Adv Biochem Eng Biotechnol
DOI: 10.1007/10_2016_45

Genome-Scale Modeling of Thermophilic Microorganisms

Sanjeev Dahal, Suresh Poudel, and R. Adam Thompson

Abstract Thermophilic microorganisms are of increasing interest for many industries as their enzymes and metabolisms are highly efficient at elevated temperatures. However, their metabolic processes are often largely different from their mesophilic counterparts. These differences can lead to metabolic engineering strategies that are doomed to fail. Genome-scale metabolic modeling is an effective and highly utilized way to investigate cellular phenotypes and to test metabolic engineering strategies. In this review we chronicle a number of thermophilic organisms that have recently been studied with genome-scale models. The microorganisms spread across archaea and bacteria domains, and their study gives insights that can be applied in a broader context than just the species they describe. We end with a perspective on the future development and applications of genome-scale models of thermophilic organisms.

Keywords Draft reconstruction, Flux balance analysis, Flux variability analysis, Gap-filling, Genome-scale models, Stoichiometric matrix

S. Dahal (✉) and S. Poudel
UT-ORNL Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN, USA

Oak Ridge National Laboratory, Oak Ridge, TN, USA

BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN, USA
e-mail: sdahal@vols.utk.edu

R.A. Thompson
BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN, USA

Contents

1 Introduction to Thermophilic Microorganisms
2 Uses of Thermophilic Microorganisms in Industry
3 Genome-Scale Modeling of Metabolism
4 Genome-Scale Modeling of Thermophilic Microorganisms
4.1 *Clostridium thermocellum*
4.2 *Thermotoga maritima*
4.3 *Thermus thermophilus*
4.4 *Sulfolobus solfataricus*
4.5 *Thermobifida fusca*
4.6 *Moorella thermoacetica*
4.7 *Streptococcus thermophilus*
4.8 *Thermoanaerobacterium saccharolyticum*
4.9 Models Deposited in BioModels Database
5 Conclusion
References

1 Introduction to Thermophilic Microorganisms

According to Brock, thermophiles are organisms that can grow and reproduce at high temperatures [1]. Generally, 50–60°C is regarded as the minimal temperature for bacteria and archaea to be considered thermophiles, because this is the known upper limit for eukaryotes. Even within thermophiles there is a distinction between hypothermophiles and hyperthermophiles based on their optimum temperature. Hypothermophiles prefer temperatures of up to 80°C whereas hyperthermophiles can have a temperature preference of up to 100°C.

Thermophilic microorganisms can be found in various habitats such as geothermal hot springs in places such as Yellowstone National Park where *Thermus aquaticus* was discovered [2]. Another major known habitat is the area around deep-sea vents from which *Methanococcus jannaschii* was found [3]. Nutritionally, thermophiles which span the metabolic range from phototrophy to chemotrophy, from autotrophy to heterotrophy, and from aerobic to anaerobic capabilities have been described in the literature [4].

Generally, studies have shown that, at optimal temperatures, thermophiles show lower growth yield compared to their mesophilic counterparts at their respective optimal temperatures. The lower yield is attributed to a higher energy for maintenance requirements such as turnover of proteins and nucleotides, cell mobility, and ionic maintenance. Because of their high temperature growth conditions, thermophiles require more energy for maintaining these conditions [5]. There are some exceptions to the rule, such as *Thermothrix thiopara* [6]. The observed reduced growth efficiency and higher maintenance requirements have made these organisms interesting in research as these organisms tend to produce various catabolic products in larger quantities than other organisms [7].

2 Uses of Thermophilic Microorganisms in Industry

Thermophilic organisms have been utilized in several industrial areas such as the fuel industry, waste management, and mining. Thermophiles and their enzymes have been widely regarded as the most efficient way to generate biofuels from lignocellulose (contains cellulose, hemicellulose, and lignin). The use of thermostable organisms and enzymes provides several advantages such as faster conversion of substrates, decreased risk of contaminations, and more compound recovery. Several thermophile-produced enzymes have been proposed for degradation of cellulosic biomass. Some of the enzymes are cellulases (which degrade cellulose) and xylanases (which degrade hemicellulose). Furthermore, thermophilic organisms have been suggested to be the microbial cell factory for consolidated bioprocessing (CBP) in which degradation of lignocellulose and fermentation of sugars are accomplished in one step. Examples of these organisms are *Clostridium thermocellum*, *Caldicellulosiruptor saccharolyticus*, and *Caldicellulosiruptor bescii*.

In addition to biofuels, thermophilic organisms also find their use in the area of waste management [8–11]. Studies have shown that the use of both mesophilic and thermophilic digesters could help recover energy from biowastes such as livestock manure and food waste. Furthermore, the use of thermophiles for the recovery of metals from industrial and municipal wastes has also been proposed [12, 13]. Bioleaching is the process through which microorganisms are used to extract metals from ores and waste products. This process has been used for the extraction of metals such as zinc, copper, gold, and molybdenum using organisms such as *Metallosphaera sedula* [14, 15], *Sulfolobus* [14, 15], *Sulfobacillus* [16], or *Ferroplasma* [17].

3 Genome-Scale Modeling of Metabolism

Genome-scale modeling is a powerful tool that has been used for many applications, such as the prediction of cellular phenotypes, elucidation of biological principles, rational strain design for metabolic engineering, simulation of co-cultures, and the interpretation of OMICs and other high-throughput datasets [18]. The most common method for analyzing a genome-scale metabolic network is called Flux Balance Analysis (FBA). In general, a metabolic network can be represented by a stoichiometric matrix $\boldsymbol{S} \in \mathbb{R}^{m \times n}$, consisting of m metabolites and n reactions, such that the entry $s_{i,j}$ is the stoichiometric coefficient of metabolite i in reaction j. A valid flux distribution vector $\boldsymbol{v} \in \mathbb{R}^{n \times 1}$ satisfies a steady-state condition

$$\boldsymbol{S} \cdot \boldsymbol{v} = 0$$

and is thus constrained by mass balance. The flux distribution vector is also constrained by thermodynamics such that

$$v_j \geq 0$$

for all irreversible reactions *j*. FBA relies on the stoichiometric and thermodynamic constraints to optimize a cellular objective, such as maximizing cell growth, maximizing product synthesis, or minimizing ATP hydrolysis [19].

Using this framework, optimization problems can be coupled in a multitude of ways to probe cellular metabolism, and multiple software packages have been developed to facilitate the construction and analysis of genome-scale models [20, 21]. In addition, many curated models of thermophiles and non-thermophiles have been deposited in the BiGG database [22]. As the metabolism of thermophiles are typically less well-understood than that of their mesophilic counterparts, and the challenges of living at higher temperatures favors alternative metabolic pathways, genome-scale models are effective tools to study thermophiles. The following section outlines several curated genome-scale models and how they have been used to study thermophilic metabolism, in particular increasing the understanding of deviations from model organisms and generating hypotheses for further study.

4 Genome-Scale Modeling of Thermophilic Microorganisms

4.1 Clostridium thermocellum

C. thermocellum is a gram-positive bacterium of great interest for consolidated bioprocessing of lignocellulose to biofuels because it exhibits one of the fastest growth rates on crystalline cellulose which it directly converts to the biofuels such as ethanol, hydrogen, and isobutanol.

The first genome-scale model of *C. thermocellum* was created for strain ATCC 27405 by Roberts et al. [23]. The model, called *i*SR432, consists of 577 reactions, 525 metabolites, and 432 genes. The model consists of the cellulosome data contained in its proteomic information. The cellulosome is a large extracellular protein complex, which is optimized for hydrolyzing cellulose into glucose oligomers of length 2–6. The draft reconstruction was based on the genome annotations from databases such as IMG, UniProt, and KEGG [24]. Additional transport reactions were added based on a reciprocal BLAST hit between *C. thermocellum* genome and the Transport Classification Database (TCDB) [25]. In the draft reconstruction, the investigators discovered that there were missing gaps, especially because of species-specific metabolism such as cellulosome production, cellulose and chitin degradation, biosynthesis of teichoic acid and peptidoglycan, steroid metabolism, and transport reactions. Therefore a manual gap-filling was carried out

on the reconstruction to fill additional gaps using literature and experimental data. Furthermore, several other gaps were resolved using reciprocal blast hit between all the genes containing the missing Enzyme Commission (EC) number and the *C. thermocellum* ATCC 27405 genome. The process was iteratively performed until a positive flux for biomass synthesis was observed. The model was tested against data for growth on minimal media containing either cellobiose or fructose in continuous or batch cultures and was able to reproduce some phenotypes, although most fermentation product flux profiles were inaccurate. However, the addition of RNA-Seq data allowed for better predictions [26].

Following the construction of the model *i*SR432, much has been learned about the metabolism of *C. thermocellum*, particularly dealing with its atypical central carbon metabolism and redox processes [27, 28]. With these updates in mind, Thompson et al. have constructed and curated a new genome-scale model of *C. thermocellum*, this time of the genetically tractable strain DSM 1313 [29–31]. This new model of *C. thermocellum* DSM 1313 also incorporates a more dynamic cellulosome component, which allowed the researchers to predict more accurately the growth on soluble versus insoluble substrates. This is a key distinction because different substrates lead to different fermentation profiles and energetic requirements for growth. Using this updated model, the authors delved into the changes in metabolism between various substrates to propose a regulatory mechanism that explains the difference. The authors also used a strain design algorithm for optimal production of ethanol, hydrogen, and isobutanol, paving the way for future metabolic engineering [31]

4.2 Thermotoga maritima

T. maritima is a hyperthermophilic anaerobic bacterium believed to be one of the most ancient of eubacteria [32]. Its metabolism is classified as chemoorganotrophic, catabolizing sugars to produce CO_2, acetate, lactate, and hydrogen [33]. For a free-living organism, it has one of the smallest genomes [34]. Zhang et al. created the first metabolic reconstruction of *T. maritima* [35]. This model integrated structural information to examine the evolution of protein folds in the metabolism, and consists of 478 genes, 503 metabolites, and 645 reactions. The model was able to reproduce experimental results for growth and secretion profiles on different substrates. The protein information was gathered through literature data first followed by homology-based annotation databases. Finally, FBA and gap-filling were iteratively carried out until the model was able to replicate experimental growth results.

One important conclusion from the integration of structural data was the strengthening of the 'patchwork' hypothesis, which states that gene duplication events result in proteins that evolve to function in a similar manner to each other but in different pathways [36]. Furthermore, this study discovered that specific folds dominate the proteins involved in central metabolism, which suggests divergent evolution of ancient proteins. The core essential proteins, however, have relatively

diverse folds because they catalyze highly specific reactions, which require particular enzymes.

Nogales et al. expanded the model created by Zhang et al. to study hydrogen production in *T. maritima* [37]. As *T. maritima* is a hyperthermophile and produces large amounts of hydrogen from various complex sugar polymers, it is an ideal candidate for microbial hydrogen bioproduction. The model expansion consists of modifying reactions involving hydrogen production and ferredoxin based on recent findings in *T. maritima*. Similarly, reactions are added for secretion of certain metabolites and in the Entner–Doudoroff (ED) pathway to improve the model. The predictive capability of the new reconstruction was found to be better than that of the original as determined by its ability to replicate more experimental results in silico. According to the model and experimental data, it was confirmed that *T. maritima* grows faster on polysaccharides than on other carbon sources. Moreover, it was determined that this organism mainly uses Embden–Meyerhoff (EM) and ED rather than the oxidative branch of the pentose phosphate pathway (OPP) to catabolize sugars for growth-coupled production of hydrogen. Furthermore, it was demonstrated that acetate production results in improved growth and hydrogen yield. The authors also concluded that sulfur is one of the important electron sinks in this organism based on model prediction and literature information.

The model was further used for developing an understanding of the redox balancing in the organism. The analysis suggested that carbon metabolism leads to surplus NADH, which is consumed in ED pathway coupled with sulfur reduction and subsequently causes the stoichiometric ratio of ferredoxin to NADH to be less than two. This result suggests the role of ED pathway in internal redox balancing. The investigators also used in silico knockouts to determine the mutants that have improved hydrogen yield on various substrates and limited sulfur condition. The simulations suggested that double mutants of acetate thiokinase (ACKr, converts acetyl phosphate to acetate) and L-lactate dehydrogenase (LDH_L, converts pyruvate to lactate) have improved hydrogen production. Mutation in either triose phosphate isomerase (TPI) or fructose bisphosphate aldolase (FBPA) also led to an increase in hydrogen yield when grown on glucose but at the expense of growth rate. Furthermore, it was determined that the addition of an efficient NADP+ regenerating enzyme could drive glucose metabolism through the OPP pathway and lead to more hydrogen production. The knock-in in combination with certain double knockouts resulted in a very high hydrogen yield. An additional confirmation from in silico simulation was that the introduction of NADP+ regenerating enzyme enabled the organism to grow on glycerol, on which the wild type cannot sustain growth. This knock-in coupled with triple mutations in FBPA, 2-dehydro-3-deoxy-phosphogluconate aldolase (EDA_R) and TPI led to improved hydrogen production of up to 5 mol per mole of glycerol [37].

4.3 Thermus thermophilus

T. thermophilus is a gram-negative organism that grows aerobically and anaerobically with the help of exogenous electron acceptors such as nitrate. It can consume a wide variety of protein substrates and carbohydrates, which is facilitated by a suite of proteases, glucosidases, and lipases to allow ideal growth between the temperatures of 65 and 72°C. Lee et al. created the first genome-scale model of this organism, which incorporates several distinct features present in thermophiles and is specific to *T. thermophilus* [38]. The reconstruction, called *i*TT548, contains 548 genes, 796 reactions, and 635 metabolites. The draft metabolic reconstruction was carried out in the usual manner, but the gap-filling required organism-specific information or knowledge from related organisms. Specifically, the pathways for carotenoid synthesis and those relating to the growth on various substrates were incorporated. One of the most distinct features of this model was that the biomass composition was determined using experimental results and literature information exclusive to thermophilic organisms.

When *i*TT548 was compared with *i*AF1260, the then most recent genome-scale model of *Escherichia coli*, it was determined that the amino acid biosynthesis pathways for lysine and methionine were different in these two organisms. Similarly, the model also clearly demonstrated that the carotenoid and polyamine biosynthesis were some of the unique characteristics of this organism compared to other gram-negative bacteria. These two groups of metabolites enable this organism to withstand high temperatures [39–42]. Furthermore, the model also predicted that the organism utilized amino acids to synthesize branched chain fatty acid. When constraint-based flux analysis was carried out on minimal and rich glucose media, the simulation results were consistent with experimental data for growth rate. The simulation demonstrated that certain amino acids were consumed at a higher rate than others to synthesize fatty acids. This pattern was distinct compared to *E. coli* because *T. thermophilus* consumed nutrients to drive fluxes more toward fatty acid synthesis rather than toward energy production.

Finally, gene essentiality was determined in this organism. It was demonstrated that genes involved in carotenoid biosynthesis were the most essential genes in both rich and minimal media. Most of the amino acid metabolism genes were found to be essential in minimal media except for genes involved in biosynthesis of tryptophan, proline, and tyrosine. Moreover, the model also showed that genes involved in oxidative phosphorylation and the citric acid cycle were also essential because *T. thermophilus* is an obligate aerobe. Finally, the simulations showed that *T. thermophilus* has a higher proportion of essential genes compared to *E. coli*. It was concluded that the more rigid network of *Thermus thermophilus* helps it to survive in higher temperatures.

4.4 Sulfolobus solfataricus

S. solfataricus is a hyperthermoacidophilic organism within the phylum *Crenarchaeota* found in volcanic hot springs [43]. It is strictly aerobic and can grow either autotrophically or heterotrophically. It also has the ability to oxidize sulfur. Moreover, its metabolism is very diverse, containing a bicarbonate fixation pathway, so it can grow chemolithoautotrophically and has the ability to grow on phenol [44, 45].

Ulas et al. created the first ever genome-scale model of *S. solfataricus* [43]. It was created by compiling the information from annotations created by EnzymeDetector software [46], various databases such as KEGG [47], MetaCyc [48], BRENDA [49], *Sulfolobus*-specific literature, and experimental data. Following this initial step, manual gap-filling was carried out and resulted in a genome-scale model *i*TU515 which contains 831 reactions involving 705 metabolites. The model was able to depict accurately the broad range of metabolism of *S. solfataricus*.

Following the reconstruction, the predictive capabilities of the model were determined. The organism was grown in silico on various carbon sources and FBA was carried out to compare with experimental results. *S. solfataricus* has a very low growth rate and utilizes only 25% of carbon toward biomass production, which the model was able to predict accurately. *S. sofataricus* has a modified ED pathway such that the glucose flux can go toward either semi-phosphorylative or non-phosphorylative branch. The model demonstrated that the carbon flux of glucose was divided in a ratio of 1:4 between the semi-phosphorylative/reverse ribulose monophosphate pathway and the non-phosphorylative/TCA cycle. In another analysis using this model, the effect of exopolysaccharide (EPS) on the growth of the organism was investigated. *S. solfataricus* produces EPS using the imported carbon flux. The model clearly demonstrated that when EPS-producing reactions were added, the growth rate decreased when grown on glucose media. Therefore, according to the model, the production of EPS causes lower biomass flux.

When flux variability analysis (FVA) was carried out on the model, it was determined that for an optimal flux toward biomass, 79 reactions showed variability. Most of the significant flux variation appeared because of the semi-phosphorylative and non-phosphorylative branches of the ED pathway. Similarly, FVA analysis for suboptimal analysis (in which up to 95% of optimal growth is considered) caused the number of reactions to increase from 79 to 352 that showed flux variability.

The ability of the organism to grow chemolithotrophically using the hydroxypropionate-hydroxybutyrate cycle under aerobic conditions was demonstrated on a bicarbonate source. FBA showed that the ED pathway was inactive and sulfur metabolism and the hydroxypropionate-hydroxybutyrate cycle was active when grown on bicarbonate. Furthermore, the model predicted the growth rate to be higher than that on glucose. In a similar analysis, growth on phenol was determined.

The FBA demonstrated some significant differences between growth on glucose and on phenol such as active phenol uptake and degradation and inactive ED pathways. Biomass production on phenol was determined to be one of the lowest possibly because of the requirement of ATP for the production of pyruvate from phenol.

Additionally, the model was used for the analysis of growth on various other carbon sources. Using FBA and normalization of carbon uptake rate for each carbon source to 1 mmol of carbon atoms per gram dry weight per hour, the model predicted growth on 35 different substrates. Around 13 of the substrates produced more biomass than on glucose. It was determined that carbon sources entering the central pathway through the TCA cycle resulted in lower biomass production, except 2-oxoglutarate which showed a higher yield. Glycerol was determined to be the source of highest biomass production because metabolism of glycerol produced twice the amount of ATP per six carbon atoms than other substrates.

Finally, a gene essentiality analysis on glucose media was carried out on this model. Around 18% of genes were determined to be essential because the in silico deletion of these genes resulted in biomass production of less than 2% of the original. The genes involved in the central metabolism such as the reverse ribulose-monophosphate pathway and gluconeogenesis were determined to be essential. The model predicted only some genes in ED pathway and TCA cycle to be crucial.

4.5 Thermobifida fusca

T. fusca is an aerobic, gram-positive bacterium of the *Actinomycetes* phylum [50]. Because of its stability at high pH and temperature, and possession of an efficient cellulolytic system consisting of several endo- and exocellulases, *T. fusca* could be useful in consolidated bioprocessing of lignocellulose for biofuel production. Deng and Fong demonstrated that *T. fusca* could be manipulated and optimized to produce propanol from untreated biomass [51]. Vanee et al. created three different genome-scale models of *T. fusca* through (1) an automated approach using Model SEED, (2) a semi-automated approach using KEGG, and (3) a proteomics-based model using proteome data for cells grown on cellobiose [52]. The Model SEED (Tfu_v1)-based reconstruction contains 1,302 reactions, but cannot predict growth on cellobiose. Similarly, the semi-automated approach (Tfu_v2) produced a model with 1,002 reactions, but it also could not predict the growth on cellobiose accurately. The proteomics-based genome-scale model (*i*Tfu296) consists of 975 reactions with 296 genes. The simulation with *i*Tfu296 predicted a growth rate similar to that experimentally derived. Between Tfu_v2 and *i*Tfu296, vast differences were observed in functions, and the study suggested that *i*Tfu296 was much closer to the in vivo phenotype. During growth on cellobiose, the *i*Tfu296

predicted 110 active reactions among which the majority of reactions were involved in carbohydrate and amino acid metabolism (Table 1).

The investigators were interested in terpenoid biosynthesis in *T. fusca*, and hence added 16 reactions to compute the feasibility of flux through the terpenoid backbone pathway. The objective function used for FBA was biomass production. Except for one reaction, all reactions in the mevalonate pathway were found to be active, and provided investigators with a hypothesis to test the presence of this pathway experimentally. The experimental results, in contrast, demonstrated that the non-mevalonate pathway is present in *T. fusca*. This further emphasizes the importance of genome-scale models because, with the help of the model, the researchers were able to investigate quickly the terpenoid biosynthesis pathway in *T. fusca*.

4.6 Moorella thermoacetica

M. thermoacetica is a strict anaerobe that can use both electron transport phosphorylation and substrate level phosphorylation to produce energy. It has the ability to convert substrates such as carbon dioxide, glucose, or fructose into acetate and produce ATP [53]. It primarily utilizes the Wood–Ljungdahl (WLD) pathway to produce acetate from CO_2 and hydrogen, and as such it has for decades been widely studied as a model organism in acetogenesis. A genome-scale reconstruction of *M. thermoacetica* was created by Islam et al. and was called *i*AI558. The model contains 558 genes and 705 reactions [53]. The highest number of active reactions was determined to be involved in the cofactor metabolism subsystem.

*i*AI558 was used for simulation of growth on various substrates such as H_2, CO_2, CO, and methanol for autotrophic growth and glucose, fructose, and xylose for heterotrophic growth. Additionally, ATP production and yield were also computed. The growth simulation was compared to experimental data. The model accurately predicted growth rates on H_2-CO_2 (syngas) and CO. The growth rate and yield for CO was determined to be the highest among autotrophic substrates. Growth on heterotrophic substrates produced higher yield, growth rate, and ATP production than that on autotrophic substrates.

Simulation of growth on syngas and glucose were also compared. The study demonstrated that for syngas, *M. thermoacetica* mainly used WLD and gluconeogenesis and conserved energy through electron transport phosphorylation (ETP) or anaerobic respiration. During growth on glucose, glycolysis was the most dominant process, but the WLD pathway was also highly active. The energy conservation/production was carried out by substrate level phosphorylation in glycolytic reactions. Because substrate level phosphorylation is more efficient in ATP generation [53], heterotrophic growth produces higher ATP yield, which was corroborated by the model simulation.

The model was further used for study of ATP generation during autotrophic growth. Reactions were added to the model based on hypotheses proposed by

Table 1 Salient features of some of the important genome scale models described in the text

Features	iSR432[a]	*T. saccharolyticum*	iTZ479[a]	iTfu296[b]	iAI558	iTU515
Total reactions	631	537	645	975	705	718
Exchange	54	22[c]	83	30[c]	60	58
Gene-associated reactions	463	461	518	NA	620	606
Genes	432	315	479	296	558	515
Unique metabolites	525	502[c]	503	734[c]	630[c]	705
Notable complex carbon substrates	Cellobiose	NA	Starch, xylan, cellulose	Cellobiose	NA	Cellulose, starch, xylan
Dominant pathways	Amino acid metabolism	Nucleotide metabolism	Amino acid metabolism Nucleotide metabolism	Amino acid metabolism	Cofactor metabolism	Nucleotide metabolism

[a]Only first model used
[b]Proteomics model used
[c]Determined by parsing the model

previous studies. The first mechanism proposed by Mock et al. was the production of ATP through formate-hydrogen lyase (FHL) and methylene-tetrahydrofolate reductase (MTHFR), which act together to generate a proton gradient for electron transport phosphorylation [54]. The bifurcating ferredoxin:NAD hydrogenase (HYDFDNr} and electron bifurcating ferredoxin:NADP oxidoreductase (FRNDPRr) are also considered important for ATP generation during autotrophic growth. To investigate this mechanism, the investigators changed exchange fluxes for CO_2 and H_2 and monitored ATP flux. The study found that there was no change in ATP flux when such changes were made. The second mechanism proposed by Schuchmann and Muller assumes that ferredoxin hydrogenase (FRHD) is the enzyme required for energy conservation [55]. For this mechanism, HYDFDNr reaction with different stoichiometries than that used in the first mechanism is proposed. Using this hypothesis, the simulations were carried out. Similar results for ATP production were observed when the exchange fluxes for CO_2 and H_2 were varied. However, when the stoichiometry of reactions catalyzed by FRHD and HYDFDNr was changed that, the simulation results were comparable to experimental results for ATP production. A linear relationship between ATP flux and CO_2 and H_2 supply was observed during the simulation of the model containing the stoichiometric changes. Hence, the investigators have proposed that for energy conservation on autotrophic substrates, the proposed second mechanism requires modification.

4.7 Streptococcus thermophilus

S. thermophilus is an important organism in the dairy industry, especially in the production of yoghurt and cheese [56]. It is a borderline thermophile with an optimum growth temperature of 45°C [56]. The genome-scale reconstruction of *S. thermophilus* LMG18311 was constructed by Pastnik et al. to study its amino acid metabolism. The reconstruction was based on closely related organisms such as *L. planatarum* and *L. lactis*. Through a manual gap-filling procedure using literature and experimental evidence, the reconstruction was completed. The model consists of 429 genes and 522 reactions. The biomass was determined in the study itself, and hence makes this model more relevant.

The model could accurately predict that the organism cannot grow on histidine because of the lack of histidine biosynthesis genes. Similarly, the model analysis also determined that *ychE*, a gene involved in the synthesis of cysteine from methionine in *L. lactis*, is truncated in *S. thermophilus*, which could explain its apparent auxotrophy to either of these amino acids. Furthermore, the model could correctly suggest that homofermentative lactic acid production is the primary metabolism in *S. thermophilus*.

4.8 Thermoanaerobacterium saccharolyticum

T. saccharolyticum is a gram-positive anaerobe that is chemoorganotrophic in nature. It is known for high ethanol yields from hexoses and pentoses. Curie et al., to study the metabolic capabilities of *T. saccharolyticum*, created a genome-scale metabolic model of the organism by parsing information from the reconstruction of *C. thermocellum* ATCC 27405 [57]. After the construction of the initial model, gap-filling was carried out using literature information and FBA-Gap [57]. This gap-filling algorithm proposes a minimal set of reactions from a curated reactome database to be added to the model to support biomass synthesis. The refined model consists of 516 reactions, 315 genes, and 528 metabolites [57].

The investigators used experimental data to constrain hydrogenase reactions that contributed less toward hydrogen production. Specifically, the energy-conserving hydrogenase (ECH) was blocked, the bifurcating hydrogenase (BIFH2) and NADH hydrogenase (NADH2) were made irreversible, and the hydrogen export was constrained to reflect the experimental data. These constraints were shown to alter fluxes dramatically, and more accurately predicted the higher ethanol production observed experimentally.

The model was tested for gene knockouts to enhance the production of ethanol. The simulation predicted that the deletion of lactate dehydrogenase (LDH) and phosphotransacetylase (PTA) leads to optimal growth and high ethanol yield. The model also suggested that the deletion of reactions catalyzed by LDH and ferredoxin hydrogenase (HFS) leads to high ethanol yield at the expense of growth rate. The predictions were consistent with experimental results. Furthermore, the model indicated that the deletion of LDH, HFS, and glutamate dehydrogenase (GDH) leads to a marginal increase in ethanol production compared to that of LDH and HFS only. Overall, the model was successful in predicting the metabolic behavior of *T. saccharolyticum* when grown on cellobiose.

4.9 Models Deposited in BioModels Database

In addition to the curated models above, there have been a number of models of thermophiles automatically generated using genome annotations and deposited in the BioModels Database [58, 59]. These include *Pyrolobus fumarii* ([60], BMID:140676), *Pyrococcus furiosus* ([61], BMID:141276), *Archaeoglobus fulgidus* ATCC 49558 ([62], BMID:140871), *Methanococcus jannaschii* ([63], BMID:140493), *Aeropyrum pernix* ([64], BMID:142009), *Aquifex aeolicus* ([65], BMID:141549), *Hyperthermus butylicus* ([66], BMID:140823), *Desulfurococcus kamchatkensis* ([67], BMID:141539), *Desulfurococcus mucosus* ([68], BMID:141869), *Staphylothermus hellenicus* ([69], BMID:140958), and *Alicyclobacillus acidocaldarius* ([70], BMID:140735). As with all automatic

reconstructions, these models need more curation, but they serve as a platform for further development and can only increase in applicability.

5 Conclusion

The range of metabolisms exemplified by thermophilic microorganisms is quite wide, considering the relatively few places on Earth where they thrive. Metabolic network modeling is an effective way to study the metabolism of thermophiles and to compare their metabolism to their mesophilic counterparts. We have chronicled several cases where thermophilic microorganisms have been studied with genome-scale models. However, there are still many challenges for expanding the scope of metabolic network models of thermophiles, such as estimating thermodynamic parameters at higher temperatures, measuring kinetic parameters of key metabolic enzymes, and fully understanding how cofactor usage changes at high temperature (e.g., the preference of ATP versus pyrophosphate as an energy carrier).

Despite the aforementioned issues, genome-scale models of thermophilic organisms are very useful tools for understanding and engineering the metabolisms of non-model strains to enhance their ability for use in the biofuel, waste management, and mining industries. As more data are acquired for thermophiles, in particular large OMICs datasets, the metabolic models continue to improve.

References

1. Brock TD (1985) Life at high temperatures. Science 230:132–138
2. Brock TD, Freeze H (1969) Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile. J Bacteriol 98(1):289–297
3. Bult CJ, White O, Olsen GJ, Zhou L (1996) Complete genome sequence of the methanogenic archaeon. Methanococcus jannaschii. Science 273:1058
4. Brock TD (1967) Life at high temperatures. Science 158:1012–1019
5. Robb F, Antranikian G, Grogan D, Driessen A (2007) Thermophiles: biology and technology at high temperatures. CRC Press
6. Caldwell D, Brannan D, Kieft T (1983) Thermothrix thiopara: selection and adaptation of a filamentous sulfur-oxidizing bacterium colonizing hot spring tufa at pH 7.0 and 74 C. Ecol Bull 38:129–134
7. Zeikus J (1979) Thermophilic bacteria: ecology, physiology and technology. Enzyme Microb Technol 1:243–252
8. Shelef G, Kimchie S, Grynberg H (1980) High-rate thermophilic anaerobic digestion of agricultural wastes. In Biotechnol Bioeng Symp (United States). Environmental and Water Resources Engineering Dept., Technion, Haifa, Israel
9. Gajalakshmi S, Abbasi S (2008) Solid waste management by composting: state of the art. Crit Rev Environ Sci Technol 38:311–400
10. Cecchi F, Pavan P, Alvarez JM, Bassetti A, Cozzolino C (1991) Anaerobic digestion of municipal solid waste: thermophilic vs. mesophilic performance at high solids. Waste Manag Res 9:305–315

11. Micolucci F, Gottardo M, Cavinato C, Pavan P, Bolzonella D (2016) Mesophilic and thermophilic anaerobic digestion of the liquid fraction of pressed biowaste for high energy yields recovery. Waste Manag 48:227–235
12. Deveci H, Akcil A, Alp I (2004) Bioleaching of complex zinc sulphides using mesophilic and thermophilic bacteria: comparative importance of pH and iron. Hydrometallurgy 73:293–303
13. Krebs W, Brombacher C, Bosshard PP, Bachofen R, Brandl H (1997) Microbial recovery of metals from solids. FEMS Microbiol Rev 20:605–617
14. Barrett J (1990) Metal extraction by bacterial oxidation of minerals. Horwood
15. Rossi G (1990) Biohydrometallurgy. McGraw-Hill
16. Bobadilla-Fazzini RA, Cortés MP, Maass A, Parada P (2014) Sulfobacillus thermosulfidooxidans strain Cutipay enhances chalcopyrite bioleaching under moderate thermophilic conditions in the presence of chloride ion. AMB Express 4:1
17. Zhang L, Wu J, Wang Y, Wan L, Mao F, Zhang W, Chen X, Zhou H (2014) Influence of bioaugmentation with Ferroplasma thermophilum on chalcopyrite bioleaching and microbial community structure. Hydrometallurgy 146:15–23
18. Oberhardt MA, Palsson BØ, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Mol Syst Biol 5:320
19. Varma A, Palsson B (1994) Metabolic flux balancing: basic concepts, scientific and practical use. Nat Biotechnol 12:994–998
20. Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. PLoS Comput Biol 9:e1002980
21. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S et al (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc 6:1290–1307
22. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE (2016) BiGG models: a platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res 44:D515–D522
23. Roberts S, Gowen C, Brooks JP, Fong S (2010) Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. BMC Syst Biol 4:31
24. Roberts SB, Gowen CM, Brooks JP, Fong SS (2010) Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production. BMC Syst Biol 4:1
25. Milton H, Reddy VJ, Tamang D, Västermark A (2014) The transporter classification database. Nucleic Acids Res 42:251–258
26. Gowen CM, Fong SS (2010) Genome-scale metabolic model integrated with RNAseq data to identify metabolic states of *Clostridium thermocellum*. Biotechnol J 5:759–767
27. Thompson RA, Layton DS, Guss AM, Olson DG, Lynd LR, Trinh CT (2015) Elucidating central metabolic redox obstacles hindering ethanol production in *Clostridium thermocellum*. Metab Eng 32:207–219
28. Zhou J, Olson DG, Argyros DA, Deng Y, van Gulik WM, van Dijken JP, Lynd LR (2013) Atypical glycolysis in Clostridium thermocellum. Appl Environ Microbiol 79:3000–3008
29. Feinberg L, Foden J, Barrett T, Davenport KW, Bruce D, Detter C, Tapia R, Han C, Lapidus A, Lucas S et al (2011) Complete genome sequence of the cellulolytic thermophile *Clostridium thermocellum* DSM1313. J Bacteriol 193:2906–2907
30. Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, Murphy DM, McCool JD, Warner AK, Rajgarhia VB, Lynd LR et al (2010) Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a pta mutant. Appl Environ Microbiol 76:6591–6599
31. Thompson RA, Dahal S, Garcia S, Nookaew I, Trinh CT (2016) Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of Clostridium thermocellum DSM 1313 implementing an adjustable cellulosome. Biotechnology Biofuels 9:194
32. Ozaki S, Fujimitsu K, Kurumizaka H, Katayama T (2006) The DnaA homolog of the hyperthermophilic eubacterium Thermotoga maritima forms an open complex with a minimal 149-bp origin region in an ATP-dependent manner. Genes Cells 11:425–438

33. Huber R, Langworthy TA, König H, Thomm M, Woese CR, Sleytr UB, Stetter KO (1986) Thermotoga maritima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. Arch Microbiol 144:324–333
34. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA et al (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. Nature 399:323–329
35. Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K, Deacon AM, Wooley J, Lesley SA, Wilson IA (2009) Three-dimensional structural view of the central metabolic network of Thermotoga maritima. Science 325:1544–1549
36. Jensen RA (1976) Enzyme recruitment in evolution of new function. Annu Rev Microbiol 30:409–425
37. Nogales J, Gudmundsson S, Thiele I (2012) An in silico re-design of the metabolism in Thermotoga maritima for increased biohydrogen production. Int J Hydrogen Energy 37:12205–12218
38. Lee N-R, Lakshmanan M, Aggarwal S, Song J-W, Karimi IA, Lee D-Y, Park J-B (2014) Genome-scale metabolic network reconstruction and in silico flux analysis of the thermophilic bacterium Thermus thermophilus HB27. Microb Cell Fact 13:1
39. Kaneda T (1991) Iso-and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. Microbiol Rev 55:288–302
40. Nordström KM, Laakso SV (1992) Effect of growth temperature on fatty acid composition of ten thermus strains. Appl Environ Microbiol 58:1656–1660
41. Pask-Hughes RA, Shaw N (1982) Glycolipids from some extreme thermophilic bacteria belonging to the genus Thermus. J Bacteriol 149:54–58
42. Oshima T (2007) Unique polyamines produced by an extreme thermophile, Thermus thermophilus. Amino Acids 33:367–372
43. Ulas T, Riemer SA, Zaparty M, Siebers B, Schomburg D (2012) Genome-scale reconstruction and analysis of the metabolic network in the hyperthermophilic archaeon Sulfolobus solfataricus. PLoS One 7:e43401
44. Teufel R, Kung JW, Kockelkorn D, Alber BE, Fuchs G (2009) 3-Hydroxypropionyl-coenzyme A dehydratase and acryloyl-coenzyme A reductase, enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in the Sulfolobales. J Bacteriol 191:4572–4581
45. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, Alber BE, Fuchs G (2010) Autotrophic carbon fixation in archaea. Nat Rev Microbiol 8:447–460
46. Quester S, Schomburg D (2011) EnzymeDetector: an integrated enzyme function prediction tool and database. BMC Bioinformatics 12:376
47. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30
48. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 42:D459–D471
49. Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Söhngen C, Stelzer M, Thiele J, Schomburg D (2010) BRENDA, the enzyme information system in 2011. Nucleic Acids Res gkq1089
50. Wilson DB (2004) Studies of Thermobifida fusca plant cell wall degrading enzymes. Chem Rec 4:72–82
51. Deng Y, Fong SS (2011) Metabolic engineering of Thermobifida fusca for direct aerobic bioconversion of untreated lignocellulosic biomass to 1-propanol. Metab Eng 13:570–577
52. Vanee N, Brooks JP, Spicer V, Shamshurin D, Krokhin O, Wilkins JA, Deng Y, Fong SS (2014) Proteomics-based metabolic modeling and characterization of the cellulolytic bacterium Thermobifida fusca. BMC Syst Biol 8:1
53. Islam MA, Zengler K, Edwards EA, Mahadevan R, Stephanopoulos G (2015) Investigating Moorella thermoacetica metabolism with a genome-scale constraint-based metabolic model. Integr Biol 7:869–882

54. Mock J, Wang S, Huang H, Kahnt J, Thauer RK (2014) Evidence for a hexaheteromeric methylenetetrahydrofolate reductase in Moorella thermoacetica. J Bacteriol 196:3303–3314
55. Schuchmann K, Müller V (2014) Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. Nat Rev Microbiol 12:809–821
56. Hols P, Hancy F, Fontaine L, Grossiord B, Prozzi D, Leblond-Bourget N, Decaris B, Bolotin A, Delorme C, Ehrlich SD (2005) New insights in the molecular biology and physiology of Streptococcus thermophilus revealed by comparative genomics. FEMS Microbiol Rev 29:435–463
57. Currie DH, Raman B, Gowen CM, Tschaplinski TJ, Land ML, Brown SD, Covalla SF, Klingeman DM, Yang ZK, Engle NL (2015) Genome-scale resources for Thermoanaerobacterium saccharolyticum. BMC Syst Biol 9:1
58. Chelliah V, Juty N, Ajmera I, Ali R, Dumousseau M, Glont M, Hucka M, Jalowicki G, Keating S, Knight-Schrijver V et al (2015) BioModels: ten-year anniversary. Nucleic Acids Res 43:D542–D548
59. Le Novère N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B et al (2006) BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res 34:D689–D691
60. Anderson I, Göker M, Nolan M, Lucas S, Hammon N, Deshpande S, Cheng J-F, Tapia R, Han C, Goodwin L et al (2011) Complete genome sequence of the hyperthermophilic chemolithoautotroph Pyrolobus fumarii type strain (1A(T)). Stand Genomic Sci 4:381–392
61. Robb FT, Maeder DL, Brown JR, DiRuggiero J, Stump MD, Yeh RK, Weiss RB, Dunn DM (2001) Genomic sequence of hyperthermophile, Pyrococcus furiosus: implications for physiology and enzymology. Methods Enzymol 330:134–157
62. Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD et al (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. Nature 390:364–370
63. Tsoka S, Simon D, Ouzounis CA (2004) Automated metabolic reconstruction for Methanococcus jannaschii. Archaea 1:223–229
64. Kawarabayasi Y, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A et al (1999) Complete genome sequence of an aerobic hyperthermophilic crenarchaeon, Aeropyrum pernix K1. DNA Res 6:83–101, 145–152
65. Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M et al (1998) The complete genome of the hyperthermophilic bacterium Aquifex aeolicus. Nature 392:353–358
66. Brügger K, Chen L, Stark M, Zibat A, Redder P, Ruepp A, Awayez M, She Q, Garrett RA, Klenk H-P (2007) The genome of Hyperthermus butylicus: a sulfur-reducing, peptide fermenting, neutrophilic Crenarchaeote growing up to 108 degrees C. Archaea (Vancouver, BC) 2:127–135
67. Ravin NV, Mardanov AV, Beletsky AV, Kublanov IV, Kolganova TV, Lebedinsky AV, Chernyh NA, Bonch-Osmolovskaya EA, Skryabin KG (2009) Complete genome sequence of the anaerobic, protein-degrading hyperthermophilic crenarchaeon Desulfurococcus kamchatkensis. J Bacteriol 191:2371–2379
68. Wirth R, Chertkov O, Held B, Lapidus A, Nolan M, Lucas S, Hammon N, Deshpande S, Cheng JF, Tapia R et al (2011) Complete genome sequence of Desulfurococcus mucosus type strain (O7/1). Stand Genomic Sci 4:173–182
69. Anderson I, Wirth R, Lucas S, Copeland A, Lapidus A, Cheng JF, Goodwin L, Pitluck S, Davenport K, Detter JC et al (2011) Complete genome sequence of Staphylothermus hellenicus P8. Stand Genomic Sci 5:12–20
70. Mavromatis K, Sikorski J, Lapidus A, Glavina Del Rio T, Copeland A, Tice H, Cheng J-F, Lucas S, Chen F, Nolan M et al. (2010) Complete genome sequence of Alicyclobacillus acidocaldarius type strain (104-IA). Stand Genomic Sci 2:9–18