Introduction

The use of microorganisms to produce compounds of commercial value enjoys a rich history. Of recent interest is the use of algae for the synthesis of nutraceuticals and biofuels. For example, high-value molecules are extracted from microalgae, such as carotenoid pigments and docosahexaenoic acid (DHA), an ω3 fatty acid [1]. Polysaccharides, sterols, and polyunsaturated fatty acids are all nutraceutical compounds extracted from algae [2]. Large-scale commercial culture of strains of Chlorella and Arthospira as a nutritious food date back to the 1960s and 1970s, respectively [1]. Table 1 lists a few of the well-defined molecules of commercial value that are purified from algal sources.

Microalgae hold promise as a source of renewable energy. Algae-derived hydrogen, methane, triacylglycerols, and ethanol all serve as potential materials for biofuels [3–6]. For example, depending on production conditions, Schizochytrium sp. and Botryococcus braunii may yield 50–77% and 25–75% oil by mass, respectively [3]. Algae oils are rich in the triacylglycerols that serve as material for conversion to biodiesel [3]. Some species of microalgae, such as Chlamydomonas reinhardtii, may pro-
duce hydrogen directly [4, 7]. Additionally, the doubling time of microalgae in the exponential growth phase is as short as 3.5 h, and they are efficient at utilizing light to produce biomass, facilitating rapid fuel production [4]. Although some algae may be capable of utilizing biomass feedstocks as other microbes do, utilizing the photosynthetic route will arguably be the most efficient means of biofuel production [5].

Microalgal biofuel cultivation promises to be highly sustainable. Importantly, microalgae are much more distant from the human food chain than plant crops, avoiding competition between agricultural and biofuel resources [3]. As shown in Table 2, biodiesel produced from photosynthetic microalgae have a much higher yield than current biofuels and can be cultured on marginal land, further reducing the diversion of agricultural resources. Additionally, some algae can be cultured with saltwater or wastewater, avoiding use of freshwater resources [5]. Since microalgal fuel yields on an area basis are higher than currently possible with crops, they are more capable of meeting fuel demand [3]. Furthermore, microalgae cultures have been demonstrated to fix carbon dioxide, and may be utilized in the bioremediation of industrial flue gases [8–10]. Algal fuels are therefore carbon neutral, or carbon negative in the case of hydrogen.

Despite the advantages of algae as a source of biofuels, there are still significant challenges that must be addressed before algal biofuels can be widely adopted. Although compatible with the existing fuel infrastructure, biodiesel from algae is not yet economically competitive with fossil fuels or corn ethanol (Table 2). For algae biodiesel production, an additional challenge will be altering the selected algae to produce triacylglycerol fatty acid constituents with the optimal length and hydrocarbon saturation [5]. In this review article, we describe a systems level metabolic modeling approach that enables the generation of hypotheses to modify algal metabolism towards more efficient

### Table 1. Selected molecularly defined products currently isolated from microalgae

<table>
<thead>
<tr>
<th>Name or family</th>
<th>Structure</th>
<th>Companies</th>
<th>Commercialized species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astaxanthin</td>
<td><img src="image" alt="Structure" /></td>
<td>Cyanotech [1]</td>
<td>Haematococcus pluvialis [1]</td>
</tr>
<tr>
<td>(food colorant, antioxidant [1])</td>
<td></td>
<td>Mera Pharmaceuticals [1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bioreal [1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parry’s Pharmaceuticals [1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Algatech [1]</td>
<td></td>
</tr>
<tr>
<td>Docosahexaenoic acid</td>
<td><img src="image" alt="Structure" /></td>
<td>Seamboitic</td>
<td>Cryptocodinium conhii [1]</td>
</tr>
<tr>
<td>(ω3 fatty acid, cardiovascular health, brain development [1])</td>
<td></td>
<td>Martek Biosciences Corporation [1]</td>
<td>Shizochtrium sp. [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OmegaTech [1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nutrinova [1]</td>
<td></td>
</tr>
<tr>
<td>Ethanol (biofuel)</td>
<td><img src="image" alt="Structure" /></td>
<td>Algenol Biofuels [70]</td>
<td>Various cyanobacteria</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Seamboitic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inventre Chemical</td>
<td></td>
</tr>
<tr>
<td>Hydrogen (biofuel)</td>
<td><img src="image" alt="Structure" /></td>
<td>Solarvest BioEnergy [70]</td>
<td>Chlamydomonas reinhardtii [71]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Solaveit BioEnergy [70]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Seamboitic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inventre Chemical</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Solazyme [72]</td>
<td></td>
</tr>
<tr>
<td>Triacylglycerols</td>
<td><img src="image" alt="Structure" /></td>
<td>Aurora Biofuels [70]</td>
<td>Haematococcus pluvialis [10, 73]</td>
</tr>
<tr>
<td>(biodiesel precursor)</td>
<td></td>
<td>Solarvest BioEnergy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Seamboitic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inventre Chemical</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Solazyme [72]</td>
<td></td>
</tr>
<tr>
<td>β-Carotene</td>
<td><img src="image" alt="Structure" /></td>
<td>Western Biotechnology [1]</td>
<td>Dunaliella salina [1]</td>
</tr>
<tr>
<td>(food colorant, provitamin A, antioxidant [1])</td>
<td></td>
<td>Betatene [1]</td>
<td></td>
</tr>
</tbody>
</table>

a) Species data from published trial results by Aquasearch [10]. Additional species are likely suitable for biodiesel production [74], but the identity of algae employed in new commercial ventures are not usually publicized.
production of desired compounds. We describe how such network models are constructed and present a number of case studies in which network modeling has been carried out.

2 In silico directed metabolic engineering approaches facilitate economical production schemes

Theoretically, the yield and synthesis rate of any metabolite could be optimized through the process of metabolic engineering. Metabolic engineering can be described as the optimization of entire metabolic or biosynthetic pathways through the manipulation of the genetic content or environmental context [11]. The advantages of utilizing in silico directed metabolic engineering to optimize microbial production processes over traditional strain improvement methods have already been demonstrated for commercially important microbes such as Saccharomyces cerevisiae [12]. Traditional methods of strain improvement include many rounds of selection, mutagenesis, mating, and hybridization [12]. Modeling approaches obviate this labor-intensive process and further minimize the potential for the introduction and accumulation of undesired mutations that may compromise production conditions [12]. Metabolic engineering can also exploit quantitative fine-tuning of gene expression to optimize product yields [12]. Due to the required efficiency of the production process and necessity to achieve high yields, metabolic models may play an essential part in making microalgal biofuel production commercially viable. Several examples where metabolic engineering guided by large-scale mathematical models have optimized the production of a desired metabolite are presented in Table 3. Many of the models now employed capture metabolic processes at a genome scale [13].

The approaches taken for the production of pharmaceutical compounds, especially biologics, offer a significant contrast to what would be expected for successful methodology for biofuel production. Pharmaceutical production often employs genetic engineering approaches to overexpress a single recombinant protein [11]. This approach cannot be used to optimize production of small molecule metabolites, such as triacylglycerols. The interconnectivity of metabolic pathways, with many metabolites feeding into multiple reactions, can make the optimization process counter-intuitive; a greater knowledge of metabolic network properties and mathematical modeling of these networks are needed to optimize bioproduction processes [14]. The production of some small molecule therapeutics may also take advantage of mathematical modeling of metabolic networks in the future. For example, metabolic modeling has been applied to investigate ways to increase penicillin production [15], and metabolic models will likely play a role in optimizing strains for the production of new antibiotics [16]. Notably, C. reinhardtii is being developed for the production of therapeutic proteins [17].

Mathematical modeling of metabolism can elucidate metabolic network properties and facilitates optimization. At the simplest level, metabolic modeling can supplement high-throughput data generation technologies, such as transcri-
tional profiling, to develop a meaningful visual representation of network function [14]. Furthermore, optimizing individual pathways can impact the utilization of global cofactors, such as NADH, NADPH, and ATP [18]. The utilization of pooled resources by different pathways is one reason a genome-scale mathematical model can be necessary to interpret phenotypic changes in metabolically engineered organisms [14]. Additionally, mathematical optimization focuses development efforts on the engineering strategies most likely to yield improvements in yield, titer, productivity, and robustness [19]. The required resources and time to commercialization can be greatly reduced compared to purely experimental development methods [19].

3 Selection of microalgae for biofuels production

In general, two approaches might be utilized to develop a metabolically engineered organism. Novel strains with perhaps less well-defined metabolisms but with unique, advantageous characteristics (e.g., ability to process a particular substrate) may be utilized and subjected to targeted genetic modification as needed [11]. Alternatively, a model organism with relatively well-defined metabolic machinery already in place could be utilized.

The advantage to utilizing microalgae strains that already produce a desired metabolite is that it may be possible to find wild-type strains that give good yields. The search for such microorganisms is called bioprospecting [11]. The disadvantage is that molecular techniques may not exist for efficiently introducing and obtaining expression of genes in novel microorganisms [11]. For example, difficulties encountered trying to engineer Clostridium acetobutylicum to increase butanol production have led some researchers to develop new butanol production microbes in place of C. acetobutylicum [20, 21]. Notably, aside from Chlamydomonas reinhardtii, methods for the genetic manipulation of algal species are not well established [5].

A more extreme case of truly de novo metabolic design would be to build an organism from scratch for the optimal production of the metabolite of interest [22]. Indeed, recent advances in synthetic biology techniques include the construction of full Mycoplasma genomes and their introduction into an organism [23, 24]. A fully synthetic approach would facilitate the design of microbial factories that would use a minimal mixture of inexpensive feedstock for growth and the optimized conversion to the desired metabolite [25]. However, although synthetic approaches have been successfully utilized to add pathways and gene networks to organisms [26], these approaches have not yet been utilized to make a minimal, fully engineered microbe capable of producing compounds of economic value. There are also additional fundamental challenges to constructing a vi-
able synthetic organism as a commercial production platform, such as making the associated regulatory networks sufficiently robust to environment perturbations, mutations, and noise in gene expression [27, 28].

Each approach has advantages, and all hold potential for microalgal biofuels production. There is substantial interest in both algal bioprospecting and developing laboratory algae such as *C. reinhardtii*. Metabolic systems analysis can play an important role in both approaches.

4 Developing systems biology of algae through metabolic network modeling

Systems biology provides the means to understand the emergent properties of biological systems and predict systems behavior under different physiological conditions. Metabolic network modeling, as a systems approach, integrates different large-scale datasets, genomic information, and mathematical equations, to model and predict the metabolic fluxes of an organism. As described in more detail below, network reconstruction is an iterative process that starts with building a draft metabolic network using the available literature and genomic evidence, the incorporation of reaction stoichiometry, gene-reaction association, and cellular localization of reactions. The next step is the conversion of the reconstructed network into a computable format. The final step is the evaluation and refinement of the network model through comparison with experimental data [29]. The iteration of these steps can improve the accuracy of the model.

4.1 Metabolic network reconstruction and analysis

The workflow for the development and refinement of a metabolic network model is illustrated in Fig. 1. Sequenced genomes serve as a starting point for the reconstruction. In addition to the *C. reinhardtii* genome [30], complete genome assemblies for several algae-related species are available (e.g., *Acaryochloris marina* [31], *Anabaena* sp., *Cyanidioschyzon merolae* [32, 33], *Ostreococcus tauri* [34], and *Synechococcus* sp. [35–37]). As additional high-quality metabolic network reconstructions emerge, metabolism in multiple algal species can be compared *in silico*. Their reconciliation may serve as an additional validation of their reconstruction and enhance understanding of microbial specialization. Comparisons will facilitate selecting an optimal species as a starting point for biofuel production. Furthermore, analysis of several metabolic networks may help to identify ideal species for modification based on the best production potential rather than optimal production in the starting strain [38].

After sequencing, the genome is structurally annotated to define genes and transcribed elements. Once open reading frames (ORFs) are delineated, molecular function can be assigned through comparison with genes associated with proteins of known functions. Functional assignments can be made through profile-based domain assignments [39] or, as a first draft, by predicting protein function based on sequence similarity with proteins of previously annotated function in a database such as Uniprot (http://www.uniprot.org/). The automated annotation pipeline results in a genome annotated with Enzyme Commission (EC) numbers which designate the putative catalytic function of the gene product [40]. The reliability of this process is improved by the availability of accurate annotation data for related organisms.

With an annotated genome in hand, a reconstruction can be generated in a structured format such as a stoichiometric matrix. The stoichiometric matrix accounts for compounds (as rows) and corresponding chemical transformations (as columns).
in which the elements of the matrix correspond to the stoichiometric coefficients. While the stoichiometry of metabolic reactions is fixed, the annotated genome enables the identification of which reactions are to be included in a given network. Reactions are assigned to the annotated genes using a metabolic database such as the Kyoto Encyclopedia of Genes and Genomes (KEGG). Reaction properties such as reversibility or localization to specific cellular compartments are also built into the network model [41]. The resulting reaction network may contain incomplete pathways or lack metabolic functions for which there is empirical evidence. In such cases, the network is curated to make it consistent with the known physiological and biochemical characteristics of the organism [42]. The model is then converted to a computable format to allow for quantitative analysis [43]. SBML formats facilitate the exchange of models between research groups and compatibility with software tools (http://sbml.org).

It is likely that the model will lack reactions that are present in the organism, as many gene functions are undetermined. It is also possible the model may include reactions which are not present [43]. Developing a metabolic network model is an iterative process in which the model is refined as hypotheses based on simulations are tested against experimental results [44]. Metabolomic and transcriptomic data from high-throughput experiments can be used to evaluate and refine the model, iteratively improving its capacity to predict phenotypes.

With a mathematically defined model, analysis can be performed to optimize or characterize the network. Because metabolic reactions occur on a fast time scale relative to other cellular processes, a reasonable assumption that enables the application of several analytical approaches is that the metabolic network operates at steady state. The steady-state assumption is inherent to flux balance analysis (FBA), a widely used metabolic modeling strategy. To analyze the network, constraints are placed on reaction fluxes, such as on the exchange reactions responsible for taking in nutrients, and the network is optimized with respect to a goal, frequently taken to be the growth of the organism (biomass production). The maximization of the objective function subject to constraints makes the linear programming problem a cornerstone of metabolic FBA. However, metabolic systems models are most frequently underdetermined: there are more reactions than metabolites, and there are frequently many solutions that give the same maximum objective. Software tools to perform constraint-based analysis on stoichiometric metabolic models are freely available (for example, the COBRA toolbox [45]). Genome-scale constraint-based models and FBA have been reviewed in more depth elsewhere [13, 46].

The constraint-based analysis approach can be applied to predict flux through metabolic pathways, optimal growth media, product yields, and other factors relevant to bioprocess design and optimization. In the context of metabolic engineering, gene knockouts are simulated by removing the corresponding reactions from the model. While the wild-type system is typically assumed to be optimized for biomass production, techniques have been developed to explore knockout combinations and gene additions that will maximize the production of a target metabolite by coupling it to cell growth [47, 48]. Interestingly, knockout phenotypes may no longer have the same biological objectives as their wild-type parents. It has been noted that the metabolic networks of mutants behave suboptimally with respect to growth, and instead more closely resemble the unperturbed network [49]. Thus, mutant phenotypes may be modeled more accurately through Minimization of Metabolic Adjustment (MOMA) rather than optimization of biomass production [49, 50]. These analytical tools may be useful for metabolic engineering strategies.

5 Case studies

As described, the process of metabolic network reconstruction naturally lends itself to an iterative approach. Subsequent rounds of model refinement facilitate the testing of hypotheses in vivo. The metabolic network model becomes a tool not just for finding the optimal solution to industrially relevant metabolic engineering challenges, but an integral part of conducting genome-scale research into the fundamental operating principles and mechanisms of organisms. To truly exploit the power of the metabolic network modeling approach, in silico research can be directly coupled to experimental verification, improving knowledge of the network components, annotation of the genome, and confidence in model predictions. We discuss three such examples where metabolic modeling has demonstrated encouraging results in the development of engineered microbial strains. A more extensive listing of model-driven metabolic engineering is shown in Table 3. While the application of these metabolic network analyses to algal systems is relatively limited to date, these examples provide an overview of the status of the field and some of the opportunities available.
5.1 Metabolic network reconstruction of *C. reinhardtii* with transcript verification

Manichaikul *et al.* [51] have described an iterative methodology for building a high-confidence, experimentally verified model of central metabolism and have applied the method to an updated *C. reinhardtii* genome sequence. Interestingly, the first round of automated functional annotation found six new enzymatic reactions involved in the production of triacylglycerols that were not present in the previous annotated genome, an enhancement potentially very relevant for future studies into biodiesel production. The reconstructed metabolic network model was initially focused on central metabolism. The network structure and reaction stoichiometry were identified by coupling the automated functional annotation with a manual review of the literature, KEGG, and the Expert Protein Analysis System proteomics server (ExPASy). Postulated transcripts encoding the enzymes mediating the network reactions were verified *in vivo* utilizing RT-PCR and rapid amplification of cDNA ends (RACE). The experimental verification improved the original structural annotation of the sequence, refining 5% of the ORFs. An additional round of expansion and verification was then applied to the network. Interestingly, two of the transcripts could not be verified experimentally, and literature evidence showed that one of the unverified transcripts is regulated by light in a genus of cyanobacteria. It is, therefore, likely the approach can be applied to account for the differential regulation of transcript expression, and thus network structure, based on growth conditions. Boyle and Morgan [52] also constructed a model of *C. reinhardtii* central metabolism and also demonstrated the utility of such network models for refining genome annotation and predicting phenotypes of the alga under defined environments. These network reconstructions can serve as a platform and starting point for more detailed metabolic engineering programs (as described below).

5.2 Optimization of ethanol production in *C. thermocellum*

Recently, a genome-scale metabolic model of *C. thermocellum* was constructed to investigate the production of ethanol from the alkaline cellulose degradation product, cellobiose [53]. The model identified several important knowledge gaps related to central metabolism. None of the existing genome annotations contained a gene for pyruvate kinase, and BLASTP identified several candidate genes that could encode the enzyme. The analysis also identified a gap in the citric acid cycle. The genome does not appear to encode for succinate dehydrogenase and enzymatic activity could not be detected. However, small amounts of succinate were detected in *C. thermocellum* culture, so it is likely there is an alternate pathway utilizing the metabolite [53, 54]. It will be important to determine the metabolic fate of succinate in future experiments and refine the model for improved predictions.

Strategies for increasing ethanol production through genetic modifications and altering the feedstock were identified. Metabolic reactions can exhibit a range of theoretical flux values while meeting the biological objective, and, notably, Roberts *et al.* [53] found this to be the case for ethanol production. Alternative solutions that result in the same optimal objective were sought in flux variability analysis (FVA). FVA predicted strains missing ferroredoxin hydrogenase and growth in media supplemented with lactate and malate results in a maximal 35-fold increase in the maximum theoretical ethanol yield, to about 140 mmol/gDW/h.

5.3 Optimization of lycopene production in *E. coli*

Alper *et al.* [50] investigated gene knockout methods to further optimize an industrial *E. coli* strain for the production of lycopene. A significant difficulty for *in silico* metabolic knockout design is that exhaustive search strategies are combinatorially complex and, therefore, not practical for designs exploiting multiple knockouts. Sequential strategies are not theoretically guaranteed to find the global optimum in the gene knockout space, especially if synergistic interactions are critical to the optimal solution. However, their sequential search method, as validated by an exhaustive pairwise search, performed excellently in identifying the best knockout combinations. Overall, *in vivo* verification of changes in microbial growth and ethanol production agreed well with predictions. However, the accuracy of the *in silico* prediction was compromised when the knockout resulted in the accumulation of 3-phosphoglycerate, a metabolite with known regulatory functions. The genome-scale stoichiometric model utilized did not incorporate regulatory effects, which may explain the discrepancy. Utilizing the sequential search strategy, a triple knockout mutant along the optimal *in silico* path was verified *in vivo* to produce 37% more lycopene than the parent industrial strain.
6 Ready for application: Algal metabolic systems analysis

To date, efforts aimed at genome-scale metabolic modeling have been primarily directed at bacterial networks. Model bacteria, such as *E. coli*, are the among the best characterized organisms, simplifying the substantial task of building a high-quality, curated model [55]. The small genomes of bacteria such as *E. coli* and *H. pylori* have also facilitated the expansion of the scope of the metabolic models to the genome scale [55], which have been iteratively tested and refined [55–57]. Additionally, industries of commercial scale, where modeling and optimization approaches have demonstrated value for other products, have a critical interest in also optimizing the production processes for products derived from microbes [19]. Notably, models of a much more ambitious scale have recently been constructed, such as multicompartmental genome-scale models of human metabolism [58] and the plant *Arabidopsis thaliana* [59].

These advances are being employed in the field of algal biotechnology, and arguably the field of algal systems biotechnology is still in an early stage of development. The sequencing of *C. reinhardtii*'s nuclear [30, 60], mitochondrial [61, 62], and chloroplast genomes [63] has enabled the few published large-scale computational models of algal metabolism. Three computational models of *C. reinhardtii* metabolism have been published. The first constraint-based model featured 484 reactions and 458 metabolites localized in the cytosol and mitochondria [52]. Shortly thereafter, an independent model was published with 259 reactions and 267 metabolites localized to the cytosol, mitochondria, chloroplast, glyoxysome, and flagellum [51]. Additionally, a relatively large kinetic model of algal metabolism has been constructed that includes 95 reactions with 38 metabolites localized to the cytosol and mitochondria [64].

Construction and validation of accurate algal models is certainly more challenging than prokaryotic organisms given the multiple organelles and genomes. However, there is some guidance available from efforts with another complex, photosynthetic organism, *A. thaliana*, and *C. reinhardtii* should be an easier organism to work with [65]. One of the fundamental difficulties with complex multicompartmental models is determining the compartments to which specific metabolic reactions are localized, as duplication of portions of biochemical pathways occurs. A recent study employing three pentose phosphate model alternatives in *A. thaliana* was not able to distinguish between the possibilities using steady-state isotope labeling data [66]. This result emphasizes the need for additional biochemical evidence to develop accurate metabolic models, especially if a metabolic design situation requires manipulating compartment-specific reaction fluxes. However, it is worth noting that networks as large as that of the human [67] and *A. thaliana* [68] have been modeled with less accounting for compartmentalization. It is accepted that model construction is an iterative process [69], and the algae field is well-situated to begin applying and refining these models to guide experimental methods to produce products of commercial value from *C. reinhardtii*.

7 Summary and conclusions

Systems-based metabolic engineering holds promise for algal bioprocess design. Genome-scale models will generate testable hypotheses that may increase understanding of algal metabolism and lead to non-intuitive optimization strategies that traditional methods are unlikely to produce. The adoption of systems-based approaches to metabolic engineering of algae may be a critical step towards making algae-derived biofuels economically competitive. Several sequencing projects are underway, and the subsequent development of in silico models will cooperatively reinforce the utility of systems analysis for the algal biotechnology industry.

We would like to thank Ani Manichaikul for insightful discussions into the applications of microalgae-derived biotechnologies. This work was supported by a grant from the Office of Science (Biological and Environmental Research), US Department of Energy, DE-FG02-07ER64496, to J.P. and K.S.-A.

The authors have declared no conflict of interest.

8 References


