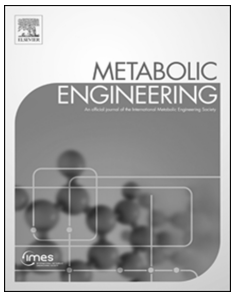

Engineering improved ethylene production: Leveraging systems Biology and adaptive laboratory evolution

Sophie Vaud, Nicole Pearcy, Marko Hanževački, Alexander M.W. Van Hagen, Salah Abdelrazig, Laudina Safo, Muhammad Ehsaan, Magdalene Jonczyk, Thomas Millat, Sean Craig, Edward Spence, James Fothergill, Rajesh Reddy Bommareddy, Pierre-Yves Colin, Jamie Twycross, Paul Dalby, Nigel Minton, Christof M. Jäger, Dong-Hyun Kim, Jianping Yu, Pin-Ching Maness, Sean Lynch, Carrie Eckert, Alex Conradie, Samantha J. Bryan

PII: S1096-7176(21)00111-7

DOI: https://doi.org/10.1016/j.ymben.2021.07.001

Reference: YMBEN 1829

To appear in: *Metabolic Engineering*

Received Date: 12 March 2021

Revised Date: 26 May 2021

Accepted Date: 2 July 2021

Please cite this article as: Vaud, S., Pearcy, N., Hanževački, M., Van Hagen, A.M.W., Abdelrazig, S., Safo, L., Ehsaan, M., Jonczyk, M., Millat, T., Craig, S., Spence, E., Fothergill, J., Bommareddy, R.R., Colin, P.-Y., Twycross, J., Dalby, P., Minton, N., Jäger, C.M., Kim, D.-H., Yu, J., Maness, P.-C., Lynch, S., Eckert, C., Conradie, A., Bryan, S.J., Engineering improved ethylene production: Leveraging systems Biology and adaptive laboratory evolution, *Metabolic Engineering* (2021), doi: https://doi.org/10.1016/j.ymben.2021.07.001.

Author Contributions

Sophie Vaud: Methodology, Validation, Conceptualisation, Investigation, Writing.

Nicole Pearcy: Software, Methodology, Writing, Conceptualisation, Writing review and editing, visualization.

Marko Hanževački: Methodology, Visualization, Writing.

Alexander M. W. Van Hagen: Investigation

Salah Abdelrazig: Investigation and Writing.

Laudina Safo: Investigation

Muhammad Ehsaan: Investigation

Magdalene Jonczyk: Investigation.

Thomas Millat: Formal analysis.

Sean Craig: Investigation.

Edward Spence: Investigation and Visualisation.

James Fothergill: Investigation.

Rajesh Reddy Bommareddy: Investigation.

Pierre-Yves Colin: Investigation

Jamie Twycross: Methodology, supervision, resources.

Paul Dalby: Methodology, supervision, resources.

Nigel Minton: Resources, funding acquisition, writing review, and editing

Christof M. Jäger: Methodology, supervision, Writing and Writing review and editing.

Dong-Hyun Kim: Resources and Methodology, supervision, writing and Writing review and editing

Jianping Yu: Conceptualisation, Writing review and editing, resources.

Pin-Ching Maness: Conceptualisation, Writing review and editing, Resources.

Sean Lynch: Investigation, Visualisation, and Conceptualisation.

Carrie Eckert: Conceptualisation, Writing, Writing review and editing, Resources, funding acquisition, Supervision, Project administration.

Alex Conradie: Conceptualisation, Writing review and editing.

Samantha Bryan: Conceptualisation, methodology, writing original draft, writing review and editing, supervision, funding acquisition, project administration.

Engineering Improved Ethylene Production: Leveraging Systems Biology and Adaptive Laboratory Evolution

Sophie Vaud[1*], Nicole Pearcy[1*], Marko Hanževački[2], Alexander M. W. Van Hagen[2], Salah Abdelrazig[3], Laudina Safo[3], Muhammad Ehsaan[1], Magdalene Jonczyk[1], Thomas Millat[1] Sean Craig[2], Edward Spence[2], James Fothergill[1], Rajesh Reddy Bommareddy[2], Pierre-Yves Colin[4], Jamie Twycross[1,5], Paul Dalby[4], Nigel Minton[1], Christof M. Jäger[2], Dong-Hyun Kim[3], Jianping Yu[6], Pin-Ching Maness[6], Sean Lynch[6,7,8], Carrie Eckert[6,7], Alex Conradie[2], and Samantha J. Bryan[2 #]

[1]BBSRC/EPSRC Synthetic Biology Research Centre, The Biodiscovery Institute, University of Nottingham, Nottingham, NG7 2RD.

[2]Department of Chemical and Environmental Engineering, Faculty of Engineering, University of Nottingham, Nottingham, NG7 2RD.

[3] Centre for Analytical Bioscience, Advanced Materials and Healthcare Technologies Division, School of Pharmacy, University of Nottingham, NG7 2RD.

[4]Department of Biochemical Engineering, Bernard Katz Building, University College London, WC1E 6BT.

[5]School of Computer Science, University of Nottingham, Nottingham, NG7 2RD.

[6] Biosciences Center, National Renewable Energy Laboratory (NREL), Golden, CO, USA.

[7] Renewable and Sustainable Energy Institute (RASEI), University of Colorado Boulder, USA

[8]Somalogic, Inc., Boulder, CO, USA.

[*]These two authors contributed equally.

[#] For correspondence. E-mail Samantha.bryan@nottingham.ac.uk.

Abstract

Ethylene is a small hydrocarbon gas widely used in the chemical industry. Annual worldwide production currently exceeds 150 million tons, producing considerable amounts of CO_2 contributing to climate change. The need for a sustainable alternative is therefore imperative. Ethylene is natively produced by several different microorganisms, including *Pseudomonas syringae* pv. *phaseolicola* via a process catalyzed by the ethylene forming enzyme (EFE), subsequent heterologous expression of EFE has led to ethylene production in non-native bacterial hosts including *E. coli* and cyanobacteria. However, solubility of EFE and substrate availability remain rate limiting steps in biological ethylene production. We employed a combination of genome scale metabolic modelling, continuous fermentation, and protein evolution to enable the accelerated development of a high efficiency ethylene producing *E. coli* strain, yielding a 49-fold increase in production, the most significant improvement reported to date. Furthermore, we have clearly demonstrated that this increased yield resulted from metabolic adaptations that were uniquely linked to the EFE enzyme (WT vs mutant). Our findings provide a novel solution to deregulate metabolic bottlenecks in key pathways, which can be readily applied to address other engineering challenges.

Ethylene is an important chemical utilized for a variety of applications, and production currently exceeds 150 million tons (230 million tons by 2030). Ethylene is currently produced from the steam cracking of ethane and naphtha, which produces vast quantities of carbon dioxide (CO_2), contributing to global warming[1,2]. The biological production of ethylene production via engineered microorganisms offers a sustainable alternative. However, the implementation of a biological ethylene platform will require significant improvements to meet commercial production yields and offer a techno economically viable option. Technoeconomic analysis predicted that a process conversion facility with an annual production rate of 10 MMGGE hydrocarbon fuel would need to aim for a price of $5.36 per GGE[3].

The ethylene forming enzyme (EFE) utilizes α-ketoglutarate (AKG) and arginine (L-Arg) as substrates for ethylene production[4-6]. Crystallographic and biochemical studies on the EFE enzyme from *Pseudomonas syringae* pv. *phaseolicola* revealed a branched mechanism, consisting of a typical non-heme FeII- and 2-oxoglutarate-dependent (2-OG) oxygenase reaction yielding succinate, guanidine, CO_2, and pyrroline-5-carboxylate (P5C) and a secondary Grob-type oxidative fragmentation of a 2-OG intermediate generating ethylene and CO_2[6-8]. Ethylene has been produced in a wide variety of different microorganisms utilizing EFE from *P. syringae*[9-15]. These approaches have relied on engineering strains with selected knockouts, increased expression, and substrate availability[9-15]. Yet, EFE solubility and the availability of AKG and L-Arg, key substrates for ethylene productivity remain rate limiting steps in ethylene productivity and thus significantly limit yield.

Here we present an alternative approach to increase ethylene productivity in *E. coli*, simultaneously targeting substrate availability (AKG and L-Arg) and solubility (EFE) through a combination of genome scale metabolic modeling (GSMM), directed evolution and adaptive laboratory evolution (ALE). Our approach resulted in a 49-fold improvement in ethylene production in *E. coli*, compared to the 3-4-fold improvements in productivity[11, 16] already cited in the literature. Using a combination of molecular dynamics, targeted metabolomics analysis and the GSMM, we unraveled the potential metabolic adaptions responsible for the high ethylene-yielding strains. This is the first example of growth coupled ethylene production and demonstrates the power of this approach for subsequent efforts to relieve metabolic bottlenecks.

Results

Integral framework for optimizing ethylene production

An integrated computational and experimental pipeline that combines directed evolution and adaptive laboratory evolution with genome scale metabolic models (GSMs) (Fig. 1, Supplementary Fig. S1) was applied with the aim of increasing ethylene yields in *E. coli*. Firstly, a GSM of *E. coli* was used to identify and evaluate, *in silico*, knockout strategies for coupling ethylene production to growth. Coupling cellular growth to productivity has become increasingly desirable, making growth the driving force of production[17]. Mutagenesis, ALE and directed evolution were then utilised with the aim of overcoming known bottlenecks, including substrate availability, and improving enzyme solubility and/or substrate binding activity. Molecular-dynamics simulations were then used to identify causative relationships between EFE structural changes and improved EFE activity. Utilizing a combination of ALE and competitive

fermentation we are able to select mutants with increased fitness, and inherently increased ethylene production. The mutations in these strains enabled metabolic rewiring that promoted substrate availability for the EFE enzyme. A combination of targeted metabolomics and genome scale model analysis was utilised to mechanistically evaluate the best ethylene-yielding mutant strains at the systems level.

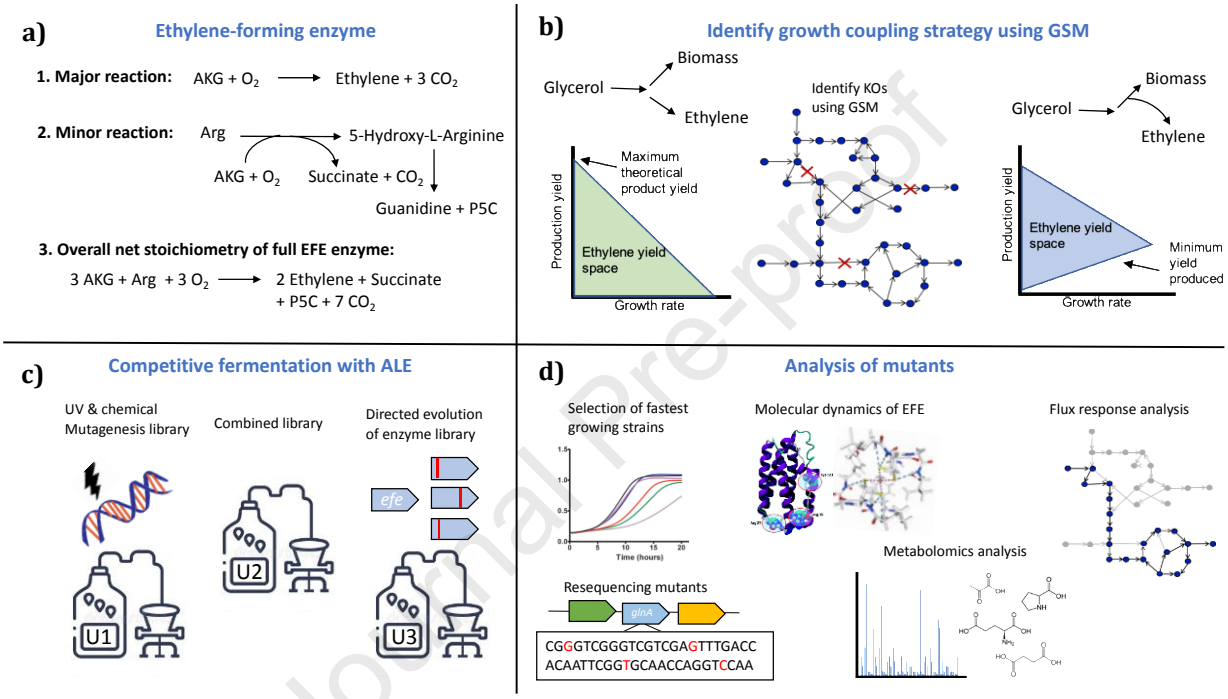

Figure 1: Illustration of the integral pipeline used in this study. (a) Major, minor, and overall reaction catalysed via the ethylene-forming enzyme (EFE). **(b)** Identification of a growth couple using the GSM of *E. coli*. The yield space initially (green graph) includes feasible solutions where ethylene yield is zero. The aim is to use the GSM to predict KOs that result in the minimum guaranteed ethylene yield increase with increasing growth rates (blue graph). **(c)** After achieving a growth-coupling *in vivo*, UV, chemical mutagenesis and directed evolution were carried out with the aim of increasing substrate availability and enzyme solubility. Competitive fermentation with adaptive laboratory evolution was applied to the UV and chemical mutagenesis library (U1), the combined library (U2) and the directed evolution library (U3). **(d)** Genome resequencing,

metabolomics analysis, metabolic flux response analysis and molecular dynamics simulations of EFE were then carried out on the best performing strains to provide a genetic and mechanistic insight into increased ethylene yield.

Genome scale model analysis identified two candidate growth-coupling strategies

A GSM is a comprehensive model of metabolism which contains all the known biochemical reactions for an organism. These models, combined with constraint-based approaches[18], provide a mechanistic platform for predicting the response of an organism following metabolic perturbations, such as gene knockouts and insertions. The available GSM of *E. coli* str K-12 MG1655, iJO1366 was used to identify novel knockout strategies for increasing ethylene production. First, the full EFE enzyme (Reaction 3 in Fig. 1) was added to the model, resulting in iJO1366-EFE[19]. Single KO analysis was then simulated in iJO1366-EFE using both Flux Balance Analysis (FBA)[18] and Minimization of Metabolic Adjustments (MOMA)[20]. FBA assumes that bacteria have evolved to maximize their growth, and therefore identifies a flux distribution that maximizes flux towards the biomass equation. MOMA is an alternative method for simulating flux distributions in mutant strains, which assumes bacteria will try and minimize their metabolic adjustments from the wild type strain for increased chances of survival. MOMA minimizes the Euclidean distance between the mutant and wild type flux distributions as the objective function to mimic this assumption in the GSM (see Materials and Methods). The flux distribution predicted using MOMA may therefore not predict a stable growth-coupling. The KO strategies that resulted in increased ethylene productivity, whilst also maintaining a high growth rate were considered.

Two promising strategies were identified. The first strategy involved knocking out the AKG dehydrogenase (AKGDH) and was predicted via MOMA to achieve 18 mmol of ethylene per mole of glycerol, whilst only reducing the growth rate by 10% (Fig. 2a, 2c). AKGDH is encoded by the *sucA* gene, and converts AKG to succinyl-CoA, generating NADH[21]. This *in silico* solution indicates that a Δs*ucA* mutant increases Arg availability for EFE compared to iJO1366-EFE (Fig. 2b-c). The second strategy involved knocking out either the glutamate 5-kinase (GLU5K) or glutamate-5-semialdehyde dehydrogenase (G5SD) and was predicted via FBA to result in 24 mmol of ethylene per mole of glycerol (Fig. 2a). GLU5K, encoded by the gene *proB*, catalyzes the transfer of a phosphate group to glutamate to form glutamate-5-phosphate[22]. The NADPH-dependent reduction of glutamate-5-phosphate into glutamate-5-semialdehyde is catalyzed by G5SD, encoded by the *proA* gene[22]. A comparison between the flux distributions of the Δ*proB* mutant and iJO1366-EFE (Fig. 2b, d), demonstrated that the only route to P5C, a precursor to proline synthesis, was by redirecting flux towards EFE. The two knockout strategies proposed here were both evaluated *in vivo* via recombineering in the MG1655 strain.

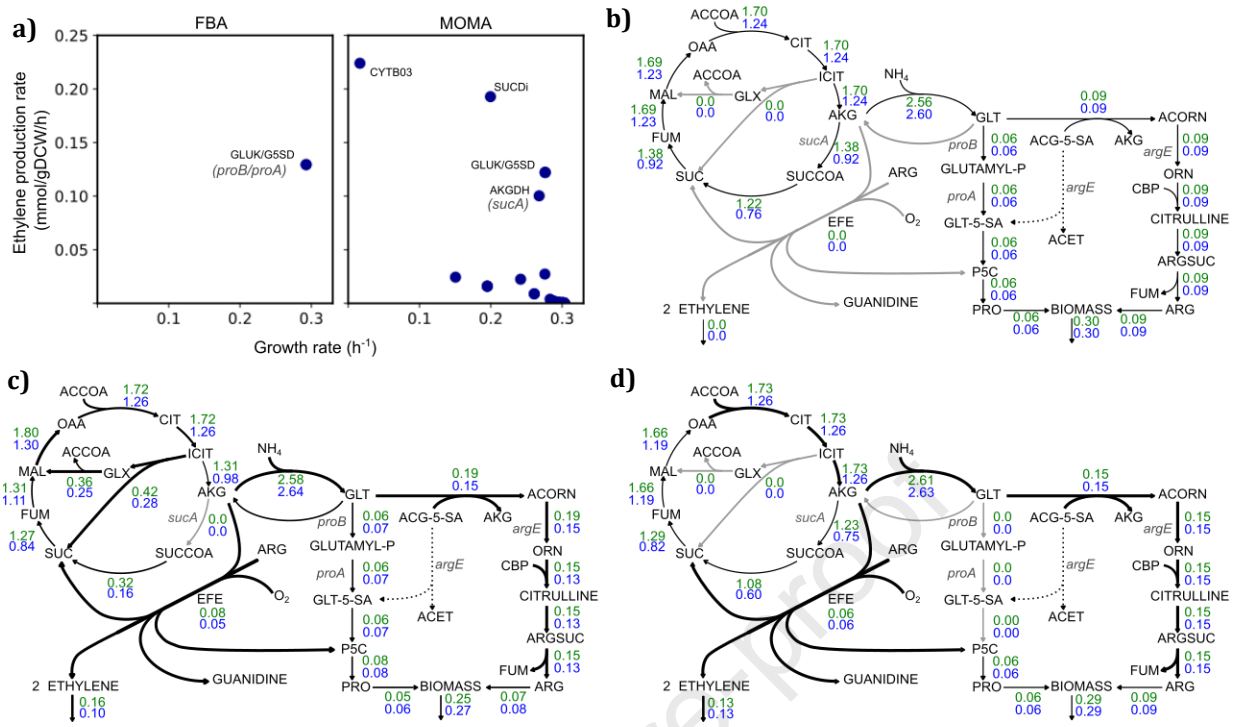

Figure 2: *In silico* knockout (KO) simulations using the genome scale model iJO1366-EFE. **(a)** *In silico* growth rate (h^{-1}) vs. ethylene production (mmol/gDCW/h) for the reaction KO candidates using flux balance analysis (left) and MOMA (right). Flux diagrams then show the ethylene forming enzyme (EFE) connected to the TCA cycle, glutamate, proline and arginine metabolism. The flux distributions shown are for **(b)** the iJO1366 wild type model simulated using FBA, **(c)** the Δ*sucA*-iJO1366-EFE mutant model simulated using MOMA and **(d)** the Δ*proB*-iJO1366-EFE mutant model simulated using FBA. Note that reaction fluxes are shown for growth on both glucose (blue) and glycerol (green). Reactions that are highlighted grey carry zero flux. The reaction fluxes that have increased in the mutant strains compared to the wild type strain are highlighted by thicker lines in (c) and (d). Reaction abbreviations in **(a)**: GLUK – Glutamate 5-kinase, G5SD- Glutamate-5-semialdehyde dehydrogenase, CYTB03- cytochrome oxidase bo3, SUCDi- succinate dehydrogenase, AKGDH- alpha-ketoglutarate dehydrogenase. Metabolite abbreviations in **(b)-(d)**: ACCOA- acetyl-CoA, CIT- citrate, ICIT- isocitrate, AKG- alpha-

ketoglutarate, SUCCOA- succinyl-CoA, SUC- succinate, FUM- fumarate, MAL- malate, OAA- oxaloacetate acid, GLT- glutamate, GLUTYAMYL-P- glutyamyl-phosphate, GLT-5-SA- glutamate-5-semialdehyde, P5C- 1-pyrroline-5-carboxylic acid, PRO- proline, ACORN- N-acetyl-L-ornithine, ACET- acetate, ORN- ornithine, ARGSUC- arginine-succinate, ARG- arginine, ACG-5-SA- N-acetylglutamate-5-semialdehyde, O_2 - oxygen. Note that the reaction represented by a dotted line, which corresponds to N-acetylornithine deacetylase, is in the iJO1366 model but was constrained to zero since it is not known capable of carrying flux in *E. coli in vivo*.

Growth Coupling ethylene to succinate decoupled succinate from ethylene production

The *sucA* gene was knocked out using recombineering in MG1655, and growth was evaluated in minimal media. Surprisingly, the *sucA* mutant could still grow without succinate, therefore growth was not coupled to succinate (Supplementary Fig. S2a). Therefore, we considered generating a double KO targeting *serA*, which encodes 3-phosphoglycerate dehydrogenase, SerA has been shown to have promiscuous activity towards AKG[23, 24]. This could explain the ability of the Δ*sucA* mutant to grow without succinate. We therefore created a double Δ*sucA ΔserA* knockout, this strain was no longer able to grow without succinate. The *efe* gene from *P. syringae* was codon optimized for *E. coli*, synthesized, and expressed in the Δ*sucA ΔserA* mutant, demonstrating a clear coupling to succinate through EFE. Given EFE solubility is a key bottleneck in the production of ethylene we used error prone PCR (epPCR) to generate an EFE library, with the aim of selecting variants with improved solubility. The library was transformed into the *ΔsucA ΔserA* mutant. Enrichment was then utilized to select the best performing strains, briefly, following growth, dilutions were plated, and 50 colonies were picked and the EFE gene was sequenced. Growth curves were generated for the most enriched variants, these included 4 strains, Q31E, Q28H, a strain with 20+

individual variants in the EFE enzyme and a strain with a stop codon which inactivated the EFE enzyme, and the four variants were compared to the wildtype (Supplementary Fig. S2b). Although the growth rate of the *ΔsucA ΔserA* mutant improved when the EFE variants were expressed, suggesting succinate was produced, ethylene production was no longer detected in any of the variants. Of the tested variants, all residues were generally conserved across the *Pseudomonas* EFE proteins (Supplementary Fig. S2c). These results agree with previous structural work that suggests 2-OG oxygenases may have evolved to favor the production of succinate, a versatile biomolecule and key component of TCA, rather than ethylene[7].

Growth coupling ethylene to proline resulted in increased ethylene productivity and EFE solubility.

The *proB* KO strain was evaluated *in vivo*, confirming growth required the addition of proline. The synthesized *efe* gene was cloned into the broad host-range plasmid pBBR1, and the pGEM vector (Promega), under the constitutive *phaC* promoter[25] and the synthetic promoter p15[26], respectively. To modulate for differences in plasmid copy-number, between pBBR1 and pGEM both promoters were exchanged and cloned into the opposite vector. Both plasmids were transformed into MG1655 and the Δ*proB* mutant generating the following strains MG1655 pBBR1 *PphaC efe,* MG1655 pGEM p15 *efe,* MG1655 pBBR1 p15 *efe,* MG1655 pGEM *PphaC efe* and Δ*proB* pBBR1 *PphaC efe,* Δ*proB* pGEM p15 *efe,* Δ*proB* pBBR1 p15 *efe* and Δ*proB* pGEM *PphaC efe.* The Δ*proB efe* strains were all subject to four rounds of selective passaging in M9 media without proline to establish the growth couple. Relative ethylene production was assessed in M9 media. Supplementary Fig. S3a-b clearly shows a correlation between plasmid copy-number and ethylene production as previously shown[11]. Importantly, in both the Δ*proB*_pBBR1 *efe* strains and

the Δ*proB*_pGEM *efe* strains, increasing ethylene production correlated to increased EFE solubility (Supplementary Fig. S4a-b). The growth rate remained similar between the individual MG1655 pBBR1 and pGEM efe and the Δ*proB* pGEM p15 *efe* strain. The Δ*proB* pBBR1 *PphaC efe* strain, however, has a slightly lower growth rate and an increased doubling time compared to the other strains (Table 1, Supplementary Fig. 3c).

Strain	MG1655	MG1655 pBBR1 *PphaC efe*	MG1655 pGEM p15 *efe*	Δ*proB* pBBR1 *PphaC efe*	Δ*proB* pGEM p15 *efe*.
μmax (h^{-1})	**0.31**	**0.28**	**0.31**	**0.2**	**0.28**
Dt (h)	**2.236**	**2.478**	**2.236**	**3.466**	**2.476**

Table 1: Growth rates and doubling times for MG1655 (WT), MG1655 pBBR1 *PphaC efe* and pGEM p15 *efe* and Δ*proB* pBBR1 *PphaC efe* and Δ*proB* pGEM p15 *efe*.

Random mutagenesis and competitive fermentation using proline coupling successfully isolated strains with improved ethylene production.

We evaluated the effect of whole cell mutagenesis on EFE solubility and substrate availability using the proline growth-couple as a selective screen. The Δ*proB*_pBBR1*efe* strains with both the P*phaC* and p15 promoters and the Δ*proB*_pGEM *efe* strains with both the *PphaC* and p15 promoters were subject to both UV and chemical mutagenesis. Mutant libraries from both UV and chemical mutagenesis, and from the epPCR-generated EFE library, were then subject to competitive fermentation on glycerol to select for the fastest growing strains with increased ethylene production. To run these experiments, three individual fermenters were set up: U1, U2 and U3, representing the UV and chemical mutagenesis libraries only (**U1**), the UV and chemical mutagenesis libraries and the epPCR library (**U2**) and the epPCR library only (**U3**). Each

fermentation also included the appropriate control plasmids (MG1655, MG1655 pBBR1 *PphaC efe,* MG1655 pGEM p15 *efe,* MG1655 pBBR1 p15 *efe* and MG1655 pGEM *PphaC efe).* Following inoculation, the dilution rate was increased steadily from 0.02 to 0.15 h-1. Cells were sampled at each steady state, this continued until the dilution rate could no longer be increased, with a total chemostat run time of 55.3 generations.

Both the U1 and U2 fermentations generated strains with improved ethylene production, and several strains which had a marked decrease in ethylene production (Supplementary Fig. S5a-b). The ethylene measurements were repeated in triplicate for the 10 highest producing strains from both the U1 (UV and chemical libraries) and U2 (UV, chemical and epPCR libraries) fermentations (Supplementary Fig. S6a). Furthermore, sequence analysis confirmed that the *efe* gene had not been mutated in these strains, implying that no mutants had been selected from the epPCR library in the U2 fermentation. Interestingly all the plasmids isolated from U1 and U2 were pBBR1-based, suggesting that high copy number plasmids were selected against. The best ethylene producer was isolated from the U2 fermentation, U2-48 EFE, which contained the pBBR1 *Pphac efe* plasmid, this strain had an ethylene production of 80 nmol/OD/ml. Western blot analysis on the soluble and insoluble fractions of U2-48 EFE confirmed there was an increase in soluble protein (Supplementary Fig. S4c). The growth rate of U248 EFE was comparable to Δ*proB*_pBBR1 *PphaC efe* (Table 2). The most marked effect on ethylene production was seen in the U3 fermentation (epPCR library only), where only one strain had improved ethylene production, U3-26 EFE (600 nmol/OD/ml) (Supplementary Fig. S6b). This strain contained the pGEM p15 plasmid with an error prone EFE variant. There was a slight decrease in the growth rate of U3-26 EFE compared to the Δ*proB*_pGEM p15 *efe* strain, (Table 2). There was also an increase in soluble

EFE protein in the U3-26 EFE strain (Supplementary Fig. S4d). The wild type EFE enzyme was cloned into pGEM p15 and transformed into the U3-26 background to generate U3-26 WT EFE. Importantly, EFE solubility also increased in this strain, demonstrating that mutations in the U3-26 genomic background could be responsible for the increased solubility of the EFE enzyme (Supplementary Fig S4e). A significant proportion of the colonies isolated from the U3 fermentation exhibited no ethylene production, similar to the succinate growth couple (Supplementary Fig. S2a-b), suggesting the enzyme is able to decouple P5C and ethylene. Nevertheless, unlike the Δ*sucA/serA* mutant, this strategy was successful in isolating strains with improved ethylene production.

Strain	MG1655	MG1655 pBBR1 *PphaC efe*	MG1655 pGEM p15 *efe*	Δ*proB* pBBR1 *PphaC efe*	Δ*proB* pGEM p15 *efe*.	U2-48 EFE	U3-26 EFE
μmax (h^{-1})	**0.31**	**0.28**	**0.31**	**0.2**	**0.28**	**0.2**	**0.24**
Dt (h)	**2.236**	**2.478**	**2.236**	**3.466**	**2.476**	**3.466**	**2.888**

Table 2: Growth rates and doubling times comparison for MG1655 (WT), MG1655 pBBR1 *PphaC efe*, MG1655 pGEM p15 *efe*, Δ*proB* pBBR1 *PphaC efe*, Δ*proB* pGEM p15 *efe*, U248 EFE and U326 EFE.

Sequence analysis of the *efe* gene in U3-26-EFE indicated the presence of several SNPs in the EFE enzyme which could play a key role in improved ethylene productivity.

The *efe* gene from U3-26 EFE was sequenced and had 5 SNPs: G105A, R184H, L339S, G108A and C837T. The SNPs were recreated in the WT EFE, both individually and in combination to assess the individual and combined effect on ethylene productivity. The mutated versions of EFE were transformed into both the WT strain MG1655 and the Δ*proB* background. The combined

SNPs decreased ethylene production in both strains when compared to the U3-26 EFE strain (Supplementary Fig. S6c). As a control, the *efe* gene from U3-26 EFE was also amplified and re-transformed into the WT background, which resulted in a significant decrease in ethylene production. A similar but smaller decrease was seen when the *efe* gene from U3-26 EFE was transformed into the Δ*proB* strain (Supplementary Fig. S6d). Therefore, the SNPs occurring in the *efe* gene may have effectively altered the stoichiometric balance between the major and minor reactions (as described in Fig. 1).

Structural and dynamic insights into WT EFE and U3-26 EFE from molecular-dynamics (MD) simulations indicate an altered balance of promiscuity to Arg and AKG

To investigate possible structural dynamic effects of the three mutations in the U3-26 variant on the balanced catalytic promiscuity of EFE molecular-dynamics (MD) simulations were applied comparing the structural characteristics of the WT and U3-26 EFE. A particular focus was set on the binding of L-Arg, which is important in different ways for both reaction pathways. In both cases, a tight L-Arg binding is responsible to stabilize AKG binding in a catalytic bidentate position. However, for the minor reaction (hydroxylation) L-Arg additionally acts as co-substrate, which demands a different binding, than when acting as a spectator for the major pathway (ethylene formation).

The newly introduced mutations in U3-26 EFE, G105, R184, and L339, are spatially separated from the active site and located at the distal regions on the protein surface. Martinez et al. (2017) demonstrated that most distal modifications of the C-terminus lead to a modest decrease in the overall activity of EFE[8]. Our analysis demonstrates that the R184 residue in the WT indirectly

interacts with the amine group of the L-Arg substrate in the active site, stabilizing the substrate through a hydrogen-bonding network via T86 and D91. Additionally, L-Arg forms a guanidinium stacking arrangement with R171 and a hydrogen bond with the sidechain of Y192 creating an unusual twist in the peptide bond. This peptide twist has been reported to have a significant role in catalysis enabling the correct orientation of L-Arg for hydroxylation while also stimulating the conversion of AKG to ethylene.[8] MD simulations show that the U3-26 EFE variant behaves differently with implications on the active site even though the mutations are distant to the substrate binding sites. Generally, both enzymes appear stable in the simulations with a slightly enhanced flexibility of the mutant. Particularly significantly increased fluctuations in U3-26 EFE are localized around H184, which indirectly influences the active site.

The WT active site showed to be more stable in comparison (Supplementary. Fig. S7a-b) with stabilizing interactions with L-Arg, namely stacking with R171 and hydrogen bonds with D191 and AKG. The R184H mutation leads to a disruption of the hydrogen bonding network and destabilization of L-Arg (Fig. 3b). This leads to L-Arg adopting a slightly more flexible binding whilst still stabilizing tight binding of AKG necessary for both reaction paths. In addition, the twisted peptide bond between D191 and Y192 previously observed in the X-ray structure remains stable in simulations of both enzyme variants (Fig. 3a). In contrast simulations without L-Arg bound in the active site showed AKG switching to monodentate and catalytically inactive chelation of central iron. The key difference in L-Arg binding between the WT and the U3-26 EFE variant observed from the simulations and induced by the more variable binding of L-Arg is that it changes its positioning slightly away from perfect binding for a key hydrogen abstraction from L-Arg essential in the hydroxylation pathway. In other words, whilst L-Arg binding remains accurate

enough in U3-26-EFE to stabilize correct AKG binding and the formation of the twisted peptide proposed to be essential for both reaction paths, its more flexible binding prevents adequate reactive positioning for hydrogen abstraction within the hydroxylation pathway. This is likely to be the main reason that this variant shifts its reactivity further away from the hydroxylation reaction.

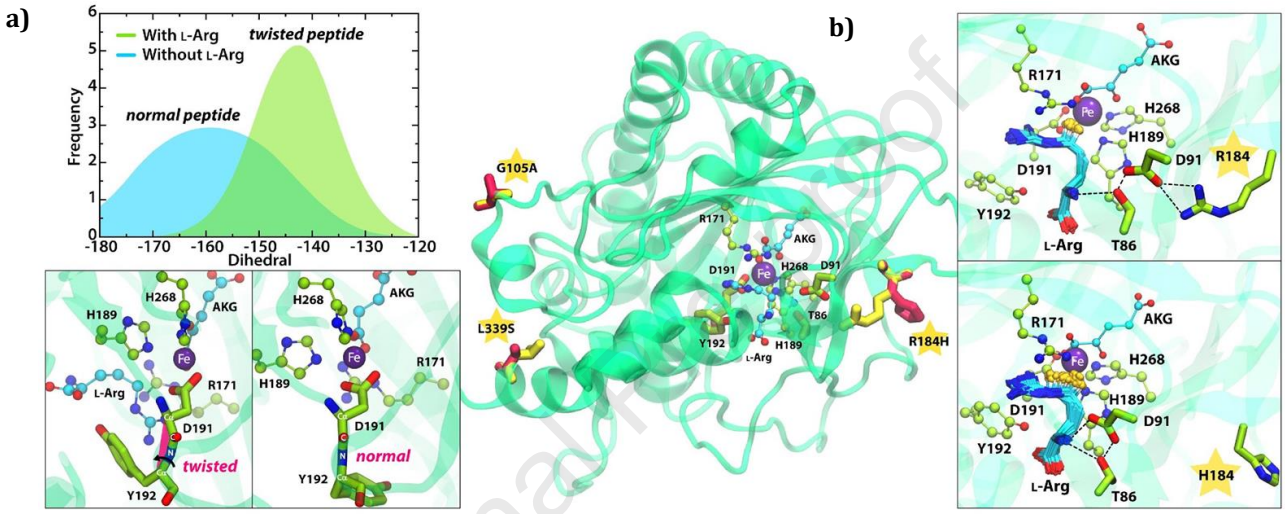

Figure 3: In silico study of the structure of EFE. Structural insights into WT EFE and the triple mutation from U3-26 EFE. Overall structure displaying mutations (yellow stars) and active site (middle); **(a)** comparison of the peptide bond histograms (upper) around D191 and Y192 showing the twisted peptide in the presence (lower left) and normal peptide in the absence (lower right) of bound L-Arg; **(b)** network of hydrogen bonds involving residue 184 and flexibility of L-Arg bound to catalytic site illustrated by overlaid structures from MD simulation of WT (upper) and U3-26 (lower).

Whole genome sequence analysis of both U3-26 EFE and U2-48 EFE identified SNPs in genes linked to nitrogen metabolism, glycerol uptake and the stringent response.

The absence of any individual mutations in the *efe* gene of U2-48 EFE and the subsequent reduction in ethylene productivity, following the expression of the *efe* gene from U3-26 EFE in the Δ*proB* strain suggested that mutations in the genomic background of both strains may be responsible for the increased ethylene productivity in these strains. Genomic DNA was extracted from MG1655, Δ*proB* pBBR1 *PphaC efe*, Δ*proB* pGEM p15 *efe*, U2-48 EFE and U3-26 EFE for Next-Generation Sequencing (NGS). Multiple chromosomal insertions and deletions were detected in the Δ*proB* pBBR1 *PphaC efe*, Δ*proB* pGEM p15 *efe*, U2-48 EFE and U3-26 EFE strains (~100), along with several SNPs. The deletions and insertions ranged in size from 20-150bp with the majority being in hypothetical and uncharacterized proteins, which were conserved across the strains. Both U2-48 and U3-26 had a small number of additional mutations in regulatory and replication genes. These included a 75bp insertion in the heat shock protein (HslR) and a 132bp deletion in the transcriptional regulator *yfjR*, which were both conserved between U2-48 EFE and U3-26 EFE. Furthermore, a small number of specific insertions and deletions were identified in U3-26 EFE, these included a 138bp insertion in *rutR*, the master regulator of pyrimidine metabolism and a 43bp deletion in *rpoN*, which plays a key role on arginine catabolism and is a stress regulator in the cell and a 33bp insertion in the replication gene *dnaE*, a core component of DNA polymerase III. Two conserved SNPs were found in the four strains, which were absent from the MG1655. These SNPs were identified in the genes *glnA* and *ptsP* (Table 3), both of which can be linked to nitrogen metabolism. The *glnA* gene, encodes glutamine synthetase, which catalyzes the conversion of glutamate and ammonia to glutamine[27]. Importantly, *glnA* has long been established to play a key role in nitrogen regulation in the cell[28]. PtsP is a component of the

phosphoenolpyruvate-dependent phosphotransferase system[29, 30], which is responsible for catalyzing the import of sugars subject to regulation by AKG, ensuring carbon and nitrogen uptake are well regulated[28].

Two further SNPs were found in U2-48 EFE, *mscK* and *nagK*, which have links to ammonia metabolism (Table 3). MscK is an ion channel activated by high external concentrations of potassium/ammonium.[31] The *nagK* gene encodes N-acetyl-D-glucosamine kinase, which is involved in the peptidoglycan-recycling pathway.[32] Unlike the other SNPs, the SNP in *nagK* predicted, using Phyre 2.0[33], resulted in a complete loss of function. A reduction in peptidoglycan recycling may have altered the cellular availability of ammonia and sugar, as previously shown under stresses such as carbon, nitrogen, or amino acid starvation.[32]

Four extra SNPs were identified in U3-26 EFE in *crl*, *gatC*, *lacZ* and *glpR* (Table 3). The sigma factor binding protein Crl modulates sigma factor binding with the RNA holoenzyme, particularly σ^{70} and σ^{38} [34]. Importantly, σ^{38} (RpoS) is triggered during the cell's stringent response to stress and has been found to be essential for cell viability. The genes *gatC* and *glpR* both have a role in the PTS system and glycerol utilization. The *gatC* gene, is part of the phosphoenolpyruvate-carbohydrate phosphotransferase system (PTS system), which is involved in the transport and phosphorylation of sugar[35]. The GlpR protein acts as a repressor of genes involved in glycerol 3-phosphate metabolism[36].

Mutated gene	Function	Mutated sequence
mscK	water and ion membrane transporter	1405C>A (Arg469Ser)
nagK	N-acetyl-D-glucosamine kinase	292_293delCGinsAA; 299delT (Val100fs)
ptspP	PEP-protein phosphotransferase enzyme I	758C>A (Ala733fs)
glnA	Glutamine synthetase	550C>T (Pro184Ser)
gatC	PTS system EIIC component	Insertion CC
lacZ	Beta-galactosidase	16delG (Asp6fs); 177_178insA (Arg60fs)
crl	Sigma factor S-binding protein	G>A
glpR	Glycerol-3-phosphate regulon repressor	_>G

Table 3. SNPs identified in the stains Δ*proB* pBBR1 *PphaC efe,* Δ*proB* pGEM p15 *efe,* U248 EFE and U3-26 EFE after NGS. SNPs in both *glnA* and *ptsP* were found in all 4 strains. SNPs in *mscK* and *nagK* were found only in U248 EFE and SNPs in *gatC, lacZ, crl* and *glpR* were only found in U326 EFE.

Metabolomic analysis suggests a common increase in AKG:glutamate concentrations despite distinct patterns of metabolic rewiring in the central carbon metabolism of both the U2-48 EFE and U3-26 EFE strains.

Targeted metabolomics analysis was carried out to try to gain a better insight into the potential metabolic rewiring occurring in the high ethylene yielding strains. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used to selectively measure the levels of 36 metabolites associated with central metabolism in U2-48 EFE, U3-26 EFE, Δ*proB* pBBR1 *PphaC efe*, Δ*proB*_pGEM p15 *efe*, MG1655 *PphaC efe* and MG1655 p15 *efe* in triplicate (Fig. 4a-b)[37]. There was a general increase in the relative abundance of all metabolite concentrations in the Δ*proB*

pBBR1 *PphaC efe* strain compared to MG1655 *PphaC efe* (Fig. 4a). The precursors of the EFE enzyme, AKG and L-Arg, had a $\log_2$-fold increase of 1.4 and 2.7, respectively, although this was not significant. Interestingly, P5C increased, yet the proline concentration decreased, however, these changes were not significant. Compared to MG1655 p15 *efe*, the Δ*proB*_pGEM p15 *efe* strain had minimal changes in glycolysis metabolites, but significant ($p<0.05$) increases in glycerol, citrate, and acetate (Fig. 4a), as well as a significant decrease ($p<0.05$) in ornithine. AKG had a 1.3 $\log_2$-fold increase, whilst L-Arg remained similar, these were not significant, but may limit substrate flux to EFE.

Differences in the metabolite levels were also seen in U2-48 EFE and U3-26 EFE when compared to their respective Δ*proB efe* strains (Fig. 4a-b). Central carbon metabolite levels significantly decreased ($p<0.05$) in U3-26 EFE, whilst proline had a 1.9 $\log_2$-fold increase (Fig. 4a). P5C levels also decreased in U3-26 EFE, although this was not significant. *In vitro* P5C whole cell enzyme assays confirmed that the level of P5C was significantly reduced in U3-26 EFE ($p<0.05$) when compared directly to the MG1655 pGEM p15 *efe* and Δ*proB* pGEM p15 *efe* (Supplementary Fig. 8). The P5C level was also significantly reduced in the U3-26 WT EFE background. U2-48 EFE had significant decreases ($p<0.05$) in glucose, phosphoenol-pyruvate (PEP), glutamine (GLN) and fructose-6-phosphate (F6P), and considerable increases in glycerol, citrate, shikimate, acetyl-AMP and proline ($p<0.05$) (Fig. 4a). Interestingly, the AKG:glutamine ratio increased in both U3-26 EFE and U2-48 EFE. This ratio is important as it regulates central and nitrogen metabolism[38]. Interestingly, L-Arg concentrations decreased 0.62 $\log_2$-fold in U2-48 EFE, whilst increased 1.6 $\log_2$-fold in U3-26 EFE.

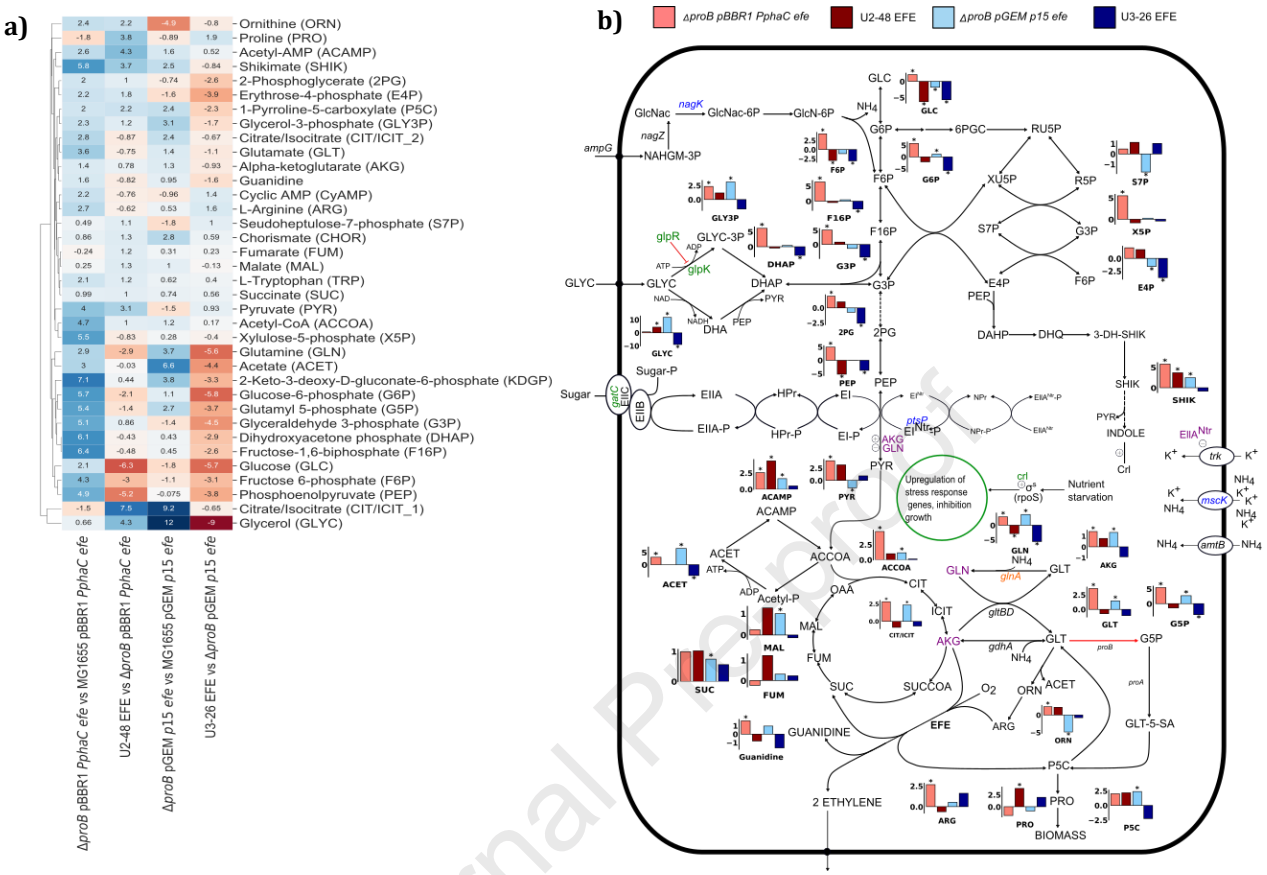

Figure 4: Metabolomics. (a) Heatmap with hierarchical clustering of the rows. Each row represents one of the 36 metabolites in the targeted metabolomics analysis and the columns represent the log2-fold changes between strains. **(b)** Network diagram between EFE, central carbon metabolism and nitrogen assimilation. Hypothetical relationships between the SNPs identified in each of the strains are also presented. SNPs found across all strains are coloured in orange, SNPs unique to U2-48 EFE are coloured in blue and SNPs related to U3-26 EFE are coloured in green. An increase in the ratio between alpha-ketogluatarate (AKG) and glutamine (GLN) is evident across all four strains and pathways or enzymes affected by the ratio variations are also highlighted in orange. Bar plots show changes in metabolite pools for a targeted set of compounds. Concentrations levels that have significantly increased or decreased ($p < 0.05$, t-test)

are indicated with an asterisk. The red bar represents the log_2-fold change between Δ*proB efe* and MG1655, the magenta bar represents the log_2-fold change between U2-48 EFE and Δ*proB*_pBBR1 *PphaC efe*, the green bar represents the log_2-fold change between Δ*proB*_pGEM p15 *efe* andMG1655, and blue represents the log_2-fold change between U3-26 EFE and Δ*proB*_pGEM p15 *efe; mscK*: water and ion membrane transporter; *nagK*: N-acetyl-D-glucosamine kinase; *ptsP*: PEP-protein phosphotransferase enzyme I; *glnA*: glutamine synthetase; *crl*: activator of σ^S-regulated genes; *glpR*: glycerol-3-phosphate regulon repressor; *gatC*: PTS system EIIC component. Metabolite abbreviations not provided in the heatmap: 3-Dehydroshikimate (3-DH-SHIK), 3-Dehydroquinate (DHQ), 3-Deoxy-D-arabino-heptulosonate 7-phosphate (DAHP), Xylulose 5-phosphate (XU5P), Ribose-5-phosphate (R5P), Ribulose 5-phosphate (RU5P), 6-Phospho D-glucono-1,5-lactone (6PGC), N-Acetyl-beta-D-glucosamine-1,6-anhydro-N-acetyl-beta-D-muramate (NAHGM-3P), N-Acetyl-D-glucosamine (GLcNac), N-Acetyl-D-glucosamine 6-phosphate (GlcNac6P), Glucosamine 6-phosphate (GlcN-6P).

In silico flux response analysis provides a mechanistic insight into the causal relationships of the SNPs.

Flux response analysis (FRA) using the GSM was carried out to predict the changes to metabolic fluxes in relation to increased ethylene production (Fig. 5a). Firstly, glycerol utilization via glycerol kinase (GLYK) was predicted as the favorable route for ethylene production. Flux via the alternative route from glycerol to glyceraldehyde-3-phosphate (G3P) and F6P, as well as the pentose phosphate pathway, is found to decrease, which may be the reason for the significant changes in F6P, glucose-6- phosphate (G6P) and erythrose-4-phosphate (E4P) found in the metabolomics data (Fig. 5d-e). Pyruvate kinase (PYK), which forms a complex with *ptsP*, is

bypassed, such that PEP is converted to oxaloacetate (OAA) and may be the reason for the significant changes in PEP levels but not pyruvate (Fig. 5d-e). This bypass from PEP to OAA also allows a reduction in flux through the latter part of the TCA, allowing increased AKG availability to EFE. A shift in glutamate metabolism is also predicted, including increased recycling of P5C towards glutamate, as well as reduced flux through glutamine synthetase (encoded by the mutated gene *glnA*). Changes to glutamate fluxes may be required for generating increased L-Arg levels for EFE, whilst also affecting glutamine concentration for regulating metabolism. The increase in proline concentration in Fig. 4b for U3-26 EFE may therefore be a result of reduced flux from P5C being required for L-Arg biosynthesis. Finally, increased flux through the L-Arg biosynthesis pathway increases acetate production, therefore inducing acetate recycling via acetyl-CoA synthetase, and may be the reason for the significant increase in acetate levels in Δ*proB*_pGEM p15 *efe* and the significant decrease ($p<0.05$) in U3-26 EFE (Fig. 5c, e).

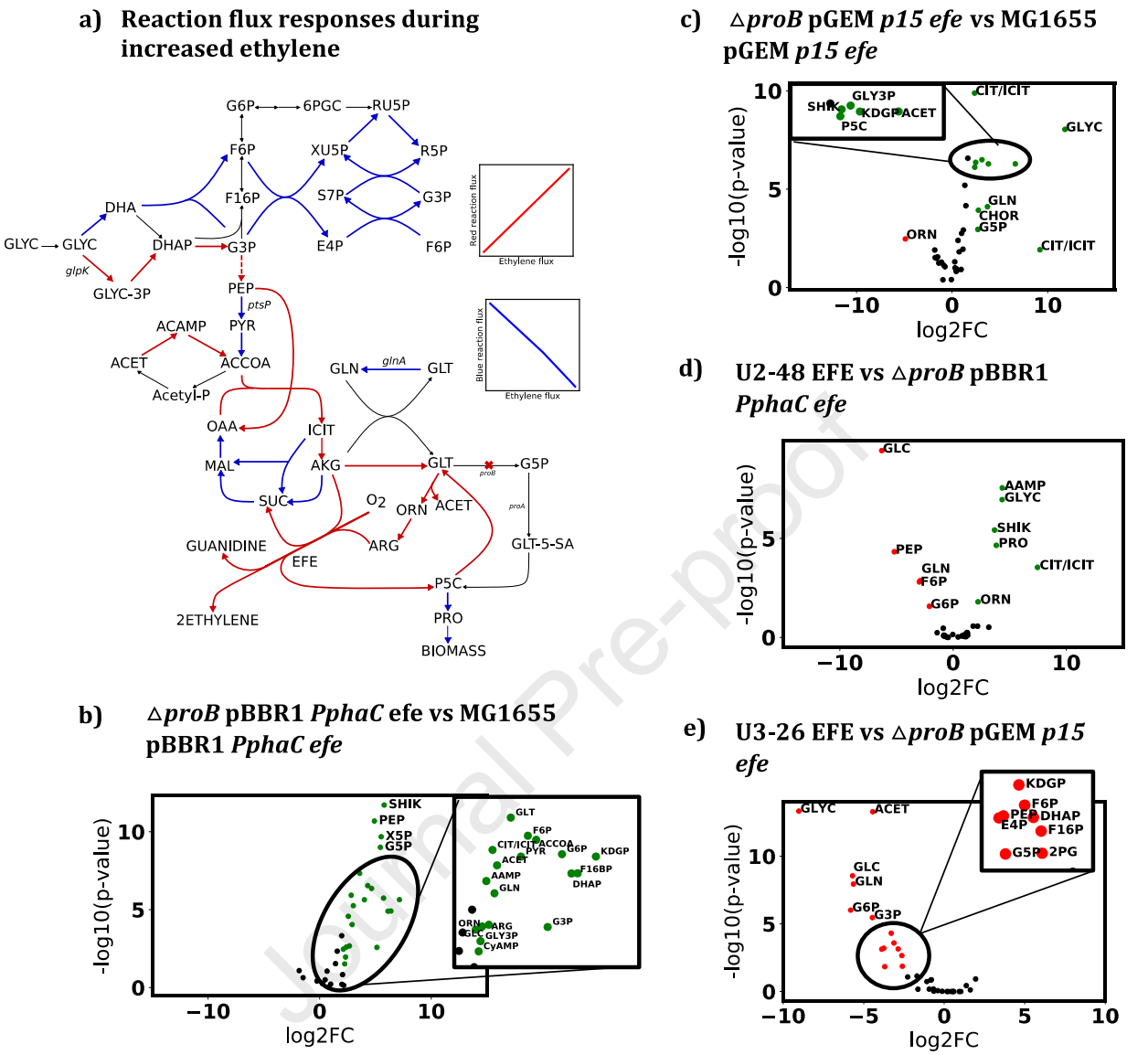

Figure 5: (a) Flux response analysis during the increase of ethylene production using the iJO1366 genome scale model (GSM). Arrows highlighted in red, blue and black indicate reactions whose flux is predicted to increase, decrease or remain unchanged, respectively, with increased ethylene productivity. Volcano plots in **(b)** – **(e)** show significant changes in metabolite concentrations of the evolved strains. The x-axis is the $\log_2$ fold change ($\log_2$-fold change) and the y-axis is the $-\log_{10}$ of the p-value calculated by univariate analysis (t-test). A 2-fold change with a p-value less than 0.05 was used as the cut-off criterion for identifying significantly changing metabolites.

Markers highlighted in green, red and black represent the metabolites whose concentrations significantly increased, significantly decreased or were not significantly changing, respectively. **(b)** metabolite concentrations in the Δ*proB*_pBBR1 *PphaC efe* strain compared to those in the MG1655 pBBR1 *PphaC efe* strain, **(c)** metabolite concentrations in the Δ*proB*pGEM p15 *efe* strain compared to those in MG1655 pGEM p15 *efe* , **(d)** metabolite concentrations in the U2-48 EFE strain compared to the Δ*proB*_pBBR1 *PphaC efe* strain, and **(e)** metabolite concentrations in the U3-26 EFE strain compared to the Δ*proB*_pGEM p15 *efe* strain. Metabolite abbreviations: 1-Pyrroline-5-carboxylate (P5C), 2-Keto-3-deoxy-D-gluconate-6-phosphate (KDPG), 2-Phosphoglycerate (2PG), Acetyl-AMP (AAMP), Acetyl-CoA (ACCOA), alpha ketoglutarate (AKG), Chorismate (CHOR), citrate/isocitrate (CIT/ICIT), Dihydroxyacetone phosphate (DHAP) (DHAP), Erythrose-4-phosphate (E4P), Fructose-1,6-biphosphate (F16BP), Glutamate (GLT), Glutamyl 5-phosphate (G5P), Glyceraldehyde 3-phosphate (G3P), Glycerol (GLYC), Glycerol-3-phosphate (GLY3P), L-Arginine (ARG), L-Tryptophan (TRP), Ornithine (ORN), Pyruvate (PYR), Shikimate (SHIK), Xylulose-5-phosphate (X5P), acetate (ACET), Glucose (GLC), Glucose-6-phosphate (G6P), Phosphoenolpyruvate (PEP), Glutamine (GLU), Proline (PRO), Fructose 6-phosphate (F6P). and Cyclic-AMP (CyAMP).

Discussion

Our study set out to improve ethylene productivity in *E. coli* MG1655, targeting the two major rate limiting steps in ethylene production, solubility of the EFE enzyme and substrate availability of AKG and Arg. Conventional approaches typically target mutagenesis of individual genes and pathways to increase substrate availability and utilize strong promoters and ribosomal binding sites (RBS) to enhance EFE solubility[39]. In this study, we have taken a multidisciplinary approach that

combines mechanistic GSMs to identify candidate growth-coupling strategies, ALE and directed evolution to improve substrate availability via deregulation of the bacteria's metabolism and improved enzyme solubility. The combination of these techniques resulted in a 49-fold improvement in ethylene production, this is the most significant fold-increase reported in the literature to date ($P \leq 0.001$) and substantially better than the 3-4-fold increases reported previously for *E. coli*.[16] (Fig. 6a). Using molecular dynamics and targeted metabolomics analysis we identified point mutations in the EFE enzyme and individual cellular metabolic adaptation strategies, which led to increased ethylene productivity in the best performing strains.

The GSM allowed us to identify and evaluate the efficacy of two growth coupling strategies, involving succinate and proline, which we hoped would lead to an increase in ethylene production. Both strategies were evaluated *in vivo* and interestingly both strategies generated mutants which were unable to generate ethylene but were still able to generate both succinate and P5C respectively, resulting in increased growth rates. This result demonstrates that the growth couple via succinate and proline was successful, in selecting variants with improved growth rates, despite not always resulting in increased ethylene production.

NMR analyses previously demonstrated that P5C formation was much lower than succinate formation, leading the authors to suggest that P5C appears to be substantially uncoupled from succinate formation *in vitro*[7]. However, we have clearly demonstrated that this is not the case *in vivo,* with both P5C and succinate being produced in all the strains generated despite ethylene productivity. Furthermore, these results suggest that the dual circuit mechanism can be uncoupled *in vivo*. It is interesting to note that the succinate couple did not generate a single variant capable

of ethylene production, while the proline couple through P5C, generated strains with both increased, decreased and no ethylene production, suggesting that simply uncoupling or minimizing one side of the reaction may not result in a straightforward increase in ethylene productivity.

Four rounds of ALE allowed both the Δ*proB* pBBR1 and pGEM *efe* strains to evolve and rewire their metabolic networks to increase intracellular metabolite concentrations, which resulted in improved ethylene productivity, regardless of the plasmid copy number (Figure 6b). Genomic sequencing confirmed that both the Δ*proB* pBBR1 *PphaC efe* strain and the Δ*proB* pGEM p15 *efe* had acquired several genetic mutations including a number of genomic insertions and deletions in hypothetical and uncharacterized genes and two SNPs in the genes *glnA* and *ptsP*. GlnA catalyzes the synthesis of glutamine from AKG and plays a key role in the central nitrogen metabolic circuit in the cell[28]. The PTS system is responsible for catalyzing the import of sugars and is subject to regulation by AKG, ensuring carbon and nitrogen uptake are regulated[28]. SNPs in these genes may have subsequently increased the intracellular AKG levels in this strain and increases in the AKG level can also act as a metabolic signal for nitrogen regulation. Changes in *ptsP* may also have affected flux through pyruvate kinase, which was negatively correlated with ethylene production in the flux response analysis. The FRA suggested the PEP carboxylase is utilised, bypassing both pyruvate kinase and pyruvate dehydrogenase, and allowing for a reduced flux through the latter part of the TCA cycle from AKG, which would allow for the increased availability of AKG seen in the Δ*proB efe* strains. A recent study by Durall *et al*., demonstrated that overexpressing PEPc (phosphoenolpyruvate carboxylase) increased the levels of TCA intermediates and subsequently ethylene productivity[39].

Metabolomics analysis on both Δ*proB* pBBR1 *PphaC efe* and the Δ*proB* pGEM p15 *efe* confirmed that the two strains differed considerably, the only difference between the two strains being plasmid copy number and EFE expression levels (Supplementary Fig. 3a-b). The Δ*proB* pGEM p15 *efe* strain had a visible increase in soluble EFE protein, probably linked to the increased expression of the EFE enzyme in this strain. The strain Δ*proB* pBBR1 *PphaC efe*, had decreased levels of proline, whilst all other metabolites generally increased (Fig. 5a-b) coupled with a reduction in the growth rate (Table 1). The Δ*proB*_pGEM p15 *efe* strain, had increased levels of glycerol degradation and the metabolites, citrate, and acetate, whilst many of the other metabolites remained unchanged and the growth was comparable to MG1655 (Table 2).

Random mutagenesis and competitive fermentation generated the strain U2-48 EFE. The U2-48 EFE strain had a wildtype copy of the EFE enzyme and generated 80 nmol/OD/ml of ethylene. U2-48 EFE also had several genomic insertions and deletions, including a 75bp insertion in the heat shock protein HslR, the heat shock response has been shown to upregulate the chaperone network in *E,coli* and 132 bp deletion in the uncharacterized transcriptional regulator, YfjR. These genomic changes may be linked to the increased soluble expression of the EFE enzyme in U2-48. Previous studies have demonstrated that overexpression of the chaperones GroEL/GroES, DnaK/DnaJ/GrpE resulted in increased yields of soluble recombinant protein[40]. However, it is not clear if the insertions and deletions render the proteins inactive, this requires further investigation, which is beyond the scope of the current study. Two extra mutagenic SNPs were also found in U2-48 EFE compared to the Δ*proB* pBBR1 *PphaC efe* strain, these SNPs were present in both *mscK* and *nagK*. The *mscK* gene is involved in ammonia uptake and therefore may have directly affected ammonia availability[31]. The *nagK* gene is involved in peptidoglycan (PG) recycling, which

provides the cell with ammonia and sugars under stresses, such as carbon, nitrogen, or amino acid starvation[32]. Protein structure prediction suggested a complete loss of function in *nagK*, which could affect ammonia availability, whilst affecting the sugar "foraging" response of the cell[32]. This may have increased the AKG:glutamine ratio, the established signal for nitrogen limitation in enteric bacteria[28, 41]. Subsequent decreases in the relative metabolite levels of glucose, fructose-6-phosphate and glucose-6-phosphate support this theory. Importantly, a shift in equilibrium of the AKG:glutamine ratio may have enabled altered flux though carbon and nitrogen metabolism, subsequently increasing the available AKG pool, and in turn, ethylene production. Flux response analysis using the GSM also suggested changes to glutamate metabolism are necessary for increased arginine availability to EFE.

The solubility of the EFE protein has been shown to be a significant bottleneck in ethylene productivity, with the vast majority of the EFE protein being in the insoluble fraction in the cell[11]. We generated an eqPCR library to generate EFE variants with improved solubility, previous attempts to generate libraries for EFE have failed, due to the need for a high throughput screen to select improved variants. We utilized both the succinate and proline growth couples to select for variants with improved solubility and productivity. However, only the proline couple resulted in the identification of a single strain with increased ethylene productivity, and protein solubility, U3-26 EFE, which produced 600nmol/OD/ml of ethylene, the most significant amount reported in *E. coli* to date. Several of the other U3 strains were unable to produce ethylene, (Supplementary Fig. 5c). Sequencing of the variant EFE enzymes from these U3 strains confirmed they were all truncated versions of the EFE enzyme, which surprisingly despite having some severe truncations still retained some functionality, given the strains could still produce P5C and succinate.

Sequencing of the *efe* gene confirmed that U3-26 EFE had five SNP's. Computational structural insights from MD simulations comparing WT EFE with U3-26 EFE demonstrated that these mutations did influence the active site, even though their positions were distant to the catalytic site. A tight bidentate binding of AKG to the metal center and the peptide twist caused by the presence of L-Arg have been proposed to be important for ethylene production in the major catalytic pathway of EFE. However, despite the benefit of the mutual binding of two substrates, there exists a certain level of competition between major and minor reactions. Therefore, slight changes in the binding pattern in the active site might influence the outcome of the balance between the promiscuous reaction in EFE. The simulations presented here show that the R184H mutation in U3-26 EFE changes a crucial H-bond network between the R184 with L-Arg through T86 and D91 altering the binding of L-Arg. This implies that the slight disruption of L-Arg interactions with the active site in U3-26 EFE could change the catalytic reaction profile in favor of the ethylene production route, since L-Arg will be in a less favorable position for hydroxylation, which has been observed experimentally[7, 8, 42].

Increased ethylene productivity could not be replicated through the addition of the U3-26 (*efe*) enzyme to other strains, including both the wildtype strain MG1655 and the Δ*proB* strain. Furthermore, subsequent reintroduction of the individual and combined SNPs also failed to increase productivity. Suggesting that improved solubility/catalytic activity of the EFE enzyme was not solely responsible for the increased ethylene production in U3-26 EFE. Whole genome sequencing of the U3-26 EFE strain revealed several genomic insertions and deletions including the 75bp insertion in the heat shock protein HslR and the 132bp deletion in the transcriptional

regulator YfjR. Furthermore, a 33bp insertion was identified in the catalytic subunit of the DNA polymerase III enzyme, DnaE, a complex, multi-chain enzyme, which is responsible for most of the replicative DNA synthesis in the cell. A 43bp deletion was also identified in the regulator of arginine catabolism and the stress response, sigma factor RpoN. Importantly, the stress response has also been shown to play a key role in increased chaperone expression. These mutations may account for the increased EFE solubility in U3-26 EFE. Interestingly, expression of the EFE (wt) enzyme in the U3-26 background also resulted in an increase in soluble protein, which confirmed that the genomic background of this strain had evolved enabling increased EFE solubility. Three extra SNPs were also identified in U3-26 EFE in *crl*, *gatC*, and *glpR*. Both *gatC* and *glpR* have roles in the PTS system and glycerol utilization, respectively[35,36]. Crl increases the competitiveness of RpoS[34]. RpoS is essential for cell viability under non-optimal growth conditions, but it can, however, impede bacterial growth in the absence of stress[34].

The metabolomic analysis demonstrated that the L-Arg concentration had increased in U3-26 EFE, while the level of AKG decreased, although not significant, this supports the idea that the stoichiometry of the minor and major reactions of the EFE enzyme has been altered in U3-26 EFE. Acetate is produced as a by-product of L-Arg biosynthesis, therefore the decrease in acetate concentration in U3-26 EFE also supports reduced L-Arg demand. Furthermore, metabolomic analysis confirmed there was a reduction in both P5C and guanidine in U3-26 EFE, which supports the notion that L-Arg is in a much less favorable position for hydroxylation in the minor reaction. The *in vitro* P5C assays also demonstrated that there was a statistically significant reduction in P5C in U3-26 EFE compared to MG1655 pGEM p15 *efe* and Δ*proB*_pGEM p15 *efe* ($P \leq 0.05$), supporting the notion that the SNPs in U3-26 EFE have altered the stoichiometric balance of the

ethylene forming reaction in favor of the major reaction (Supplementary Fig. 8). Interestingly expression of the wildtype EFE enzyme in the U3-26 background also resulted in a significant decrease in P5C *in vitro* compared to MG1655 pGEM p15 *efe* and Δ*proB*_pGEM p15 *efe* ($P \leq 0.05$). This could be due to substrate availability in the U3-26 background. Furthermore, the proline concentration surprisingly increased in U3-26 EFE despite the reduction in P5C, however the flux response analysis predicts that P5C is directed towards glutamate thus increasing L-Arg biosynthesis and therefore available flux towards proline may have increased with reduced L-Arg demand.

Additionally, shikimate levels were significantly increased ($p<0.05$) in all the strains, except for U3-26 EFE, and therefore this may play a role in cellular growth regulation. Furthermore, GlpR, regulates glycerol degradation via glycerol-3-phosphate kinase. The flux response analysis found increased flux via glycerol-3-phosphate kinase was favorable for ethylene production. Coupled with the significant decreases ($p<0.05$) seen in the metabolite levels of both glycerol and glycerol-3-phosphate, this suggests *glpR* mutations have improved glycerol utilization inU3-26 EFE.

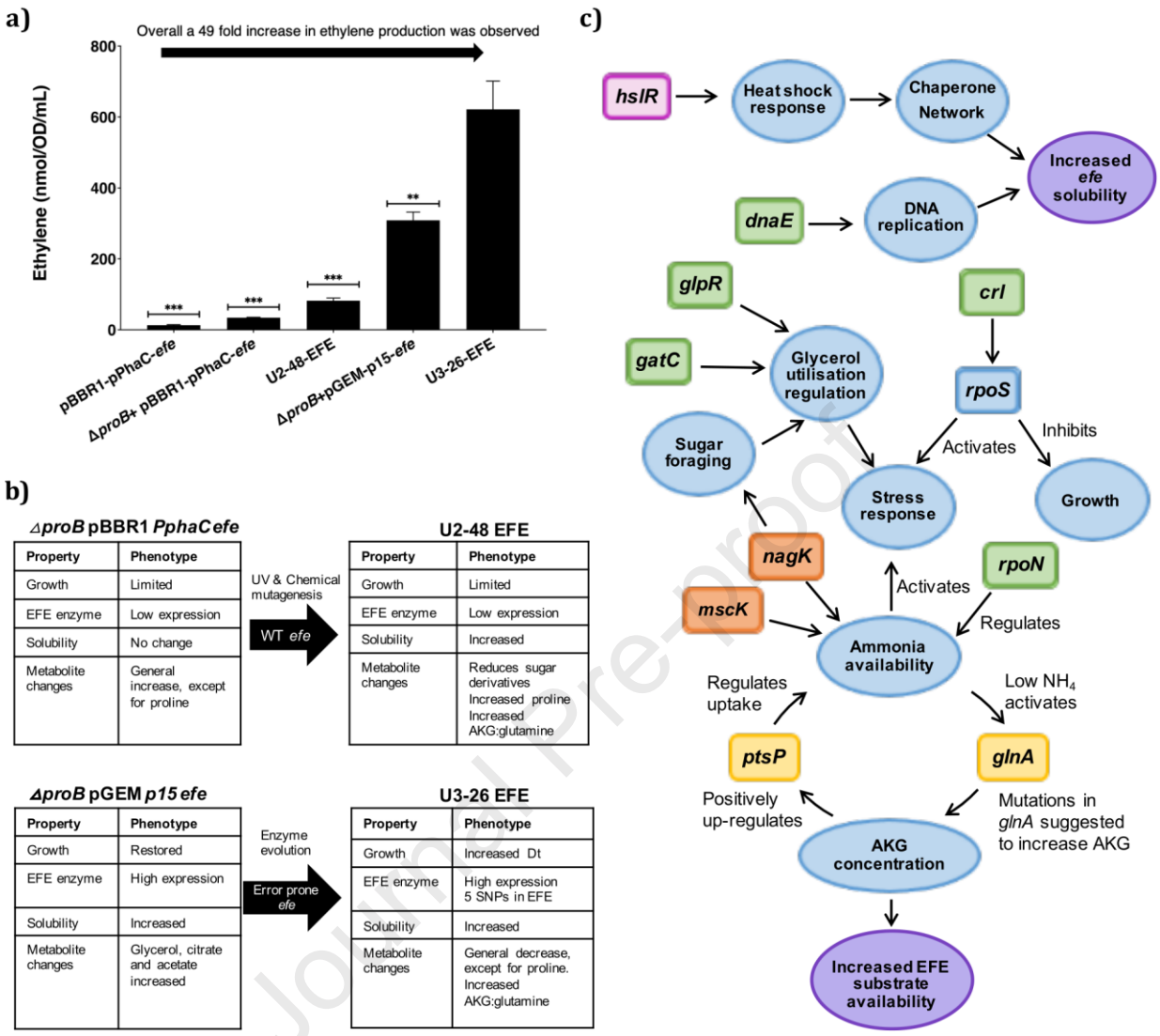

Figure 6. Conclusion. (a) Stepwise improvement in ethylene productivity along the engineering process of MG1655 expressing the *efe* gene. Promoter optimisation, growth coupling, adaptation of the origin of replication, random mutagenesis and adaptive evolution in fermentation resulted in 49-fold increase. (b) Tables showing the phenotypes of the strains Δ*proB* pBBR1 PphaC *efe* (top left), U2-48 EFE (top right), Δ*proB* pGEM p15 *efe* (bottom left) and U3-26 EFE (bottom right). Dn: double time. **(c)** Overview showing the suggested metabolic adaptions inferred from the genome sequencing analysis, metabolomics and flux response analysis for the evolved strains

for increasing *efe* solubility and substrate availability. *mscK*: water and ion membrane transporter; *nagK*: N-acetyl-D-glucosamine kinase; *ptsP*: PEP-protein phosphotransferase enzyme I; *glnA*: glutamine synthetase; *crl*: activator of σS-regulated genes; *glpR*: glycerol-3-phosphate regulon repressor; *gatC*: PTS system EIIC component; *dnaE*: DNA polymerase; *rpoN*: RNA polymerase sigma factor; *hslR*: heat shock protein. Genes highlighted in yellow were identified in all four strains, genes highlighted in orange were identified in the U2-48 EFE, genes highlighted in green were identified in U3-26 EFE and genes highlighted in pink were identified in U2-48 and U3-26.

Conclusion

This study clearly demonstrated that genetic rewiring of central carbon metabolism can be achieved through a combination of genome scale metabolic modelling, ALE and directed evolution. The strains generated through this study all adapted and evolved to channel flux to the EFE enzyme to maintain growth (Fig 6b, c). Interestingly each of the four strains Δ*proB*_pBBR1 *PphaC efe*, Δ*proB*_pGEM p15 *efe*, U2-48 EFE and U3-26 EFE had a different metabolic strategy to improve growth and subsequently ethylene productivity and this strategy was directly coupled to the EFE enzyme. Subsequent uncoupling of the EFE enzyme from its designated strain resulted in a decrease in ethylene productivity. Thus, the close interplay between enzyme function and strain evolution can be utilized to generate enhanced strains with improved flux and ethylene production.

Materials and Methods

Genome scale model modifications and constraints

The *Escherichia coli* strain. K-12 MG1655 model iJO13666[19] was obtained from the BiGG database[43] via the Computer Assisted Metabolic Engineering and Optimization (CAMEO) toolbox[44] v0.11.15. Model iJO1366 was modified to include the reaction associated with the ethylene-forming enzyme (EFE) was added to the model, together with transporter reactions for ethylene and guanidine. The reaction N-acetylornithine deacetylase, encoded by the gene *argE* (b3957), was also constrained to zero as there is no experimental evidence to suggest this reaction carries flux in *E. coli*[45].

All simulations were carried out using a maximum glycerol or glucose uptake rate of 5.5 and 3.15 mmol $gDCW^{-1}h^{-1}$, respectively. which corresponds to the experimental growth rate of 0.3 h^{-1} in the wild type. A maximum uptake of 18.5 mmol $gDCW^{-1}h^{-1}$ of oxygen was also set in the model. All other M9 minimal medium compounds could freely enter the system. Cplex v12.9.0.0 was used as the solver for all GSM simulations.

In silico knockout analysis

Single reaction knockouts were simulated in the model in a sequential manner by constraining each reaction bounds to zero. Fluxes were determined for the *in silico* mutant strains using two approaches from CAMEO: parsimonious flux balance analysis (pFBA) and minimisation of metabolic adjustments (MOMA). pFBA identifies the optimal solution in the mutant that maximises growth rate, whilst also minimising the total sum of flux[46]. Knockouts that result in

ethylene production using this approach are predicted growth coupling candidates. MOMA, however, was also used to search for alternative strategies for redirecting flux towards ethylene, which finds the solution that minimises the Euclidean distance between a reference metabolic state and the mutant's flux distribution[20]. We used the pFBA solution of the iJO1366-efe model as the reference metabolic state.

In silico Flux response analysis

Flux response analysis[47, 48] was carried out to predict changes to central carbon metabolic fluxes to increased ethylene productivity. The maximum (0.57 mmol/gDCW/h) flux through the EFE reaction was calculated for the *proB*-iJO1366-efe model, whilst biomass was constrained between 0.2 h-1 and 0.3 h^{-1} in accordance with the experimentally determined values of the evolved strains. We then ran FBA repeatedly on the model using maximisation of the growth rate as the objective, whilst incrementally increasing the EFE flux bounds from the EFE value found by pFBA (0.06 mmol/gDCW.h) to the maximum EFE value. The flux profiles of each reaction to increasing ethylene production were generated by plotting reaction flux (y-axis) against the perturbed EFE reaction flux (x-axis).

For full experimental details, see *SI Appendix*.

Supplementary Information is linked to the on-line version of the paper.

Acknowledgments

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC; grant number BB/L013940/1) and the Engineering and Physical Sciences Research

Council (EPSRC) under the same grant number and the Green Chemicals Beacon of Excellence, University of Nottingham.

Author contributions

SJB, NP, SV, AC, NM, SL, PCM, JY and CE conceived the project. NP did the modelling and computational analysis. SJB, AVH, MJ and SV generated all the constructs. SJB, AVH and SV performed the ethylene measurements. SV, ME, AVH, RRB performed the fermentation. MH and CJ performed the structural analysis on EFE. PD, PYC and AVH performed the error prone PCR. SL performed the mutagenesis and error prone PCR for the *sucA/serA* growth couple. SA and DK performed the LC-MS/MS. SA, NP, and TM performed the metabolic analysis. SJB, SV, NP, MH, CJ, PCM, JY, SL and CE wrote the paper with contributions from all authors.

.

Author Information

 Correspondence and requests for materials should be addressed to Samantha.bryan@nottingham.ac.uk.

References

1. Eckert, C. et al. Ethylene-forming enzyme and bioethylene production. *Biotechnol Biofuels* **7**, 33 (2014).
2. Zhao, Z., et al. Low-carbon roadmap of chemical production: A case study of ethylene in China. *Renewable and Sustainable Energy Reviews* **97(C)**, 580-591 (2018).
3. Markham, J.N. et al. Techno-economic analysis of a conceptual biofuel production process from bioethylene produced by photosynthetic recombinant cyanobacteria. *Green Chemistry* **18**, 6266-6281 (2016).
4. Goto, M., Ishida, Y., Takikawa, Y. & Hyodo, H. Ethylene production by the Kudzu strains of *Pseudomonas syringae* pv. *phaseolicola* causing Halo Blight in *Pueraria lobata* (Willd) Ohwi. *Plant and Cell Physiology* **26**, 141-150 (1985).

5. Goto, M.H., H. Ethylene production by cell-free extract of the Kudzu strain of *Pseudomonas syringae* pv. *phaseolicola*. *Plant and Cell Physiology* **28**, 405-414 (1987).
6. Fukuda, H. et al. Two reactions are simultaneously catalyzed by a single enzyme: the arginine-dependent simultaneous formation of two products, ethylene and succinate, from 2-oxoglutarate by an enzyme from Pseudomonas syringae. *Biochem Biophys Res Commun* **188**, 483-489 (1992).
7. Zhang, Z. et al. Structural and stereoelectronic insights into oxygenase-catalyzed formation of ethylene from 2-oxoglutarate. *Proc Natl Acad Sci U S A* **114**, 4667-4672 (2017).
8. Martinez, S. et al. Structures and Mechanisms of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme: Substrate Binding Creates a Twist. *J Am Chem Soc* **139**, 11980-11988 (2017).
9. Pirkov, I., Albers, E., Norbeck, J. & Larsson, C. Ethylene production by metabolic engineering of the yeast Saccharomyces cerevisiae. *Metab Eng* **10**, 276-280 (2008).
10. Johansson, N., Persson, K.O., Larsson, C. & Norbeck, J. Comparative sequence analysis and mutagenesis of ethylene forming enzyme (EFE) 2-oxoglutarate/Fe(II)-dependent dioxygenase homologs. *BMC Biochem* **15**, 22 (2014).
11. Lynch, S., Eckert, C., Yu, J., Gill, R. & Maness, P.C. Overcoming substrate limitations for improved production of ethylene in E. coli. *Biotechnol Biofuels* **9**, 3 (2016).
12. Sakai, M., Ogawa, T., Matsuoka, M., & Fukuda, H. Sakai, Miho, et al. "Photosynthetic conversion of carbon dioxide to ethylene by the recombinant cyanobacterium, *Synechococcus sp.* PCC 7942, which harbors a gene for the ethylene-forming enzyme of *Pseudomonas syringae*. *Journal of fermentation and bioengineering* **84**, 434-443.
13. Veetil, V.P., Angermayr, S.A. & Hellingwerf, K.J. Ethylene production with engineered Synechocystis sp PCC 6803 strains. *Microb Cell Fact* **16**, 34 (2017).
14. Durall, C., Lindberg, P., Yu, J. & Lindblad, P. Increased ethylene production by overexpressing phosphoenolpyruvate carboxylase in the cyanobacterium Synechocystis PCC 6803. *Biotechnol Biofuels* **13**, 16 (2020).
15. Tao, L., Dong, H.J., Chen, X., Chen, S.F. & Wang, T.H. Expression of ethylene-forming enzyme (EFE) of Pseudomonas syringae pv. glycinea in Trichoderma viride. *Appl Microbiol Biotechnol* **80**, 573-578 (2008).
16. Digiacomo, F. et al. Ethylene-producing bacteria that ripen fruit. *ACS synthetic biology* **3**, 935-938 (2014).
17. von Kamp, A. & Klamt, S. Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nat Commun* **8**, 15956 (2017).
18. Orth, J.D., Thiele, I. & Palsson, B.O. What is flux balance analysis? *Nat Biotechnol* **28**, 245-248 (2010).
19. Orth, J.D. et al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Mol Syst Biol* **7**, 535 (2011).
20. Segre, D., Vitkup, D. & Church, G.M. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* **99**, 15112-15117 (2002).
21. Yu, B.J. et al. sucAB and sucCD are mutually essential genes in Escherichia coli. *FEMS Microbiol Lett* **254**, 245-250 (2006).
22. Joyce, A.R. et al. Experimental and computational assessment of conditionally essential genes in Escherichia coli. *J Bacteriol* **188**, 8259-8271 (2006).

23. Kim, J. et al. Hidden resources in the Escherichia coli genome restore PLP synthesis and robust growth after deletion of the essential gene pdxB. *Proc Natl Acad Sci U S A* **116**, 24164-24173 (2019).
24. Zhao, G. & Winkler, M.E. A novel alpha-ketoglutarate reductase activity of the serA-encoded 3-phosphoglycerate dehydrogenase of Escherichia coli K-12 and its possible implications for human 2-hydroxyglutaric aciduria. *J Bacteriol* **178**, 232-239 (1996).
25. Fukui, T., Ohsawa, K., Mifune, J., Orita, I. & Nakamura, S. Evaluation of promoters for gene expression in polyhydroxyalkanoate-producing Cupriavidus necator H16. *Appl Microbiol Biotechnol* **89**, 1527-1536 (2011).
26. Alagesan, S. et al. Functional Genetic Elements for Controlling Gene Expression in Cupriavidus necator H16. *Appl Environ Microbiol* **84** (2018).
27. Kumar, R. & Shimizu, K. Metabolic regulation of Escherichia coli and its gdhA, glnL, gltB, D mutants under different carbon and nitrogen limitations in the continuous culture. *Microb Cell Fact* **9**, 8 (2010).
28. Doucette, C.D., Schwab, D.J., Wingreen, N.S. & Rabinowitz, J.D. alpha-Ketoglutarate coordinates carbon and nitrogen utilization via enzyme I inhibition. *Nat Chem Biol* **7**, 894-901 (2011).
29. Pfluger-Grau, K. & Gorke, B. Regulatory roles of the bacterial nitrogen-related phosphotransferase system. *Trends Microbiol* **18**, 205-214 (2010).
30. Gosset, G. Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. *Microb Cell Fact* **4**, 14 (2005).
31. Martinac, B., Saimi, Y. & Kung, C. Ion channels in microbes. *Physiol Rev* **88**, 1449-1490 (2008).
32. Uehara, T. & Park, J.T. The N-acetyl-D-glucosamine kinase of Escherichia coli and its role in murein recycling. *J Bacteriol* **186**, 7273-7279 (2004).
33. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858 (2015).
34. Typas, A., Barembruch, C., Possling, A. & Hengge, R. Stationary phase reorganisation of the Escherichia coli transcription machinery by Crl protein, a fine-tuner of sigmas activity and levels. *EMBO J* **26**, 1569-1578 (2007).
35. Meza, E., Becker, J., Bolivar, F., Gosset, G. & Wittmann, C. Consequences of phosphoenolpyruvate:sugar phosphotranferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in Escherichia coli. *Microb Cell Fact* **11**, 127 (2012).
36. Eppler, T., Postma, P., Schutz, A., Volker, U. & Boos, W. Glycerol-3-phosphate-induced catabolite repression in Escherichia coli. *J Bacteriol* **184**, 3044-3052 (2002).
37. Schatschneider, S. et al. Quantitative Isotope-Dilution High-Resolution-Mass-Spectrometry Analysis of Multiple Intracellular Metabolites in Clostridium autoethanogenum with Uniformly (13)C-Labeled Standards Derived from Spirulina. *Anal Chem* **90**, 4470-4477 (2018).
38. Xiao, D. et al. The glutamine-alpha-ketoglutarate (AKG) metabolism and its nutritional implications. *Amino Acids* **48**, 2067-2080 (2016).
39. Durall, C., Lindberg, P., Yu, J. & Lindblad, P. Increased ethylene production by overexpressing phosphoenolpyruvate carboxylase in the cyanobacterium Synechocystis PCC 6803. *Biotechnology for biofuels* **13**, 16 (2020).

40. de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. & Bukau, B. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. *BMC biotechnology* **7**, 1-9 (2007).
41. Huergo, L.F. & Dixon, R. The Emergence of 2-Oxoglutarate as a Master Regulator Metabolite. *Microbiol Mol Biol Rev* **79**, 419-435 (2015).
42. Martinez-Gomez, K. et al. New insights into Escherichia coli metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. *Microb Cell Fact* **11**, 46 (2012).
43. Norsigian, C.J. et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res* **48**, D402-D406 (2020).
44. Cardoso, J.G.R. et al. Cameo: A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories. *ACS Synth Biol* **7**, 1163-1166 (2018).
45. Javid-Majd, F. & Blanchard, J.S. Mechanistic analysis of the argE-encoded N-acetylornithine deacetylase. *Biochemistry* **39**, 1285-1293 (2000).
46. Lewis, N.E. et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* **6**, 390 (2010).
47. Lee, K.H., Park, J.H., Kim, T.Y., Kim, H.U. & Lee, S.Y. Systems metabolic engineering of Escherichia coli for L-threonine production. *Mol Syst Biol* **3**, 149 (2007).
48. Poolman, M.G., Kundu, S., Shaw, R. & Fell, D.A. Responses to light intensity in a genome-scale model of rice metabolism. *Plant Physiol* **162**, 1060-1072 (2013).

Table 1

Strain	MG1655	MG1655 pBBR1 *PphaC efe*	MG1655 pGEM p15 *efe*	Δ*proB* pBBR1 *PphaC efe*	Δ*proB* pGEM p15 *efe*
µmax (h^{-1})	0.31	0.28	0.31	0.2	0.28
Dt (h)	2.236	2.478	2.236	3.466	2.476

Table 1: Growth rates and doubling times for MG1655 (WT), MG1655 pBBR1 *PphaC efe* and pGEM p15 *efe* and Δ*proB* pBBR1 *PphaC efe* and Δ*proB* pGEM p15 *efe*.

Table 2

Strain	MG1655	MG1655 pBBR1 *PphaC efe*	MG1655 pGEM p15 *efe*	Δ*proB* pBBR1 *PphaC efe*	Δ*proB* pGEM p15 *efe*	U2-48 EFE	U3-26 EFE
µmax (h^{-1})	0.31	0.28	0.31	0.2	0.28	0.2	0.24
Dt (h)	2.236	2.478	2.236	3.466	2.476	3.466	2.888

Table 2: Growth rates and doubling times comparison for MG1655 (WT), MG1655 pBBR1 *PphaC efe*, MG1655 pGEM p15 *efe*, Δ*proB* pBBR1 *PphaC efe*, Δ*proB* pGEM p15 *efe*, U2-48 EFE and U3-26 EFE.

Table 3

Mutated gene	Function	Mutated Sequence
mscK	Water and ion membrane transporter	1405C>A (Arg469Ser)
nagK	N-acetyl-D-glucosamine kinase	292_293delCGinsAA;299delT (Val100fs)
ptsP	PEP-protein phosphotransferase enzyme I	758>A (Ala733fs)
glnA	Glutamine synthetase	550C>T (Pro184Ser)
gatC	PTS system EIIC component	Insertion CC
lacZ	Beta-galactosidase	16delG (Asp6fs);177_178insA (Arg60fs)
crl	Sigma factor S-binding protein	G>A
glpR	Glycerol-3-phosphate regulon repressor	_>G

Table 3. SNPs identified in the stains Δ*proB* pBBR1 *PphaC efe,* Δ*proB* pGEM p15 *efe*, U248 EFE and U3-26 EFE after NGS. SNPs in both *glnA* and *ptsP* were found in all 4 strains. SNPs in *mscK* and *nagK* were found only in U2-48 EFE and SNPs in *gatC, lacZ, crl* and *glpR* were only found in U3-26 EFE.

- First example of growth coupled ethylene production.
- Increased ethylene productivity in *E. coli* 49-fold.
- Increased ethylene forming enzyme (EFE) solubility *in vivo*.
- Distal mutations in the EFE enzyme influence the active site.