A Genome Scale Model of *Geobacillus thermoglucosidasius* (C56-YS93) reveals its biotechnological potential on rice straw hydrolysate

Ahmad Ahmad[a], Hassan B. Hartman[b], S. Krishnakumar[a], David A. Fell[b], Mark G. Poolman[b,*], Shireesh Srivastava[a,**]

[a] *DBT-ICGEB Center for Advanced Bioenergy Research, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, 110067, New Delhi, India*

[b] *Department of Biological and Medical Sciences, Oxford Brookes University, Gipsy Lane, OX3 0BP, Oxford, UK*

Abstract

Rice straw is a major crop residue which is burnt in many countries, creating significant air pollution. Thus, alternative routes for disposal of rice straw are needed. Biotechnological treatment of rice straw hydrolysate has potential to convert this agriculture waste into valuable biofuel(s) and platform chemicals. *Geobacillus thermoglucosidasius* is a thermophile with properties specially suited for use as a biocatalyst in lignocellulosic bioprocesses, such as high optimal temperature and tolerance to high levels of ethanol. However, the capabilities of *Geobacillus thermoglucosidasius* to utilize sugars in rice straw hydrolysate for making bioethanol and other platform chemicals have not been fully explored. In this work, we have created a genome scale metabolic model (denoted iGT736) of the organism containing 736 gene products, 1159 reactions and 1163 metabolites. The model was validated both by purely theoretical approaches and by comparing the behaviour of the model to previously published experimental results. The model was then used to determine the yields of a variety of platform chemicals from glucose and xylose — two primary sugars in rice straw hydrolysate. A comparison with results from a model of *Escherichia coli* shows that *G. thermoglucosidasius* is capable of producing a wider range of products, and that for the products also produced by *E. coli*, the yields are comparable. We also discuss strategies to utilise arabinose, a minor component of rice straw hydrolysate, and propose additional reactions to lead to the synthesis of xylitol, not currently produced by *G. thermoglucosidasius*. Our results provide additional motivation for the current exploration of the industrial potential of *G. thermoglucosidasius* and we make our model publicly available to aid the development of metabolic engineering strategies for this organism.

Keywords:
Metabolic modelling, rice straw, thermophile, fermentation, flux balance analysis,

1. Background

1.1. Geobacillus thermoglucosidasius

Geobacillus thermoglucosidasius C56-YS93 (hereafter, *G. thermoglucosidasius*) is a thermophilic, gram positive, facultative anaerobic bacterium, capable of producing ethanol (which it can tolerate at concentrations up to 10% v/v) and grows at an optimal temperature of 65°C (Brumm et al., 2015). The ability to grow at high temperature obviates the cooling costs otherwise associated with mesophilic fermentation, significantly reduces the risk of microbial contamination (Cripps et al., 2009), and potentially aids the ethanol

*Corresponding author
**Co-corresponding author
Email addresses: ahmadbioinfo@gmail.com (Ahmad Ahmad), Hassan.Hartman@phe.gov.uk (Hassan B. Hartman), skrishnabio@gmail.com (S. Krishnakumar), dfell@brookes.ac.uk (David A. Fell), mgpoolman@brookes.ac.uk (Mark G. Poolman), shireesh@icgeb.res.in (Shireesh Srivastava)

recovery process (Ennis et al., 1986, Gong et al., 1998, Taylor et al., 1998), resulting in higher productivity and reduction of downstream processing costs (Cripps et al., 2009, Niu et al., 2015). Its ability to ferment at high temperatures also makes it a viable candidate for simultaneous saccharification and fermentation (SSF) as well as possibly for consolidated bioprocessing (Lin and Tanaka, 2006, Tang et al., 2009).

In addition, it is able to ferment C5, as well as C6 sugars, to produce ethanol (Fong et al., 2006, Cripps et al., 2009, Tang et al., 2009), and other *Geobacillus* species are reported as capable of producing ethanol from lignocellulosic feedstocks (Hartley and Shama, 1987). More recently Bartosiak-Jentys et al. (2013) reported that *G. thermoglucosidasius* is capable of utilising xylans, by excretion of xylanases, potentially reducing the need for hydrolytic pretreatment of the feedstock.

1.2. Lignocellulose and rice straw

The global production of rice straw is estimated at 7×10^8 tons per year (Wi et al., 2013). Its utility as animal feed is limited owing to its high ($\approx$ 13% w/w) silica content, compared to barley straw (2%) or wheat straw (1–5%), resulting in a digestibility of 45% (Soest, 2006). Although it can be ploughed back into the soil and serve as fertiliser for the next growth season, this is associated with spread of plant disease (Gadde et al., 2009) and the plant material is therefore more often incinerated as waste (Binod et al., 2010). However, open field burning of rice straw results in emission of hazardous compounds (e.g. various dioxins) with adverse effects on human health and the environment (Gadde et al., 2009), and has the additional drawback potentially useful biomass is lost. Hence there is an obvious interest in both reducing the negative impact of rice straw incineration and making full use of the plant biomass.

Lignocellulosic biomass, including rice straw, is primarily composed of a mixture of lignin, cellulose and hemicellulose requiring pretreatment and hydrolysis to release embedded polymerized sugar molecules, and a variety of methods are used to this end. Dilute acid hydrolysis solubilizes hemicellulose with greater efficiency than cellulose, resulting in a liquor with greater concentration of C5 sugars compared to C6 sugars. On the other hand, alkaline hydrolysis solubilizes the lignin as well as the hemicellulosic fractions. The biomass remaining can be further hydrolysed using enzyme cocktails to increase the sugar yield from biomass. The enzymatic hydrolysis produces a mixture of C5 and C6 sugars. Depending on the treatment used, rice straw yields approximately 42–85% glucose, 6–24% xylose and 0–4% arabinose (A. M. Lali, personal communication).

1.3. Target products

There is increasing interest in replacing the current fossil-based petrochemical production routes for high value chemicals with biotechnological processes using biomass as feedstock (Choi et al., 2015). Many chemicals of commercial interest are synthesised from a small number of platform compounds, including succinate, citrate, ethanol, lactate, xylitol, and sorbitol (which are also useful in their own right). For example, 1,4-butanediol, tetrahydrofuran, gamma-butyrolactone are produced from succinate; itaconic acid (which can be further refined to methyl methacrylate) from citrate; polyethylene, polypropylene, and polyethylene terephthalic acid (PET[1]) from ethanol; polylactic and acrylic acids from lactate; xylaric acid, ethylene glycol and propylene glycol from xylitol (in addition to its use as an additive in food, cosmetics and pharmaceuticals), and isosorbide from sorbitol (Choi et al., 2015). As these platform compounds are (with the exception of sorbitol) ubiquitous in metabolism there is obvious commercial potential in engineering organisms to produce them from low value feedstocks such as rice straw. The goal of this study is, therefore, to investigate the potential of the *G. thermoglucosidasius* metabolic network to grow and generate these products from the sugars found in rice straw hydrolysate.

1.4. Metabolic Modelling

Except for the most trivial of examples, it is not possible to reason about the properties of a system of interconnected reactions via intuitive inspection of a network diagram: a formalised and objective framework

[1]**Abbreviations** GPR: Gene-protein-reaction relationship, GSM: Genome-scale model, PET: Polyethelene terephthalic acid, PGDB:Pathway/genome database.

is required, and such a framework is provided by the field of metabolic modelling. The mathematical basis of this is to treat the system as a set of ordinary differential equations in terms of concentration, and, once the system is so defined, to bring to bear a variety of well established techniques.

Metabolic models can be broadly divided into two categories, *structural* and *kinetic*. In the former category, the system is defined in terms of reaction stoichiometry and direction, while in the latter explicit rate equations are specified for each reaction. For practical purposes, studies of large scale ($>\approx 50-100$ reactions tend to be restricted to the structural category.

It is now routine to generate structural models derived from publicly available gene annotations, accounting for all reactions catalysed by the enzymes encoded in an organism's genome (hence described as Genome-Scale Models (GSMs)). Such models require substantial theoretical validation (Poolman et al., 2006, Gevorgyan et al., 2008) before they can be deemed informative or applied to real-world problems. It should be emphasised that GSMs only involve metabolites and reactions, not the association between genes, proteins, enzymes and reactions (GPR relationships), although this can be included in the model description as metadata.

In this contribution we have kept separate the model and the database from which it was derived. By defining the model in terms of unique reaction and metabolite identifiers from the database, the GPR associations of the model can be extracted from the database, without being constrained to use any specific version of it. A benefit of this approach is that GPR associations stored in the database can be kept up to date by the wider scientific community as new data is made available.

Once theoretically validated to ensure that the network does not violate fundamental physical laws, the model can then be analysed in a number of ways to gain insight into the network's properties in the context of both basic and applied science. As far as this paper is concerned, we seek to identify: utilisable carbon substrates for growth; the range of economically important products; their yield from these substrates, and additional reactions that, if added to the network, would broaden the range of utilisable carbon substrates.

1.5. Analysis of Structural Models

The analysis of structural metabolic models is predicated on the steady-state assumption, i.e., that over the time period under investigation concentrations of internal metabolites remain stable. This can be expressed mathematically as:

$$\mathbf{Nv} = \mathbf{0} \tag{1}$$

where

N is the stoichiometry matrix, whose rows represent metabolites, columns represent reactions, and whose elements represent the stoichiometric coefficients of reactions with respect to reactants.

v is the (unknown, or incompletely known)vector of reaction rates.

0 is the vector of rates of change of metabolite concentration with zero defining steady state.

Equation 1 represents an underdetermined set of simultaneous equations, and the structural analysis of metabolic networks therefore revolves around the identification of properties that must be true for *any* solution, and the identification of individual *potential* solutions (Heinrich and Schuster, 1996).

The former of these activities depend, in the main, upon the analysis of the *kernel* (or null-space) of **N**. These analyses are, amongst other things, invaluable for the initial validation of a model, but are less relevant to the primary goal of this paper and hence not elaborated here (but see Heinrich and Schuster (1996), Schuster and Fell (2007) for basic theory and Poolman et al. (2007), Pfeiffer et al. (1999), Gagneur and Klamt (2004) for examples of applications).

The identification of potential individual solutions depends upon the application of the technique of linear programming (LP). This was first proposed by Fell and Small (1986) and Watson (1986) and later popularised under the name Flux Balance Analysis (FBA) (Varma and Palsson, 1993a,b, Orth et al., 2010). Linear programming (Chvátal, 1983) is an algorithm that seeks to optimise (maximise or minimise) one or

more variables in a solution of a system of linear equations (the *objective*), subject to a number of constraints, where both the variables to be optimised and the constraints are defined by the context of the problem.

A generic linear program for FBA may be specified as:

$$\begin{array}{lll} \min|\max & \mathbf{v}_{\text{targs}} & \text{Objective} \\ \text{subject to} & \begin{cases} \mathbf{Nv} = \mathbf{0} \\ \min(v_i) \leq v_i \leq \max(v_i) \end{cases} & \text{Constraints} \end{array} \tag{2}$$

The objective specifies a set of target reactions whose fluxes are to be minimised or maximised while complying with the steady-state assumption ($\mathbf{Nv} = \mathbf{0}$) and that all fluxes must be kept within maximum and minimum limits. The steady-state constraint means that any solution must be a vector in the null-space of $\mathbf{N}$. Specifying lower and upper limits on fluxes serves two purposes: firstly to ensure that irreversible reactions only carry positive flux and secondly to maintain given fluxes (typically inputs or outputs) to observed or desired values.

Examples of objectives include maximisation of biomass synthesis (Varma and Palsson, 1993b), and minimising the sum of reaction fluxes (Holzhütter, 2004, Poolman et al., 2009) as a proxy for optimal protein investment in enzyme machinery. Although in reality the 'true' biological objective is not known, and must depend to a certain extent on cell type, these two classes of objective function prove to be extremely useful tools with which to explore the capabilities of large metabolic networks (Feist et al., 2007, Liu et al., 2010, Lee et al., 2014).

In addition to standard FBA, numerous methods have been suggested where FBA is applied to metabolic engineering. Examples of these include: OptKnock - identification of optimal reaction knock outs in order to maximise the yield of a given set of metabolites, using bilevel optimisation (Burgard et al., 2003); OptGene - similar purpose to OptKnock, but based on evolutionary programming (Patil et al., 2005); OptORF - identifies optimal gene knock outs or overexpressions, including regulatory genes, for a given yield to optimise (Kim and Reed, 2010); OptStrain - given a universal metabolic database and a species-specific metabolic model, the procedure identifies the minimal, required number of gene additions to the original strain and computes the optimal set of knock outs using OptKnock (Pharkya et al., 2004); OptReg - extension of OptKnock to allow optimal up- or down regulation of reactions, given a reference flux solution (Pharkya and Maranas, 2004); OptForce - given a reference solution and an optimal solution w.r.t some yield, variability analysis is used to identify which reactions need to change in the reference solution to obtain the optimal solution, after which bilevel optimisation is then used to suggest the required genetic modifications (Ranganathan et al., 2010).

1.5.1. Models of thermophilic organisms

Despite the large number of published genome-scale models (more than 100 (Monk et al., 2014)), thermophilic organisms are notably neglected, with only 5 reported to date (Tang et al., 2009, Zhang et al., 2009, Ulas et al., 2012, Lee et al., 2014, Cordova et al., 2015). The model published in Tang et al. (2009) was of the closely related *G. thermoglucosidasius* M10EXG strain. This model is however not based on the genome sequence of the strain, but is a modified version of a GSM of *Geobacter sulfurreducens* (Mahadevan et al., 2006, Tang et al., 2009), nor is the model publicly available (but see the discussion section of this paper). The model published in Cordova et al. (2015) of *Geobacillus* LC300 is a central carbon metabolism core model based on the annotated genome sequence. Liberal et al. (2015) describe software for comparing and curating metabolic models and used it to analyse a draft model of *G. thermoglucosidasius* NCIMB 11955 obtained from the SEED database (Henry et al., 2010), however, the resulting model is not publicly available.

It is therefore not possible to attempt to replicate, or expand, those analyses without constructing a new model. Consequently we describe the construction, curation and analysis of a new GSM of *G. thermoglucosidasius* directly derived from its BioCyc Pathway/Genome DataBase (PGDB) (`http://biocyc.org` Karp et al. (2011)), make it publicly available, and evaluate its potential as a biocatalyst for conversion of sugars

present in rice straw hydrolysate into higher value products that are of industrial interest. We believe that this work is of interest, not only to the metabolic modelling community, but also to researchers involved in all areas of microbial biotechnology.

2. Methods

2.1. Model construction

The model was constructed and analysed using the approach described for a number other organisms *Salmonella Typhimurium* (Hartman et al., 2014), *Oryza sativa* L. *japonica* (Poolman et al., 2013) and *Arabidopsis thaliana* (Poolman et al., 2009), using the 'ScrumPy' metabolic modelling software package (Poolman, 2006). The model is derived from the BioCyc (Karp et al., 2011) Pathway/Genome DataBase (PGDB) for *G. thermoglucosidasius* (C56-YS93) and is comprised of a number of modules, viz:

1. A set of reactions directly imported from the BioCyc database, subject to revision as required;
2. Transport reactions describing the export of biomass precursors;
3. Transport reactions describing the import and export of compounds from/to the medium;
4. The electron transport chain, and
5. Additional reactions required for the generation of biomass precursors, not reported in BioCyc.

2.1.1. Reactions from BioCyc

The flatfiles for the *G. thermoglucosidasius* PGDB, (GeoCyc, v 18.5), were downloaded from the BioCyc ftp site and the 'PyoCyc' component of the ScrumPy metabolic modelling package (Poolman, 2006) was used to extract the reactions in the database.

Reactions involved with non-metabolic species, or species to which empirical formulae cannot be assigned (e.g. 'DNA-Combined-With-Exogenous-DNA' or 'Protein-Disulfides') were omitted. All remaining reactions were ensured to be atomically balanced in terms of the elements C, N, P and S. With the exception of reactions of the electron transport chain (see section 2.1.4), this requirement was not imposed for H and O in reactions involved with protons, hydroxide ions or water, as the stoichiometric coefficients of reactions with respect to these species varies according to pH, and water is anyway assumed to be present to excess.

Reactions were recorded in the ScrumPy model description format in terms of their BioCyc unique identifiers (UIDs) thereby facilitating cross-referencing of subsequent results to the parent database (for example to establish GPR relationships).

2.1.2. Biomass transport

Transporters were defined for precursors of polymeric biomass compounds: amino acids, fatty acids, nucleotides, cell wall components, and cofactors. Macromolecular composition (relative protein, DNA, RNA, lipid and cell wall) was assumed to be similar to that of the closely related organism *Bacillus subtilis* as reported by Oh et al. (2007). A variety of literature sources were used in order to define the monomeric composition of the macromolecular components: amino acid distribution (Oh et al., 2007, Tang et al., 2009); nucleotide distribution (Oh et al., 2007, Coorevits et al., 2012); fatty acids and lipid (Rosa et al., 2000, Nazina et al., 2001), and cell wall composition (Grant and Wicken, 1970, Oh et al., 2007). See supplementary data for details.

2.1.3. Import/export of media components

Transporters were defined to account for the import of glucose, xylose and arabinose (the major sugar components of rice-straw hydrolysate), along with transporters for the uptake of NH_3, SO_4, O_2 and P_i.

Transporters were also defined for the export of CO_2 fermentation products (formate, acetate, ethanol, pyruvate, lactate and succinate) and the additional industrial target products sorbitol, xylitol and citrate (Belal, 2013).

2.1.4. Electron transport chain

The electron transport chain (ETC) involves the translocation of protons over the cell membrane which are subsequently re-imported via ATP synthase and, in order to maintain conservation of energy, it is essential that these protons are properly conserved. This is achieved by defining the stoichiometries of ETC reactions in terms of intra and extra-cellular proton species which are distinct from those in the rest of the model, i.e.:

$$\begin{aligned}
&\text{Complex I} && \text{NADH + Q + 4 } \mathrm{H}^+_{int} && \longrightarrow && \text{NAD + QH + 3 } \mathrm{H}^+_{ext} \\
&\text{Complex IV} && \tfrac{1}{2}\ \mathrm{O_2} + \text{QH + 2 } \mathrm{H}^+_{int} && \longrightarrow && \text{2 } \mathrm{H}^+_{ext} + \text{Q + } \mathrm{H_2O} \\
&\text{ATP Synthase} && \text{ADP + } \mathrm{P_i} + \text{ 3 } \mathrm{H}^+_{ext} && \longrightarrow && \text{ATP + 3 } \mathrm{H}^+_{int} + \mathrm{H_2O}
\end{aligned}$$

Yielding a net stoichiometry of:

$$\text{4 NADH + 9 } \mathrm{H}^+_{int} + \text{2 } \mathrm{O_2} + \text{5 ADP + 5 } \mathrm{P_i} \longrightarrow \text{4 NAD + 9 } \mathrm{H_2O} + \text{5 ATP}$$

This corresponds to a P/O ratio of 1.25 compared to a value of 1.33 for *Escherichia coli* (Maloney, 1987).

2.1.5. Extra reactions

In addition to the modules listed above we also included a module defining extra reactions that are either essential for the production of particular biomass components though absent from the *G. thermoglucosidasius* PGDB, or else lumped biosynthetic reactions involving metabolites with defined composition, but whose empirical formulae were missing in the original PGDB. Identification of extra reactions needed for biomass synthesis was based on comparisons between biosynthetic routes (as defined in BioCyc) for the synthesis of these components.

Candidate genes encoding these extra reactions were subsequently identified using the PRIAM bioinformatics tool (February 2014 release) (Clotilde Claudel-Renard et al., 2003), which assigns putative enzyme functions to amino-acid sequences. In this analysis the translated genome sequence of *G. thermoglucosidasius* was used for identification of possible enzyme functions.

2.1.6. Energetic requirements

Consumption of metabolic energy for cellular maintenance and growth related polymerisation was modelled by fixing the flux of a generic ATP hydrolysis reactions (ATPase: ATP + $\mathrm{H_2O}\longrightarrow$ADP + $\mathrm{P_i}$). Total ATP demand for maintenance was estimated using:

$$r_{\mathrm{ATP}} = Y_{\mathrm{ATP}} \cdot \mu + m_{\mathrm{ATP}} \tag{3}$$

where Y_{ATP} is the yield coefficient of ATP costs associated with biomass synthesis, μ is a growth rate, and m_{ATP} is the ATP cost associated with non-growth maintenance. The parameters used for determining r_{ATP} were $Y_{\mathrm{ATP}} = 41 \cdot 10^{-3}$ (mol ATP (g DW)$^{-1}$) (Niu et al., 2015), μ was assumed to be $\mu_{max} = 0.3$ h^{-1} (Tang et al., 2009), and $m_{\mathrm{ATP}} = 6.1 \cdot 10^{-3}$ mol (g DW)$^{-1}$h^{-1} (Niu et al., 2015).

2.2. Model Analysis

2.2.1. LP/FBA

The model was used to define the linear program:

$$\begin{aligned}
&\text{min.} \quad : \quad \textstyle\sum |v_i| \\
&\text{s.t.} \quad \begin{cases} \mathbf{Nv} = \mathbf{0} & \\ v_i \geq 0, & \forall i \in \{\text{irrevs}\} \\ v_t = \mathbf{t}_t, & \forall t \in \{\text{targs}\} \end{cases}
\end{aligned} \tag{4}$$

where the objective is to minimise the sum of all steady-state fluxes subject whilst maintaining non-negative flux in irreversible reactions ($v_i \geq 0, \forall i \in \{\text{irrevs}\}$) and a specified fixed flux in a set of 'target' reactions ($v_t = \mathbf{t}_t, \forall t \in \{\text{targs}\}$).

An additional weighting factor was included in this formulation for the lumped fatty-acid synthesis reactions in order to account for the fact that one flux unit in these reactions would require unit flux in approximately 30 reactions *in vivo*.

To investigate biomass production, target fluxes were assigned to the transporters of biomass precursors with values calculated from the assumed biomass composition, described above, and growth rate. To investigate the ability of the model to generate useful products, a single target was set to represent unit flux for the product's transporter and the objective changed to minimise total substrate uptake with other outputs left unconstrained.

2.2.2. Theoretical validation

In addition to ensuring that all reactions individually conserved mass, the model was determined to be globally consistent in terms of energy and stoichiometry. Energetic consistency was validated by setting a target flux value of 1 to the generic ATPase reaction and zero to all transporters and attempting to solve equation 4. The existence of any feasible solution would indicate that the model describes a metabolism in which ATP could be regenerated without any net mass flow, violating the conservation of energy; no such solution can be obtained with the final curated model. Likewise, redox consistency was also verified with respect to the oxidation of NADH and NADPH in the absence of net mass flow.

Stoichiometric consistency was validated by identifying unconserved metabolites, as defined by Gevorgyan et al. (2008). Unconserved metabolites were limited to those consisting entirely of hydrogen and oxygen atoms (protons, molecular hydrogen, molecular oxygen, hydrogen peroxide, water and superoxide), as would be expected since protons and water were not balanced in the model.

2.3. Biomass production from rice straw hydrolysate and other potential nutrients

To verify biomass production, flux values for the biomass component transporters were set according to the molar ratio of observed biomass composition as target values in equation 4. Solutions for single carbon sources were obtained by setting target values of zero to the other carbon substrate importers, and solutions for anaerobic conditions were obtained by setting a target value of zero for the O_2 transporter.

2.4. Determination of biomass yields

To calculate biomass yields, the specific growth rates (μ) for *G. thermoglucosidasius* reported by Tang et al. (2009) were used to fix the fluxes in the biomass transporters and the substrate uptake (v_s) required to support these fluxes determined while minimising total flux in the system. Yield (g biomass per g substrate consumed) is then calculated as:

$$Y_{\mathrm{x/s}} = \frac{1000 \cdot \mu}{M \cdot v_{\mathrm{s}}} \tag{5}$$

where $Y_{\mathrm{x/s}}$ is the yield, M is the molecular mass of the substrate ($\mathrm{g.mol^{-1}}$, and v_s is the substrate flux $\mathrm{mmol.(g\ DW)^{-1}.h^{-1}}$.

2.5. Generation and yields of industrially useful products from rice straw hydrolysate

For each of the products considered (ethanol, lactate, citrate, xylitol, succinate, and sorbitol), the corresponding transport reaction flux was set to unit production with the objective modified to minimise total carbon uptake. The ratios of carbon atom flux in the product to the substrate gives the % yield. For comparison with the capabilities of *E. coli* these calculations were repeated after importing the GSM *i*AF1260 (Feist et al., 2007) into ScrumPy.

3. Results

3.1. General properties and biomass production from rice straw hydrolysate and other Carbon Sources

The final *G. thermoglucosidasius* GSM (hereafter referred to as iGT736) has a total of 1159 reactions (of which 75 are transporters), and 1163 internal and 78 external metabolites. Table 1 compares iGT736 with other available genome-scale models of thermophiles. iGT736 is significantly larger than these models, representing a greater coverage of the metabolic network.

[Table 1 about here.]

With these external metabolites, a total of 636 reactions can carry steady-state flux. The complete list of genes, reactions, enzymes, and GPR assignments, along with the model in SBML and ScrumPy formats, can be found in the supplementary data.

Solutions to equation 4 demonstrate that the model can account for the simultaneous production of the 54 biomass precursors and 6 fermentation by-products under aerobic or anaerobic conditions with either glucose or xylose assumed to be the sole carbon source. Under anaerobic conditions, if both glucose and xylose are assumed to be simultaneously available, the optimal solution is obtained at an uptake ratio of about 10:1.

We also tested whether the model could account for growth or no growth on another 18 carbon sources for which experimental data is available (Nazina et al., 2001, 2005, Coorevits et al., 2012). There was agreement in all but two cases. The model did not show growth on lactose, but the experimental accounts were not in agreement with one another for this substrate. The experiments show growth on turanose, a glucose–fructose disaccharide, but the model does not as there is no suitable glucosidase explicitly annotated in the genome. Agreement between model and experiment was observed for growth on acetyl–glucosamine, glucose, glycerol, amylose, mannitol, methylglucose, ribose, salicin, sucrose and xylose, and for lack of growth on arabinose, galactose, amylopectin, myo–inositol, melezitose, melibiose, and raffinose.

The model results agreed with the experimental report (Nazina et al., 2001) that the minor rice straw sugar arabinose cannot be utilised. However, the addition of just two reactions to the model, arabinose isomerase (E.C. 5.3.1.4) and ribulokinase (E.C. 2.7.1.16) along with a transporter, would allow arabinose to be used as the sole carbon source for growth under aerobic and anaerobic conditions. Furthermore, with these two reactions included, then under anaerobic conditions and with all three sugars available as carbon sources, glucose and arabinose are consumed in an approximate 10:1 ratio.

[Table 2 about here.]

[Table 3 about here.]

Calculated and experimental biomass yields for growth on glucose and xylose are in good agreement and are presented in Table 3. The model results suggest that the biomass yield under anaerobic conditions for *G. thermoglucosidasius* is in fact slightly greater for xylose (0.14 g/g) compared to glucose (0.11 g/g).

3.2. Generation of industrially useful products from rice straw hydrolysate

With the addition of arabinose isomerase and ribulokinase, for reasons described above, the model was able to account for the synthesis of all the target compounds, except xylitol, under aerobic or anaerobic conditions and using any of the three sugars as the sole carbon source. The addition of a single reaction to the network, xylulose reductase (E.C. 1.1.1.9), enables the model to account for the synthesis of xylitol.

Optimal yields for each product were determined by modifying equation 4 to minimise total carbon uptake, using each of the three sugars in turn as the sole carbon source, under aerobic and anaerobic conditions. The results obtained are presented in table 4. Solutions obtained when all three sugars were assumed to be simultaneously available were equivalent to those for single carbon sources (not shown).

For all products, the optimal yield was unaffected by the carbon source (consistent with the observation that there was no improvement in yield for mixed substrates), with the minor exception of sorbitol under

aerobic conditions, for which the yield was 1% greater when glucose was the carbon source rather than the pentoses.

Under aerobic conditions, O_2 was not utilised for the optimal production of ethanol, lactate or xylitol, the solutions being identical to those in the anaerobic situation. More surprising are the results obtained for optimal anaerobic yields of succinate and citrate in which methionine or cysteine (and ethanol) are produced as by-products with a concomitant consumption of NH_3 and SO_4, reflecting the requirement for an alternative electron acceptor when generating more oxidised products. However, if NH_3 uptake is blocked the same yield is achieved with ethanol as the only by-product, indicating the existence of multiple optima for these products. Furthermore, if the objective function is changed to minimise total flux in the system, although different solutions are obtained (result not shown) the yields remain substantially unaffected.

[Table 4 about here.]

[Table 5 about here.]

4. Discussion and conclusions

4.1. Comparison with experiment

A previous study has investigated flux distributions in *G. thermoglucosidasius* using ^{13}C MFA (Tang et al., 2009). Although the underlying model used in that study is not directly comparable (nor publicly available), a comparison of the reported input-output fluxes with those calculated here is not without interest and is presented in table 6.

The most notable deviation between the two sets of fluxes is the lack of lactate production by the model under anaerobic conditions. However, we note firstly that the network defined by the model remains viable (ie produces biomass) if the experimentally observed values in Table 6 are imposed as constraints in 2, and secondly, that the distributions of fermentation products calculated from this model are sensitive to the exact biomass composition, which, as noted above, is approximate and may reasonably be expected to vary according to the environment. On this basis we may conclude that the model described here defines a reaction network that is a realistic representation of *G. thermoglucosidasius*.

[Table 6 about here.]

4.2. Geobacillus thermoglucosidasius *as a platform for generating industrially useful compounds*

As shown in Table 4, the model analysis suggests that *G. thermoglucosidasius* is capable of producing high yields of a broad range of industrially important chemicals, under both aerobic and anaerobic conditions. The yields of the compounds are similar between the three sugars - the only deviation from the common pattern is the negligibly higher yield of sorbitol from glucose (89%), compared to the other two sources (88%). This would indicate that *G. thermoglucosidasius* has the capacity to convert the vast majority of the carbohydrate in rice straw hydrolysate into high yields of valuable products.

The initial model was unable to consume arabinose, consistent with previous experimental reports (Nazina et al., 2001). However, the addition of two enzymes, arabinose-isomerase and ribulokinase (EC 5.3.1.4 and 2.7.1.16), to the network allow both growth and target product synthesis from this compound, suggesting that a genetically engineered organism will be able to utilise all the sugar components of rice straw hydrolysate. However, this model result was predicated on arabinose being able to enter the cells, and we could not find any protein sequences with homology to arabinose transporters from *E. coli* and *B. subtilis* in the genome, so expression of a heterologous transporter may be necessary as well. Similarly, the initial model was unable to produce xylitol unless xylulose reductase was added to the model.

The analysis used here for *G. thermoglucosidasius* was also applied to an *E. coli* GSM (iAF1260) (Feist et al., 2007). The *E. coli* model analysis suggested that only ethanol and succinate could be produced from the three sugars. *E. coli* has been reported to be able to produce lactate, albeit at a very low rate (Mazumdar et al., 2010). To the best of our knowledge, wild-type *E. coli* has not been reported to produce any of the other products. As with *G. thermoglucosidasius*, the *E. coli* yields are almost identical for

the three sugars, with a slightly higher yield of succinate on glucose (99%), compared with the other two (97%). Comparing the modelling results for two organisms shows that *E. coli* seems to be able to produce a higher yield of succinate (97-99%) under anaerobic conditions compared to *G. thermoglucosidasius* (80%). The higher succinate yield of *E. coli* can be explained by the presence of a formate-hydrogen lyase in *E. coli*, which is absent in *G. thermoglucosidasius*. If the formate-hydrogen lyase reaction is added to the *G. thermoglucosidasius* model the succinate yield increases to 100% for all three sugars.

It is noteworthy that the analysis, when applied to *G. thermoglucosidasius*, suggests that amino acids (methionine and cysteine) are co-produced with succinate and citrate under anaerobic conditions. Further analysis shows that solutions resulting in co-production of methionine, cysteine or ethanol are completely equivalent. A similar result was obtained for the *E. coli* model when constrained to produce succinate. Unlike *G. thermoglucosidasius*, the *E. coli* analysis indicated production of hydrogen sulphide, but like *G. thermoglucosidasius* the solution was not unique and a large number of alternative solutions were possible, reflecting numerous ways in which the redox balance can be adjusted.

4.3. Concluding remarks

In conclusion, we have generated a publicly available genome-scale model of *G. thermoglucosidasius* that demonstrates the potential of this organism to grow and generate useful products from the carbon sources found in rice-straw hydrolysate. Further more we have been able to use it to propose interventions to utilise substances that it can not currently consume. This suggests that it is a valid and useful platform for further development and investigation by researchers in the biotechnology/metabolic engineering arenas. Moreover, we have demonstrated that a range of industrially interesting compounds can be produced by native *G. thermoglucosidasius* (ethanol, succinate, citrate, sorbitol, lactate) or after the addition of a single reaction in the case of xylitol. This can be compared with results obtained with a GSM of *E. coli*, a popular mesophilic biocatalyst, which indicated that it was only able to produce succinate and ethanol.

The model has been submitted to the biomodels database (`https://www.ebi.ac.uk/biomodels-main` (Chelliah et al., 2015)) with the unique identifier MODEL170306000, and is also available from the author's (MP & DF) website `http:mudshark.brookes.ac.uk/Models/Geobac`

5. Acknowledgements

We thank Professor Arvind Lali, ICT, Mumbai, for providing us with data on the composition of rice straw hydrolysate before publication.

Funding: This work was supported by a collaborative DBT–BBSRC grant (RICEFUEL; BBSRC reference BB/K020358/1) and a BBSRC grant to DAF (BB/J019712/1). Ahmad is supported by the DBT-ICGEB Bioenergy Center grant.

6. References

J. Bartosiak-Jentys, A. H. Hussein, C. J. Lewis, and D. J. Leak. Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. *Microbiology*, 159:1267–1275, 2013.

E. B. Belal. Bioethanol production from rice straw residues. *Braz J Microbiol*, 44:225–234, 2013.

P. Binod, R. Sindhu, R. R. Singhania, S. Vikram, L. Devi, S. Nagalakshmi, N. Kurien, R. K. Sukumaran, and A. Pandey. Bioethanol production from rice straw: An overview. *Bioresour. Technol.*, 101:4767–4774, 2010.

P. J. Brumm, M. L. Land, and D. A. Mead. Complete genome sequence of *Geobacillus thermoglucosidasius* C56-YS93, a novel biomass degrader isolated from obsidian hot spring in Yellowstone National Park. *Standards in Genomic Sciences*, 10:doi: 10.1186/s40793–015–0031–z, 2015.

A. P. Burgard, P. Pharkya, and C. D. Maranas. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.*, 84:647–57, 2003.

V. Chelliah, N. Juty, I. Ajmera, R. Ali, M. Dumousseau, M. Glont, M. Hucka, G. Jalowicki, S. Keating, V. Knight-Schrijver, A. Lloret-Villas, K. N. Natarajan, J.-B. Pettit, N. Rodriguez, M. Schubert, S. M. Wimalaratne, Y. Zhao, H. Hermjakob, N. Le Novère, and C. Laibe. BioModels: ten-year anniversary. *Nucleic Acids Research*, 43(D1):D542, 2015. doi: 10.1093/nar/gku1181. URL `http://dx.doi.org/10.1093/nar/gku1181`.

S. Choi, C. W. Song, J. H. Shin, and S. Y. Lee. Biorefineries for the production of top building block chemicals and their derivatives. *Metabolic Engineerin*, 28:223–239, 2015.

V. Chvátal. *Linear Programing*. W.H. Freeman, 1983.

Clotilde Claudel-Renard, Claude Chevalet, Thomas Faraut, and Daniel Kahn. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Research*, 31(22):6633–6639, 2003.

A. Coorevits, A. E. Dinsdale, G. Halket, L. Lebbe, P. D. Vos, A. V. Landschoot, and N. A. Logan. Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus*, *G. stearothermophilus*, *G. jurassicus*, *G. toebii*, *G. thermodenitrificans* and *G. thermoglucosidans* (nom. corrig., formerly *'thermoglucosidasius'*); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. *Int. J. Syst. Evol. Microbiol.*, 62:1470–1485, 2012.

L. T. Cordova, C. P. Long, K. P. Venkataramanan, and M. R. Antoniewicz. Complete genome sequence, metabolic model construction and phenotypic characterization of *Geobacillus* LC300, an extremely thermophilic, fast growing, xylose-utilizing bacterium. *Metab. Eng*, 32:74–81, 2015.

R. E. Cripps, K. Eley, D. Leak, B. Rudd, M. Taylor, M. Todd, S. Boakes, S. Martin, and T. Atkinson. Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metabolic Engineering*, 11:398–408, 2009.

B. M. Ennis, C. T. Marshall, I. S. Maddox, and A. H. J. Paterson. Continuous product recovery by in-situ gas stripping/condensation during solvent production from whey permeate using *Clostridium acetobutylicum*. *Biotechnology Letters*, 8:725–730, 1986.

A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. O. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3(121), 2007.

D. A. Fell and J. R. Small. Fat synthesis in adipose tissue. *Biochem. J*, 238:781–786, 1986.

J. C. N. Fong, C. J. Svenson, K. Nakasugi, C. T. C. Leong, J. P. Bowman, B. Chen, D. R. Glenn, B. A. Neilan, and P. L. Rogers. Isolation and characterization of two novel ethanol-tolerant facultative-anaerobic thermophilic bacteria strains from waste compost. *Extremophiles*, 10:363 – 372, 2006.

B. Gadde, S. Bonnet, C. Menke, and S. Garivait. Air pollutant emissions from rice straw open field burning in India, Thailand and the Philippines. *Environ. Pollut.*, 157:1554–1558, 2009. doi: 10.1016/j.envpol.2009.01.004.

J. Gagneur and S. Klamt. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, 5, 2004. doi: 10.1186/1471-2105-5-175.

A. Gevorgyan, M. G. Poolman, and D. A. Fell. Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics*, 24(19):2245–2251, October 2008. ISSN 1460-2059. URL `http://view.ncbi.nlm.nih.gov/pubmed/18697772`.

C. Gong, N. Cao, J. Du, and G. Tsao. Ethanol production from renewable resources. *Adv. Biochem. Eng. Biotechnol.*, 65: 207–241, 1998.

W. D. Grant and J. Wicken. Autolysis of cell walls of *Bacillus stearothermophilus* B65 and the chemical structure of the peptidoglycan. *Biochem. J.*, 118:859 – 868, 1970.

B. S. Hartley and G. Shama. Novel Ethanol Fermentations from Sugar Cane and Straw. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 321:555 – 568, 1987.

H. B. Hartman, D. A. Fell, S. Rossell, P. R. Jensen, M. J. W. L. Thorndahl, L. Jelsbak, J. E. Olsen, A. Raghunathan, S. Daefler, and M. G. Poolman. Identification of potential drug targets in *Salmonella enterica* sv. Typhimurium using metabolic modelling and experimental validation. *Microbiology*, 160:1252–1266, 2014.

R. Heinrich and S. Schuster. *The Regulation of Cellular Systems*. Chapman & Hall, London, England, 1996.

C. S. Henry, M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay, and R. L. Stevens. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol.*, 28:977–982, 2010. doi: 10.1038/nbt.1672.

H.-G. Holzhütter. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur. J. Biochem.*, 271(14):2905–22, July 2004. doi: 10.1111/j.1432-1033.2004.04213.x.

P. D. Karp, M. Latendresse, and R. Caspi. The Pathway Tools Pathway Prediction Algorithm. *Stand Genomic Sci*, 5:doi: 10.4056/sigs.1794338, 2011.

J. Kim and J. L. Reed. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol.*, 4, 2010. doi: 10.1186/1752-0509-4-53.

N.-R. Lee, M. Lakshmanan, S. Aggarwal, J.-W. Song, I. A. Karimi, D.-Y. Lee, and J.-B. Park. Genome-scale metabolic network reconstruction and *in silico* flux analysis of the thermophilic bacterium *Thermus thermophilus* HB27. *Microb. Cell Fact.*, 13:doi: 10.1186/1475–2859–13–61, 2014.

R. Liberal, B. K. Lisowska, D. J. Leak, and J. W. Pinney. PathwayBooster: a tool to support the curation of metabolic pathways. *BMC Bioinformatics*, 16, 2015. doi: 10.1038/nbt.1672.

Y. Lin and S. Tanaka. Ethanol fermentation from biomass resources: current state and prospects. *Appl. Microbiol. Biotechnol.*, 69:627 – 642, 2006.

L. Liu, R. Ågren, S. Bordel, and J. Nielsen. Use of genome-scale metabolic models for understanding microbial physiology. *FEBS Lett.*, 584:2556 – 2654, 2010.

R. Mahadevan, D. R. Bond, J. E. Butler, A. Esteve-Nuñez, M. V. Coppi, B. Ø. Palsson, C. H. Schilling, and D. R. Lovley. Characterization of metabolism in the Fe(III)-reducing organism *Geobacter sulfurreducens by constraint-based modeling. Appl Environ Microbiol.*, 72:1558–68, 2006.

P. C. Maloney. *Coupling to an energized membrane: role of ion-motive gradients in the tranduction of metabolic energy. Escherichia coli* and *Salmonella typhimurium*. Cellular and Molecular Biology. American Society for Microbiology, 1987.

S. Mazumdar, J. M. Clomburg, and R. Gonzalez. *Escherichia coli* Strains Engineered for Homofermentative Production of d-Lactic Acid from Glycerol. *Appl. Environ. Microbiol.*, 76:4327–4336, 2010.

J. Monk, J. Nogales, and B. O. Palsson. Optimizing genome-scale network reconstructions. *Nat. Biotechnol*, 32(5):447–452, 2014.

T. N. Nazina, T. P. Tourova, A. B. Poltaraus, E. V. Novikova, A. A. Grigoryan, A. E. Ivanova, A. M. Lysenko, V. V. Petrunyaka, G. A. Osipov, S. S. Belyaev, and M. V. Ivanov. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. *Int. J. Syst. Evol. Microbiol.*, 52 (Pt. 2):433 – 446, 2001.

T. N. Nazina, D. S. Sokolova, A. A. Grigoryan, N. M. Shestakova, E. M. Mikhailova, A. B. Poltaraus, T. P. Tourova, A. M. Lysenko, G. A. Osipov, and S. S. Belyaev. *Geobacillus jurassicus* sp. nov., a new thermophilic bacterium from a high–temperature reservoir, and the validation of the *Geobacillus* species. *Sys. Appl. Microbiol.*, 28:43–53, 2005.

H. Niu, D. Leak, N. Shah, and C. Kontoravdi. Metabolic characterization and modeling of fermentation process of an engineered *Geobacillus thermoglucosidasius* strain for bioethanol production with gas stripping. *Chemical Engineering Science*, 122:138 – 149, 2015.

Y.-K. Oh, B. O. Palsson, S. M. Park, C. H. Schilling, and R. Mahadevan. Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data. *J. Biol. Chem.*, 282:28791–28799, 2007.

J. D. Orth, I. Thiele, and B. Ø. Palsson. What is flux balance analysis? *Nat Biotechnol.*, 28, 2010. doi: 10.1038/nbt.1614.

K. R. Patil, I. Rocha, J. F'orster, and Nielsen. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics*, 23, 2005. doi: 10.1186/1471-2105-6-308.

T. Pfeiffer, I. Sanchez-Valdenebro, J. Nuno, F. Montero, and S. Schuster. METATOOL: for studying metabolic networks. *Bioinformatics*, 15(3):251–257, 1999.

P. Pharkya and C. Maranas. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng.*, 8:1–13, 2004.

P. Pharkya, A. P. Burgard, and C. Maranas. OptStrain: A computational framework for redesign of microbial production systems. *Genome Res.*, 14:2367–2376, 2004.

M. G. Poolman. ScrumPy - metabolic modelling with Python. *IEE Proceedings Systems Biology*, 153(5):375–378, 2006.

M. G. Poolman, B. K. Bonde, A. Gevorgyan, H. H. Patel, and D. A. Fell. Challenges to be faced in the reconstruction of metabolic networks from public databases. *IEE Proceedings Systems Biology*, 153(5):379–384, 2006.

M. G. Poolman, C. Sebu, M. K. Pidcock, and D. A. Fell. Modular decomposition of metabolic systems via null-space analysis. *Journal of Theoretical Biology*, 249:691–705, 2007.

M. G. Poolman, L. Miguet, L. J. Sweetlove, and D. A. Fell. A genome-scale metabolic model of *Arabidopsis thaliana* and some of its properties. *Plant Physiol.*, 151(3):1570–1581, 2009.

M. G. Poolman, S. Kundu, R. Shaw, and D. A. Fell. Responses to Light Intensity in a Genome-Scale Model of Rice Metabolism. *Plant Physiol.*, 162:1060–1072, 2013.

S. Ranganathan, P. F. Suthers, and C. D. Maranas. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol.*, 6:e1000744, 2010. doi: 10.1371/journal.pcbi.1000744.

S. B. Roberts, C. M. Gowen, J. P. Brooks, and S. S. Fong. Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production. *BMC Systems Biology*, 4:31, 2010.

S. M. L. J. Rosa, M. d. C. Antunes-Madeira, M. J. Matos, A. S. Jurado, and V. M. C. Madeira. Lipid composition and dynamics of cell membranes of *Bacillus stearothermophilus* adapted to amiodarone. *Biochim. Biophys. Acta*, 1487:286 – 295, 2000.

S. Schuster and D. Fell. *Bioinformatics - From Genomes to Therapies*, volume 2, chapter 20, pages 755–805. Wiley-VCH, 2007.

P. V. Soest. Rice straw, the role of silica and treatments to improve quality. *Anim. Feed Sci. Technol.*, 130:137–171, 2006. doi: 10.1016/j.anifeedsci.2006.01.023. URL `http://dx.doi.org/10.1016/j.anifeedsci.2006.01.023`.

Y. Tang, R. Sapra, D. Joyner, T. Hazen, S. Myers, D. Reichmuth, H. Blanch, and J. Keasling. Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain. *Biotechnol Bioeng.*, 102:1377–1386, 2009.

F. Taylor, M. Kurantz, N. Goldberg, and J. Craig Jr. Continuous Fermentation and Stripping of Ethanol. *Biotechnology Letters*, 20:67 –72, 1998.

T. Ulas, S. A. Riemer, M. Zaparty, B. Siebers, and D. Schomburg. Genome-scale reconstruction and analysis of the metabolic network in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *PLoS ONE*, 7:doi: 10.1371/journal.pone.0043401, 2012.

A. Varma and B. O. Palsson. Metabolic Capabilities of *Escherichia coli*: I. Synthesis of Biosynthetic Precursors and Cofactors. *J. Theor. Biol.*, 165:477–502, 1993a.

A. Varma and B. O. Palsson. Metabolic Capabilities of *Escherichia coli*: II. Optimal Growth Patterns. *J. Theor. Biol.*, 165: 503–522, 1993b.

M. R. Watson. A discrete model of bacterial metabolism. *Comp. Appl. Biosci.*, 2(1):23–27, 1986.

S. G. Wi, I. S. Choi, K. H. Kim, H. M. Kim, and H.-J. Bae. Bioethanol production from rice straw by popping pretreatment. *Biotechnology for Biofuels*, 6(1):1–7, 2013. ISSN 1754-6834. doi: 10.1186/1754-6834-6-166. URL `http://dx.doi.org/10.1186/1754-6834-6-166`.

Y. Zhang, I. Thiele, D. Weekes, Z. Li, L. Jaroszewski, K. Ginalski, A. M. Deacon, J. Wooley, S. A. Lesley, I. A. Wilson, B. Palsson, A. Osterman, and A. Godzik. Three-Dimensional Structural View of the Central Metabolic Network of *Thermotoga maritima*. *Science*, 325:doi: 10.1126/science.1174671, 2009.

List of Tables

Tables

Table 1: Comparison of iGT736 with other published thermophile GSMs

Organism	Genes	Reactions	Metabolites	Reference
G. thermoglucosidasius iGT736	736	1159	1163	This work
Thermus Thermophilus	548	796	635	Lee et al. (2014)
Clostridium thermocellum	432	577	525	Roberts et al. (2010)
Sulfolobus solfataricus	515	776	705	Ulas et al. (2012)

Table 2: By-products and number of reactions used, calculated for growth on different carbon sources. The solution for aerobic growth with mixed carbon sources was the same as that for aerobic growth with glucose as the sole carbon source.

	Aerobic		**Anaerobic**	
Source	Reactions	By-products	Reactions	By-products
Glc	328	Ac, CO_2	325	Suc, EtOH, Ac, Form
Xyl	338	Ac, CO_2	320	Suc, EtOH, Ac, Form
Ara	321	Ac, CO_2	324	EtOH, Ac, Form
Mixed				
Glc	329	Ac, CO_2		
Glc, Ara			329	Suc, EtOh, Ac, Form

Table 3: Comparison of calculated and experimental biomass yields of *G. thermoglucosidasius* on Glucose and Xylose. Experimental values are calculated from Tang et al. (2009)

	Aerobic		**Anaerobic**	
Source	Calculated	Observed	Calculated	Observed
Glc	0.40	0.37	0.11	0.08
Xyl	0.33	0.23	0.14	0.10

Table 4: Optimal (%C/C) yields of industrially important compounds, determined by minimising total carbon input to produce 1 mole of product, under aerobic and anaerobic conditions, using arabinose, glucose or xylose as the sole carbon substrate. Results presented here were obtained after reactions enabling growth on arabinose and production of xylitol had been added; see main text for details.

		Aerobic		**Anaerobic**		
Substrate	Product	Yield	By-product	Co-substrates	Yield	By-product
Arabinose	Suc	100	None		80	EtOH
	Citrate	100	None	NH_3, SO_4	66	Met, EtOH
	EtOH	As anaerobic			66	CO_2
	Lact	As anaerobic			100	None
	Sorbitol	88	CO_2		80	Ac, CO_2
	Xylitol	As anaerobic			90	CO_2
Glucose	Suc	100	None	NH_3, SO_4	80	Met
	Citrate	100	None	NH_3, SO_4	66	Cys, EtOH
	EtOH	As anaerobic			66	CO_2
	Lact	As anaerobic			100	None
	Sorbitol	89	CO_2		80	Ac, CO_2
	Xylitol	As anaerobic			90	CO_2
Xylose	Suc	100	None		80	EtOH
	Citrate	100	None	NH_3, SO_4	66	Met, EtOH
	EtOH	As anaerobic			66	CO_2
	Lact	As anaerobic			100	None
	Sorbitol	88	CO_2		80	Ac, CO_2
	Xylitol	As anaerobic			90	CO_2

Table 5: Analysis repeated with E. coli (iAF1260) GSM. Note that only succinate and ethanol were producible.

		Aerobic		**Anaerobic**		
Substrate	Product	Yield	By-product	Co-substrates	Yield	By-product
Arabinose	Suc	100	None	SO_4	97	EtOH, H_2S, H_2
	EtOH	As anaerobic			66	CO_2
Glucose	Suc	100	None		99	EtOH,H_2S, H_2
	EtOH	As anaerobic			66	CO_2
Xylose	Suc	100	None	SO_4	97	EtOH, H_2S, H_2
	EtOH	As anaerobic			66	CO_2

Table 6: Calculated (this study) and experimental (Tang et al., 2009) rates of product formation for aerobic and anaerobic growth with glucose or xylose as the sole carbon source and with no constraints on the transport of fermentation by-products (NR: not reported).

Substrate	Growth h^{-1}	Uptake mmol (g DW)$^{-1}h^{-1}$		Product	Output mmol (g DW)$^{-1}h^{-1}$	
		Calculated	Observed		Calculated	Observed
				Aerobic		
Glucose	0.35	5.2	5	Acetate	5.4	3.2
				CO_2	7.2	NR
				Lactate	0	0.1
				Ethanol	0	0.05
Xylose	0.27	6.1	5	Acetate	8.1	5.0
				CO_2	4.0	NR
				Ethanol	0	0.05
				Anaerobic		
Glucose	0.1	4.2	5	Formate	3.2	5.2
				Acetate	7.8	3.1
				Lactate	0.0	4.5
				Ethanol	1.2	1.9
				Succinate	0.09	Trace
Xylose	0.15	6.4	7	Formate	4.8	6.5
				Acetate	8.5	5.0
				Lactate	0	6.2
				Ethanol	1.9	2.7
				Succinate	0.14	Trace