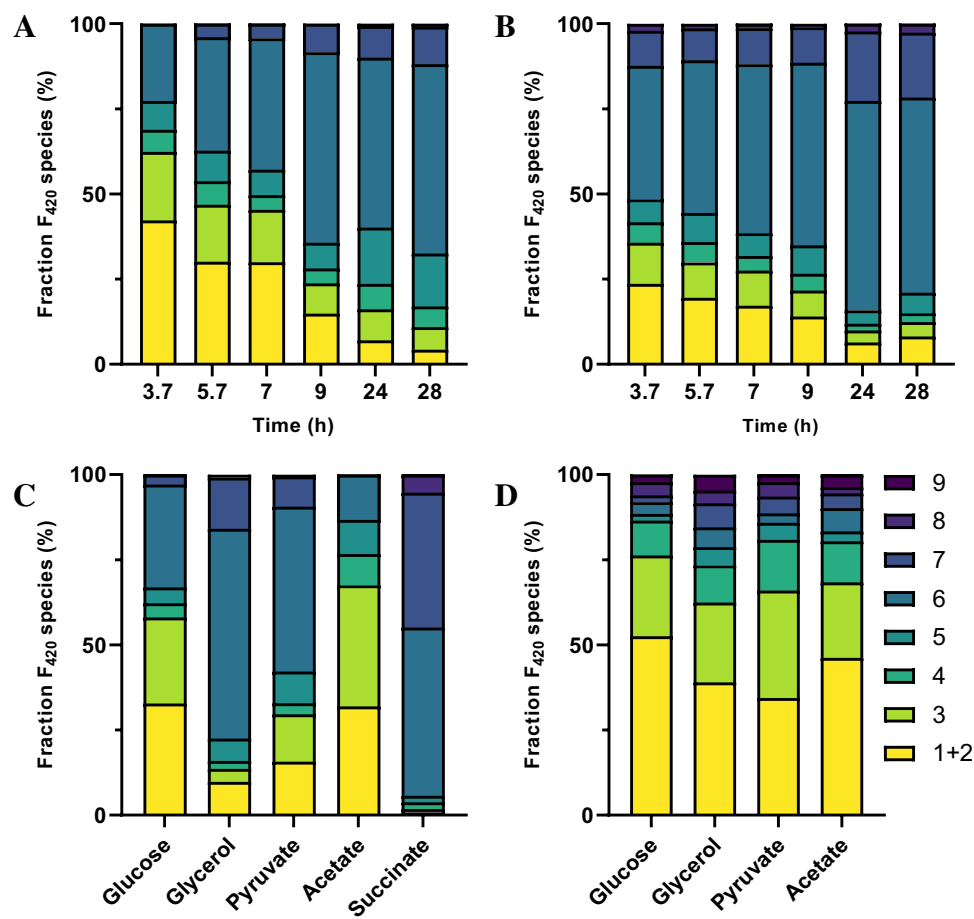


Figure 8. HPLC-FLD data for different glutamate residues attached to F_{420} at different time points after induction of F_{420} pathway using Glucose (A) and Glycerol (B). HPLC-FLD data for different glutamate residues attached to F_{420} (C) and $3PG-F_{420}$ (D) using different carbon sources at the end of exponential phase.

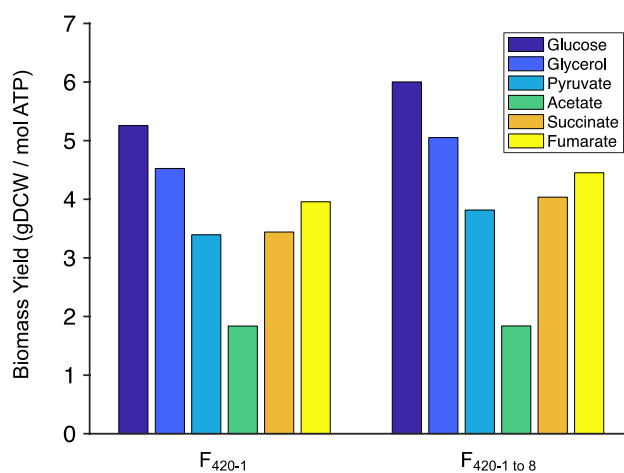


Figure 9. Theoretical biomass yields (at 30% maximum growth) with respect to total energy produced in the form of ATP predicted by the iEco-F420 metabolic model of *E. coli* BL21 simulated with different carbon sources. F_{420-1} indicates the yields for cells synthesizing cofactor F_{420} with only one glutamate tail, whereas $F_{420-1 \text{ to } 8}$ indicate those for cells synthesizing a mixture of F_{420} molecules with varying number of glutamate tails. The uptake of C-source was fixed to 60 C-mol in all simulations to account for differences in number of carbon atoms in C-sources.

network including the central carbon metabolism using various carbon sources. This allowed us to identify, test and rationalize a number of approaches to improve F_{420} yield. Table 1 summarises the F_{420} yield, concentration, and productivity obtained using different strains and conditions in this study. In terms of productivity, glycerol was found to be the best carbon source for F_{420} production amongst those tested as it allowed the best balance between optimizing PEP concentration (compared to glucose) vs. slower growth (compared with acetate, succinate, etc.). We also examined whether alternative substrates for FbiD could be used to remove the reliance on PEP, showing that although 3PG is a viable substrate and is more abundant in *E. coli*, the use of the *P. rhizoxinica* FbiD did not result in higher 3PG- F_{420} yields (compared with F_{420}), presumably because of the downstream enzymes, which are sourced from *M. smegmatis*, have a preference for F_{420} -metabolites over 3PG- F_{420} metabolites. Replacing other enzymes in the pathway with those from *P. rhizoxinica* may be a worthwhile strategy to improve 3PG- F_{420} yield, as the theoretical yield of 3PG- F_{420} is similar to that of PEP-derived F_{420} and strategies have already been developed to improve 3PG availability by increasing glycolytic flux by knocking out the *zwf* gene involved in first step of pentose phosphate pathway⁵⁵. Other strategies that we tested to increase F_{420} yield included increasing PEP production through over expression of PPS; this produced the highest productivity with pyruvate. Our results also indicate that F_{420} composition and concentration in *E. coli* is comparable to some of the best natural producers such as *M. smegmatis*. The F_{420} yield obtained with *M. smegmatis* is approximately 0.30 $\mu\text{mol/g}$ DCW in wild type²⁵ and 3.0 $\mu\text{mol/g}$ DCW in engineered *M. smegmatis*³². However, cultivation of *M. smegmatis* in shake flasks takes up to 96 hours³², while engineered *E. coli* cultivation takes 16 h (this study). This results in space-time yield of 114 nmol/h/gDCW for the recombinant *E. coli* system, vs. 31 nmol/h/gDCW for the recombinant *M. smegmatis* system³². This increased yield is in addition to the many other advantages of using *E. coli*, including less expensive antibiotics (ampicillin vs. hydromycin), reduced safety risks and thus greater accessibility to the technology, and the fact that *E. coli* is a much more widely used strain for protein engineering and DNA modification. By systematically optimizing growth and production conditions in *E. coli*, we have created a system that should make production of F_{420} more economical at an industrial scale and the study of F_{420} -dependent enzymes more accessible, improving on the previous attempts for F_{420} production through the use of more exotic and challenging bacterial species, such as *M. smegmatis*.

Materials and methods

Bacterial strain, vector and media composition. *E. coli* BL21 DE3 (New England Biolabs) was used for protein expression and DH5 α (New England Biolabs) strain was used for plasmid propagation. LB media consisting of 10 g/L tryptone, 5 g/L yeast extract and 10 g/L of NaCl was used for the plasmid propagation and cloning, 15 g/L of agar was added to prepare LB agar plates. *E. coli* protein expression and F_{420} production studies were done in M9 minimal media consisting of 6.78 g/L Na_2HPO_4 , 3 g/L KH_2PO_4 , 1 g/L NH_4Cl , 2.5 g/L NaCl, 100 \times trace elements solution (1.667 g/L $\text{FeCl}_3\cdot 6\text{H}_2\text{O}$, 0.018 g/L $\text{ZnSO}_4\cdot 7\text{H}_2\text{O}$, 0.012 g/L $\text{CuCl}_2\cdot 2\text{H}_2\text{O}$, 0.012 g/L $\text{MnSO}_4\cdot \text{H}_2\text{O}$, 0.018 g/L $\text{CoCl}_2\cdot 6\text{H}_2\text{O}$, and 2.225 g/L $\text{Na}_2\text{EDTA}\cdot 2\text{H}_2\text{O}$), 0.246 g/L $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$, 0.011 g/L CaCl_2 , and 5 g/L of glucose or 0.167 C-mol of different carbon sources with the final pH of 7.3. Chloramphenicol (25 $\mu\text{g/ml}$), kanamycin (50 $\mu\text{g/ml}$) and ampicillin (100 $\mu\text{g/ml}$) were added where appropriate.

E. coli was cultivated in shake flasks at 30 $^\circ\text{C}$ with shaking at 200 rpm. The F_{420} biosynthesis genes in pSB1C3 plasmid were induced with 200 ng/ml of tetracycline. *E. coli* with pRSF duet plasmid was induced with IPTG (0.1 mM) for the expression of PPS and PEP carboxykinase (PPCK) enzymes. The biomass was measured using optical density at 600 nm.

Plasmid construction. The synthesis of the plasmid expressing the F_{420} biosynthesis pathway operon was described in Bashiri et al.¹⁸. The pathway consists of CofD (Accession number: Q8PVT6), CofE (A0QTG1), CofC (A0QUZ4) and CofGH (NC_008596) genes under the control of the tetracycline-inducible promoter BBa_R0040⁵⁶ and the artificial terminator BBa_B1006⁵⁷. The F_{420} biosynthesis operon had been previously synthesized by GenScript and cloned in to pSB1C3 containing the constitutive tetracycline repressor cassette BBa_K145201 with *EcoRI/XbaI* and *PstI/SpeI* restriction enzymes, plasmid construction is explained in Bashiri et al.¹⁸. This construct, hereafter referred to as pF420, enables production of F_{420} to be induced by the addition of tetracycline. All the gene sequences except CofD were obtained from *Mycobacterium smegmatis*, CofD gene sequence was obtained from *Methanosarcina mazei*. All the genes were codon optimized for expression in *E. coli* BL21 DE3 strain. All F_{420} proteins were FLAG tagged and soluble expression of F_{420} pathway proteins was confirmed using western blot¹⁸. In order to produce 3PG- F_{420} , the *M. smegmatis* CofC homologue was replaced by the homologue from *Paraburkholderia rhizoxinica* (E5ASS2). This F_{420} biosynthesis operon was synthesised (Biomatik) in two fragments (P_rhizo_3PG-F420-1 and P_rhizo_3PG-F420-2), P_rhizo_3PG-F420-1 was a 3.2 kb flanked with *EcoRI* and *BamHI* restriction sites in pUC57, P_rhizo_3PG-F420-2 was a 3.6 kb flanked with *BamHI* and *PstI* restrictions sites in vector pUC57. Three-way ligation was performed to ligate P_rhizo_3PG-F4201 (cut with *EcoRI* and *BamHI*), P_rhizo_3PG-F4202 (*BamHI* and *PstI*) and pF420 (cut with *EcoRI* and *PstI*). The ligation resulted in plasmid pF420-3PG (Supplementary Table S2).

The *ppsA* gene encoding PPS from *E. coli* (P23538) and *pck* gene encoding PPCK from *E. coli* (B5YTV3) were synthesized and cloned into pETCC2⁵⁸ plasmid flanked by *NdeI* and *BamHI* restriction sites (Twist Bioscience) to produce PPS-pETCC2 and PPCK-pETCC2 plasmids (Supplementary Table S2). pF420 have mutated pUC57 origin of replication which causes high copy number⁵⁹ and pETCC2 have pBR322 origin of replication, which are incompatible for co-transformation. Therefore, genes encoding PPS and PPCK were cloned into pRSF duet in order to have replicative compatibility with pF420 vector. pETCC2 and pRSF duet plasmids were digested with restriction enzymes *NdeI* and *BamHI*, the genes were gel purified from pETCC2 digestion and were ligated to pRSF duet plasmid. The pRSF duet plasmid containing the genes encoding PPS (PPS-pRSF) or PPCK (PPCK-pRSF) were used to transform BL21 DE3 containing pF420 plasmid using electroporation. Transformed cells

were selected on LB agar plates supplemented with chloramphenicol and kanamycin. Supplementary Table S2 summarises the list of plasmids used in this study.

Expression of F_{420} pathway proteins in pF420 had been confirmed previously using immunoblotting. The procedure is explained in detail by Bashiri et al.¹⁸. Soluble expression of PPS, PPCK and F_{420} -3PG operon proteins was confirmed through SDS-PAGE gel (Supplementary Fig. S3).

Analytical methods

F_{420} detection, quantification and chain length measurement. Production of F_{420} in *E. coli* and its detection was confirmed using LC-MS, as previously reported¹⁸. For the quantification of F_{420} from *E. coli*, 1 ml of sample from shake flask cultivation was taken and centrifuged at 10,000g for 2 min. The pellet was resuspended in 120 μ l of 75% ethanol, boiled for 3 min at 94 °C to lyse the cells, resuspension was centrifuged at 10,000g for 2 min and the fluorescence of 100 μ l supernatant was measured in a SpectraMax M3 (Molecular Devices) 96-well plate spectrofluorometer (excitation at 420 nm and emission at 480 nm). Fluorescence correlated directly with F_{420} concentration in the cell lysate. F_{420} amount relative to the biomass concentration was obtained in fluorescence (Units)/Biomass (OD_{600}). The correlation between Biomass (OD_{600}) and dry cell weight (DCW) was 0.56 g/L DCW for OD_{600} of 1.0. There was a linear correlation between absorbance at 420 nm and fluorescence (420 nm excitation and 480 nm emission). The fluorescence values of the cell lysate were converted to absorbance, and the extinction coefficient of 41.4 mM^{-1}/cm was used⁶⁰ to convert fluorescence unit of cell lysate to mM of F_{420} in cell lysate. Using these parameters fluorescence (units)/Biomass (OD) was converted to μ mol F_{420}/g DCW.

Analytical separation of F_{420} species based on the length of its glutamate tail is shown in S Supplementary Fig. S4 and was achieved with an ion-paired reverse phase HPLC-FLD protocol as reported previously⁶¹. The supernatant was run on an Agilent 1200 series HPLC system equipped with an Agilent fluorescence detector and an Agilent Poroshell 120 EC-C18 2.1 \times 50 mm 2.7 mm column. The system was run at a flow rate of 0.5 ml/min and the samples were excited at 420 nm and emission was detected at 480 nm. A linear gradient of two buffers were used: Buffer A, containing 20 mM ammonium phosphate, 10 mM tetrabutylammonium phosphate, pH 7.0. Buffer B, 100% acetonitrile. A gradient was run from 25 to 40% buffer B as follows: 0–1 min 25%, 1–10 min 25%–35%, 10–13 min 35%, 13–16 min 35–40%, 16–19 min 40%–25%.

Intracellular PEP was measured using single ion monitoring method (SIM) in single quad (Agilent 6120). Two buffers were used, Buffer A containing 10 mM ammonium formate with pH adjusted to 4.0 and Buffer B containing 100% acetonitrile, isocratic flow with 25% buffer B was used with flow rate of 0.5 ml/min using Agilent Poroshell 120 EC-C18 2.1 \times 50 mm 2.7 mm column. 10 μ l of sample was injected and SIM method was used to detect mass of 167 m/z in negative mode in order to estimate PEP concentration in cell lysate.

Metabolic network model of *E. coli* BL21. The iHK1487 metabolic model of *E. coli*³¹, containing 1487 genes, 2701 reactions, and 1164 metabolites, was used as the scaffold to integrate F_{420} biosynthesis pathway in this study. 360 reactions were found to be mass or charge imbalanced in the model, of which 338 reactions were demand or exchange reactions and one was biomass reaction, and therefore, were excluded from further corrections. Of the remaining 21 imbalanced reactions, 12 were involved in the lipopolysaccharide and cell envelope biosynthesis/recycling pathways, four were in alternate carbon metabolism pathway, three were in capsular polysaccharide biosynthesis/recycling pathway, and two were transporters (Supplementary File 1; Table S14). 14 imbalanced reactions were corrected by balancing protons, correcting chemical formula, or modifying participating metabolite(s) (see Supplementary File 1 for details). For example, four new metabolites were amended to the metabolic model to represent balanced core oligosaccharide lipid A molecules. The remaining seven imbalanced reactions could not be further resolved due to the lack of metabolite and/or enzyme specificity, rendered non-essential after flux analysis.

The bacterial F_{420} biosynthesis pathway proposed by Bashiri et al.¹⁸ was added to the original metabolic model to enable prediction of F_{420} production by engineered *E. coli*. This updated model was named iEco-F420. F_O synthase in the iEco-F420 model was modified by adding two molecules of *S*-adenosyl-L-methionine to the reactant side, and adding two molecules of L-methionine, one molecule of ammonia, and two molecules of 5'-deoxyadenosine to the product side of the catalysing reaction. Moreover, the updated model allows for analysing the biosynthesis of cofactor F_{420} with up to eight glutamate residues (through L-glutamate:coenzyme F_{420} ligase, CofE). For that, either an F_{420} molecule with one glutamate residue or a stoichiometric combination of F_{420} molecules with varying glutamate tails can be set as the target to analyse their flux and biosynthesis profile. Furthermore, the iEco-F420 model contains F_{420} -dependent formate dehydrogenase⁶², F_{420} -dependent G6P dehydrogenase⁶³, F_{420} -NADP oxidoreductase⁶⁴, F_{420} -dependent oxidoreductase⁵⁰, and F_{420} -reducing hydrogenase⁶⁵ allowing for the analysis of cofactor F_{420} recycling and regeneration within the metabolic network of engineered F_{420} -producing *E. coli*.

Integrating the F_{420} biosynthesis pathway along with correcting imbalances resulted in the iEco-F420 metabolic model of *E. coli* with 26 new metabolites and 43 new reactions. The updated metabolic model is available in Excel format in Tables S15 and S16 of Supplementary File 1. All reaction fluxes are in mmol/gDCW-h except for the reaction representing cell biomass formation that is expressed in h^{-1} . The M9 minimal medium composition was used to constrain the input of nutrients in the updated model. Independent simulations were run using glucose, glycerol, pyruvate, acetate, fumarate, or succinate as sole carbon sources (60 C-mol of carbon source). The uptake of C-source was fixed to 60 C-mol in all simulations to account for differences in number of carbon atoms in C-sources. The objective function was to maximize F_{420} production, while maintaining growth at 30% of its maximum to represent in vivo growth conditions. Maintenance ATP requirements were fixed at 5.17 mmol/g DCW and the minimum oxygen uptake was set to 18.5 mmol/g DCW/h. For simulating in vivo

over-expression of PPS and PPCK experiments, flux through PPS or PPCK was fixed in each run by constraining their lower and upper bounds to a value between zero and 70 mmol/g DCW/h, whereas zero represents no over-expression. Only one carbohydrate transporter was allowed to be active in each of these simulation runs. The model was assembled in a format compatible for flux balance analysis⁶⁶. FBA optimization problems were solved by GNU Linear Programming Kit (GLPK) (<http://www.gnu.org/software/glpk/>) solver in MATLAB using COBRA toolbox⁶⁷. Flux variability analysis (FVA) was performed to obtain range of fluxes under optimal growth conditions as described previously⁶⁸.

Data availability

The datasets generated during and/or analysed during the current study are available either as supplementary files or from the corresponding author on reasonable request.

Received: 30 May 2021; Accepted: 21 October 2021

Published online: 05 November 2021

References

- Daniels, L., Fuchs, G., Thauer, R. K. & Zeikus, J. G. Carbon monoxide oxidation by methanogenic bacteria. *J. Bacteriol.* **132**, 118–126 (1977).
- Eirich, L. D., Vogels, G. D. & Wolfe, R. S. Proposed structure for coenzyme F₄₂₀ from *Methanobacterium*. *Biochemistry* **17**, 4583–4593. <https://doi.org/10.1021/bi00615a002> (1978).
- Eirich, L. D., Vogels, G. D. & Wolfe, R. S. Distribution of coenzyme F₄₂₀ and properties of its hydrolytic fragments. *J. Bacteriol.* **140**, 20–27 (1979).
- Ferry, J. G. *Methanogenesis: Ecology, physiology, biochemistry and genetics*. (Springer, Berlin, 2012).
- Ferry, J. G. Biochemistry of methanogenesis. *Crit. Rev. Biochem. Mol.* **27**, 473–503. <https://doi.org/10.3109/10409239209082570> (1992).
- Knittel, K. & Boetius, A. Anaerobic oxidation of methane: Progress with an unknown process. *Annu. Rev. Microbiol.* **63**, 311–334. <https://doi.org/10.1146/annurev.micro.61.080706.093130> (2009).
- Nazem-Bokaei, H., Yan, Z., Maranas, C. D. & Ferry, J. G. in *Methane Biocatalysis: Paving the Way to Sustainability* 183–197 (Springer, Berlin, 2018).
- Wang, F. P. *et al.* Methanotrophic archaea possessing diverging methane-oxidizing and electron-transporting pathways. *ISME J.* **8**, 1069–1078. <https://doi.org/10.1038/ismej.2013.212> (2014).
- Thauer, R. K. Anaerobic oxidation of methane with sulfate: On the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO₂. *Curr. Opin. Microbiol.* **14**, 292–299. <https://doi.org/10.1016/j.mib.2011.03.003> (2011).
- Reuke, B., Korn, S., Eisenreich, W. & Bacher, A. Biosynthetic precursors of deazaflavins. *J. Bacteriol.* **174**, 4042–4049. <https://doi.org/10.1128/jb.174.12.4042-4049.1992> (1992).
- Neusser, D. *et al.* The genes lmbB1 and lmbB2 of *Streptomyces lincolnensis* encode enzymes involved in the conversion of L-tyrosine to propylproline during the biosynthesis of the antibiotic lincomycin A. *Arch. Microbiol.* **169**, 322–332. <https://doi.org/10.1007/s002030050578> (1998).
- Jacobson, F. & Walsh, C. Properties of 7,8-didemethyl-8-hydroxy-5-deazaflavins relevant to redox coenzyme function in methanogen metabolism. *Biochemistry* **23**, 979–988. <https://doi.org/10.1021/bi00300a028> (1984).
- Walsh, C. Naturally-occurring 5-deazaflavin coenzymes—Biological redox roles. *Acc. Chem. Res.* **19**, 216–221. <https://doi.org/10.1021/ar00127a004> (1986).
- Taylor, M. C. *et al.* Identification and characterization of two families of F₄₂₀H₂-dependent reductases from *Mycobacteria* that catalyze aflatoxin degradation. *Mol. Microbiol.* **78**, 561–575. <https://doi.org/10.1111/j.1365-2958.2010.07356.x> (2010).
- Shah, M. V. *et al.* Cofactor F₄₂₀-dependent enzymes: An under-explored resource for asymmetric redox biocatalysis. *Catalysts* **9**, 868 (2019).
- Yan, Z., Wang, M. & Ferry, J. G. A Ferredoxin- and F₄₂₀H₂-dependent, electron-bifurcating, heterodisulfide reductase with homologs in the domains bacteria and archaea. *mBio*. <https://doi.org/10.1128/mBio.02285-16> (2017).
- Nazem-Bokaei, H. & Maranas, C. D. A prospective study on the fermentation landscape of gaseous substrates to biorenewables using *Methanosarcina acetivorans* metabolic model. *Front. Microbiol.* **9**, 1855. <https://doi.org/10.3389/fmicb.2018.01855> (2018).
- Bashiri, G. *et al.* A revised biosynthetic pathway for the cofactor F₄₂₀ in prokaryotes. *Nat. Commun.* **10**, 1558. <https://doi.org/10.1038/s41467-019-09534-x> (2019).
- DePas, W. H., Bergkessel, M. & Newman, D. K. Aggregation of nontuberculous mycobacteria is regulated by carbon-nitrogen balance. *mBio* **10**, e01715–01719. <https://doi.org/10.1128/mBio.01715-19> (2019).
- Kimachi, T., Kawase, M., Matsuki, S., Tanaka, K. & Yoneda, F. First total synthesis of coenzyme factor 420. *J. Chem. Soc. Perkin Trans. 1*, 253–256. <https://doi.org/10.1039/P19900000253> (1990).
- Kimachi, T., Tanaka, K. & Yoneda, F. Synthesis of a proposed isomer of F₄₂₀ having a-glutamyl bonding. *J. Heterocycl. Chem.* **28**, 439–443. <https://doi.org/10.1002/jhet.5570280244> (1991).
- Hossain, M. S. *et al.* Convenient synthesis of deazaflavin cofactor F₀ and its activity in F₄₂₀-dependent NADP reductase. *Org. Biomol. Chem.* **13**, 5082–5085. <https://doi.org/10.1039/C5OB00365B> (2015).
- Drenth, J., Trajkovic, M. & Fraaije, M. W. Chemoenzymatic synthesis of an unnatural deazaflavin cofactor that can fuel F₄₂₀-dependent enzymes. *ACS Catal.* **9**, 6435–6443. <https://doi.org/10.1021/acscatal.9b01506> (2019).
- Decamps, L. *et al.* Biosynthesis of F₀, precursor of the F₄₂₀ cofactor, requires a unique two radical-SAM domain enzyme and tyrosine as substrate. *J. Am. Chem. Soc.* **134**, 18173–18176. <https://doi.org/10.1021/ja307762b> (2012).
- Isabelle, D., Simpson, D. R. & Daniels, L. Large-scale production of coenzyme F₄₂₀-5,6 by using *Mycobacterium smegmatis*. *Appl. Environ. Microbiol.* **68**, 5750–5755. <https://doi.org/10.1128/aem.68.11.5750-5755.2002> (2002).
- Braga, D. *et al.* Metabolic pathway rerouting in *Paraburkholderia rhizoxinica* evolved long-overlooked derivatives of coenzyme F₄₂₀. *ACS Chem. Biol.* <https://doi.org/10.1021/acscchembio.9b00605> (2019).
- Braga, D. *et al.* Metabolic pathway rerouting in *Paraburkholderia rhizoxinica* evolved long-overlooked derivatives of coenzyme F₄₂₀. *ACS Chem. Biol.* **14**, 2088–2094. <https://doi.org/10.1021/acscchembio.9b00605> (2019).
- Bashiri, G. & Baker, E. N. Convergent pathways to biosynthesis of the versatile cofactor F₄₂₀. *Curr. Opin. Struct. Biol.* **65**, 9–16. <https://doi.org/10.1016/j.sbi.2020.05.002> (2020).
- Grochowski, L. L., Xu, H. & White, R. H. Identification and characterization of the 2-phospho-L-lactate guanylyltransferase involved in coenzyme F₄₂₀ biosynthesis. *Biochemistry* **47**, 3033–3037. <https://doi.org/10.1021/bi702475t> (2008).
- Braga, D. *et al.* Metabolic pathway rerouting in *Paraburkholderia rhizoxinica* evolved long-overlooked derivatives of coenzyme F₄₂₀. *bioRxiv*. <https://doi.org/10.1101/670455> (2019).
- Kim, H., Kim, S. & Yoon, S. H. Metabolic network reconstruction and phenome analysis of the industrial microbe, *Escherichia coli* BL21 (DE3). *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0204375> (2018).

32. Bashiri, G., Rehan, A. M., Greenwood, D. R., Dickson, J. M. J. & Baker, E. N. Metabolic engineering of cofactor F₄₂₀ production in *Mycobacterium smegmatis*. *PLoS ONE* **5**, e15803. <https://doi.org/10.1371/journal.pone.0015803> (2010).
33. Lowry, O. H., Carter, J., Ward, J. B. & Glaser, L. The effect of carbon and nitrogen sources on the level of metabolic intermediates in *Escherichia coli*. *J. Biol. Chem.* **246**, 6511–6521 (1971).
34. Andersen, K. B. & von Meyenburg, K. Charges of nicotinamide adenine nucleotides and adenylate energy charge as regulatory parameters of the metabolism in *Escherichia coli*. *J. Biol. Chem.* **252**, 4151–4156 (1977).
35. Gosset, G. Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. *Microb. Cell Fact.* **4**, 14. <https://doi.org/10.1186/1475-2859-4-14> (2005).
36. Gottlieb, K., Albermann, C. & Sprenger, G. A. Improvement of L-phenylalanine production from glycerol by recombinant *Escherichia coli* strains: The role of extra copies of *glpK*, *glpX*, and *tktA* genes. *Microb. Cell Fact.* **13**, 96. <https://doi.org/10.1186/s12934-014-0096-1> (2014).
37. Graupner, M. & White, R. H. Biosynthesis of the phosphodiester bond in coenzyme F₄₂₀ in the methanoarchaea. *Biochemistry* **40**, 10859–10872. <https://doi.org/10.1021/bi0107703> (2001).
38. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* **5**, 593–599. <https://doi.org/10.1038/nchembio.186> (2009).
39. Braga, D., Hasan, M., Kröber, T., Last, D. & Lackner, G. Redox coenzyme F₄₂₀ biosynthesis in *Thermomicrobia* involves reduction by stand-alone nitroreductase superfamily enzymes. *Appl. Environ. Microbiol.* **86**, e00457–e420. <https://doi.org/10.1128/aem.00457-20> (2020).
40. Hansen, E. J. & Juni, E. Two routes for synthesis of phosphoenolpyruvate from C₄-dicarboxylic acids in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **59**, 1204–1210. [https://doi.org/10.1016/0006-291X\(74\)90442-2](https://doi.org/10.1016/0006-291X(74)90442-2) (1974).
41. Ranjan Patnaik, William D. Roof, Ryland F. Young & Liao, J. C. Stimulation of glucose catabolism in *Escherichia coli* by a potential futile cycle. *J. Bacteriol.* **174**, 7527–7532 (1992).
42. Patnaik, R. & Liao, J. C. Engineering of *Escherichia coli* central metabolism for aromatic metabolite production with near theoretical yield. *Appl. Environ. Microbiol.* **60**, 3903–3908 (1994).
43. Yi, J., Li, K., Draths, K. M. & Frost, J. W. Modulation of phosphoenolpyruvate synthase expression increases shikimate pathway product yields in *E. coli*. *Biotechnol. Prog.* **18**, 1141–1148. <https://doi.org/10.1021/bp020101w> (2002).
44. Yakandawala, N., Romeo, T., Friesen, A. D. & Madhyastha, S. Metabolic engineering of *Escherichia coli* to enhance phenylalanine production. *Appl. Microbiol. Biotechnol.* **78**, 283–291. <https://doi.org/10.1007/s00253-007-1307-z> (2008).
45. Liao, W. R. F. a. J. C. Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nat. Biotechnol.* **18**, 533–537 (2000).
46. Kao, K. C., Tran, L. M. & Liao, J. C. A global regulatory role of gluconeogenic genes in *Escherichia coli* revealed by transcriptome network analysis. *J. Biol. Chem.* **280**, 36079–36087. <https://doi.org/10.1074/jbc.M508202200> (2005).
47. Nocek, B. *et al.* Structure of an amide bond forming F₄₂₀:g-glutamyl ligase from *Archaeoglobus fulgidus*—A member of a new family of non-ribosomal peptide synthetases. *J. Mol. Biol.* **372**, 456–469. <https://doi.org/10.1016/j.jmb.2007.06.063> (2007).
48. Li, H., Graupner, M., Xu, H. & White, R. H. CofE catalyzes the addition of two glutamates to F₄₂₀-0 in F₄₂₀ coenzyme biosynthesis in *Methanococcus jannaschii*. *Biochemistry* **42**, 9771–9778. <https://doi.org/10.1021/bi034779b> (2003).
49. Purwantini, E., Loganathan, U. & Mukhopadhyay, B. Coenzyme F₄₂₀-dependent glucose-6-phosphate dehydrogenase coupled polyglutamylation of coenzyme F₄₂₀ in mycobacteria. *J. Bacteriol.* <https://doi.org/10.1128/jb.00375-18> (2018).
50. Ney, B. *et al.* Cofactor tail length modulates catalysis of bacterial F₄₂₀-dependent oxidoreductases. *Front. Microbiol.* **8**, 1902. <https://doi.org/10.3389/fmicb.2017.01902> (2017).
51. Wei, Z. *et al.* Folate polyglutamylation eliminates dependence of activity on enzyme concentration in mitochondrial serine hydroxymethyltransferases from *Arabidopsis thaliana*. *Arch. Biochem. Biophys.* **536**, 87–96. <https://doi.org/10.1016/j.abb.2013.06.004> (2013).
52. Matthews, R. G., Ghose, C., Green, J. M., Matthews, K. D. & Bruce Dunlap, R. Folylpolylglutamates as substrates and inhibitors of folate-dependent enzymes. *Adv. Enzyme Regul.* **26**, 157–171. [https://doi.org/10.1016/0065-2571\(87\)90012-4](https://doi.org/10.1016/0065-2571(87)90012-4) (1987).
53. Besson, V., Rebeille, F., Neuburger, M., Douce, R. & Cossins, E. A. Effects of tetrahydrofolate polyglutamates on the kinetic parameters of serine hydroxymethyltransferase and glycine decarboxylase from pea leaf mitochondria. *Biochem. J.* **292**, 425–430. <https://doi.org/10.1042/bj2920425> (1993).
54. Peck, M. W. Changes in concentrations of coenzyme F₄₂₀ analogs during batch growth of *Methanosarcina barkeri* and *Methanosarcina mazei*. *Appl. Environ. Microbiol.* **55**, 940–945 (1989).
55. Hoque, M. A. *et al.* Comparison of dynamic responses of cellular metabolites in *Escherichia coli* to pulse addition of substrates. *Biologia* **66**, 954. <https://doi.org/10.2478/s11756-011-0136-9> (2011).
56. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I₁-I₂ regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210. <https://doi.org/10.1093/nar/25.6.1203> (1997).
57. Cambray, G. *et al.* Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* **41**, 5139–5148. <https://doi.org/10.1093/nar/gkt163> (2013).
58. Peat, T. S. *et al.* Cyanuric acid hydrolase: Evolutionary innovation by structural concatenation. *Mol. Microbiol.* **88**, 1149–1163 (2013).
59. Lin-Chao, S., Chen, W.-T. & Wong, T.-T. High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. *Mol. Microbiol.* **6**, 3385–3393. <https://doi.org/10.1111/j.1365-2958.1992.tb02206.x> (1992).
60. Purwantini, E., Mukhopadhyay, B., Spencer, R. W. & Daniels, L. Effect of temperature on the spectral properties of coenzyme F₄₂₀ and related compounds. *Anal. Biochem.* **205**, 342–350. [https://doi.org/10.1016/0003-2697\(92\)90446-e](https://doi.org/10.1016/0003-2697(92)90446-e) (1992).
61. Ney, B. *et al.* The methanogenic redox cofactor F₄₂₀ is widely synthesized by aerobic soil bacteria. *ISME J.* **11**, 125–137. <https://doi.org/10.1038/ismej.2016.100> (2017).
62. Schauer, N. L., Ferry, J. G., Honek, J. F., Orme-Johnson, W. H. & Walsh, C. Mechanistic studies of the coenzyme F₄₂₀ reducing formate dehydrogenase from *Methanobacterium formicicum*. *Biochemistry* **25**, 7163–7168. <https://doi.org/10.1021/bi00370a059> (1986).
63. Nguyen, Q.-T., Trinco, G., Binda, C., Mattevi, A. & Fraaije, M. W. Discovery and characterization of an F₄₂₀-dependent glucose-6-phosphate dehydrogenase (Rh-FGD1) from *Rhodococcus jostii* RHA1. *Appl. Microbiol. Biotechnol.* **101**, 2831–2842. <https://doi.org/10.1007/s00253-016-8038-y> (2017).
64. Kumar, H., Nguyen, Q.-T., Binda, C., Mattevi, A. & Fraaije, M. W. Isolation and characterization of a thermostable F₄₂₀:NADPH oxidoreductase from *Thermobifida fusca*. *J. Biol. Chem.* **292**, 10123–10130. <https://doi.org/10.1074/jbc.M117.787754> (2017).
65. Hendrickson, E. L. & Leigh, J. A. Roles of coenzyme F₄₂₀-reducing hydrogenases and hydrogen- and F₄₂₀-dependent methylene-tetrahydromethanopterin dehydrogenases in reduction of F₄₂₀ and production of hydrogen during methanogenesis. *J. Bacteriol.* **190**, 4818–4821. <https://doi.org/10.1128/jb.00255-08> (2008).
66. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis?. *Nat. Biotechnol.* **28**, 245–248. <https://doi.org/10.1038/nbt.1614> (2010).
67. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols* **14**, 639–702. <https://doi.org/10.1038/s41596-018-0098-2> (2019).
68. Mahadevan, R. & Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264–276. <https://doi.org/10.1016/j.ymben.2003.09.002> (2003).

Acknowledgements

We would like to thank the CSIRO Synthetic Biology Future Science Platform for funding this work and providing Fellowships for MVS and HNB, and PhD Top-up funding to SWK and JA. We would also like to thank Adjunct Professor Claudia Vickers and Dr. Andrew Warden for their helpful comments during the preparation of this manuscript.

Author contributions

M.V.S., conceptualization, methodology, formal analysis, investigation, writing - original draft; H.N.-B., conceptualization, methodology, formal analysis, investigation, simulations, writing - original draft, figures; J.A., investigation; S.W.K., investigation; C.J.J., conceptualization, formal analysis, supervision, writing - review & editing, funding acquisition; C.S., conceptualization, formal analysis, supervision, writing - review & editing, project administration, funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-01224-3>.

Correspondence and requests for materials should be addressed to C.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021