Applying statistical design of experiments (DOE) to understanding the effect of growth media components on *Cupriavidus necator* H16 growth

Christopher C. Azubuike\textsuperscript{a,b}, Martin G. Edwards\textsuperscript{a}, Angharad M. R. Gatehouse\textsuperscript{a}, Thomas P. Howard\textsuperscript{#}

\textsuperscript{a}School of Natural and Environmental Sciences, Faculty of Science, Agriculture and Engineering, Newcastle University, Newcastle-upon-Tyne, NE1 7RX, United Kingdom
\textsuperscript{b}Department of Microbiology, Faculty of Science, University of Port Harcourt, East-West Road, P.M.B. 5323, Choba, Port Harcourt, Rivers State, Nigeria

\#corresponding author, Thomas P. Howard, email: Thomas.Howard@newcastle.ac.uk

Tel: +44(0)1208 48 54

ORCID iD

Christopher C. Azubuike: 0000-0002-2376-3057
Martin G. Edwards: 0000-0002-4863-653X
Angharad M. R. Gatehouse: 0000-0002-6406-2927
Thomas P. Howard: 0000-0002-5546-4043

Running Title

Data-driven modelling of *Cupriavidus necator* growth
Abstract

*Cupriavidus necator* H16 is gaining significant attention as a microbial chassis for a range of biotechnological applications. Whilst the bacterium is a major producer of bioplastics, its lithoautotrophic and versatile metabolic capabilities make the bacterium a promising microbial chassis for biofuels and chemicals using renewable resources. It remains necessary to develop appropriate experimental resources to permit controlled bioengineering and system optimization of this microbe. Here we employed statistical design of experiments to gain understanding of the impact of components of defined media on *C. necator* growth and built a model that can predict the bacterium’s cell density based on media components. This highlighted media components, and interaction between components, having the most effect on growth: fructose, amino acids, trace elements, CaCl$_2$ and Na$_2$HPO$_4$ contributed significantly to growth ($t < -1.65$ or $> 1.65$); copper and histidine were found to interact and must be balanced for robust growth. Our model was experimentally validated and found to correlate well, $r^2 = 0.85$. Model validation at large culture scales showed correlations between our model predicted growth ranks and experimentally determined ranks at 100 mL in shake flasks ($\rho = 0.87$) and 1 L in bioreactor ($\rho = 0.90$). Our approach provides valuable and quantifiable insights on the impact of media components on cell growth and can be applied to model other *C. necator* responses that are crucial for its deployment as a microbial chassis. This approach can be extended to other non-model microbes of medical and industrial biotechnological importance.

Importance

Chemically defined media (CDM) for cultivation of *C. necator* vary in components and compositions. This lack of consensus makes it difficult optimizing new processes for the
bacterium. This study employed statistical design of experiments (DOE) to understand how basic components of defined media affect *C. necator* growth. Our growth model predicts that *C. necator* can be cultivated to high-cell-density with components held at low concentrations, arguing that CDM for large scale cultivation of the bacterium for industrial purposes will be economically competitive. Although existing CDM for the bacterium are without amino acids, addition of few amino acids to growth medium shortened lag phase of growth. The interactions highlighted by our growth model shows how factors can interact with each other during a process to positively or negatively affect process output. This approach is efficient, relying on few well-structured experimental runs to gain maximum information on a biological process, growth.

**Keywords**

*Cupriavidus necator* H16, Design of experiments (DOE), chemically defined media

**Introduction**

*Cupriavidus necator* H16 is of biotechnological importance largely due to its ability to accumulate >80% of its dry cell weight as polyhydroxyalkanoate (PHA)—a biodegradable polymer and an alternative to petroleum-based polymers (1, 2). PHA (specifically 3-polyhydroxybutyrate (3-PHB)) accumulation by the bacterium is a carbon conservation mechanism. When nitrogen, oxygen, or phosphorus becomes limiting to cell growth the
bacterium diverts excess carbon to 3-PHB. This carbon store supports growth when conditions improve (3–6). *C. necator* is also chemolithoautotrophic, with the ability to use CO₂, and formate or H₂ as carbon and energy sources to support cellular metabolism (7). In the absence of oxygen it can use alternative electron acceptors (NO³⁻ and NO₂⁻) to carry out anaerobic respiration by denitrification (8). More widely, the genus is known to encode genes facilitating metabolism of environmental pollutants such as aromatics and heavy metals making it a potential microbial remediator (9–12). *C. necator* is central to efforts to construct systems for the electromicrobial conversion of CO₂ to higher alcohols (7). Metabolic engineering of *C. necator* has been demonstrated through the introduction of pathways for the biosynthesis of alcohols (6, 7, 13, 14), fatty acids (15–17), alkanes (18) and enzymes (19) under both heterotrophic and autotrophic growth conditions. The production of branched chain alcohols by *C. necator* through electricity powered cellular synthesis (i.e. electrosynthesis), demonstrates the value of this bacterium as a chassis capable of exploiting renewable feedstock for the biosynthesis of valuable products (7). As with any bacterium, the development of *C. necator* as an industrial chassis requires appropriate tools for studying and engineering the organism. One of these resources is the availability of a characterised, chemically defined growth medium. Chemically defined media are important to enable experimental reproducibility, to reliably characterise the genetics of the organism, to determine genotype by environment interactions, and to facilitate fundamental research of bacterial physiology that underpins bioengineering efforts. While different chemically defined media have been deployed for the cultivation of *C. necator* (3, 13, 14, 16, 20–22) there is no consensus regarding the components that are required, the concentration of each component, or how each component interacts to affect growth of the bacterium (Table S1). This is not surprising given the range of industrial applications with the bacterium.
Design of Experiments (DOE) is an iterative, empirical approach that systematically explores the relationship between input variables (factors) and output variables (responses). The approach yields a structured set of data that can be used to build statistical models employed in understanding or optimizing system performance. These statistical models can be validated against prior knowledge, internal statistical methods or ultimately by their ability to predict responses from new combinations of factors. The increasing availability of laboratory automation and high throughput technologies may therefore be resulting in a greater appreciation for its application in bioengineering. DOE has found use in the optimization of metabolic pathways (23, 24), cell-free protein synthesis reactions (25) and codon-use algorithms (26). It has been applied to the study of genotype-by-genotype and genotype-by-environment interactions in yeast (27) and in re-purposing enzyme activities (28).

Here we employ a statistical engineering approach to build a data driven model that can accurately predict *C. necator* growth responses to a range of media formulations. The model highlighted different formulations that allow reproducible and robust growth of *C. necator* with the minimal concentration of each component and allowed us to identify and understand interactions between components of the media. Additionally, the model allowed the learning from the small scale (1 mL) to be applied at larger volumes (100 mL and 1 L). Understanding the impact of each component of a chemically defined medium on *C. necator* growth is a fundamental tool for controlled exploration of the biotechnological potential of this important bacterium.

**Results**

**Identifying main ingredients in chemically-defined media that affect the growth of *C. necator***
The preliminary phase of the experiment was to determine whether OD$_{600nm}$ is appropriate for determining *C. necator* growth. To establish this, the bacterium was cultivated in a rich medium (LB), and correlation between the colony-forming units (cfu/mL) and OD$_{600nm}$ determined (Fig. SI). Growth obtained from optical density measurements at OD$_{600nm}$ correlated well with the number of viable cells (cfu/mL). Next, an initial scoping trial was carried out using fructose, glucose, glycerol or sucrose to identify a principle carbon source to be used for subsequent work, and to determine the range of concentrations to be tested. Four scoping trials were performed one each at low and high concentrations of media components, and two trials at the midpoint values between the two extremes (Table 1). At the ranges tested glucose, glycerol and sucrose supported little or no growth of *C. necator*, whilst fructose supported high growth except at low concentrations (Fig. 1). The OD$_{600nm}$ for the two midpoint experiments demonstrated a peak at 72 h followed by a plateau. From these scoping trails it was established that fructose would be our principle carbon source, OD$_{600nm}$ was an appropriate measure of cell growth, growth assays in 1 mL volumes in a 48-well plate format were appropriate for subsequent experiments, and recording OD$_{600nm}$ at 72 h provides a good balance between measuring growth rate and peak culture density.

We next identified key factors that might influence OD$_{600nm}$ at 72 h. An initial definitive screening design (DSD1) array was built based on 10 media components (Table S2). The composition of each media in the DSD1 are within the ranges indicated in Table 1. DSDs are highly efficient experimental designs in which all main effects can be estimated independently of other main effects and all possible two-way interactions. The requisite variant media compositions were assembled in a 48-well plate and growth was monitored. Both analyses indicated that high concentrations of fructose, CaCl$_2$ and amino acids contributed positively to growth, while high concentrations of Na$_2$HPO$_4$ and trace elements contribute negatively to growth. Factors such as NaH$_2$PO$_4$, K$_2$SO$_4$, MgSO$_4$, NH$_4$Cl and...
vitamins were not found to have statistically significant effects under the conditions tested (Fig. 2A). These analyses also highlighted several two-way interactions that may influence growth responses (Fig. S2).

Beside the main effect of each component (factor), and interactions between some components of media, the mean growth distribution between factor settings (low, mid and high) showed interesting responses (Fig. 2B). For example, the mean growth distribution for fructose showed a non-linear (quadratic) effect, arguing that the optimum fructose concentration that will result in robust and less variable growth lies above 0.5 g/L and below 40 g/L. Disodium phosphate (Na$_2$HPO$_4$) showed a non-linear effect, with growth at low setting reaching higher OD$_{600nm}$ than at high setting. Similarly, CaCl$_2$ showed a non-linear effect; the optimum concentration range for robust growth appears to lie above the low setting (0.01 g/L) and below the high setting (0.8 g/L). Trace elements also displayed a non-linear effect. Amino acid data suggested that the optimum concentration was yet to be tested. Nevertheless, we avoided increasing amino acid concentration given that this will contribute to extra carbon and nitrogen to the media. The non-significant factors displayed linear effect, except MgSO$_4$. Therefore, this initial DSD needs augmentation to establish an optimum concentration range for each component, especially components having the most significant effect on growth.

**Augmentation of the data set**

A definitive screening design can force many of the sampling points collected to the edges of the design space. For this reason, we ran a second definitive screening design (DSD2) array (Table S3) in which the concentration ranges of the components were guided by the data from DSD1, and the factors under investigation were restricted to those that were highlighted as significant in DSD1 (Fig. 2). Examining the combined data for DSD1 and DSD2 did
indeed confirm that the extreme concentrations of some of the components were detrimental to cell growth. For example, when the concentrations of fructose were at the highest and lowest values (40 and 0.5 g/L, respectively) the OD_{600nm} 72 h were both lower and more variable than when fructose was restricted to between 5 and 25 g/L (Fig. 3). This indicates that maintaining fructose between 5 and 25 g/L is key to establishing robust and reliable growth. Likewise, adjustments were made to the concentration ranges of the amino acids (5 and 20 mL/L), CaCl_2 (0.1 and 0.459 g/L) and Na_2HPO_4 (0.1 and 3.05 g/L). Other factors not identified as statistically significant were kept at the midpoint from DSD1, except NH_4Cl which was set at the lowest value because there was some evidence of a negative impact from the 2-level factor analysis (Fig. 2A).

Examination of the distributions for DSD2 revealed that the media compositions that generated the highest cell densities (OD_{600nm} > 2.0) were associated with low trace element concentrations and high amino acid content (Fig. S3). More so, given that both components are potential growth supplements, and each contained multiple ingredients, it was necessary to investigate the interaction between these two components to determine which ingredients are responsible for the observed interaction. To determine this, a medium that permits robust cell growth over 72 h was formulated and used as a control. It was observed that the absence of trace elements did not adversely affect growth of *C. necator* but that the absence of amino acids did (Fig. 4A). Interestingly, simultaneous exclusion of both amino acids and trace elements resulted in cell densities comparable with the control. We hypothesized that in the absence of one or more of the amino acids (methionine, histidine, leucine and/or arginine), the presence of one or more of the components of the trace elements (CuSO_4, FeSO_4, MnSO_4 and/or ZnSO_4) inhibits growth of *C. necator*. This was tested first by formulating media without amino acids and withdrawing each trace element in turn (Fig. 4B). Under these conditions, media without CuSO_4 but containing other trace elements resulted in growth
comparable to the control, while all three formulations that contained CuSO$_4$ had reduced cell densities. These observations support the hypothesis that CuSO$_4$—in the absence of amino acids—inhibits *C. necator* growth. To determine which amino acid(s) interacts with CuSO$_4$, a series of experiments were performed in which each amino acid was excluded in formulations with and without CuSO$_4$. The first observation was that at high concentrations of amino acids (20 mL/L) the presence or absence of CuSO$_4$ did not affect growth (Fig. 4C). The second observation was that at medium concentrations of amino acids (10 mL/L) growth was partially suppressed in the presence of CuSO$_4$ but that this was exacerbated in the absence of CuSO$_4$. Copper sulfate is therefore an important medium component and cannot simply be excluded from formulations. Similar growth responses were seen in experiments in which either methionine or leucine were excluded (Fig. 4D and E). If arginine, or most notably histidine, were removed from the medium, then the presence of CuSO$_4$ impaired growth of *C. necator* (Fig. 4F and G). For media lacking histidine this effect was also observable when all other amino acid levels were kept high (Fig. 4G). From this data we established two points: first, that histidine protects against the inhibitory effects of copper, and second, CuSO$_4$ is an important component of the media, whose absence retards growth when amino acid content is restricted. Thus, the interaction between CuSO$_4$ and histidine is a concentration dependent positive interaction. Both CuSO$_4$ and histidine concentrations must therefore be balanced for robust growth. Although formulations containing higher concentration of amino acid (> 10 mL/L) can result in higher OD$_{600nm}$, such concentration contributed ~12% to growth in the absence of fructose as a carbon source, whereas under the same formulation, concentration ≤ 10 mL/L amino acid contributed < 5% to growth. Hence, in subsequent experiments, amino acid concentration was set at 10 mL/L and trace at ≤ 1 mL/L.
**Final data augmentation**

With a greater understanding of which components and concentrations are required to formulate a medium supporting robust growth, we re-investigated the roles that NaH$_2$PO$_4$, K$_2$SO$_4$, MgSO$_4$ and NH$_4$Cl have on the system. In the original DSD these factors were not identified as being statistically significant; but those experiments were conducted under conditions in which key components (e.g. fructose and amino acids) were at settings that have since been identified as resulting in poor or unreliable growth responses. We therefore re-evaluated these factors under conditions in which NaH$_2$PO$_4$, CaCl$_2$, trace elements and amino acid concentrations were not disruptive to cell growth (Table S4). Fructose was set at either 5 g/L or 20 g/L. The results indicated that these components did indeed affect growth rate when primary factors are not restricting growth (Fig. 5). Some of the suggestions from the original DSD were confirmed. Most notable was the observation that increasing NH$_4$Cl had a detrimental effect on culture cell density. Sodium dihydrogen phosphate (NaH$_2$PO$_4$) and K$_2$SO$_4$ had some detrimental impact if concentrations were low, whilst MgSO$_4$ concentrations were not significant. These results were true at both high and low concentrations of fructose.

**Modelling the media formula-growth response landscape**

At this stage, 64 different variant formulations have been experimentally assessed in duplicate across three different experiments (25, 13 and 26 experimental runs respectively). We then built a statistical model, trained against this data set, that could describe our understanding of how the cell cultures respond to changes in media composition and predict performance in novel formulations. We performed two-level screening on all 128 runs. This identified a number of factors and factor interactions deemed significant for model projection. The screening did not highlight fructose or K$_2$SO$_4$ (as these had been at held...
concentrations that did not significantly impact growth during much of DSD2 and DSD3) but these terms were included manually in the model as we knew they were important factors from the first DSD. These terms were used to construct a standard least squares model. The least squares model was able to describe the relationships within the data with good accuracy (Fig. 6A). Although the model was internally consistent, it was important to know if it could predict OD$_{600nm}$ at 72 h in formulations it had not encountered during model training. We assessed 16 new formulations with sampling biased towards media formulations that were predicted to be in the top 25% of media performance. Each of these was assessed in triplicate and the resulting OD$_{600nm}$ compared to predicted OD$_{600nm}$ (Fig. 6B). As predicted by the model, all of the new formulations fell within the upper quartile of formula performance.

The model allowed us to visualize phenomenon observed during the data-collection phase of the investigation. For example, it visualizes the interaction between amino acids (specifically histidine) and trace elements (specifically copper) that was elucidated in Fig. 4. It shows that high concentrations of trace elements are detrimental to growth and that increasing the amino acid concentration can help mitigate the inhibitory effects of high concentrations of trace elements (Fig. 6C). The greater the concentration of trace elements included the higher the amino acid concentration needs to be. Nevertheless, increasing the concentration of amino acids impacts in other areas. A monotonic relationship was observed for fructose and amino acids; increasing the concentration of both increase cell density (Fig. 6D). Although low concentration of amino acids ($\leq$ 5 mL/L) may result in low cell density, this negative impact appears to be greater under high fructose concentration. It follows that at 5 mL/L of amino acid, reducing fructose concentration to 5 g/L will likely support higher cell density. Given that the amino acids present in the media are the only alternative source of carbon for growth, increasing the concentration of amino acids too high will impact on interpretation of experiments designed to examine carbon utilization. A balance therefore
needs to be established between fructose, amino acids and trace element concentrations. The model also visualizes interactions between Na₂HPO₄ and fructose (Fig. 6E). Higher concentrations of Na₂HPO₄ results in lower cell densities—an effect that can be partially offset by decreasing fructose concentrations. The model also suggests that if CaCl₂ is to be increased, fructose concentration must be increased and vice versa (Fig. 6F). Lastly, we had previously observed a negative effect of increasing NH₄Cl concentrations (Fig. 5D). The model indicates that this can be mitigated by increasing the K₂SO₄ and/or NaH₂PO₄ concentration (Fig. 6G and H). Understanding the interactions between media components is vital for predictions of culture performance and allows the experimenter to alter media formulations for different experimental goals.

**Distinguishing between statistically significant and essential media components**

The growth model highlighted components and interactions that contribute significantly to growth. However, it is important to draw a distinction between what is statistically significant and what is essential. Determining the essential nature of components involves the conventional approach of withdrawing one factor individually during media preparation. To ensure that the concentrations at which factors are held while individually withdrawing a component did not introduce bias on the essential nature of any component, two media were used as controls (Fig. 7A). One of these controls had every component maintained at low concentration (low concentration formula), while the other had every component maintained at high concentration (high concentration formula). Both controls were predicted to support robust growth of *C. necator* (Fig. 7A). When each component was individually withdrawn in both formulae, similar growth was observed indicating that the concentration at which components were held did not impact on the essential nature of any components of the media (Fig. 7B and C). Little growth was observed in media lacking fructose or MgSO₄ whereas...
growth was significantly impaired in media lacking amino acid or K\(_2\)SO\(_4\). The remainder components when individually withdrawn had little or no impact on growth. Therefore, it is established that fructose and MgSO\(_4\) are essential for cultivation of *C. necator* H16. Despite this, MgSO\(_4\) is not deemed statistically significant by the model (Fig. 2A). This is because under our experimental conditions, all concentrations of MgSO\(_4\) tested are likely in excess and do not limit growth. There is therefore the opportunity to reduce or increase MgSO\(_4\) concentrations if this were desirable. Lastly, amino acid and K\(_2\)SO\(_4\) are considered important, while the remainder components are considered non-essential for growth. Interestingly, for both formulae, simultaneously withdrawing some of the non-essential components (Na\(_2\)HPO\(_4\), CaCl\(_2\), and NH\(_4\)Cl) resulted in similar growth pattern, and gave OD\(_{600\text{nm}}\) comparable to that of the controls (Fig. 7D and E), arguing that cultivation of *C. necator* to high-cell-density is not nutritionally demanding—provided the main components of media are balanced.

Finally, colony-forming unit per millilitre (cfu/mL) was correlated against OD\(_{600\text{nm}}\) using control media predicted to support robust growth (Fig 7A). This was performed to further confirm that OD\(_{600\text{nm}}\) is indeed appropriate for measuring *C. necator* cell density in defined media and that the OD\(_{600\text{nm}}\) observed throughout this study were ascribed to growth (viable cells) and not as a result of precipitation of components in media. Optical density (OD\(_{600\text{nm}}\)) and cfu/mL showed similar growth trends over 120 h cultivation (Fig. S4).

**Model validation at greater volumes**

To determine whether the growth model built using data collected from microtiter plate provides understanding of *C. necator* nutritional responses at larger culture volumes, we conducted two further model validation tests: first, at 100 mL shake flasks and secondly, at 1 L using a bioreactor. Shake flask cultivations were carried out under identical conditions in
baffled and non-baffled flasks, using formulations randomly selected from an L32 fractional factorial design of experiment (Fig. S5). Growth (OD_{600nm}) for each formula in both types of flasks were strikingly similar at every interval throughout the cultivation period. Most importantly, the growth rank for baffled and non-baffled flasks were identical. Spearman’s correlation showed a significant (p < 0.05) relationship between predicted and actual growth ranks for both flask types; the correlation coefficient is \( \rho = 0.87 \) (Fig. 8A). Next, two formulae from shake flask cultivation together with three additional formulae were cultivated in 1 L bioreactor (Fig. S6). Similarly, growth rank was predictable, with significant relationship between predicted rank and actual growth rank (\( \rho = 0.90 \) (Fig. 8B)). During bioreactor cultivations, there were no significant changes in bioprocess parameters. Constant agitation at 200 rpm, with 1vvm airflow rate was sufficient to maintain dissolved oxygen (dO\(_2\)) above 20%. Although no base was added in all cultivations, the pH of the media did not drop below 4.5, the set point. Additionally, there was no correlation between pH of media (\( \rho = 0.23 \) and \( \rho = 0.52 \) for shake flask cultivation and bioreactor cultivation, respectively) prior to inoculation or after growth at 72 h (Fig. S7). Further, the predictive capability of the model was maintained irrespective of the aeration, whether the cultures were grown in baffled or non-baffled flasks or in a bioreactor maintained at high oxygen concentration.

Discussion

We developed a model trained against a structured data set for the cultivation of *C. necator* H16 in a chemically defined medium with fructose as the primary carbon source. Our approach identified significant growth factors and their effects on culture density at 72 h. Fructose, glucose, glycerol and sucrose were used in the preliminary phase with fructose considered as the best carbon source supporting robust growth under heterotrophic conditions. While *C. necator* has been reported to have broad substrate range, its ability to
utilize carbohydrates as a carbon source during heterotrophic growth appears to be limited to fructose and N-acetylglucosamine (8, 11, 29). Utilization of fructose by C. necator is most likely to occur via substrate import by an ATP-binding cassette (ABC-type) transporter, followed by catabolism via 2-keto-3-deoxy-6-phosphogluconate (KDPG), the Entner-Doudoroff pathway. The responsible genes, notably a putative regulator (frcR), ribosome transporters (i.e. frcA, frcC and frcB orthologs in Escherichia coli and Ralstonia solanacearum) and other essential genes facilitating such metabolism are located on chromosome 2 inside gene clusters for glucose, 2-ketogluconate, and glucosamine catabolism (11). In contrast, phosphofructokinase and 6-phosphogluconate dehydrogenase, key enzymes of the Embden-Meyerhoff-Parnas and oxidative pentose phosphate pathways, respectively, appear to be absent from the C. necator genome. It is therefore not surprising that glucose supported little or no growth; yet it has been reported that prolonged cultivation (> 70 h) with glucose as the sole carbon source resulted in a mutant that was able to utilize glucose (30). The mutant acquired such ability by mutating the N-acetylglucoasmine phosphotransferase system, which when deleted led to inability to utilize glucose (30, 31).

Glycerol supported low growth of C. necator. Such low growth was attributed to oxidative stress resulting from high levels of reactive oxygen species (ROS) formed as a result of elevated level of hydrogenases, leading to cell destruction by damaging DNA and other cell components (12).

Formulations of chemically defined media previously described for the cultivation of C. necator are typically prepared without amino acids (3, 7, 14, 16, 22, 32). In this study, bacterial growth is improved when a small number of amino acids (arginine, histidine, leucine and methionine) are included in the medium. The data suggest that they serve as preferred sources of nitrogen, compared to NH₄Cl. Their inclusion resulted in a shorter lag phase compared to media lacking amino acids. Although the effect of histidine is greater than
methionine, arginine and leucine, their action was synergistic. Interestingly, the model indicated that there was an interaction between fructose and amino acid content that indicates the carbon-to-nitrogen balance must be maintained, irrespective of the actual values.

Metal ions, especially divalent cations of d-block transition metals, are important for bacterial growth where they act as metalloenzyme cofactors in living cells. Their presence in high concentrations, however, tends to be detrimental to cells (33). The Cu-histidine interaction highlighted by our model has been well studied across many forms of life (34–40). Copper is known to play diverse structural and catalytic functions owing to its ability to exist either in a reduced (Cu\(^{+}\)) state with affinity for thiol and thioether groups, or an oxidized (Cu\(^{2+}\)) state with more likely coordination for oxygen or imidazole nitrogen group of amino acids including histidine (41). The toxicity of Cu ions can range from catalysis of harmful redox reaction (when bound in weak sites), to disruption of enzyme functions (when bound in strong adventitious sites) (37). Bacteria respond to high levels of copper using different metalloregulators (39, 41). Under high concentration, an integral membrane Cu\(^{+}\) transporter (P\(_{1B}\)-type ATPase)—characterised by histidine-rich domains, and conserved cytosine/histidine motifs within specific transmembrane domains—exports copper from the cytoplasm into the periplasm where in Gram-negative bacteria further detoxification and exportation are carried out by other related enzymes such as Cu\(^{+}\) oxidase, Cu chaperones and cation transporter (39, 41, 42). Further, studies have reported that copper toxicity affected sugar (glucose) utilization in different microorganisms, and its growth inhibition on fungi (Neurospora crassa and Saccharomyces cerevisiae) was due to impaired/suppression of histidine biosynthesis (43, 44).

Does copper stress affect intracellular histidine availability or biosynthesis, or does the presence of sufficient histidine diminish the effect of copper stress? To address this, two important scenarios from the bacterium’s perspective need to be considered. First, is the
bacterium responding to copper toxicity or secondly, is the bacterium responding to low availability of histidine—which is more detrimental to growth? Our results suggest both scenarios are linked and are detrimental to growth. However, given that the bacterium can synthesize some of its amino acids sufficient to sustain growth (11), and copper is established as an important trace metal for the bacterium’s growth (Fig. 4), it is more likely that something else increases the importance of histidine under copper stress making histidine limitation detrimental to growth. Histidine is a strong metal coordinating tridentate ligand with three potential metal-binding sites: the carboxylate oxygen group, the imidazole imido group and the amino nitrogen group (35). The binding to these three sites depends on the Cu ions and the complexes formed. It appears copper (Cu⁺ or Cu²⁺) stress does not impair histidine biosynthesis, rather it increases the demand of histidine in the cell. This is supported by increased concentration of histidine and other amino acids in the xylem sap of Brassica carinata under condition of excess copper (38). It is unclear which of the Cu ions toxicity is diminished by histidine addition. Answering this question will require robust metallomic studies of C. necator, which will help shed light on the mechanism involved in copper export and trafficking in the microbe, a copper-loving bacterium. Further, it will help establish whether the toxicity is inside the cell (i.e. Cu⁺ toxicity) or outside the cell (i.e. Cu²⁺ toxicity). It is still unclear the exact role histidine plays, and with which form of Cu ions, and where in the cell this role is performed. What is clear is that histidine, and some other amino acids, acts like enhancer or coordinator of Cu ions exportation to maintain copper homeostasis.

This study provides insight into the impact of media formulations on growth and cell density of C. necator. The information provided, the small-scale automated experimentation and the statistical approaches undertaken in this study, will inform further efforts aimed at optimizing other C. necator responses. This includes the biosynthesis of polyhydroxyalkanoate, platform chemicals, proteins and other products as well as future
optimization of growth under lithoautotrophic conditions with CO\(_2\) and molecular hydrogen serving as carbon and energy sources, respectively.

**Materials and Methods**

**Bacterial strains.** *Cupriavidus necator* H16 (DSM 428) was obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, German Collection of Microorganisms and Cell Cultures (DSMZ). The bacterial strain was resuscitated on nutrient agar (peptone 5 g/L, and meat extract 3 g/L) according to the supplier’s instructions, and incubated at 30 °C for 48 h.

**Chemicals.** Carbon sources (glucose, fructose, glycerol and sucrose), salts (with exception of MgSO\(_4\).H\(_2\)O and NH\(_4\)Cl), trace metals, amino acids (histidine, leucine and arginine) and some vitamins (thiamine, niacin, pantothenic acid) were obtained from Sigma-Aldrich. The remainder of the vitamins, MgSO\(_4\).H\(_2\)O, NH\(_4\)Cl and methionine were obtained from Duchefa Biochemie B.V., BDH chemicals and Formedium, respectively.

**Consideration of factors.** A comparison of the literature for the use of defined media for *C. necator* growth identified variety in both the nature and concentrations of macroelements and trace elements required for robust cell growth (Table S1). Therefore, components that served as factors for the DOE were carefully selected based on knowledge on general nutritional requirement for bacteria. The macroelements considered were C, N, O, P, Ca, Mg, S, whilst trace elements were Cu, Zn, Fe and Mn. These basic components correspond to the components of one of the existing CDM for *C. necator* (32). This medium (in g/L: 20 fructose, 4 NaH\(_2\)PO\(_4\), 4.6 Na\(_2\)HPO\(_4\), 0.45 K\(_2\)SO\(_4\), 0.39 MgSO\(_4\), 0.062 CaCl\(_2\), 0.5 NH\(_4\)Cl and 1 mL of trace solution containing in g/L: 15 FeSO\(_4\).7H\(_2\)O, 2.4 MnSO\(_4\).H\(_2\)O, 2.4
ZnSO\(_4\).7H\(_2\)O and 0.48 CuSO\(_4\).5H\(_2\)O) served as a starting point for the investigation. Additionally, few amino acids and vitamins (45, 46) were added as factors to determine their effect on *C. necator* growth. Because adjusting the pH of media might result in the addition of extra ions, which in turn might impact on interpretation of how components contribute to growth, pH of each medium was not adjusted but was measured when appropriate. *C. necator* has broad range of applications, ranging from enzymes, platform chemicals and biopolymer syntheses; however, optical density (OD\(_{600\text{nm}}\)) was chosen as the output to gain detailed understanding on how the bacterium responds to different nutritional conditions. Such understanding would guide other process improvements involving *C. necator* as a production host.

**Preparation of media.** Stock solutions of glucose, fructose, sucrose, vitamins, amino acid and each trace metal solution were filter sterilized through 0.22 µm filter, while stock solutions of glycerol, NaH\(_2\)PO\(_4\), Na\(_2\)HPO\(_4\), MgSO\(_4\).H\(_2\)O, NH\(_4\)Cl, K\(_2\)SO\(_4\), CaCl\(_2\).2H\(_2\)O were autoclaved. The trace element working solution contained (g/L): 0.48 CuSO\(_4\).5H\(_2\)O (dissolved in 0.1M HCl), 15 FeSO\(_4\).7H\(_2\)O (freshly prepared during each trace reconstitution from individual stock), 2.4 each of MnSO\(_4\).H\(_2\)O, and ZnSO\(_4\).7H\(_2\)O. A 100 × amino acid stock solution contained (g/L): 12.9 arginine, and 10 each of histidine, leucine, and methionine, while 1000 x vitamin stock solutions contained (g/L): 0.1 pyridoxine, 0.02 folic acid, and 0.05 each of niacin, nicotinamide, pantothenic acid, riboflavin and thiamine. Subsequently, medium components were added from individual stock solutions except for trace elements which were added from reconstituted working solution. Forty-eight well plates were used in all trials, and medium reconstitution in each well was carried out using an automated liquid handling system (Eppendorf epMotion M5073). All stock solutions were prepared using water as a solvent and were further diluted in sterile distilled water unless stated otherwise.
Inoculum preparation and growth measurement. For each experiment, 48 h colonies from LB agar were washed twice in sterile distilled water and diluted to a working inoculum concentration in the range of $10^8$ cfu/mL. The inoculate was further diluted 1:100 in wells containing medium. Forty-eight well microtiter plates were incubated in a rotary incubator at 30 ºC, 180 rpm. Optical density (OD) at 600 nm was measured every 24 h using a Varioskan LUX™ Multimode Microplate reader (Thermo Scientific).

Batch cultivation in a bioreactor system. Large-scale cultivations were carried out in a batch mode using 1 L chemically defined media contained in 2 L capacity fermentors (Applikon ADI fermentation system). During the cultivations, pH was maintained above 4.5, temperature at 30 ºC, agitation at 200 rpm. Dissolved oxygen (dO$_2$) was maintained above 20 % (1, 6) with airflow at 1 vvm (volume of air per volume of medium). No anti-foam agents or base were added during cultivation. Starter culture media were of the same formulations with that used in fermentors, and were prepared by growing 100 mL culture in 250 mL non-baffled flasks to late exponential growth at 30 ºC, 200 rpm for 48 h. Following polarization and calibration of the dO$_2$ probe, fermentors were inoculated with 10 mL (48 h) starter culture and cultivations were monitored on-line and off-line over 72 h. Samples were taken every 24 h for off-line OD$_{600nm}$ measurement.

Data analyses. Experimental designs were created using JMP Pro statistical software (version 13.0) and data from each experiment was analysed using this same software. Graphics were generated using GraphPad Prism 7.0.

Acknowledgements
Christopher C. Azubuike is a Commonwealth Scholar (NGCA-2016-060) funded by Department for International Development (DFID), UK government.

**Competing Interests**

The authors declare no competing interests.

**References**


Hydrogen-driven asymmetric reduction of hydroxyacetone to (R)-1,2-propanediol by Ralstonia eutropha transformant expressing alcohol dehydrogenase from Kluyveromyces lactis. Microb Cell Fact 12:2.


38. Irtelli B, Petrucci WA, Navari-Izzo F. 2009. Nicotianamine and histidine/proline are, respectively, the most important copper chelators in xylem sap of Brassica carinata under conditions of copper deficiency and excess. J Exp Bot 60:269–277.


Table 1 Scoping experiment. A scoping trial was developed to assess the impact of 10 basic components of chemically defined media.

FIG 1 Scoping trials. Growth of *C. necator* H16 in 48-well format with different carbon sources. Each experiment was carried out at low, medium (*n* = 2) or high concentrations of each media component (details can be found in Table 1).

FIG 2 Definitive Screening Design array analysis. A. Definitive screening and 2-Level screening of data within DSD1 were performed. The comparative lengths of the t-ratios for each factor and factor interaction are shown. At high concentration, bars extending to the right have a positive impact on growth, while those extending to the left have a negative impact on growth. Terms deemed significant for model projection are shaded (blue). Abbreviations: Fru., fructose; T.E., trace element mixture; A.A., amino acid mixture; Vit., vitamin mixture. The broken vertical lines indicate threshold (*t* < -1.65 or > 1.65) level at 90% confidence level. B. Summary of main effect plot for each factor. The middle-broken lines are connecting mean growth for each factor settings. Additionally, for significant factors the broken lines above and below that of the mean growth are standard deviation bands. Error bars are S.E. Mean. Analysis is based on two biological replicated arrays (Table S2).

FIG 3 Combined data for DSD1 and DSD2 for key media components. DSD1 (blue boxes) and DSD2 (white boxes). Settings were both DSDs were held at the same concentration are coloured teal. Error bars are S.E. Mean, *n* = 2 biological replicate arrays for DSD1 and DSD2.
FIG 4 Interactions between trace elements and amino acids. A. *C. necator* grown in a complete, chemically defined medium with either amino acids, trace elements or both amino acids and trace elements excluded. B. *C. necator* grown in a complete medium with each of the four trace elements excluded. C-G. *C. necator* grown in a complete, chemically defined medium with or without CuSO$_4$ and with each of the four amino acids excluded. Amino acids solution was added from 100 × stock at 20 mL/L (High) and 10 mL/L (Medium). Trace solution was added from working concentration at 2.4 mL/L. The concentration of the ingredients in the amino acids and trace element solutions are as shown in Table 1. Error bars are S.E. Mean, $n = 2$ biological replicates.

FIG 5 Re-examination of non-significant media components. The impact of components not deemed statistically significant in DSD1 were re-examined under less extreme conditions. A. NaH$_2$PO$_4$; B. K$_2$SO$_4$; C. MgSO$_4$; D. NH$_4$Cl. Experiments were conducted at 5 g/L (open circles) or 20 g/L (closed circles) fructose. Error bars are S.E. Mean, $n \geq 6$.

FIG 6 Experimental validation of model predictions in 48-well plate format and interactions of components of the defined media described by least square model. A. Model predicted values plotted against experimental data for least squares model ($r^2 = 0.87$). B. New experimental data for 1 mL culture volumes overlaid against the original least squares model ($r^2$ for new data only = 0.85). Open circles show the training data set, closed red circles are new data. Three biological replicates were assessed for each prediction. C. Amino acid and trace element interactions. D. Fructose and amino acid interactions. E. Disodium phosphate and fructose interactions. F. Calcium chloride and fructose interactions. G. Ammonium chloride and potassium sulphate. H. Ammonium chloride and monosodium phosphate. Each panel represents a two way-interaction. Red (pink) zone is an area where
OD<sub>600nm</sub> at 72 h fails to reach 1.9, and white zone is where it surpasses 1.9. Each contour grid line represents an OD<sub>600nm</sub> of 0.1 increment. All other media components were kept at concentrations permitting high growth.

**FIG 7** Determining the essential nature of each medium component. A. Prediction profiler of growth at high and low concentration media formulation. The model predicted rank and OD<sub>600nm</sub> 72 h (values in brackets) are shown on the left, media settings for each component are shown in red underneath. All concentrations are in g/L except T.E. and A.A, which are in mL/L. Abbreviations: Fru., fructose; T.E., Trace element mixture; A.A., amino acid mixture. The output is n = 2 biological replicated arrays of definitive screening designs (DSDs). The impact of individually withdrawing each component on growth at: B. High concentration media formulation; C. Low concentration media formulation. The impact of simultaneously withdrawing some components (Na<sub>2</sub>HPO<sub>4</sub>, CaCl<sub>2</sub> and NH<sub>4</sub>Cl) deemed non-essential on growth at: D. High concentration media formulation; E. Low concentration media formulation. Error bars are S.E. Mean, n = 3 biological replicates.

**FIG 8** Experimental validation of model predictions at shake-flask and bioreactor scale. A. Spearman's rank correlations (ρ = 0.87) between model predicted rank and experimental data for 100 mL culture volumes in baffled and non-baffled flasks. B. Spearman's rank correlations (ρ = 0.90) between model predicted rank and experimental data for 1 L bioreactor batch flask cultures.
Table 1 Scoping experiment. A scoping trial was developed to assess the impact of 10 basic components of chemically defined media.

<table>
<thead>
<tr>
<th>Media</th>
<th>Carbon</th>
<th>NaH$_2$PO$_4$</th>
<th>Na$_2$HPO$_4$</th>
<th>K$_2$SO$_4$</th>
<th>MgSO$_4$</th>
<th>CaCl$_2$</th>
<th>NH$_4$Cl</th>
<th>T.E.</th>
<th>A.A.</th>
<th>Vit.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.5</td>
<td>0.1</td>
<td>0.1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>20.25</td>
<td>3.05</td>
<td>3.05</td>
<td>2.41</td>
<td>1.41</td>
<td>0.41</td>
<td>2.41</td>
<td>9.53</td>
<td>2.41</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>40</td>
<td>6</td>
<td>6</td>
<td>4.8</td>
<td>2.8</td>
<td>0.8</td>
<td>4.8</td>
<td>4.8</td>
<td>4.8</td>
<td>4.8</td>
</tr>
</tbody>
</table>

All concentrations are in g/L except trace elements, amino acids and vitamins, which are mL/L. Abbreviations: T.E., Trace element mixture; A.A., amino acid mixture; Vit., vitamin mixture. Trace element working concentration contained (g/L): 15 FeSO$_4$.7H$_2$O, 2.4 MnSO$_4$.H$_2$O, 2.4 ZnSO$_4$.7H$_2$O, and 0.48 CuSO$_4$.5H$_2$O. A 100 × stock amino acid mix contained (g/L): 12.9 arginine, and 10 each of histidine, leucine and methionine. A 1000 × vitamin stock contained (g/L): 0.1 pyridoxine, 0.02 folic acid, 0.05 each of thiamine, riboflavin, niacin, pantothenic acid and nicotinamide. Carbon: fructose, glucose, glycerol or sucrose. The medium trial was performed in duplicate.
Figure 1

Comparison of growth rates under different sugar conditions:
- **Glucose**
  - Low
  - Medium
  - High

- **Glycerol**
  - Low
  - Medium
  - High

- **Sucrose**
  - Low
  - Medium
  - High

- **Fructose**
  - Low
  - Medium
  - High

OD600 nm vs. Time (h)
Figure 2

A

B

Fructose

Na$_2$HPO$_4$

CaCl$_2$

Trace element

Amino acids

Na$_2$HPO$_4$

K$_2$SO$_4$

MgSO$_4$

NH$_4$Cl

Vitamin

OD$_{600nm}$ 72 h

Concentration (g/L)

Concentration (g/L)

Concentration (g/L)

Concentration (g/L)

Concentration (m/L)

Concentration (m/L)

Concentration (m/L)

Concentration (m/L)
Figure 3

- **Fructose**
  - Concentration (g/L) vs. OD$_{600nm}$ 72 h

- **Na$_2$HPO$_4$**
  - Concentration (g/L) vs. OD$_{600nm}$ 72 h

- **CaCl$_2$**
  - Concentration (g/L) vs. OD$_{600nm}$ 72 h

- **Trace elements**
  - Concentration (mL/L) vs. OD$_{600nm}$ 72 h

- **Amino acid**
  - Concentration (mL/L) vs. OD$_{600nm}$ 72 h

Fructose, Na$_2$HPO$_4$, CaCl$_2$, Trace elements, and Amino acid are shown in the figures.
Figure 4

A) Complete - T.E.

B) Amino acids

C) Complete - T.E.

D) Methionine

E) Leucine

F) Arginine

G) Histidine

OD₆₀₀ nm 72 h

Graphs showing the effect of different amino acids on cell growth.

- Complete
- T.E.
- Amino acids
- Methionine
- Leucine
- Arginine
- Histidine

Concentration levels:
- High
- Med

OD₆₀₀ nm 72 h values are indicated for each condition.
Figure 5

A) Concentration of NaH$_2$PO$_4$ vs. OD$_{600nm}$ after 72 hours.
B) Concentration of K$_2$SO$_4$ vs. OD$_{600nm}$ after 72 hours.
C) Concentration of MgSO$_4$ vs. OD$_{600nm}$ after 72 hours.
D) Concentration of NH$_4$Cl vs. OD$_{600nm}$ after 72 hours.
Figure 6 and Figure 7 see separate TIFF
Figure 8

A

Experimen tally
determined rank

0 2 4 6 8 10
Model predicted rank

B

Experimen tally
determined rank

0 2 4 6
Model predicted rank

A B
### A

#### High formula

<table>
<thead>
<tr>
<th>Component</th>
<th>Fru.</th>
<th>NaH₂PO₄</th>
<th>Na₂HPO₄</th>
<th>K₂SO₄</th>
<th>MgSO₄</th>
<th>CaCl₂</th>
<th>NH₄Cl</th>
<th>T.E.</th>
<th>A.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td>OD (600 nm)</td>
<td>1.950071</td>
<td>1.8258</td>
<td>2.07434</td>
<td>2.405</td>
<td>1.405</td>
<td>0.459</td>
<td>0.01</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

#### Low formula

<table>
<thead>
<tr>
<th>Component</th>
<th>Fru.</th>
<th>NaH₂PO₄</th>
<th>Na₂HPO₄</th>
<th>K₂SO₄</th>
<th>MgSO₄</th>
<th>CaCl₂</th>
<th>NH₄Cl</th>
<th>T.E.</th>
<th>A.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td>OD (600 nm)</td>
<td>1.867495</td>
<td>1.5953</td>
<td>2.13969</td>
<td>1.2075</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>10</td>
</tr>
</tbody>
</table>

### B

#### High formula

<table>
<thead>
<tr>
<th>Component</th>
<th>OD₆₀₀nm 72 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>2.0</td>
</tr>
<tr>
<td>Fru.</td>
<td>1.9</td>
</tr>
<tr>
<td>NaH₂PO₄</td>
<td>1.8</td>
</tr>
<tr>
<td>Na₂HPO₄</td>
<td>1.7</td>
</tr>
<tr>
<td>K₂SO₄</td>
<td>1.6</td>
</tr>
<tr>
<td>MgSO₄</td>
<td>1.5</td>
</tr>
<tr>
<td>CaCl₂</td>
<td>1.4</td>
</tr>
<tr>
<td>NH₄Cl</td>
<td>1.3</td>
</tr>
<tr>
<td>T.E.</td>
<td>1.2</td>
</tr>
<tr>
<td>A.A.</td>
<td>1.1</td>
</tr>
</tbody>
</table>

### C

#### Low formula

<table>
<thead>
<tr>
<th>Component</th>
<th>OD₆₀₀nm 72 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>2.0</td>
</tr>
<tr>
<td>Fru.</td>
<td>1.9</td>
</tr>
<tr>
<td>NaH₂PO₄</td>
<td>1.8</td>
</tr>
<tr>
<td>Na₂HPO₄</td>
<td>1.7</td>
</tr>
<tr>
<td>K₂SO₄</td>
<td>1.6</td>
</tr>
<tr>
<td>MgSO₄</td>
<td>1.5</td>
</tr>
<tr>
<td>CaCl₂</td>
<td>1.4</td>
</tr>
<tr>
<td>NH₄Cl</td>
<td>1.3</td>
</tr>
<tr>
<td>T.E.</td>
<td>1.2</td>
</tr>
<tr>
<td>A.A.</td>
<td>1.1</td>
</tr>
</tbody>
</table>

### D

#### High formula

<table>
<thead>
<tr>
<th>Time (h)</th>
<th>OD₆₀₀nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>2.0</td>
</tr>
<tr>
<td>48</td>
<td>2.1</td>
</tr>
<tr>
<td>72</td>
<td>2.2</td>
</tr>
<tr>
<td>96</td>
<td>2.3</td>
</tr>
<tr>
<td>120</td>
<td>2.4</td>
</tr>
</tbody>
</table>

### E

#### Low formula

<table>
<thead>
<tr>
<th>Time (h)</th>
<th>OD₆₀₀nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>2.0</td>
</tr>
<tr>
<td>48</td>
<td>2.1</td>
</tr>
<tr>
<td>72</td>
<td>2.2</td>
</tr>
<tr>
<td>96</td>
<td>2.3</td>
</tr>
<tr>
<td>120</td>
<td>2.4</td>
</tr>
</tbody>
</table>