Genome-scale model of *C. autoethanogenum* reveals optimal bioprocess conditions for high-value chemical production from carbon monoxide

Rupert O.J. Norman1,2✉, Thomas Millat3, Sarah Schatschneider1,3, Anne M. Henstra1, Ronja Breitkopf4, Bart Pander1, Florence J. Annan1, Pawel Piatak1, Hassan B. Hartman4,5, Mark G. Poolman4, David A. Fell4, Klaus Winzer1, Nigel P. Minton1, Charlie Hodgman1,2

1Synthetic Biology Research Centre, University of Nottingham, University Park, Nottingham NG7 2RD, UK
2School of Biosciences, University of Nottingham, Sutton Bonington Campus, Sutton Bonington, Leicestershire LE12 5RD, UK
3Evonik Nutrition and Care GmbH, Kantstr. 2, 33798 Halle-Kinsbeck, Germany
4Department of Biology and Medical Sciences, Oxford Brooks University, Oxford OX3 0BP, UK
5Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK
✉E-mail: rupert.norman@nottingham.ac.uk

Abstract: *Clostridium autoethanogenum* is an industrial microbe used for the commercial-scale production of ethanol from carbon monoxide. While significant progress has been made in the attempted diversification of this bioprocess, further improvements are desirable, particularly in the formation of the high-value platform chemicals such as 2,3-butanediol (2,3-BD). A new, experimentally parameterised genome-scale model of *C. autoethanogenum* predicts dramatically increased 2,3-BD production under non-carbon-limited conditions when thermodynamic constraints on hydrogen production are considered.

1 Introduction

The adverse environmental and societal consequences of continued fossil-fuel dependence represent arguably the defining challenge for scientific research in the 21st century [1, 2]. One component of the solution to this problem is carbon recycling, which can be achieved through the application of bioprocesses enabling conversion of industrial waste-gas components into commodity chemicals [3]. For example, the Chicago-based company LanzaTech (www.lanzatech.com) uses *Clostridium autoethanogenum* to convert carbon monoxide (CO) into ethanol on a commercial scale [4–6]. Diversification of the product portfolio of this sustainable technology will help to secure its continued success.

The fundamental metabolic pathway enabling *C. autoethanogenum* to grow on CO is the Wood–Ljungdahl pathway (WLP) [7, 8]. This ancient biochemical pathway is split into two ‘branches’: the methyl branch and the carbonyl branch [9]. The methyl branch proceeds by constructing a methyl group from carbon dioxide (CO2) via a series of biochemical conversions including an adenine triphosphate (ATP)-consuming reaction [catalysed by formyl-tetrahydrofolate (THF) ligase, FtfL] and three redox reactions collectively requiring nicotinamide adenine dinucleotide (NADH), NAD phosphate (NADPH) and reduced ferredoxin (Fdred). The carbonyl branch is simply the reduction of CO2 to an enzyme-bound carbonyl group as catalysed by the acetyl-CoA synthase/CO dehydrogenase complex (ACS/CODH). This same enzyme complex is responsible for the final step of the pathway, in which the methyl group and enzyme-bound carbonyl group are combined with COA to form one molecule of acetyl-CoA.

Clearly, in the case of CO-fed growth, sources of CO2 reducing power and the means to generate ATP are needed for the WLP to function. As with all known acetogens [10], acetate and ATP are produced from acetyl-CoA via phosphotransacetylase and acetate kinase, thus fulfilling the energetic requirement of FtfL but not enabling a positive ATP yield [5]. The CO2 and reducing power requirements are met by two additional monofunctional CODH enzymes encoded in the *C. autoethanogenum* genome which generate Fdred and CO2 from CO (and water). Fdred is then oxidised to generate NADH and subsequently NADPH, which in turn satisfy the redox requirements of the WLP. Finally, electrons are transferred from Fdred to NAD by a membrane-bound Rnf complex [11], which produces a transmembrane proton gradient and enable the generation of ATP by F1F0 ATP synthase [12, 13]. The regeneration of NAD for this essential process is achieved through further redox reactions leading to excreted by-products, e.g. ethanol [5].

One such product of particular importance is the platform chemical 2,3-butanediol (2,3-BD) – a trace component of the native *C. autoethanogenum* product profile, the downstream products of which have an estimated global market value of $43 billion in sales [14]. Other compounds in the native product profile of *C. autoethanogenum* include acetate, ethanol and, in smaller amounts, lactate [14, 15]. Hydrogen gas is another common product in carboxydrotrophic microorganisms [16]. While the potential for 2,3-BD synthesis from industrial waste gas is well-established [14], a clear bioprocess/metabolic engineering strategy for its increased production has not been published despite the construction and analysis of a genome-scale model (GSM) of *C. autoethanogenum* [17] – an approach which has achieved success in guiding metabolic engineering through the investigation of optimal steady states [18]. Previous attempts to model *C. autoethanogenum* use parameters from GSMS of other organisms [17, 19, 20], a potential weakness identified by Dash et al. [21]. The availability of a new, manually annotated genome sequence of *C. autoethanogenum* [22], therefore, provides an opportunity to construct an improved, experimentally parameterised GSM and pursue novel strategies for enhanced 2,3-BD production.

Experimental studies aimed at manipulating the product profile of *C. autoethanogenum* have so far focused on carbon-limited growth regimes [17, 19], i.e. the scenario, in which rate of growth of a bacterial population is limited by the supply of its carbon source. Such studies tend to neglect non-carbon limitation – the case where some other nutrient becomes limiting and the carbon source is in excess. We argue that non-carbon-limited growth
Table 1  Genome sequences for Clostridia spp. referred to throughout this work

| Organism       | Year | Reference | Annotation | Structure      |
|----------------|------|-----------|------------|----------------|
| C. autoethanogenum | 2013 | [27]      | automated  | 100 contigs    |
|                | 2014 | [24]      | automated  | whole genome   |
|                | 2015 | [22]      | manual     | whole genome   |
| C. ljungdahlii  | 2010 | [28]      | automated  | whole genome   |

regimes are worth investigating because an excess supply of energy (i.e. electrons from CO) may result in the diversion of flux toward increasingly reduced compounds (e.g. 2,3-BD), especially in strict anaerobes such as C. autoethanogenum whose product profile is constrained by the redox balance of internal metabolites [23].

This work presents the construction of a new GSM of C. autoethanogenum ("MetaCLAU") including parameterisation and validation with data from continuous-culture chemostat experiments. The subsequent model is used to test the hypotheses that 2,3-BD is an optimal sink for excess reducing power supplied under non-carbon-limited conditions before and after a thermodynamic limit on hydrogen production is imposed.

2 Methods

2.1 Data

This section describes the data used for model construction, parameterisation and validation. Data used in the interpretation of model analyses are also specified.

2.1.1 Genome sequences: The task of constructing a GSM requires an annotated genome sequence. Table 1 summarises relevant genomic data available for use in this work. The ScrumPy-readable pathway genome database (PGDB) generated from the genome sequence described in [24] is available from www.metacyc.org. Flatfiles for the PGDB derived from the genome annotation described in [22] were generated with Pathway Tools [25] and can be found in [26].

2.1.2 Fermentation data: Optical density (OD), gas chromatography and cell-mass data collected at steady state during fermentation experiments were used to derive model parameters in this work. Bioreactor data were collected and processed using in-house software (see 2.7). Details of fermentation experiments are provided in the supplementary materials (S4).

The data for the estimation of growth and non-growth-associated maintenance costs are shown in the supplementary materials (S6).

2.1.3 Biomass: Proportions of macromolecular biomass components were measured experimentally with continuous-culture samples taken at steady state. Details of experimental procedures for the measurement of total DNA, RNA, protein, lipid and polysaccharide are given in the supplementary materials. Liquid chromatography-mass spectrometry (LC-MS) was used to estimate trace metabolite concentrations as described in [29].

2.2 Model construction

In the context of this paper, ‘construction’ refers to the selection of a set of reaction stoichiometries which form the metabolic network, whereas ‘curation’ refers to the identification and correction of errors and omissions in the definition of these reactions, e.g. errors in mass-balance and thermodynamic favourability. ScrumPy models are formed of a top-level ‘module’ and several sub-modules containing either automatically generated or manually defined reaction stoichiometries (see Fig. 1). This helps to organise the model components during construction and curation.

2.2.1 Draft network: Construction of the GSM began with the Tier 3 BioCyc [25, 30] database for C. autoethanogenum JA1-1 (strain DSM 10061) generated from the genome sequence published by Brown et al. [24]. This database (referred to here as the ‘CAETHG database’) formed the foundation of a draft reconstruction of the organism’s metabolic network. Automatically generated Tier 3 BioCyc databases are not fully curated, thus an additional manual genome annotation [22] was required to complete construction in line with methods described by Fell et al. [31] and Hartman et al. [32]. This additional annotation was used to create a second genome database (referred to as the ‘CLAU database’) with the PathoLogic algorithm as implemented in Pathway Tools [25, 30]. Information contained in this second database formed the basis for the continued curation of the model, the aim being to establish gene-protein-reaction relationships (GPRs) based solely on the manual annotation. For a detailed, step-by-step model construction methodology see supplementary materials (S1).

2.2.2 Curation: Initial curation steps involved removing chemically undefined metabolites from the automatically generated network. Subsequently, reactions without gene associations were investigated and removed if evidence for the necessary encoded enzymes could not be found [22]. Atomically unbalanced reactions were also corrected or removed. The thermodynamic consistency of the model, i.e. its adherence to energy conservation, was regularly checked using a specific linear programming (LP) problem detailed in Section 2.2.4. Inconsistent enzyme subsets containing erroneous reaction reversibility constraints [33] were identified using ScrumPy [34] and removed from the model (see Section 2.5).

2.2.3 Electron transport chain: A model of the electron transport chain (ETC) of C. autoethanogenum was manually constructed based on biochemical literatures [9, 11, 14, 35, 36]. The separate construction of an ETC sub-network enables the computation of its elementary modes (i.e. the set of minimal steady-state flux distributions across the network [37, 38]), since smaller networks avoid combinatorial difficulties encountered with larger networks. Inspection of the subsequent elementary modes ensures that legitimate routes for the generation of energy are available in model simulations (see Fig. 2).
balance analysis (FBA) was computed with fixed positive ATP reductase (MetFV) is unknown [40]. The definition given above
3.66 H + ATP by ATP synthase is not known. In previous studies, a value of
methylene-THF dehydrogenase, the ETC must include a
two key parameters are required: biomass
culture of ‘parameter borrowing’. In the
also derived from established physiology of Gram-positive bacteria [43].
aminotransferase capacity consistent with established physiology of Gram-positive bacteria [43].
where transport reactions are denoted vi and t is a vector of length
Xtrans containing all transport reactions in v. Any subsequent solution included a thermodynamic error, meaning
case of bacterial cells, two key parameters are required: biomass
Ribosomes [49]. The ‘Other’ class of biomass component	
47 gene associations for these reactions were mapped to
selecting ribosomal RNA and
where possible, species-specific parameters should be derived
amino acid ratios were estimated from transcribed rRNA and
characterisation of Gram-positive bacteria [43].
with experiments including protein, DNA, RNA, lipid and
detailed physiological information suitable for deriving cellular biomass composition is available for a small range of species only [46–48].
were estimated from the full genome sequence [22], whereas nucleotide estimated from the full genome sequence [22], whereas nucleotide
transcribed ribonucleoprotein sequences, respectively [see
macromolecular biomass composition was measured experimentally including protein, DNA, RNA, lipid and polysaccharide [see supplementary materials (S4) for experimental methods]. The relative abundances of various lipids in
The stoichiometry of the reaction catalysed by methylene-THF reductase (MetFV) is unknown [40]. The definition given above
CODH: CO + Fdred + H2O → 2H+ + CO2 + Fdred
MetFV: methylene – THF + 2NADH + Fdred
Ftn: 3H+ + NAD+ + Fdred → 2H+ + NADH + Fdred
2H+ + NAD+ + Fdred → 2NADH + NAD+ + Fdred
Hy×A − E: 2H+ + NAD+ + Fdred → 3H+ + NADH + Fdred
Fig. 2 Elementary mode of the ETC. Green arrows represent active reactions, grey arrows represent inactive reactions. Reversible reactions are represented as bi-directional arrows. Relative flux values and enzyme names are written next to arrow lines. Flux carried by irreversible reactions proceed in the direction indicated by the arrowhead, whereas reversible reactions carry flux in the ‘downward’ direction (in terms of the figure’s orientation)
reductase (MetFV) is unknown [40]. The definition given above
2.2.5 Transporters: The module ‘Transporters.spy’ defines all reactions involving transfer between the organism and the environment. Transporters were added based on known carbon/energy sources and products of C. autoethanogenum [15]. Transporters were also added for individual biomass precursors [31] and components of the growth media used to culture C. autoethanogenum in this paper, see supplementary materials (S4).
2.3 Model characteristics
A GSM of C. autoethanogenum consisting of 755 reactions and 772 metabolites has been constructed through the refinement of a draft network derived from a Pathway Tools database consisting of 1429 reactions and 1097 metabolites. About 73 model reactions in the curated model were unique to the initial CAETHG database. About 47 gene associations for these reactions were mapped to CLAU locus tags using EDGAR [42]. About 15 of the remaining model reactions unique to the CAETHG database were defined
transporters were added based on known carbon/energy sources and products of C. autoethanogenum [15]. Transporters were also added for individual biomass precursors [31] and components of the growth media used to culture C. autoethanogenum in this paper, see supplementary materials (S4).

\[ \begin{align*}
\text{Nfn:} & 2\text{NADP}^+ + \text{NADH} + \text{H}^+ + \text{Fd}_{\text{red}} \rightarrow 2\text{NADPH} + \text{NAD}^+ + \text{Fd}_{\text{red}} \\
\end{align*} \]
### Table 2. Overall biomass composition of C. autoethanogenum DSM 10061

| Component    | Percentage ± | Reference |
|--------------|--------------|-----------|
| protein      | 26.3 ± 2.3   |           |
| DNA          | 14.6 ± 7.5   |           |
| RNA          | 17.9 ± 4.2   |           |
| lipid        | 21.0 ± 1.7   |           |
| polysaccharide | 10.2 ± 7.8  |           |
| teichoic acid | 0.6 ± 0.1    |           |
| other        | 9.4 —        |           |

### Table 3. Maintenance costs ($r_{ATP}$) calculated at $\mu = 0.028 \text{ h}^{-1}$ with GAM and NGAM values from a range of GSMs/organisms.

| Organism          | GAM    | NGAM | $r_{ATP}$ | Reference |
|-------------------|--------|------|----------|-----------|
| C. autoethanogenum | 111.2  | 2.2  | 5.3      | —         |
| C. autoethanogenum | 41.3   | —    | 1.2      | [17]      |
| C. ljungdahlii    | 46.7   | 0.45 | 1.8      | [50]      |
| Clostridium beijerinckii | 40.0 | 8.5  | 9.6      | [51]      |
| Acetobacter woodii | 150.0  | 0.8  | 5.0      | [52]      |
| Bacillus subtilis  | 105.0  | 9.0  | 11.9     | [53]      |
| Escherichia coli  | 97.0   | 18.9 | 21.6     | [54]      |
| Saccharomyces cerevisiae | 71–91 | <1   | <3.3     | [55]      |

Units for growth-associated and non-growth-associated maintenance are mmol gDCW$^{-1}$ and mmol gDCW$^{-1}$ h$^{-1}$, respectively.

THF, CoA, flavin adenine dinucleotide, flavin mononucleotide, NAD and NADP, see supplementary materials (S4).

The overall biomass composition including essential trace metabolite pools is given in Table 2. Tables showing the estimated relative proportion of each biomass precursor within the macromolecular divisions are given in the supplementary materials (S4).

#### 2.4.2 ATP maintenance costs

The following experimental procedure was conducted to measure growth and non-growth ATP maintenance costs (growth-associated ATP maintenance cost (GAM) and non-growth-associated ATP maintenance cost (NGAM), respectively) in CO-fed continuous culture of C. autoethanogenum: The set dilution rate $D$ was increased incrementally from 0.01 to 0.028 h$^{-1}$. Steady state was reached after each change, as determined by stable culture OD. For each steady-state period, the average CO uptake rate was calculated and plotted against the corresponding specific growth rate $\mu$ which was taken as equivalent to the set dilution rate according to (8). The slope of a linear fit to this data is the growth-associated CO uptake rate; the $y$-intercept is the non-growth-associated CO uptake rate [see supplementary materials (S6)].

\[
\frac{d(OD)}{dt} = (OD)(\mu - D) = 0 \quad (8)
\]

GAM and NGAM were calculated by multiplying growth and non-growth-associated CO uptake rates by the model-calculated maximal ATP yield ($Y_{ATP}$) from CO. The growth and non-growth-associated CO uptake rates are 415 ± 92 and 5.7 ± 1.4, respectively, and the optimal $Y_{ATP}$ value is 0.375 [see (12)]. These values give a GAM of 155.7 ± 34.6 and an NGAM value of 2.2 ± 0.5 (The units for GAM and NGAM are mmol gDCW$^{-1}$ and mmol gDCW$^{-1}$ h$^{-1}$, respectively).

Since a proportion of GAM is accounted for by biomass precursor production in the metabolic network, this proportion must be subtracted from the total GAM before incorporation as a flux constraint. After subtracting the proportion of growth-associated ATP maintenance incorporated in the model's metabolic network (44.5 mmol gDCW$^{-1}$) from the experimentally derived GAM parameter, the actual GAM value used as a constraint in FBA simulations was 111.2 mmol gDCW$^{-1}$ h$^{-1}$. Total ATP maintenance ($r_{ATP}$) resulting from GAM/NGAM parameters sourced from published GSMs and an assumed growth rate of 0.028 h$^{-1}$ are shown in Table 3.

#### 2.5 Steady-state metabolism

The fundamental concept underlying the analyses presented in this work is metabolic steady state. This occurs when the concentrations of metabolites involved in a metabolic system do not change over time [56, 57]. The steady-state constraint is enforced with a set of linear differential equations representing the production and consumption of $m$ metabolites by $n$ chemical reactions. These equations can be represented as the dot product of a matrix of reaction stoichiometries $N$ and a vector of reaction fluxes ($i.e.$ net reaction rates) $v$ [58]. The steady-state condition is fulfilled when

\[
\frac{dm}{dt} = Nv = 0, \quad (9)
\]

### 2.6 Flux balance analysis

FBA is a fundamental technique for the analysis of GSMs which formulates steady-state metabolism as an LP problem given some assumed, biologically relevant objective function [61–63].

Since (9) only specifies constraints on the relative values of fluxes, further constraints are required to achieve a solution space which is sufficiently bounded for LP [64]. For example, transport ‘reactions’ responsible for the influx (and efflux) of metabolites into the system are often subject to an upper-bound constraint. Following this, an objective function is specified which maximises or minimises a selection of the flux variables. An objective is usually chosen which mimics an assumed biological objective [65] such as maximising ATP generation. The subsequent solution to the LP problem is thus an optimal flux distribution.

Results obtained through the application of FBA provide insight into the capabilities of a metabolic network, thus helping to investigate the steady-state behaviour of an organism. In this paper, FBA is used to calculate the growth rate of C. autoethanogenum as a means of model validation. Following this, FBA is used to investigate the model's response to changes in industrially relevant culture conditions including non-carbon limitation and constraints on hydrogen production.

#### 2.6.1 Network flux minimisation

The preferred objective function in this paper is the minimisation of absolute flux through all enzyme-catalysed reactions (10). For a mass-balanced metabolic network, total network flux minimisation reduces the occurrence of thermodynamically infeasible energy-generating cycles, while providing a reasonable approximation of enzymatic economy [31, 32, 34, 66, 67].
where $N$ is the stoichiometric matrix and $v$ is a vector of all reaction flux values included in the metabolic network; $v_{\text{kat}}$ is the set of network reactions catalysed by enzymes, $i.e.$ the subset of $v$ which excludes transport by diffusion (in this paper: CO, CO$_2$, H$_2$, and H$_2$O) and spontaneous reactions. This objective requires information on whether or not a reaction can occur spontaneously in the context of metabolism, as provided in Pathway Tools. $v_{\text{ATP}}$ and $m_{\text{ATP}}$ are equivalent to the growth and non-growth-associated ATP maintenance costs (GAM and NGAM), respectively, and $\mu$ is the specific growth rate. $v_{\text{ATP ass}}$ is the flux through the network's ATPase reaction (ATP $\rightarrow$ ADP + P$^+$ + H$^+$). Reaction fluxes representing the production of biomass precursors are denoted $v_{\gamma X}$, where $k$ is a vector containing the corresponding indices of these reactions in $v$. The vector $b$ contains relative abundance values for each biomass precursor which, when multiplied by $\mu$, determine the value of each $v_{\gamma X}$. Reaction fluxes representing the transport of non-biomass-associated metabolites not required as growth substrates ($i.e.$ potential products) are denoted $v_{\gamma p}$, where $p$ is a vector of length $X_{\text{prod}}$, containing the corresponding indices in $v$.

2.6.2 Growth yield/ATP maximisation: An alternative objective function is the maximisation of biomass precursor production, an approximation of growth rate optimisation [44]

$$\begin{align*}
\text{maximise: } & \mu \\
\text{subject to: } & Nv = 0 \\
& v_{\text{ATP ass}} = m_{\text{ATP}} \\
& v_{\gamma X} = g_i; \quad i \in \{1, ..., X_{\text{growth}}\} \\
& v_{\gamma p} \leq 0; \quad j \in \{1, ..., X_{\text{prod}}\}
\end{align*}$$

Here, $\mu$ is represented as a lumped chemical equation including all biomass precursors as reactants with their relative abundance given as stoichiometric coefficients. $v_{\text{ATP}}$ (GAM) is incorporated in the lumped expression as the stoichiometric coefficient on ATP. Reaction fluxes representing each of $X_{\text{growth}}$ substrate-uptake reactions are denoted $v_{\gamma X}$, where $k$ is a vector containing the corresponding indices in $v$. The vector $g$ contains fixed-flux values for each of $X_{\text{growth}}$ substrate-uptake reactions. In this paper, the fixed substrate-uptake rate is taken as the average experimentally observed uptake rate over the time period at which the culture is at steady state.

An alternative form of this LP problem maximises ATP dissipation, $i.e.$ flux carried by ATPase ($v_{\text{ATP ass}}$) in the ATP hydrolysing direction. Solutions to this problem achieve optimal ATP yields. Furthermore, a complete hierarchy of products ordered in terms of $Y_{\text{ATP}}$ can be generated by iteratively computing the $Y_{\text{ATP}}$-optimal solution while blocking ($i.e.$ constraining to 0) flux across the transporter of the product formed in the previous iteration until ATP production is infeasible.

2.7 Software

Reconstruction steps and model analyses were carried out using the ScrumPy package ([http://mudshark.brookes.ac.uk/ScrumPy](http://mudshark.brookes.ac.uk/ScrumPy)) [34]. Analysis of the previously published iCLAU786 [17] model was carried out in COBRApy [68]. Bioreactor data were analysed using an automated python/MATLAB-based software pipeline designed for the BioCommand software (BioCommand®, New Brunswick Scientific), in which data were transferred to a web server, allowing real-time updates before calibration and post-processing using a suite of MATLAB® scripts. Further details of this toolset including source code and descriptions are available on request.

3 Results

3.1 Comparison of MetaCLAU with published models

Genome-scale metabolic models have been published for both *Clostridium ljungdahlii* [50] and *C. autoethanogenum* [17, 20]. The *C. autoethanogenum* model, iCLAU786, was first made available in [17] and then updated in [20]. Additionally, the repository of metabolic models automatically generated by CarveMe [69] contains a draft network for *C. autoethanogenum*. This section shows the results of a general comparison between MetaCLAU and these published models.

Differences in the general model statistics are shown in Table 4. Importantly, the more recent iCLAU786 version [20] fails the essential condition set-out in Section 2.2.4 that no ATP can be generated without the uptake and excretion of metabolites. For this reason, this version of iCLAU786 is excluded from further analysis. Similarly, the CarveMe draft model is excluded from further analysis as it is unable to generate ATP from CO and water. An important difference is the number of dead reactions; iCLAU786 has 966 (97%) dead reactions, whereas MetaCLAU has 327 (43%). MetaCLAU also accounts for the transport of more metabolites. Finally, unlike the other models, MetaCLAU contains no inconsistent enzyme subsets [33]. Details of this comparison are provided in the externally hosted files [26].

3.2 Validation

Validation of key behaviours against experimental data was carried out to demonstrate the model’s good reliability. Substrate utilisation, the growth rate on CO and product spectrum are the validating behaviours tested in this paper.

3.2.1 Substrate utilisation: The ability of *C. autoethanogenum* to use different carbon/energy sources for the production of biomass (and maintenance) was calculated with FBA. Each compound was used as the sole carbon source in the *in silico* minimal medium by allowing influx across the corresponding transporter while applying the min [$\mu$] objective. The simulation results were compared with experimental data (Table 5), showing good agreement with known carbon/energy sources.

3.2.2 Growth rate: An average experimentally observed CO uptake rate of 16.52 $\pm$ 0.02 mmol gDCW$^{-1}$ h$^{-1}$ was applied as a fixed-flux constraint on CO transport, which, with the objective function set to maximise biomass production, resulted in a calculated optimal specific growth yield of 0.027 h$^{-1}$. This value is within the error margin of the measured specific growth rate of 0.028 $\pm$ 0.001 h$^{-1}$. A comparison of model-simulated versus reported experimental growth rates including fructose and syngas is shown in Table 6.

Table 4 Comparison of genome-scale metabolic models

| Model          | Reactions | Dead reactions | Trans. | Mets | Genes | GP Rs | EISs* | Energy conserved? |
|----------------|-----------|---------------|--------|------|-------|-------|-------|------------------|
| MetaCLAU       | 755       | 327           | 80     | 772  | 699   | 600   | 0     | YES              |
| iCLAU786 [17]  | 1001      | 810           | 28     | 1046 | 802   | 841   | 21    | YES              |
| iCLAU786 [20]  | 1109      | 966           | 66     | 1162 | 784   | 862   | 22    | NO               |
| CarveMe        | 1202      | 30            | 129    | 922  | 809   | 844   | 15    | YES              |
| iHN837         | 785       | 186           | 95     | 698  | 637   | 615   | 20    | YES              |

*This high proportion of dead reactions is due to the inclusion of external metabolites in the stoichiometric matrix of iCLAU786. See externally hosted files [26].

*Inconsistent enzyme subsets (EISs).
3.2.3 Product spectrum: The ratio of product fluxes in optimal solutions with CO as the sole source of carbon and energy was sensitive to the choice of the objective function. Maximisation of growth yield (11) resulted in acetate and CO₂ production, with no flux to ethanol. In contrast, when minimisation of network flux was the objective function (10), acetate and ethanol formed the main products. Optimal flux distributions computed with published models of *C. autoethanogenum* (iCLAU786 [17, 19] and *C. ljungdahlii* iHN637 [50]) did not include simultaneous production of ethanol and acetate without additional constraints (see Table 7). The min \( \rho_{\text{cell}} \) objective could not be applied to iCLAU786 or iHN637 since reactions that are known to occur spontaneously (in the context of metabolism) are not identified in these models.

The optimal ATP yield \( (Y_{\text{ATP}}) \) per m mole substrate uptake with CO as the sole carbon and energy source was 0.375 with acetate as the sole product. A sub-optimal \( Y_{\text{ATP}} \) value (0.342) is associated with ethanol production. A hierarchy of products based on their associated ATP yield is shown below, demonstrating that the model can return flux distributions associated with the expected range of liquid products:

\[
Y_{\text{acetate}}^{\text{ATP}} = 0.375: CO + 1/2H_2O \\
\rightarrow 1/4CH_2COO^- + 1/4H^+ + 1/2CO_2 \\
Y_{\text{ethanol}}^{\text{ATP}} = 0.341: CO + 1/2H_2O \\
\rightarrow 1/6CH_3CH_2OH + 2/3CO_2
\]

3.3 Gas shift

A range of CO uptake rates exceeding the uptake rate required to support a fixed growth rate \( (v_{\text{CO}} = 16.97, \mu = 0.028) \) was applied to simulate non-carbon-limited growth. FBA with CO uptake in the range 17–42.5 mmol gDCW⁻¹ h⁻¹ and a fixed growth rate of 0.028 h⁻¹ results in a shift from majority acetate production to ethanol production (Fig. 3) and subsequent hydrogen production in optimal solutions (Fig. 4). When the model was constrained to allow no production (or uptake) of hydrogen, 2,3-BD and lactate efflux appear in optimal solutions after the initial switch from acetate to ethanol (where 2,3-BD efflux increases with CO uptake but lactate efflux remains constant, see Fig. 5). The constraint on hydrogen efflux was applied to mimic possible limitations on hydrogen production caused by increasing hydrogen partial pressure in the external environment [73].

### Discussion

A GSM of *C. autoethanogenum* has been constructed and parameterised with experimentally derived values representing ATP maintenance costs and cellular biomass composition. The experimental parameterisation was designed to make the model more accurately represent *C. autoethanogenum* metabolism, as confirmed by the validation in Section 3.2 above. Model-calculated yields for growth on CO agree with experimentally observed growth rate data, while the production of acetate and ethanol in optimal flux distributions supporting biomass production and maintenance costs is consistent with the established native product profile of *C. autoethanogenum*.

Oversupply of CO to the metabolic network leads to excess reducing equivalents, forcing flux through pathways (see Fig. 6) allowing the removal of superfluous reducing power (electrons) from the system [14]. This has been tested using FBA by increasing the ratio between energy-substrate supply (in this case, CO) and growth rate over a fixed range (Fig. 3) representing non-carbon-limited conditions. Results show that increasing the influx

---

### Table 5 Carbon-source validation results

| Carbon source | Model predicts growth? | Growth observed? | Reference |
|---------------|------------------------|------------------|-----------|
| CO            | YES                    | YES              | [15, 17, 35] |
| CO₂ (H₂)      | YES                    | YES              | [15, 35]   |
| fructose      | YES                    | YES              | [15, 35]   |
| xylose        | YES                    | YES              | [15, 70]   |
| arabinose     | YES                    | YES              | [15]       |
| rhamnose      | YES                    | YES              | [15]       |
| pyruvate      | YES                    | YES              | [15]       |
| glutamate     | YES                    | YES              | [15]       |
| formate       | YES                    | NO               | [15]       |
| fumarate      | YES                    | YESᵃ              | [15, 71]  |
| glucose       | NOᵇ                   | NO               | [15]       |
| lactate       | NO                    | —                | [15]       |
| acetate       | NO                    | —                | —         |
| ethanol       | NO                    | —                | —         |
| 2,3-BD        | YES                    | —                | —         |

ᵃAbrini et al. [15] and Breitkopf [71] disagree, see discussion.
ᵇ*C. autoethanogenum* does not encode a glucose transporter.

---

### Table 6 Growth rate comparison: model simulation versus experimental data

| Carbon/energy source | Uptake (measured) | Growth rate (measured) | Growth rate (simulated) | Data source |
|----------------------|-------------------|------------------------|-------------------------|-------------|
| CO                   | 16.5              | 0.028                  | 0.027                   | (this study) |
| CO₂, H₂              | 22.4, 13.9        | 0.041                  | 0.063                   | [20]        |
| fructose             | 1.4               | 0.05                   | 0.027                   | [72]        |

---

### Table 7 Ethanol, acetate and CO₂ effluxes in optimal flux distributions for CO growth computed with *C. autoethanogenum* GSMS

| Model         | Objective | \( V_{\text{acetate}} \) | \( V_{\text{ethanol}} \) | \( V_{\text{CO}_2} \) | CO uptake |
|---------------|-----------|--------------------------|--------------------------|------------------------|-----------|
| MetaCLAU      | min \( \rho_{\text{cell}} \) | 0.06 | 2.7 | 12 | 18.4 |
|               | min \( \mu \) | 0.07 | 2.7 | 12 | 18.4 |
|               | max \( \mu \) | 2.3 | 0 | 5 | 10 |
| iCLAU786 [17] | min \( \mu \) | 0 | 2.4 | 10.9 | 16.7 |
|               | max \( \mu \) | 2.3 | 0 | 5.0 | 10 |
| iHN637 [50]   | min \( \mu \) | 3.1 | 0 | 8.0 | 15.6 |
|               | max \( \mu \) | 3.5 | 0 | 10.4 | 20 |

Acetate and ethanol efflux values are emboldened in flux distributions where both are produced. iCLAU786 max \( \mu \) values taken from [17].

Bold values emphasise the inclusion of both acetate and ethanol in flux distributions computed with MetaCLAU.

---

*Eng. Biol., 2019, Vol. 3 Iss. 2, pp. 32-40*

This is an open access article published by the IET under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/).
of CO beyond the rate necessary to achieve a fixed growth rate causes a switch from acetate to ethanol production over the uptake range 17–18.5 mmol gDCW⁻¹ h⁻¹. This provides a new strategy for increasing ethanol production.

It has also been hypothesised that bacterial hydrogen production will be restricted by external constraints during continuous culture. More specifically, hydrogen production may become thermodynamically unfavourable if its partial pressure in the external environment exceeds some critical value. This will cause metabolism to favour the production of other electron sinks (such as lactate) [73–75]. To test this, the CO uptake scan was repeated with hydrogen transport restricted to zero. Optimal flux distributions resulting from this simulation showed 2,3-BD production increasing monotonically with CO uptake after the acetate/ethanol switch (Fig. 5). This result is significant since 2,3-BD production has not been reported from the analysis of previous models [17, 20]. Furthermore, lactate production appeared in optimal flux distributions but was unchanged with increasing CO uptake, suggesting an association with biomass production. Thus, the analysis has uncovered a testable set of bioprocess conditions for which 2,3-BD is included in optimal metabolic flux distributions, increasing with the gas inflow.

4.1 Model availability

The model is available to download as an SBML, JSON (COBRA readable) or ScrumPy-formatted file from the Biomodels database [76] using accession number MODEL1810120001. All model and database files are available in the supplementary materials including an ipython notebook demonstrating all presented analyses (requires installation of ipython and COBRApy).

5 Acknowledgments

This work was supported by the Biotechnology and Biological Sciences Research Council, as part of the BBSRC Longer and Larger Grant GASChEm [grant number BB/K00283X/1]; the BBSRC/EPSCR Synthetic Biology Research Centre Nottingham [grant number BB/L013940/1]; and the industrial partner LanzaTech Inc. Furthermore, the authors thank their colleagues at the SBRC for their support during the preparation of the paper. MP gratefully acknowledges support from Oxford Brookes University. Finally, we thank C1net for facilitating both the collaboration between Oxford Brookes University and the University of Nottingham and for funding and organising the C1net metabolic modelling workshops. The responsibility for the content of this paper lies with the authors.
6 References

[1] Armario, N., Balzarini, V.: ‘The legacy of fossil fuels’, Chem. Asian J., 2011, 6, (13), pp. 2645–2650

[2] Pachauri, R.K., Allen, M.R., Barros, V.R., et al.: ‘Climate change 2014: synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change’ (IPCC, Switzerland, 2014)

[3] Doran, P.M.: ‘Biprocess engineering principles’ (Academic Press, Waltham, MA, 1996)

[4] Heijstra, B.D., Leang, C., Juminaga, A.: ‘Gas fermentation: cellular scale production of low carbon-fuels and chemicals from waste and renewable feedstocks’, Front. Microbiol., 2016, 7, p. 694

[5] Ljungdahl, L.G., Wood, H.G.: ‘Total synthesis of acetate from CO2 by Clostridium autoethanogenum’, Trans. R. Soc. Lond. B, 2000, 355 (1382), pp. 1873–1898

[6] References
