Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism

Mol, Viviënne; Bennett, Martyn; Sánchez, Benjamín J.; Lisowska, Beata K.; Herrgård, Markus J.; Nielsen, Alex Toftgaard; Leak, David J.; Sonnenschein, Nikolaus

Published in:
Metabolic Engineering

Link to article, DOI:
10.1016/j.ymben.2021.03.002

Publication date:
2021

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):

---

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism

Viviënne Mol, Martyn Bennett, Benjamín J. Sánchez, Beata K. Lisowska, Markus J. Herrgård, Alex Toftgaard Nielsen, David J. Leak, Nikolaus Sonnenschein

PII: S1096-7176(21)00038-0
DOI: https://doi.org/10.1016/j.ymben.2021.03.002
Reference: YMBEN 1774

To appear in: *Metabolic Engineering*

Received Date: 2 February 2021
Accepted Date: 1 March 2021


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Inc. on behalf of International Metabolic Engineering Society.
Genome-scale metabolic modelling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism

Viviënne Mol\(^a\), Martyn Bennett\(^bc\), Benjamín J. Sánchez\(^ad\), Beata K. Lisowska\(^b\), Markus J. Herrgård\(^a,e\), Alex Toftgaard Nielsen\(^a\), David J. Leak\(^e,\text{bc}\), Nikolaus Sonnenschein\(^e,d\)

* These authors contributed equally to this work.

\(^\#\) Joint senior authors

\(^a\) The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark
\(^b\) The Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom
\(^c\) The Centre for Sustainable Chemical Technologies (CSCT), University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom;
\(^d\) Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark
\(^e\) BioInnovation Institute, Copenhagen N, Denmark

† Corresponding authors: Alex Toftgaard Nielsen (atn@biosustain.dtu.dk), David J. Leak (djl36@bath.ac.uk) and Nikolaus Sonnenschein (niso@dtu.dk)

**Author contributions**

Vivienne Mol: Methodology, validation, investigation, formal analysis, data curation, writing. Martyn Bennett: Methodology, investigation, data curation, writing. Benjamín J. Sánchez: Methodology, validation, data curation, editing. Beata K. Lisowska: Conceptualization, methodology, investigation. Markus J. Herrgård: Conceptualization. Alex Toftgaard Nielsen: Conceptualization, editing, supervision. David J. Leak: Conceptualization, editing, supervision. Nikolaus Sonnenschein: Methodology, editing, supervision

All authors read and approved the final manuscript.

Declarations of interest: none
Highlights

- A validated genome-scale metabolic model for *Parageobacillus thermoglucosidasius* NCIMB 11955 is presented.
- The model recapitulates strain engineering designs and predicts carbon utilization.
- Internal flux simulations closely match experimental values.
- The model reveals previously unknown bottlenecks in anaerobic metabolism.
- Iron(III), biotin and thiamin represent a minimal supplied nutrient set required for anaerobic growth.

Abstract

*Parageobacillus thermoglucosidasius* represents a thermophilic, facultative anaerobic bacterial chassis, with several desirable traits for metabolic engineering and industrial production. To further optimize strain productivity, a systems level understanding of its metabolism is needed, which can be facilitated by a genome-scale metabolic model. Here, we present *p-thermo*, the most complete, curated and validated genome-scale model (to date) of *Parageobacillus thermoglucosidasius* NCIMB 11955. It spans a total of 890 metabolites, 1175 reactions and 917 metabolic genes, forming an extensive knowledge base for *P. thermoglucosidasius* NCIMB 11955 metabolism. The model accurately predicts aerobic utilization of 22 carbon sources, and the predictive quality of internal fluxes was validated with previously published $^{13}$C-fluxomics data. In an application case, *p-thermo* was used to facilitate more in-depth analysis of reported metabolic engineering efforts, giving additional insight into fermentative metabolism. Finally, *p-thermo* was used to resolve a previously uncharacterised bottleneck in anaerobic metabolism, by identifying the minimal required supplemented nutrients (thiamin, biotin and iron(III)) needed to sustain anaerobic growth. This highlights the usefulness of *p-thermo* for guiding the generation of experimental hypotheses and for facilitating data-driven metabolic engineering, expanding the use of *P. thermoglucosidasius* as a high yield production platform.

Keywords

Genome-scale metabolic model, systems metabolic engineering, thermophile, *in silico* strain design, anaerobic metabolism, flux balance analysis
1. Introduction

As the global transition away from petroleum-derived feedstocks continues, the need to produce commodity and fine chemicals using sustainable feedstocks has accelerated the interest in establishing microbial bioprocesses with lower environmental footprints. The microbial ‘chassis’ organisms of these bioprocesses have been developed through modern metabolic engineering strategies. Such strategies have enabled the redirection of carbon flux in metabolic pathways of the corresponding microbes towards target products, in what are commonly termed ‘microbial cell factories’.

Without an accurate picture of how cellular metabolism operates as a whole, metabolic engineering strategies can produce flux imbalances, resulting in the accumulation of carbon intermediates, metabolic bottlenecks and/or imbalances in the overall cellular redox ratio. As a result, there can be large upfront costs in microbial strain engineering to ensure economically viable biochemical product yields. To bolster traditional metabolic engineering efforts and help elucidate genotype-phenotype relationships, systems metabolic engineering aims to describe a more holistic representation of cellular metabolism through the integration of stoichiometric modelling and -omics data analyses.

In particular, the advent of cheaper DNA sequencing has given rise to genome-scale metabolic models (GEMs), in silico reconstructions of the metabolic reaction networks of a given organism, derived from its annotated genome sequence. In addition to operating as a knowledge base of metabolic information for a particular organism, GEMs can be used via constraint-based flux balance analysis to simulate carbon flux through metabolic reaction networks, enabling the rapid screening of metabolic behaviours under a range of environmental variables and biological contexts. Through comprehensive in silico predictions of metabolic phenotypes under target conditions, GEMs can also identify potential cellular redox imbalances and metabolic bottlenecks and generate hypotheses for rational, targeted genetic modifications for improved performance. GEMs can even guide...
the construction and optimization of carbon flux for either endogenous or novel heterologous microbial strain pathways towards high yields of desired products\textsuperscript{12,13}.

\textit{Parageobacillus thermoglucosidasius} NCIMB 11955 represents a Gram-positive, facultative anaerobic, thermophilic bacterial chassis with several advantageous traits for industrial bioprocesses when compared to many model bacterial chassis such as \textit{Escherichia coli} and \textit{Bacillus subtilis}\textsuperscript{14,15}. Firstly, the thermophilicity of \textit{Parageobacillus} spp. enables fermentations between 48-70°C\textsuperscript{16,17} at growth rates surpassing other thermophilic organisms\textsuperscript{18} and comparable to that of \textit{E. coli}\textsuperscript{19}. Compared to equivalent mesophilic fermentations, these process temperatures enable a reduction in both the cooling costs of large-scale exothermic fermentations, and a reduction in the risk of contamination from mesophilic microbes\textsuperscript{8,20}.

Furthermore, for industrial bioprocesses aiming for simultaneous saccharification and fermentation (SSF), the thermophilicity of \textit{P. thermoglucosidasius} is complemented by a catabolic versatility. Through extracellular secretions of thermostable amylases\textsuperscript{14}, xylanases\textsuperscript{21–23} and other hemicellulases\textsuperscript{24–26}, \textit{Parageobacillus} spp. are able to metabolise a wide range of C\textsubscript{5} and C\textsubscript{6} sugar monomers. Notably, they are able to transport then metabolize complex hemicellulosic\textsuperscript{26} and cellulosic\textsuperscript{14} polysaccharides derived from hydrolysates of lignocellulosic biomass, potentially reducing the reliance on externally supplied hydrolases involved in lignocellulosic pre-treatment.

A number of synthetic biology tools applicable to \textit{P. thermoglucosidasius} have been devised including: shuttle vectors for reliable transformation\textsuperscript{21,27}, chromosomal integration strategies\textsuperscript{28} promoter and RBS libraries to enable tuneable gene expression and validated reporter genes\textsuperscript{29–32}. Such tools have enabled \textit{P. thermoglucosidasius}, and genetically similar (\textit{Par})\textit{geobacillus} spp., to be used in the production of fuels such as bioethanol\textsuperscript{33,34}, isobutanol\textsuperscript{35} and hydrogen gas\textsuperscript{36,37} and also in fine chemicals including 2-3 butanediol\textsuperscript{38,39}, riboflavin\textsuperscript{40} and isoprenoids\textsuperscript{41}. \textit{Parageobacillus} spp. and \textit{Geobacillus} spp. have also been the source of thermostable variants of industrially useful proteases\textsuperscript{42}, carboxyl esterases\textsuperscript{43–45},
lipases along with a thermostable DNA polymerase I from *G. stearothermophilus* GIM1.543. In spite of these advances, (with the exception of natural end-products of glycolytic metabolism, such as ethanol) none of these engineered pathways have approached their potential maximum yields. In general, they have relied on natural flux to their metabolic precursors and its inherent control. The availability of a reliable GEM would enable a systems metabolic engineering approach of *P. thermoglucosidasius* to address the optimisation of flux through central metabolic pathways to balance the requirements of both production and growth. At present, only one publicly available GEM of a *P. thermoglucosidasius* exists, the related strain *P. thermoglucosidasius* C56-YS93 (denoted iGT736). While comprising 1159 reactions and 1163 metabolites, analysis of iGT736 using the GEM assessment tool Memote developed by Lieven et al. suggests that it currently lacks some fundamental features, including a biomass equation, transport reactions and stoichiometric balance (Supplementary File 1), preventing meaningful application for quantitative analysis. Additionally, a few examples exist of smaller central carbon metabolism scale models derived from experimental ^13^C isotopic tracer experiments. This includes models representing *P. thermoglucosidasius* M10EXG under aerobic and anaerobic growth conditions, and similar *Geobacillus* spp. *G. icigianus* and *Geobacillus* LC300. However, they are less useful for illustrating the scale and complexity of whole cell metabolism.

The newly constructed genome-scale metabolic model of *P. thermoglucosidasius* NCIMB 11955 presented herein (named hereafter as *p-thermo*) represents 917 genes and comprises of 890 metabolites and 1175 reactions across two compartments: cytosolic and extracellular space (representing the medium). After iterative cycles of manual curation, model refinement and analysis with Memote, *p-thermo* exhibits a 100% stoichiometric consistency, 100% charge balance and a 99.9% mass balance. It accurately captures experimentally determined utilization of 22 carbon sources using the sole input of measured
production and consumption rates and is represented in the Systems Biology Markup Language (SBML) compliant format, making it compatible with commonly used constraint-based modeling software such as COBRApy and the COBRA Toolbox v3.0 as well as more specialised software facilitating systems metabolic engineering. Validation of the predictive quality of p-thermo under aerobic, oxygen limited and anaerobic conditions is demonstrated through mapping the resulting in silico fluxes to experimentally determined 13C-flux data obtained from 13C-isotopic labelling experiments of the genetically and metabolically similar P. thermoglucosidasius M10EXG strain. The predictive power of p-thermo is further demonstrated through recapitulation of a metabolically engineered homoethanologenic strain of P. thermoglucosidasius. Lastly, p-thermo was used to investigate the fundamental requirements and metabolic bottlenecks of P. thermoglucosidasius during anaerobic growth. Currently, p-thermo represents the most complete, curated and experimentally validated genome-scale metabolic model for a Parageobacillus sp, and will be a foundational platform for guiding rational metabolic engineering strategies, -omic data integration, and strain optimization to further the potential of P. thermoglucosidasius NCIMB 11955 to operate as microbial chassis for sustainable bioprocesses.
2. Results

2.1 Model reconstruction

The presented genome-scale metabolic reconstruction of *P. thermoglucosidasius* NCIMB 11955 is based on genome sequencing by ERGO™ Integrated Genomics and Sheng et al.\(^5^8\). Genome annotation was performed through the ERGO™ Integrated Genomics suite\(^5^7\) and the RAST annotation server\(^5^9\), followed by gap filling with Pathway Booster\(^6^0\). The reconstruction was extensively manually curated using available literature and databases (KEGG, BRENDA, MetaCyc, MetaNetX and EC2PDB), according to benchmark approaches\(^6^1\). Detailed manual curation and refinement can be followed in Lisowska\(^1^4\) and in the GitHub repository. Specific attention was given to ensure nucleotide, amino acid and co-factor metabolism was accurate based on the current available knowledge. Additionally, extra care was paid to the selection and stoichiometry of the biomass precursors and their respective biosynthetic pathways, with particular focus on a more accurate representation of fatty acid and lipid biosynthesis\(^6^2\).

This metabolic model consists of 890 metabolites, involved in a total of 1175 reactions, encoded for by 917 genes, across two compartments: cytosolic and extracellular space (representing the medium). Manual curation was critical to ensure complete consistency of the model (Supplementary report 1). Central carbon metabolism of the model resembles that of previously reported *Parageobacillus* spp\(^7^,^6^3\) (Figure 1A). Of all reactions, 9.3% are involved in transport or exchange, highlighting the flexibility of the strain to grow on various carbon sources (Figure 1B). Predominantly manual gap filling, based on available literature, was used to ensure correct active metabolic pathways, accounting for 20.6% of the final *p-thermo* reactions. After gap filling, 17 dead end and 22 orphan metabolites remained, resulting in 148 universally blocked reactions (12.6% of total reactions). All dead end and orphan metabolites were manually reviewed, but a lack of conclusive evidence surrounding them with respect to *Parageobacillus* sp. prevented these gaps from being filled with sufficient accuracy and so have been deliberately unmodified until further knowledge is accrued.
The model as well as scripts used in the reconstruction and manual curation are made publicly available through Github, at [https://github.com/biosustain/p-thermo/releases/v1.0](https://github.com/biosustain/p-thermo/releases/v1.0). The model is stored using the community-standard SMBL format (Level 3, FBC Version 2) and can additionally be accessed as Supplementary File 2.

Figure 1: A) Central carbon metabolism map, with several reaction IDs highlighted. For a more detailed overview, see Supplementary File 3. B) The number of reactions in the model for several reaction class types. See section 5 for abbreviations used.

### 2.2 Biomass composition and growth energetics

To capture biological growth in stoichiometric models, a demand reaction referred to as a biomass pseudo-reaction, was added. An overview of how the biomass pseudo-reaction was defined is explained in Materials & Methods, with the final reaction components and associated stoichiometry given in Supplementary Table 1. Energetic parameters were fitted from aerobically grown chemostat experiments. The energy required to maintain cellular homeostasis is reflected in the non-growth associated maintenance (NGAM) and was found to be $3.141 \text{ mmol}_{\text{ATP}}/\text{g}_{\text{DW}}\cdot\text{h}^{-1}$ in *p-thermo*. The growth associated maintenance, (GAM), was estimated as $152.3 \text{ mmol}_{\text{ATP}}/\text{g}_{\text{DW}}$ and reflects the energy needed for cell replication, including macromolecule synthesis. The contribution of polymerization energy, required for
macromolecule synthesis, to the obtained GAM was estimated to be approximately 20% (Supplementary table 2); relatively low compared to previously reported mesophiles (30-40%)\textsuperscript{65-68}. It was previously observed that thermophilic organisms tend to require higher levels of energy for growth and homeostasis at elevated temperatures and thus have a reduced growth efficiency, shown in the high maintenance estimated\textsuperscript{18,69}. This trait of thermophiles makes them valuable hosts for bioproduction as it leads to higher production rates of catabolic products compared to other organisms.

### 2.3 Overview of metabolism

To provide a comprehensive overview, two pathway maps of the model were drawn using Escher\textsuperscript{70} corresponding to central carbon and amino acid metabolism (Supplementary Files 3 and 4) and deposited in the GitHub repository at p-thermo/maps. Traits specific to \textit{Geobacillus} spp. and \textit{P. thermoglucosidasius} NCIMB 11955, presented in the literature, were used to validate the model’s metabolism. Detailed step-by-step decisions that were made, can be followed in the GitHub repository at “p-thermo(notebooks)”. As an example, in central carbon metabolism research has shown that \textit{Geobacillus} spp., unlike many mesophilic \textit{Bacillus} species, lack genes for a 6-phosphogluconolactonase (6PGL), responsible for part of the oxidative pentose phosphate pathway (PPP)\textsuperscript{14}. Instead, the reaction can occur spontaneously and, at thermophilic temperatures, has been hypothesized to be sufficiently rapid to maintain the requisite PPP flux\textsuperscript{71}. The absence of 6PGL was captured in the model, but to reflect the active PPP pathway, a pseudo-reaction was added to allow the complete oxidative PPP to function.

\textit{(Para)geobacillus} species are known to be capable of growth on a wide range of carbohydrates, and have been shown to secrete various polysaccharide degrading enzymes such as xylanases and other hemicellulose degrading enzymes\textsuperscript{14,24–26}. To assess the metabolic capacity of the model, growth on various carbon sources was simulated (Figure 2). The choice of carbon sources was made based on previous qualitative growth experiments which demonstrate a range of sole carbon sources which allow aerobic growth
of *P. thermoglucosidasius* NCIMB 11955\textsuperscript{51}. Additionally, anaerobic growth on these substrates was computationally predicted. In both cases, carbon supply was normalized to 30 Cmol/g \textsubscript{dw}h, to accommodate different polymeric substrate forms being present in the data set. Initially, the model showed no aerobic growth on arbutin, salicin and rhamnose, due to dead-end metabolites being formed as side products in the first steps of their breakdown. Available literature was used to fill the gaps in the catabolic pathways, which enabled aerobic growth on all three carbon sources. Anaerobically, *in silico* growth on arbutin and salicin was unfeasible, as current knowledge suggests that their catabolism is oxygen dependent. Both arbutin and salicin are non-conventional carbon sources and are glycosides, composed of either a hydroquinone or salicyl alcohol functional group attached to glucose, respectively. It is known that metabolism of these glycosides occurs through splitting of the glycosidic bond, with the two functional groups being catabolized individually. With currently available knowledge, the further breakdown of the salicyl alcohol and hydroquinone functional groups is dependent on oxygen, deeming the *in silico* prediction of anaerobic growth unfeasible. As there is little knowledge about microbial catabolism of these carbon substrates, this hypothesis would warrant experimental validation.
Figure 2: Anaerobic (left) and aerobic (right) predicted biomass yields for 22 different carbon sources, for which aerobic growth has been experimentally confirmed\textsuperscript{51}. Carbon substrates were all supplied in the model at 30 Cmol/g\textsubscript{DW}h to account for differences in composition between the carbon sources.

2.4 Assessment of predictive power through \textsuperscript{13}C-flux fitting

Prior to using a genome-scale model for metabolic analyses or \textit{ab initio} predictions, it is critical to validate its predictive power based on previously attained experimental data. This was done by analysis of how well the simulated fluxes of \textit{p-thermo} match published flux distributions of \textit{P. thermoglucosidasius}. \textsuperscript{13}C-isotopic labelling is a standard tool used to elucidate intracellular fluxes in central carbon metabolism, through extensive experimental work and data analysis. Flux variability analysis (FVA) is an \textit{in silico} approach that can allow \textit{ab initio} analysis of metabolism without the need for laborious experimental data\textsuperscript{72}.
Comparing the two data types can give insights into metabolism and allow the generation of hypotheses for metabolic engineering purposes.

To make this comparison, $^{13}$C-flux data from *P. thermoglucosidasius* M10EXG subject to varying oxygen conditions was used to qualitatively assess the predictive quality of *p-thermo*\textsuperscript{7}. Whole proteome analysis (on a sequence basis) of the *P. thermoglucosidasius* M10EXG and NCIMB 11955 strains shows that the ORFs between the two strains are highly similar (Supplementary figure 1, Supplementary table 3). Importantly, considering the metabolic genes that would be captured as reactions in a metabolic model, there are only 11 and 12 unique reactions in *P. thermoglucosidasius* NCIMB 11955 and *P. thermoglucosidasius* M10EXG respectively (Supplementary tables 3, 4 and 5). Therefore, based on the overall metabolic similarity between the two strains, we assume that the $^{13}$C-flux data from *P. thermoglucosidasius* M10EXG can be utilized for a qualitative assessment of *p-thermo*.

In order to test if *p-thermo* can predict intracellular fluxes close to the $^{13}$C-flux data, the measured production and consumption rates were fixed in the model as exchange rates, and internal fluxes were predicted in a sensitivity analysis with FVA. Parsimonious enzyme usage flux balance analysis (pFBA), which has previously been shown to predict fluxes that correlate with experimental measurements, was also performed\textsuperscript{73}. The *in silico* fluxes and pFBA results were mapped to experimentally determined fluxes in aerobic, oxygen limited and anaerobic conditions (Figure 3A, B and C respectively). pFBA showed good correlation to the measured data for each condition (Supplementary Figure 2), and together with FVA showed accurate predictions of the internal central carbon fluxes (Figure 3A, B, and C). Biomass yields (Figure 3D) and oxygen consumption rates (Supplementary Figure 3) were adequately predicted as well. These analyses validate the predictive quality of the created model and highlight the power of using metabolic models for understanding intracellular fluxes when only extracellular consumption or production rates are available.
Figure 3: Results of fixing experimentally measured exchange rates and predicting intracellular flux distributions in aerobic (A), oxygen limited (B) and anaerobic (C) conditions, normalized to the glucose uptake rate. Figure 1A shows the stoichiometry of all the reactions shown on the x-axis. See Section 5 for abbreviations. FVA sensitivity analysis is shown in line ranges. Predicted and measured maximum biomass yields, for a FVA threshold set at 99% of optimum biomass, are shown in (D).
2.5 Recapitulating & interpreting knockout physiology

To evaluate the utility of *p-thermo* for metabolic engineering applications, we recreated previously reported homoethanologenic mutants of *P. thermoglucosidasius* NCIMB 11955 *in silico*. Cripps *et al.* engineered lactate dehydrogenase (*ldh*) and pyruvate formate lyase (*pfl*) knockouts *in vivo*, and supplemented their ethanol yields with an upregulation of pyruvate dehydrogenase expression (PDH<sub>up</sub>). Using *p-thermo*, the wild type (WT), Δ*ldh* and Δ*ldh*Δ*pfl*(PDH<sub>up</sub>) strains were recreated, as stoichiometric modeling cannot distinguish between upregulated expression levels (i.e. between Δ*ldh*Δ*pfl* and Δ*ldh*Δ*pfl* PDH<sub>up</sub>).

Exchange rates of the main fermentation metabolites were predicted using *p-thermo* and their accuracy evaluated based on measured data (Figure 4). In performing the analysis, two distinct thresholds for Flux Variability Analysis (FVA) were selected, 95% and 99% of optimum biomass production<sup>72</sup>, to assess the flexibility of exchange rates to the simulated conditions.

The performed simulations show a substantial discrepancy between predicted and measured yields in the WT and Δ*ldh* strains, whereas simulations tightly match the measured yields in the Δ*ldh*Δ*pfl* PDH<sub>up</sub> strain. Still, the mismatch between the experimental and *in silico* data, *p-thermo* can be used to understand metabolic branch points. The main discrepancy observed lies in the lactate and formate yields for the WT and Δ*ldh* strains, which can be traced to the cellular decision of what to do with the synthesized pyruvate (Figure 4). In this regard, there are three options: 1) conversion into lactate by lactate dehydrogenase (LDH), 2) anaerobic conversion into acetyl-CoA by pyruvate formate lyase (PFL) or 3) aerobic conversion into acetyl-CoA by pyruvate dehydrogenase (PDH). In both the WT and Δ*ldh* strain, *p-thermo* showed flux from pyruvate to acetyl-CoA to be exclusively carried through PFL, fitting with experimental expectations under anaerobic conditions due to high [NADH]<sup>74</sup>. Additionally, the conversion of pyruvate into acetate and ethanol results in one additional ATP per glucose, compared to converting pyruvate into lactate<sup>75</sup>. Therefore, from a stoichiometric perspective, *p-thermo* predicts this to be the most optimal pathway for
growth, explaining the high concentrations of formate, ethanol and acetate predicted in the simulation.

However, this was not observed in the experimental yields, presumably because of subtle differences in dissolved oxygen availability in the experimental setup that influence multiple levels of regulation in vivo, intrinsically not accurately captured by stoichiometric models. In the experimental dataset, undefined oxygen limited conditions were used in which a gradual decline in available dissolved oxygen concentration would have occurred during growth, whereas simulations were performed under anaerobic conditions. Under oxygen-limited conditions, PDH is expressed in the wild type P. thermoglucosidasius, where PFL is typically only active under completely anaerobic conditions. The transition of physiological states in response to decreasing oxygen availability results in excess NADH and creates a redox imbalance in the cell which is alleviated through production of lactate as the production of formate by PFL is restricted. This could explain the discrepancy between the experimentally measured low formate and high lactate production in the WT strain and the prediction by p-thermo. In the LDH knockout at low dissolved oxygen conditions, which prevents PFL activity, PDH instead predominantly carries flux to acetyl-CoA. In this instance, in order to maintain cellular redox balance, the Δldh cells increase the produced ethanol/acetate ratio. This picture highlights the complexity of cellular and enzymatic regulation that is poorly captured in stoichiometric models, as well as the difficulty in simulating uncontrolled environments accurately.

However, the performed simulations can still be used to visualize and understand the burden that lactate production can have on cellular growth. The inability to induce PFL at moderate levels of oxygen limitation, puts a larger reliance on fermentative metabolism to lactate, providing less energy. We used p-thermo to investigate the possible impact this has. First, all measured exchange rates for the three strains were fitted to the model and used in subsequent determination of predicted biomass yields. This showed that the model is physiologically capable of capturing the measured data, albeit with a lower predicted
biomass yield than was experimentally measured, suggesting that stoichiometrically sub-optimal fermentation pathways were active in vivo (Supplementary Figure 4A). Finally, the effect of increasing lactate production on biomass yield was computed, showing the energetic loss that occurs from lactate production (Supplementary Figure 4B). Overall, this highlights the importance of complex regulation in dictating in vivo metabolism, over pure stoichiometric optima per se.

Figure 4: Comparison of in silico predictions of fermentation product yields in three engineered strains with experimentally determined data from Cripps et al.\textsuperscript{33}, when solely the carbon uptake rate and knockouts were fixed in the model. Yield (g_/g_s) is shown for predicted and measured exchange rates. Each panel highlights a different strain: wild type (WT), Δldh and ΔldhΔpfl (PDHup). Two varying thresholds for FVA were run: 95% and 99% of the optimum biomass production. The in silico predicted biomass yield (Y_m) for 99% of the optimum biomass production is shown for each condition.

2.6 Genome-scale metabolic modeling allows the elucidation of metabolic bottlenecks

The availability of a comprehensive GEM can also facilitate the elucidation of metabolic bottlenecks and identification and optimization of chemically defined growth media\textsuperscript{77}. Thus, \textit{p-thermo} was used to help resolve known issues of the anaerobic metabolic physiology of \textit{P. thermoglucosidasius}. Although \textit{P. thermoglucosidasius} is clearly capable of classical mixed acid fermentation and shows elements of a regulated aerobic-anaerobic switch as revealed by transcriptomic analysis\textsuperscript{78} (although the aerobic respiratory electron transport chain
remained active in an oxygen-scavenging state under fermentative conditions), it has long
been known that growth under anaerobic conditions on existing minimal defined growth
medias requires additional supplements in comparison to growth under aerobic conditions.
Typically, this was resolved by supplementation with a small amount of oxygen or yeast
extract. Therefore, here we used simulations of p-thermo to find a minimal set of
defined nutrients that can achieve anaerobic growth of P. thermoglucosidasius. As a first
observation, when fed true minimal, anaerobic medium, the model predicted no growth, in
accordance with experimental observations. However, fermentative energy generation was
observed, which highlights that oxygen requirement comes from critical secondary
metabolites or cofactors that cannot be synthesized anaerobically, which is corroborated by
previous observations. By minimizing the oxygen uptake in the model, a critical reaction set
requiring oxygen was generated (Table 1). This analysis highlighted a complex combination
of components that cannot be synthesized anaerobically: thiamine, biotin, folate, vitamin
B12, spermine, spermidine and hemin. Additionally, iron(III) must be available in the medium
to allow porphyrin biosynthesis.

As expected, in silico supplementation of these components rescued anaerobic growth,
providing a combination of candidates for experimental validation. P. thermoglucosidasius
DSM 2542 was obtained from the DSM stock center. While DSM 2542 carries several
mutations (11 SNPs and 2 indels) compared to the NCIMB 11955 sequence, the only
differences found in coding sequences are annotated as hypothetical proteins. Thus, it is
expected that they would not impact the predictive outcome of the model between the two
strains. The simulated essential components were experimentally added together in trace
amounts to form a supplementation mix (see Materials and Methods); to assess if it would
allow anaerobic growth. This was compared to Wolfe’s vitamin solution, a commonly used
mix of vitamins in base thermophilic minimal medium (TMM). It should be noted that
Wolfe’s vitamin solution contains thiamin, biotin, folate and vitamin B12, amongst other
nutrients, and that TMM contains trace amounts of iron(III). To uncover the minimal sets of
components needed to rescue anaerobic growth, eight different conditions were tested, all composed of base TMM with 10 g/l glucose: 1) no added nutrients, 2) 0.2% yeast extract, 3) biotin, 4) thiamin, 5) biotin and thiamin, 6) Wolfe’s vitamins, 7) Supplementation mix and 8) Wolfe’s vitamins plus the unique components of the supplementation mix (spermine, spermidine and heme) (Figure 5, Supplementary figure 5).

Table 1: Overview of critical reactions that require oxygen to allow growth in p-thermo.

<table>
<thead>
<tr>
<th>Reaction ID</th>
<th>EC-code</th>
<th>Reaction</th>
<th>In silico requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>1.11.1.6 &amp; 1.11.1.21</td>
<td>Catalase-peroxidase</td>
<td>Glyoxylate and dicarboxylate metabolism</td>
</tr>
<tr>
<td>BLUB</td>
<td>1.13.11.79</td>
<td>5,6-dimethylbenzimidazole synthase</td>
<td>Vitamin B12 biosynthesis</td>
</tr>
<tr>
<td>GLYCTO1</td>
<td>1.1.3.15</td>
<td>(S)-2-hydroxy-acid oxidase</td>
<td>Folate biosynthesis</td>
</tr>
<tr>
<td>PPPGO_1</td>
<td>1.3.3.4</td>
<td>Protoporphyrinogen oxidase</td>
<td>Porphyrin (heme) biosynthesis</td>
</tr>
<tr>
<td>FERO</td>
<td>1.16.3.1</td>
<td>Ferroxidase</td>
<td>Porphyrin (Fe$^{3+}$) metabolism</td>
</tr>
<tr>
<td>ACD0</td>
<td>1.13.11.54</td>
<td>Acireductone dioxygenase</td>
<td>Aliphatic polyamine biosynthesis</td>
</tr>
<tr>
<td>GLYHOR</td>
<td>1.4.3.19</td>
<td>Glycine oxidase</td>
<td>Thiamine biosynthesis</td>
</tr>
</tbody>
</table>

Experimental observations suggested that a combination of thiamin, biotin and iron(III) were the minimal required supplementation set needed to sustain anaerobic growth, as no difference was observed when additional defined supplementation was added (Figure 5). Yeast extract also contains significant amounts of amino acids and other components and so provides an additional growth advantage, as expected. However, this highlights a discrepancy with the model predictions, as a larger minimal supplementation set was originally predicted (Table 1). Finally, as expected, base TMM can support aerobic growth, at a maximum rate of 0.267± 0.021 h$^{-1}$ (Supplementary figure 5A), confirming the synthesis of the critical components in the presence of oxygen.

There are several reasons that can explain the differences between the in silico and experimentally determined minimal supplementation set. The incomplete understanding of thermophilic life introduces additional levels of complexity that are typically not captured by automatic annotation pipelines dependent on predominantly mesophilic datasets leading to errors in the annotation of thermophilic traits. For example, genomes of thermophilic
organisms show a correlation with higher G/C content, less intergenic regions and a higher functional stability (reflected by the lower ratio of non-synonymous to synonymous substitutions over time). Additionally, thermostable proteins can have significantly altered structure compared to their mesophilic counterparts performing the same reaction, confounding homology-based annotation. Through the observed discrepancies, we can unveil additional insights of anaerobic metabolism of *P. thermoglucosidasius*.

In the first place, the *in silico* dependence on vitamin B12 highlights the inaccuracy of annotation pipelines. Vitamin B12 synthesis is classically divided into two routes: canonical (aerobic) and non-canonical (anaerobic). Although the genome annotation of *P. thermoglucosidasius* NCIMB 11955 reveals parts of either pathway, neither is complete. It has been proposed that possibly a novel, blended pathway may be present; however, this may arise from incorrect annotations based on lacking knowledge of thermophilic vitamin B12 biosynthesis genes. The possibility to grow without vitamin B12 supplementation does highlight both an aerobic and anaerobic functional pathway in *P. thermoglucosidasius* NCIMB 11955. To further understand the *de novo* biosynthesis of vitamin B12, experimental validation would be required.

In *p-thermo*, the *in silico* oxygen requirement for spermine and spermidine biosynthesis comes from the downstream recycling of a biosynthetic by-product: 5'-methylthioadenosine (5-MTA). 5-MTA recycling is also important in a novel, oxygen independent MTA-isoprenoid shunt, involved in the methionine salvage pathway. This pathway has been characterized in *Rhodospirillum rubrum* and orthology analysis highlights the possible presence of parts of this pathway in various facultative anaerobic Bacillus spp. This presents the possibility of an alternate 5-MTA recycling pathway, explaining the independence of anaerobic growth to spermine or spermidine addition.

Similarly to spermine and spermidine, the oxygen requirement in the *in silico* folate biosynthesis pathway stems from the formation of glycolaldehyde as a side product, which is further oxidized to glyoxylate. The *in silico* oxidase responsible for glyoxylate formation
requires oxygen (EC 1.1.3.15). However, reports show that *Moorella thermoacetica*, a thermophilic obligate anaerobe, can grow on glycolate through the formation of glyoxylate, highlighting the possibility for a (to date) unknown alternate electron acceptor\textsuperscript{92,93}

Finally, heme is suggested to be synthesized in *P. thermoglucosidasius* from glycine using a 5-aminolevulinic acid synthase\textsuperscript{51} and notably using an oxygen-dependent protoporphyrinogen oxidase. Both *E. coli* and *B. subtilis* have an oxygen independent coproporphyrinogen-III oxidase (hemN), known to be responsible for anaerobic heme biosynthesis, using other electron acceptors such as fumarate, or nitrate over oxygen\textsuperscript{94–97}. While the current genome annotation of *P. thermoglucosidasius* NCIMB 11955\textsuperscript{58} suggests that only the oxygen dependent path is present, the data presented herein suggest that supplementation with hemin is not required for growth (Figure 5). One possible explanation for this discrepancy can be found when performing a tBLASTn with the *B. subtilis* hemN (NCBI accession CAB61616) against the *P. thermoglucosidasius* NCIMB 11955 genome. This highlighted a significant hit (CP016622 region 3448674..3449762, 52% identity, E-value: $10^{-115}$). This suggests the possibility that some form of this oxygen independent heme biosynthesis route could also be present highlighting the need to better understand the metabolism of non-model organism chassis.

This identification of the minimal, defined anaerobic medium highlights how GEMs can be used to facilitate experimental hypotheses, where previous hypotheses have failed. The result, a defined minimal anaerobic medium is valuable for further investigation into anaerobic metabolism through \textsuperscript{13}C-characterization studies, where defined media are critical. Additionally, this identification of a series of components which support anaerobic growth of *P. thermoglucosidasius* at a minimal medium level can further help inform the development of industrial growth media for other microbial chassis used in anaerobic bioprocesses improving growth and chemical product yields.
Figure 5: Experimental growth rates calculated and maximum observed absorbance values when *P. thermoglucosidasius* NCIMB 11955 was grown anaerobically in a microtiter plate reader in TMM base medium, supplemented with various nutrients, as indicated. Dashed line indicates inoculation absorbance, when an inoculation optical density of 0.05 was used.
3. Discussion

*Parageobacillus* spp. represent valuable microbial chassis for metabolic engineering and fermentative bioproduction. Many advantages derive from their thermophilic character, with additional advantages coming from species specific traits. However, to further develop *Parageobacillus* spp. into fully optimized microbial cell factories, additional in-depth and systems level understanding of metabolism is required, for which omic analyses and genome scale metabolic models are critical. Currently, various automatic pipelines exist for generating metabolic models on the sole basis of a genome sequence. Yet, for thermophilic organisms, significant faults resulting from automatic annotation pipelines are evident, as these are based on predominantly mesophilic datasets. Thermophilic genomes show different characteristics, the effect of which on metabolism is still poorly understood, making translation into a predicted function difficult. Thus, significant manual curation is needed in the generation of GEMs for thermophilic organisms, which is limited by the availability of knowledge on thermophilic metabolism. This is reflected by the relatively high need for manual gap-filling, and a high resulting percentage of blocked reactions in *p-thermo*. To increase the understanding of genotype-phenotype relationships in thermophilic hosts, the availability of a GEM acts as a considerable step facilitating systems level studies. As a result, with more knowledge arising, iterative rounds of model improvement are possible.

Therefore, in this work, we developed *p-thermo*, to date the most complete, curated and validated genome-scale metabolic model for a facultative anaerobic *Parageobacillus* sp. In it, genomic and biochemical knowledge were combined into a single powerful knowledge base, providing a critical tool for data-driven metabolic engineering, -omic data integration, process design and optimization. The model accurately captured substrate usage *in silico*, showing the metabolic flexibility of the strain for production with alternative carbon sources (Lisowska, 2016). Furthermore, $^{13}$C-isotopic data verified the quality of *p-thermo* for predicting central internal fluxes of the model, when solely production and consumption rates are measured, a common practice when evaluating metabolic engineering designs.
Going beyond validation, p-thermo was used to provide more in-depth analysis of previously reported metabolic engineering approaches. Initial p-thermo simulations did not completely match experimental data, as stoichiometric models are incapable of capturing complex levels of regulation that play a dominant role in in vivo metabolism. Nonetheless, p-thermo was used to investigate the pyruvate branch point in central metabolism and allowed additional insights into metabolic flux distributions in various genetic backgrounds. With p-thermo further insights into metabolism can be gained, allowing improved targeted metabolic engineering in subsequent designs.

Finally, we used p-thermo to generate hypothesis driven experiments to alleviate a bottleneck in anaerobic metabolism, where previous experimental design was unsuccessful. Doing so revealed fundamental insights into the metabolism of P. thermoglucosidasius, and also demonstrated that significant knowledge gaps still exist. This analysis, in combination with the $^{13}$C-based verification, highlights an additional obstacle in working with thermophilic GEMs, where annotation pipelines are less precise: information on central carbon metabolism can be inferred with relative accuracy, whereas peripheral metabolic pathways are not significantly understood and require further systems-level investigation.

Overall, p-thermo, together with other systems level and omics based approaches, act as a tool to improve our understanding of genotype-phenotype relationships. Genome-scale metabolic models are in this way a critical part of an iterative cycle, and are essential to the use and efficacy of thermophilic hosts for metabolic engineering and industrial bioproduction.
4. Materials & Methods

4.1 Model construction & curation

Genome sequencing of Parageobacillus thermoglucosidasius NCIMB 11955 was initially performed by ERGOTM Integrated Genomics (funded by TMO Renewables Ltd) and subsequently updated using the published P. thermoglucosidasius NCIMB 11955 genome sequence (NCBI accession CP016622[chromosome], CP016623[pNCI001], and CP016624[pNCI002]). Genome annotation was performed through the ERGOTM Integrated Genomics suite and the RAST server. Pathway Booster was used for gap filling, resolving gaps through comparisons with evolutionarily-related genomes. Upon base construction of the model, further manual curation was done following standard procedures. To do so, missing information was primarily obtained from literature and using various databases: BRENDA, EC2PDB, KEGG, MetaNetX or MetaCyc. Whenever information on P. thermoglucosidasius was lacking, available references from other (Para)geobacillus spp or Bacillus spp were added. All further manual curation and refinement can be found in the GitHub repository. Model improvement was ensured by running Memote after each modification.

4.2 Biomass composition and growth energetics

To model growth, a biomass pseudo-reaction was added to the model. The reaction pools metabolites needed for growth into a biomass metabolite. Base biomass composition was previously determined experimentally according to reported practices. Lipid composition was obtained from previous reports, and was incorporated into the model according to a restrictive approach, in which a determined acyl chain length is assumed for all lipid species. Further fine-tuning of biomass composition was performed based on available enzymatic and metabolic requirements of the strain, with case-by-case justification given in the GitHub repository. Critical metabolites known to be required for catabolic functions of essential enzymes, such as heme, were added at trace stoichiometries based on knowledge from related organisms and scaled to ensure that all biomass components added up to 1
g/g_{dw}^{99}. Growth energetics (ATP cost of growth-associated maintenance and ATP requirement for non-growth associated maintenance) were estimated by minimizing the prediction error of the specific substrate consumption rate and the specific growth rate of glucose fed, aerobic chemostats\textsuperscript{51,61}. The P/O ratios were obtained from the given data for \textit{B. subtilis}\textsuperscript{100}. The contribution of polymerization of each metabolite type to the total growth associated maintenance was estimated based on previously reported polymerization energies\textsuperscript{65}.

4.3 Transport reactions

The model has two compartments: extracellular and intracellular. Transport reactions were inferred from genome annotations and homology to known transporters. Additionally, knowledge about growth on various substrates was used to validate the presence of the corresponding transporters.

4.4 Stoichiometric modeling & applied constraints

In traditional flux balance analysis\textsuperscript{10}, reaction stoichiometries are converted into a stoichiometric matrix (\(S\)), with \(m \times n\) dimensions, where \(m\) represents the various metabolites and \(n\) represents the number of reactions. Coefficients in the matrix are either positive or negative, reflecting production and consumption, respectively. Stoichiometric modeling works under the assumption of a pseudo-steady-state, represented as:

\[ S \cdot v = 0 \]

Where the vector \(v\) contains the fluxes of all reactions, given in units of mmol/g_{dw}h. As there are more metabolites than reactions, to solve this underdetermined system, linear programming is used by formulating an objective function (\(z\)), per default set as biomass accumulation. Reversibility of reactions is set based on thermodynamic prediction and a default medium is defined; in the case of \textit{p-thermo}, minimal medium, with D-glucose as default carbon source is used.
Quantification of metabolic fluxes was performed using flux variability analysis (FVA)\(^7\). When running FVA, a threshold below the optimum is used to represent the metabolic freedom that is given to a model. In this study, a sensitivity analysis was run with FVA thresholds from 90 to 99\%, to evaluate at what sensitivity level the model better matches the experimental data. Additionally, parsimonious flux balance analysis (pFBA) was run, by conducting a bilevel linear programming optimization that computes the optimum (growth) solution of the network, whilst minimizing the sum of all fluxes\(^7\). In doing so, this optimization predicts the most stoichiometrically efficient pathway set, and captures the maximum biomass per unit flux objective that has previously been described to be well supported by proteomic and transcriptomic data\(^7,10\).

4.5 Genome comparison

For an unbiased genome comparison, the *P. thermoglucosidasius* NCIMB 11955 genome was obtained from NCBI (Accession: CP016622), and the *P. thermoglucosidasius* M10EXG genome was obtained from the Integrated Microbial Genome database (ID 2501416905). Genome annotation was performed using RASTk\(^9\), after which the two proteomes were compared through blast bi-directional best hits to create a homology matrix between the strains, based on a published pipeline\(^10\). To filter for metabolic genes, any ORF associated to a predicted EC code was considered metabolic. The exact workflow can be followed in the GitHub repository.

4.6 Experimental procedures

The *P. thermoglucosidasius* NCIMB 11955 (DSM2542) strain was obtained from DSMZ\(^7\). The strain was grown in either 2SPY medium or base thermophile minimal medium (TMM), modified from Fong *et al.*\(^8\). 2SPY was used for a first preculture, and contains per liter, 16g soy peptone, 10 g yeast extract and 5 g NaCl, adjusted to pH 6.8. Base TMM contains, per liter: 930 ml Six salts solution (SSS), 40 ml 1M MOPS (pH 8.2), 10 ml 1 mM FeSO\(_4\) in 0.4 M tricine, 10 ml 0.132 M K\(_2\)HPO\(_4\), 10 ml 0.953 M NH\(_4\)Cl, 0.5 ml 1M CaCl\(_2\) and trace element
solution, adjusted to a final pH of 6.8. SSS contains, per 930 ml: 4.6 g NaCl, 1.35 g Na$_2$SO$_4$, 0.23 g KCl, 0.037 g KBr, 1.72 g MgCl$_2$·6 H$_2$O and 0.83g NaNO$_3$. The trace element solution contained, per liter, 1g FeCl$_3$·6 H$_2$O, 0.18 g ZnSO$_4$·7 H$_2$O, 0.12 g CuCl$_2$·2 H$_2$O, 0.12 g MnSO$_4$·H$_2$O and 0.18 g CoCl$_2$·6 H$_2$O. D-glucose to a final concentration of 10g/L was added to the base TMM.

When indicated, the base TMM was supplemented with one of the following, to the indicated final concentrations: 0.2% (w/v) yeast extract, 2 µg/L biotin, 5 µg/L thiamine-HCl, 1x Wolfe’s vitamins, or 1x supplementation mix. 1000x Wolfe’s vitamins consist of, per liter, 10 mg pyridoxine hydrochloride, 5.0 mg thiamine-HCl, 5.0 mg riboflavin, 5.0 mg nicotinic acid, 5.0 mg calcium D-(+)-pantothenate, 5.0 mg p-aminobenzoic acid, 5.0 mg thioctic acid, 2.0 mg biotin, 2.0 mg folic acid and 0.1 mg vitamin B12. The 1000x supplementation mix contained, per liter, 2.0 mg biotin, 5.0 mg thiamine-HCl, 2.0 mg folic acid, 0.1 mg vitamin B12, 127 µg/L spermidine, 174 µg/L spermine tetrahydrochloride and 0.7 mg hemin.

The aerobic cultures were inoculated to a starting OD$_{600}$ of around 0.05, after an overnight culture on the base TMM medium. Growth was monitored through OD$_{600}$ measurements, during growth at 60°, 200 rpm in baffled shake flasks. Anaerobic medium was prepared similarly to aerobic medium, but 1 µg/L resazurin was added to ensure complete anaerobic conditions. The medium was flushed with nitrogen gas prior to use. All anaerobic work was performed in an anaerobic chamber. Overnight cultures were run in anaerobic serum flasks at 60°C, 200 rpm and used to inoculate a microtiter plate to a final OD of 0.05 in 200 µL volume. After sealing, the OD$_{600}$ was measured every 15 minutes for 10 hours in a Biotek Epoch2 microplate spectrophotometer, placed in an anaerobic chamber (run at 60°C, with linear shaking).
5. Abbreviations

Metabolites

13dpg 3-Phosphoglyceroyl phosphate
2pg 2-phosphoglycerate
3pg 3-phosphoglycerate
6pgl 6-phosphogluconolactone
ac acetate
acald acetaldehyde
accoa acetyl-CoA
actp acetyl phosphate
akg α-ketoglutarate
asp aspartate
cit citrate
dhp dihydroxyacetone phosphate
e4p erythrose 4-phosphate
etoh ethanol (etoh)
f6p fructose-6-phosphate
fdp fructose 1,6-bisphosphate
for formate
fum fumarate
g3p glyceraldehyde-3-phosphate
g6p glucose-6-phosphate
glc glucose
gly glycine
icit iso-citrate
lac L-lactate
mal malate
oaa oxaloacetate
pep phosphoenolpyruvate
phe phenylalanine
pyr pyruvate
r5p ribose-5-phosphate
ru5p ribulose-5-phosphate
s7p sedoheptulose 7-phosphate
ser serine
succ succinate
succoa succinyl-CoA
udpg uridine diphosphate glucose
xu5p xylulose-5-phosphate
<table>
<thead>
<tr>
<th>Reaction Code</th>
<th>Enzyme Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKr</td>
<td>Acetate kinase</td>
</tr>
<tr>
<td>ACONTa</td>
<td>Aconitase</td>
</tr>
<tr>
<td>ALCD2x</td>
<td>Alcohol dehydrogenase</td>
</tr>
<tr>
<td>ASPTA</td>
<td>Aspartate transaminase</td>
</tr>
<tr>
<td>CS</td>
<td>Citrate synthase</td>
</tr>
<tr>
<td>DDPA</td>
<td>3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase</td>
</tr>
<tr>
<td>ENO</td>
<td>Enolase</td>
</tr>
<tr>
<td>FBA</td>
<td>Fructose-bisphosphate aldolase</td>
</tr>
<tr>
<td>FUM</td>
<td>Fumarate hydratase</td>
</tr>
<tr>
<td>G6PDH2r</td>
<td>Glucose 6-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GALUi</td>
<td>UTP-glucose-1-phosphate uridylyltransferase</td>
</tr>
<tr>
<td>GHMT</td>
<td>Glycine hydroxymethyltransferase</td>
</tr>
<tr>
<td>GLCptts</td>
<td>Glucose phosphotransferase transporter</td>
</tr>
<tr>
<td>GLUSy</td>
<td>Glutamate synthase</td>
</tr>
<tr>
<td>ICL</td>
<td>Isocitrate lyase</td>
</tr>
<tr>
<td>LDH_L</td>
<td>L-lactate dehydrogenase</td>
</tr>
<tr>
<td>MDH</td>
<td>Malate dehydrogenase</td>
</tr>
<tr>
<td>PC</td>
<td>Pyruvate carboxylase</td>
</tr>
<tr>
<td>PDH</td>
<td>Pyruvate dehydrogenase</td>
</tr>
<tr>
<td>PFL</td>
<td>Pyruvate formate lyase</td>
</tr>
<tr>
<td>PGCD</td>
<td>Phosphoglycerate dehydrogenase</td>
</tr>
<tr>
<td>PGI</td>
<td>Glucose-6-phosphate isomerase</td>
</tr>
<tr>
<td>PGK</td>
<td>Phosphoglycerate kinase</td>
</tr>
<tr>
<td>PPCK</td>
<td>Phosphoenolpyruvate carboxykinase</td>
</tr>
<tr>
<td>PRPPS</td>
<td>Phosphoribosylpyrophosphate synthetase</td>
</tr>
<tr>
<td>PSCVT</td>
<td>3-phosphoshikimate 1-carboxyvinyltransferase</td>
</tr>
<tr>
<td>SUCDi</td>
<td>Succinate dehydrogenase</td>
</tr>
<tr>
<td>SUCOAS</td>
<td>Succinyl-CoA synthetase</td>
</tr>
<tr>
<td>TALA</td>
<td>Transaldolase</td>
</tr>
<tr>
<td>TKT1</td>
<td>Transketolase</td>
</tr>
<tr>
<td>TKT2</td>
<td>Transketolase</td>
</tr>
</tbody>
</table>

701
**Availability of data and materials**

The metabolic model, scripts and corresponding datasets generated during the study are all freely available, under an Apache 2.0 license, at the GitHub repository: https://github.com/biosustain/p-thermo/releases/v1.0.

**Funding**

VM was funded by the Novo Nordisk Foundation through NNF18CC0033664; MB was supported by the EPSRC EP/L016354/1; BJS, MJH and NS acknowledge funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 686070; NS furthermore acknowledges support from the Novo Nordisk Foundation (NNF17SA0031362); BKL was funded by a BBSRC-CASE studentship (Project code 1100372); ATN was supported through NNF16OC0021814 and NNF20CC0035580; DJL also acknowledges support from BBSRC (BB/J001120/2).

**Acknowledgements**

We thank Dr. Shyam Maskapalli for help with biomass composition analysis.
6. References


85. Lusk, B. G. Thermophiles; or, the Modern Prometheus: the importance of extreme


Supplementary Figures

Supplementary figure 1: Whole proteome comparison between *P. thermoglucosidasius* NCIMB 11955 and *P. thermoglucosidaius M10EXG*, for all ORFs (A) and when filtered for metabolic genes (B, i.e. genes with an EC number associated to them). Supplementary tables 4 and 5 list the unique metabolic ORFs between the two strains.
Supplementary figure 2: Correlation between pFBA analysis of the model and experimentally derived data, normalized to glucose uptake rate, in aerobic, oxygen limited and anaerobic conditions. Errors for measured fluxes and variability in pFBA fluxes are shown. A linear fit has been applied to assay correlation, with the R value indicated per condition.
Supplementary figure 3: Predicted oxygen consumption rates for the three conditions, when measured exchange rates of fermentation products were fit to the model.
Supplementary figure 4: A) result of computing predicted in silico biomass yield, when measured exchange rates, carbon uptake rates and genetic manipulations (i.e. knockouts) have been introduced. B) Effect of lactate production on biomass yield when all other measured exchange rates are fixed for the WT strain. The dotted line indicates the measured lactate production in these conditions.
Supplementary figure 5: A) Aerobic shake flask experiment of *P. thermoglucosidasius* NCIMB 11955 on TMM base medium. Shaded area shows standard deviation between three biological replicates. B) Anaerobic growth curves of *P. thermoglucosidasius* NCIMB 11955 grown on base TMM supplemented with various nutrients. Experiment was performed in a microtiter plate reader, and shaded area represents standard deviation of measurements over quadruplicates. (YE = yeast extract)
### Supplementary Table 1: Stoichiometry of the biomass reaction in \( p \)-thermo.

<table>
<thead>
<tr>
<th>Compound</th>
<th>mmol/gDCW</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proteins</strong></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>0.5142</td>
</tr>
<tr>
<td>Arg</td>
<td>0.1831</td>
</tr>
<tr>
<td>Asp</td>
<td>0.2219</td>
</tr>
<tr>
<td>Asn</td>
<td>0.2219</td>
</tr>
<tr>
<td>Cys</td>
<td>0.1079</td>
</tr>
<tr>
<td>Glu</td>
<td>0.322</td>
</tr>
<tr>
<td>Gln</td>
<td>0.322</td>
</tr>
<tr>
<td>Gly</td>
<td>0.4077</td>
</tr>
<tr>
<td>His</td>
<td>0.0778</td>
</tr>
<tr>
<td>Ile</td>
<td>0.2728</td>
</tr>
<tr>
<td>Leu</td>
<td>0.3475</td>
</tr>
<tr>
<td>Lys</td>
<td>0.3172</td>
</tr>
<tr>
<td>Met</td>
<td>0.0832</td>
</tr>
<tr>
<td>Phe</td>
<td>0.1451</td>
</tr>
<tr>
<td>Pro</td>
<td>0.1665</td>
</tr>
<tr>
<td>Ser</td>
<td>0.1811</td>
</tr>
<tr>
<td>Thr</td>
<td>0.2688</td>
</tr>
<tr>
<td>Trp</td>
<td>0.1026</td>
</tr>
<tr>
<td>Tyr</td>
<td>0.1036</td>
</tr>
<tr>
<td>Val</td>
<td>0.3761</td>
</tr>
<tr>
<td><strong>Nucleic Acids</strong></td>
<td></td>
</tr>
<tr>
<td>Ribonucleic acid (RNA)</td>
<td></td>
</tr>
<tr>
<td>AMP</td>
<td>0.1193</td>
</tr>
<tr>
<td>CMP</td>
<td>0.0915</td>
</tr>
<tr>
<td>GMP</td>
<td>0.0915</td>
</tr>
<tr>
<td>UMP</td>
<td>0.1193</td>
</tr>
<tr>
<td>Deoxyribonucleic acid (DNA)</td>
<td></td>
</tr>
<tr>
<td>dAMP</td>
<td>0.0077</td>
</tr>
<tr>
<td>dCMP</td>
<td>0.0059</td>
</tr>
<tr>
<td>dGMP</td>
<td>0.0059</td>
</tr>
<tr>
<td>dTMP</td>
<td>0.0077</td>
</tr>
<tr>
<td><strong>Lipids</strong></td>
<td></td>
</tr>
<tr>
<td>Phospholipids</td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>0.0290</td>
</tr>
<tr>
<td>PG</td>
<td>0.0277</td>
</tr>
<tr>
<td>CLPN</td>
<td>0.0296</td>
</tr>
<tr>
<td><strong>Carbohydrates</strong></td>
<td></td>
</tr>
<tr>
<td>D-Fructose</td>
<td>0.1048</td>
</tr>
<tr>
<td>GDP-Mannose</td>
<td>0.0057</td>
</tr>
<tr>
<td>UDP-D-Galactose</td>
<td>0.1895</td>
</tr>
<tr>
<td>UDP-D-Xylose</td>
<td>0.1825</td>
</tr>
<tr>
<td>UDP-D-Glucose</td>
<td>0.0096</td>
</tr>
<tr>
<td>UDP-L-Arabinose</td>
<td>0.0407</td>
</tr>
<tr>
<td>Compound</td>
<td>mmol/gDCW</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Salts</strong></td>
<td></td>
</tr>
<tr>
<td>Phosphorous</td>
<td>0.0420</td>
</tr>
<tr>
<td>Calcium</td>
<td>0.0028</td>
</tr>
<tr>
<td>Potassium</td>
<td>0.6323</td>
</tr>
<tr>
<td>Magnesium</td>
<td>0.0875</td>
</tr>
<tr>
<td>Iron</td>
<td>0.0304</td>
</tr>
<tr>
<td><strong>Vitamins &amp; cofactors</strong></td>
<td></td>
</tr>
<tr>
<td>Thiamin B1</td>
<td>0.0002</td>
</tr>
<tr>
<td>Riboflavin B2</td>
<td>0.0002</td>
</tr>
<tr>
<td>Vitamin B12</td>
<td>0.0002</td>
</tr>
<tr>
<td>Vitamin B6</td>
<td>0.0002</td>
</tr>
<tr>
<td>10-Formyltetrahydrofolate</td>
<td>0.0004</td>
</tr>
<tr>
<td>Biotin</td>
<td>1.79E-06</td>
</tr>
<tr>
<td>Bacillithiol</td>
<td>6.23E-05</td>
</tr>
<tr>
<td>Bacillithiol disulfide</td>
<td>1.56E-07</td>
</tr>
<tr>
<td>Chorismate</td>
<td>0.0002</td>
</tr>
<tr>
<td>FAD</td>
<td>0.0002</td>
</tr>
<tr>
<td>FMN</td>
<td>0.0002</td>
</tr>
<tr>
<td>Heme</td>
<td>0.0002</td>
</tr>
<tr>
<td>Proto-heme</td>
<td>0.0002</td>
</tr>
<tr>
<td>Siroheme</td>
<td>0.0002</td>
</tr>
<tr>
<td>Menaquinol</td>
<td>0.0003</td>
</tr>
<tr>
<td>NAD+</td>
<td>0.0145</td>
</tr>
<tr>
<td>NADH</td>
<td>0.0267</td>
</tr>
<tr>
<td>NADP</td>
<td>0.0042</td>
</tr>
<tr>
<td>NADPH</td>
<td>0.0027</td>
</tr>
<tr>
<td>S-Adenosyl-L-methionine</td>
<td>0.0002</td>
</tr>
<tr>
<td>Spermidine</td>
<td>0.0011</td>
</tr>
<tr>
<td>Spermine</td>
<td>0.0025</td>
</tr>
<tr>
<td>Putrescine</td>
<td>0.0054</td>
</tr>
<tr>
<td>CoA</td>
<td>0.0002</td>
</tr>
<tr>
<td><strong>Intracellular Metabolites</strong></td>
<td></td>
</tr>
<tr>
<td>Acetyl-CoA</td>
<td>0.0002</td>
</tr>
<tr>
<td>Succinyl-CoA</td>
<td>8.75E-05</td>
</tr>
<tr>
<td><strong>Energy Requirement</strong></td>
<td></td>
</tr>
<tr>
<td>ADP</td>
<td>- 152.28</td>
</tr>
<tr>
<td>Pi</td>
<td>-152.27</td>
</tr>
<tr>
<td>Ppi</td>
<td>0.0011</td>
</tr>
<tr>
<td>H</td>
<td>- 152.28</td>
</tr>
<tr>
<td>H2O</td>
<td>152.28</td>
</tr>
<tr>
<td>ATP</td>
<td>152.31</td>
</tr>
</tbody>
</table>
**Supplementary table 2:** Estimation of polymerization energy needed to form biomass from the different monomer classes present in the biomass reaction. This energy fraction constitutes part of the growth associated maintenance that is present in the biomass reaction. Polymerization energy per molecule was obtained from literature\(^6\).

<table>
<thead>
<tr>
<th>Cellular content (w/w)</th>
<th>Polymerization energy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Per molecule (mmATP/g polymer)</td>
</tr>
<tr>
<td>Protein</td>
<td>0.52</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>0.1</td>
</tr>
<tr>
<td>RNA</td>
<td>0.16</td>
</tr>
<tr>
<td>DNA</td>
<td>0.1</td>
</tr>
<tr>
<td>Lipids</td>
<td>0.09</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
</tr>
</tbody>
</table>

**Supplementary table 3:** Overview of an analyses of filtering the unique ORFs detected in the genome analyses between two *Para*geobacillus strains, when various levels of filtering are applied to elucidate how many reactions would be unique in models made of each strain, and finally which would be connected to any pre-existing metabolites in the network. Supplementary table 4 and 5 highlight the unique metabolic ORFs identified.

<table>
<thead>
<tr>
<th>P. thermoglucosidasius NCIMB 11955</th>
<th>P. thermoglucosidosiadius M10EXG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique ORFs</td>
<td>259</td>
</tr>
<tr>
<td>Unique metabolic ORFs(^1)</td>
<td>40</td>
</tr>
<tr>
<td>Unique ORFs as reactions(^2)</td>
<td>18</td>
</tr>
<tr>
<td>Unique reactions(^3)</td>
<td>11</td>
</tr>
<tr>
<td>Connected unique reactions(^4)</td>
<td>8</td>
</tr>
</tbody>
</table>

\(^1\) Unique metabolic ORFs are ORFs with an E.C. code associated to them
\(^2\) Subset of unique metabolic ORFs that would actually be captured as a reaction in the model
\(^3\) Subset of unique ORFs that would cause new reactions to the model
\(^4\) Subset of unique reactions that are connected by a main metabolite to one or more pre-existing metabolites in the model. Note, all these reactions are still blocked (i.e. no two main metabolites found in model).
**Supplementary table 4**: Metabolic ORFs unique to *P. thermoglucosidasius* NCIMB 11955, detected in the genome comparison.

<table>
<thead>
<tr>
<th>Gene annotation</th>
<th>EC</th>
<th>Annotated Kegg Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>abfA</td>
<td>3.2.1.55</td>
<td>Alpha-L-arabinofuranosidase</td>
</tr>
<tr>
<td>adk</td>
<td>2.7.4.3</td>
<td>Adenylate kinase</td>
</tr>
<tr>
<td>araA</td>
<td>5.3.1.4</td>
<td>L-arabinose isomerase</td>
</tr>
<tr>
<td>cocE</td>
<td>3.1.1.84</td>
<td>Carboxylesterase</td>
</tr>
<tr>
<td>ecfA2</td>
<td>3.6.3.2</td>
<td>Hydrolase, involved in transmembrane movement</td>
</tr>
<tr>
<td>glf</td>
<td>5.4.99.9</td>
<td>UDP-galactopyranose mutase</td>
</tr>
<tr>
<td>glnX</td>
<td>6.1.1.17</td>
<td>Glutamyl-tRNA synthetase</td>
</tr>
<tr>
<td>hdi IVa</td>
<td>3.8.1.2</td>
<td>2-haloacid dehalogenase</td>
</tr>
<tr>
<td>helD_2</td>
<td>3.6.4.12</td>
<td>DNA helicase</td>
</tr>
<tr>
<td>ispD</td>
<td>2.7.7.60</td>
<td>2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase</td>
</tr>
<tr>
<td>ispF</td>
<td>4.6.1.12</td>
<td>2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase</td>
</tr>
<tr>
<td>kpsU</td>
<td>2.7.7.38</td>
<td>3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase)</td>
</tr>
<tr>
<td>lytC_3</td>
<td>3.5.1.28</td>
<td>N-acetylmuramoyl-L-alanine amidase</td>
</tr>
<tr>
<td>mazF</td>
<td>3.1.27</td>
<td>Esterase (component of type II toxin-antitoxin system)</td>
</tr>
<tr>
<td>mcsB</td>
<td>2.7.14.1</td>
<td>Protein arginine kinase</td>
</tr>
<tr>
<td>mngB</td>
<td>3.2.1.170</td>
<td>Mannosylglycerate hydrolase</td>
</tr>
<tr>
<td>mrmC</td>
<td>3.1.26</td>
<td>Esterase (ribonuclease)</td>
</tr>
<tr>
<td>mtlD</td>
<td>1.1.1.17</td>
<td>Mannitol-1-phosphate 5-dehydrogenase</td>
</tr>
<tr>
<td>ppk</td>
<td>2.7.4.1</td>
<td>Polyphosphate kinase</td>
</tr>
<tr>
<td>ppx</td>
<td>3.6.1.11</td>
<td>Exopolyphosphatase</td>
</tr>
<tr>
<td>pseG</td>
<td>3.6.1.57</td>
<td>UDP-2,4-diacetamido-2,4,6-trideoxy-beta-L-altropyranose hydrolase</td>
</tr>
<tr>
<td>rapA_1</td>
<td>3.6.4.2</td>
<td>RNA Polymerase associated hydrolase</td>
</tr>
<tr>
<td>rbsD</td>
<td>5.4.99.62</td>
<td>D-ribose pyranase</td>
</tr>
<tr>
<td>rhpA</td>
<td>3.6.4.13</td>
<td>ATP-dependent helicase</td>
</tr>
<tr>
<td>rpoA</td>
<td>2.7.7.6</td>
<td>DNA-directed RNA polymerase</td>
</tr>
<tr>
<td>rpoB</td>
<td>2.7.7.6</td>
<td>DNA-directed RNA polymerase</td>
</tr>
<tr>
<td>rpoC</td>
<td>2.7.7.6</td>
<td>DNA-directed RNA polymerase</td>
</tr>
<tr>
<td>srIB</td>
<td>2.7.1.198</td>
<td>Glucitol/sorbitol PTS system EIIA component</td>
</tr>
<tr>
<td>srIE</td>
<td>2.7.1.198</td>
<td>Glucitol/sorbitol PTS system EIIA component</td>
</tr>
<tr>
<td>tagD</td>
<td>2.7.7.39</td>
<td>Glycerol-3-phosphate cytidylyltransferase</td>
</tr>
<tr>
<td>trmB</td>
<td>2.1.1.33</td>
<td>tRNA (guanine-N7-)-methyltransferase</td>
</tr>
<tr>
<td>truA</td>
<td>5.4.99.12</td>
<td>tRNA pseudouridine synthase</td>
</tr>
<tr>
<td>wecA</td>
<td>2.7.8.40</td>
<td>UDP-GalNAc:undecaprenyl-phosphate GalNAc-1-phosphate transferase</td>
</tr>
<tr>
<td>xylA_1</td>
<td>3.2.1.37</td>
<td>Xylan 1,4-beta-xylosidase</td>
</tr>
<tr>
<td>xynA</td>
<td>3.2.1.8</td>
<td>Endo-1,4-beta-xylanase</td>
</tr>
<tr>
<td>xynB</td>
<td>3.2.1.37</td>
<td>Xylan 1,4-beta-xylosidase</td>
</tr>
<tr>
<td>YDAf_2</td>
<td>2.3.1.4</td>
<td>Acyltransferase</td>
</tr>
<tr>
<td>yeeF</td>
<td>3.1.2.3</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>yjjG</td>
<td>3.1.3.5</td>
<td>5’-nucleotidase</td>
</tr>
<tr>
<td>yobL</td>
<td>3.1.2.3</td>
<td>Ribonuclease</td>
</tr>
</tbody>
</table>
Supplementary table 5: Metabolic ORFs unique to *P. thermoglucosidasius* M10EXG, detected in the genome comparison. Interestingly, it appears the M10EXG strain encodes a complete myo-inositol utilizing operon (*iol*), where some components are lacking in NCIMB 11955\textsuperscript{103}. Data shows NCIMB 11955 indeed lacks the capability to grow on inositol\textsuperscript{51}, but data on M10EXG is lacking.

<table>
<thead>
<tr>
<th>Gene annotation</th>
<th>EC</th>
<th>Annotated Kegg Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>cocE</td>
<td>3.1.1.84</td>
<td>Carboxylesterase</td>
</tr>
<tr>
<td>cwlK</td>
<td>3.4.--.</td>
<td>Peptidase</td>
</tr>
<tr>
<td>cynS</td>
<td>4.2.1.104</td>
<td>Cyanate lyase</td>
</tr>
<tr>
<td>derK</td>
<td>2.7.1.210</td>
<td>D-erythrulose 4-kinase</td>
</tr>
<tr>
<td>dhaL</td>
<td>2.7.--.</td>
<td>Phosphotransferase</td>
</tr>
<tr>
<td>fdtA</td>
<td>5.3.2.3</td>
<td>TDP-4-oxo-6-deoxy-alpha-D-glucose-3,4-oxoisomerase</td>
</tr>
<tr>
<td>hsdM_1</td>
<td>2.1.1.72</td>
<td>Site-specific DNA-methyltransferase (adenine-specific)</td>
</tr>
<tr>
<td>hsdM_2</td>
<td>2.1.1.72</td>
<td>Site-specific DNA-methyltransferase (adenine-specific)</td>
</tr>
<tr>
<td>hsdR_1</td>
<td>3.1.21.3</td>
<td>Type I restriction enzyme</td>
</tr>
<tr>
<td>hsdR_2</td>
<td>3.1.21.3</td>
<td>Type I restriction enzyme</td>
</tr>
<tr>
<td>iolB</td>
<td>5.3.1.30</td>
<td>5-deoxy-glucuronate isomerase</td>
</tr>
<tr>
<td>iolE</td>
<td>4.2.1.44</td>
<td>Inosose dehydratase</td>
</tr>
<tr>
<td>ioll</td>
<td>5.3.99.11</td>
<td>2-keto-myo-inositol isomerase</td>
</tr>
<tr>
<td>iolX</td>
<td>1.1.1.370</td>
<td>Scylo-inositol 2-dehydrogenase (NAD+)</td>
</tr>
<tr>
<td>lhgO</td>
<td>1.1.3.15</td>
<td>Glycolate oxidase</td>
</tr>
<tr>
<td>lipA</td>
<td>2.8.1.8</td>
<td>Lipoyl synthase</td>
</tr>
<tr>
<td>lsrF</td>
<td>2.3.1.245</td>
<td>3-hydroxy-5-phosphonoxy pentane-2,4-dione thiolase</td>
</tr>
<tr>
<td>mutT4</td>
<td>3.6.1.--</td>
<td>Esterase</td>
</tr>
<tr>
<td>NA</td>
<td>2.4.1.161</td>
<td>Oligosaccharide 4-alpha-D-glucosyltransferase</td>
</tr>
<tr>
<td>radD</td>
<td>3.6.4.12</td>
<td>DNA repair helicase</td>
</tr>
<tr>
<td>recD</td>
<td>3.1.11.5</td>
<td>Exodeoxyribonuclease V</td>
</tr>
<tr>
<td>rfbC</td>
<td>5.1.3.13</td>
<td>dTDP-4-dehydrorhamnose 3,5-epimerase</td>
</tr>
<tr>
<td>sqhC</td>
<td>4.2.1.137</td>
<td>Sporulenol synthase</td>
</tr>
<tr>
<td>sunS</td>
<td>2.4.1.--</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>tatD</td>
<td>3.1.21.--</td>
<td>Endodeoxyribonucleases</td>
</tr>
<tr>
<td>uvrD1</td>
<td>3.6.4.12</td>
<td>ATP-dependent DNA helicase</td>
</tr>
<tr>
<td>wapA_1</td>
<td>3.1.--.</td>
<td>tRNA(Glu)-specific nuclease</td>
</tr>
<tr>
<td>wapA_2</td>
<td>3.1.--.</td>
<td>tRNA(Glu)-specific nuclease</td>
</tr>
<tr>
<td>wapA_3</td>
<td>3.1.--.</td>
<td>tRNA(Glu)-specific nuclease</td>
</tr>
</tbody>
</table>