C4GEM - Genome-Scale Metabolic Model to study C4 plant metabolism

Corresponding author
Lars Keld Nielsen
Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Corner College and Cooper Roads (Bldg 75), Brisbane Qld 4072 Australia.
Phone: +617 3346 3986
Email address: lars.nielsen@uq.edu.au
C4GEM – a Genome-Scale Metabolic Model to study C4 plant metabolism

Cristiana Gomes de Oliveira Dal’Molin, Lake-Ee Quek, Robin William Palfreyman, Stevens Michael Brumbley & Lars Keld Nielsen

Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane Qld 4072 Australia.
Abstract

Leaves of C4 grasses (such as maize, sugarcane and sorghum) form a classical Kranz leaf anatomy. Unlike C3 plants, where photosynthetic CO2 fixation proceeds in the mesophyll, the fixation process in C4 plants is distributed between two cell types, the mesophyll (M) cell and the bundle sheath (BS) cell. Here we develop a C4 genome-scale model (C4GEM) for the investigation of flux distribution in M and BS cells during C4 photosynthesis. C4GEM is the first large-scale metabolic model that encapsulates metabolic interactions between two different cell types. C4GEM is based on the Arabidopsis model (AraGEM), but has been extended by adding reactions and transporters responsible to represent 3 different C4 subtypes (NADP-ME, NAD-ME and PCK. C4GEM has been validated for its ability to synthesize 47 biomass components and consists of 1588 unique reactions, 1755 metabolites, 83 inter-organelle transporters and 29 external transporters (including transport through plasmodesmata). Reactions in the common C4 model have been associated with well annotated C4 species (NADP-ME subtypes): 3557 genes in *Sorghum bicolour*, 11623 genes in *Zea mays* and 3881 genes in *Saccharum officinarum*. The number of essential reactions not assigned to genes is 131, 135, and 156 in sorghum, maize and sugarcane, respectively. Flux balance analysis was used to assess the metabolic activity in M and BS cells during C4 photosynthesis. Our simulations were consistent with chloroplast proteomic studies and C4GEM predicted the classical C4 photosynthesis pathway and its major effect in organelle function in M and BS. The model also highlights differences in metabolic activities around photosystem I and photosystem II for 3 different C4 subtypes. Effects of CO2 leakage were also explored. C4GEM is a viable framework for *in silico* analysis of cell cooperation between M and BS cells during photosynthesis, and can be used to explore C4 plant metabolism.

**Key words**: metabolic reconstruction; genome-scale model, C4 plant; systems biology, flux analyses.
Introduction

Many of the most productive crops in agriculture, such as maize, sorghum and sugarcane, possess the Kranz leaf anatomy directly and causally associated with the C4 photosynthetic pathway (Laetsch, 1974; Hatch and Kagawa, 1976). Unlike C_3 plants where photosynthetic CO_2 fixation proceeds in a single tissue, the mesophyll, in C_4 plants this process is distributed between mesophyll (M) and bundle sheath (BS) (Moore, 1984). In the Kranz anatomy, a ring of M, containing a few small chloroplasts concerned with the initial fixing of carbon dioxide, surrounds a sheath of parenchyma cells (BS), which have large chloroplasts involved in the Calvin-Benson cycle. BS cells have thick cell walls and contain centrifugally arranged chloroplasts with large starch granules and unstacked thylakoid membranes, whereas the M cells contain randomly arranged chloroplasts with stacked thylakoids and little or no starch (Kennedy et al., 1977; Moore, 1984; Spilatro and Preiss, 1987). The fully differentiated BS and M chloroplasts each accumulate a distinct set of photosynthetic enzymes and proteins that enable them to cooperate in carbon fixation.

All C4 species operate on the same basic theme of pumping CO_2 via C4 acids from mesophyll tissue when PEPCase activity is enhanced to BS layer where Rubisco is localized and C4 acids are decarboxilated. Other than this, the only common feature shared by C4 plants is a reduction in the ratio of M to BS cells when compared to C3 plants. Because diffusion of organic acids between the M and BS must be relatively rapid, M cells in the C4 plants are rarely more than one cell distant from BS cells. As a result, M to BS ratios are between 1 and 2 in C4 plants, while they are over 4 in most laminate C3 leaves. Beyond these common features, C4 plants exhibit substantial variation in how they accomplish CO_2 concentration. These variations result from three distinct decarboxylation modes and multiple patterns of anatomical modification.
Biochemically, the major distinguishing feature of C4 subtypes is the enzyme used for
decarboxylation step in BS (Hatch, 1974; Hatch and Kagawa, 1976; Hatch, 1992). The three
carboxilation modes are: NADP-malic enzyme (NADP-ME: present for example in maize,
sugarcane and sorghum), NAD-malic enzyme (NAD-ME: present in Panicum miliaceum and
Amaranthus spp) and PEP carboxykinase (PKC: present in Panicum maximum and Aamaranthus
spp). These different carboxilation modes are represented in Figure 1. It is important to note that
these different mechanisms lead to different patterns of intercellular movements of C3 and C4
compounds. During C4 photosynthesis, atmospheric CO2 is initially fixed in M cells by
phosphoenolpyruvate carboxylase (PEPC) to form C4 dicarboxylic acids. The movement of
photosynthetic intermediates between mesophyll and sheath cells is restricted largely or entirely
to the plasmodesmata and transpirational water movement through the cell walls (Evert et al.,
1977). The C4 dicarboxylic acids are transferred to Kranz cells (BS cells), and decarboxylated
there. The released CO2 is refixed via ribulose 1,5 biphosphate carboxylase of Kranz cell
chloroplasts and incorporated into the Calvin-Benson cycle (Hatch and Kagawa, 1973, 1974,
1976). The transfer of atmospheric CO2 to BS cells is unique to C4 photosynthesis since it leads
to concentration of CO2 at the site of the reaction of Rubisco (Furbank et al., 1989, 1990). This
CO2 concentrating mechanism, together with specific properties of photosynthetic enzymes,
configurations of intracellular and intercellular transporters and modifications in leaf anatomy,
allows C4 plants to achieve higher photosynthetic efficiency than C3 plants (Bull, 1969). C4
photosynthesis allows fast biomass accumulation with high nitrogen and water use efficiency
(Smith, 1976; Sage, 2004).
Attempts have been made to engineer C4 metabolism in C3 plant species, such as rice, by
overexpressing C4 enzymes (from maize) (Suzuki et al., 2000; Ku et al., 2001; Agarie et al.,
2002; Suzuki et al., 2006). However, significant increase in photosynthetic efficiency in the C3 mutants is yet to be achieved.

With the genome sequence available for sorghum (*Sorghum bicolour*) (Paterson et al., 2009) and maize (*Zea mays*) (Maize Genome Sequencing Project, 2009; Schnable et al., 2009), it is possible to perform a whole-genome-level exploration of C4 pathway evolution by comparing key photosynthetic genes in sorghum and maize (Wang et al., 2009; Yilmaz et al., 2009). These genome projects also facilitate the annotation of other important C4 crops for which incomplete genome sequences are available, such as sugarcane (also a NADP-ME type) (Dufour et al., 1997; Guimaraes et al., 1997). Along with the increase in genomic information, diverse datasets, including genomic, transcriptomic, proteomic and metabolomic data, are becoming readily available for plants (Morgenthal et al., 2006; Thompson and Goggin, 2006; Dai et al., 2007; Yonekura-Sakakibara et al., 2008; Minic et al., 2009; Sawada et al., 2009), including C4 plant species (Majeran et al., 2005; Papini-Terzi et al., 2005; Majeran et al., 2008; Vicentini and Menossi, 2009).

The annotated genomes, along with literature data, define the known biological components of organisms. From these components, biological networks can be reconstructed and the properties of the system explored. *In silico* genome-scale metabolic reconstruction remains the most tangible and applicable systems biology approach for metabolic engineering (Blazeck and Alpert, 2010; Dietz and Panke, 2010; Tyo et al., 2010). A well curated genome-scale network reconstruction is a common denominator for those studying systems biology of an organism. For these metabolic networks, a stoichiometric formalism (Edwards and Palsson, 2000; Thiele and Palsson, 2010) is natural and underpins a powerful set of analytical tools for exploring relationship between genotype and phenotype, predicting product yields and growth rates under
different environment, and discovering global effects of genetic manipulations (Edwards and Palsson, 2000; Famili et al., 2003; Price et al., 2003; Quek and Nielsen, 2008; Oberhardt et al., 2009; Thiele and Palsson, 2010).

The most represented domain for genome-scale models is bacteria, with 25 species reconstructed (Oberhardt et al., 2009). Genome scale metabolic networks have also been reconstructed for mouse (Sheikh et al., 2005; Quek and Nielsen, 2008), human (Mo et al., 2007; Sigurdsson et al., 2009) and more recently for Arabidopsis (Poolman et al., 2009; de Oliveira Dal'Molin et al., 2010). These reconstructions vary both in detail and in scope. For example, Poolman et al.’s Arabidopsis reconstruction (Poolman et al., 2009) considered two organelles (cytosol and mitochondria) to represent the plant metabolic network and hence was limited to describing heterotrophic cell culture. In contrast, our model (AraGEM) (de Oliveira Dal'Molin et al., 2010) describes metabolism across cytosol, plastids, mitochondria, peroxisomes and vacuole, which enabled exploration of metabolic differences in photosynthetic and non-photosynthetic cells, and to examine in greater metabolic detail cells undergoing photosynthesis, photorespiration, beta-oxidation or respiration. While both models covered the full range of known Arabidopsis reactions, validation was limited to primary metabolism, i.e., the synthesis of major biomass components (amino acids, nucleotides, lipid, starch, cellulose, some vitamins). Nevertheless both models are important contributions redressing the conspicuous lack of plant metabolic reconstructions (Oberhardt et al., 2009).

Genome-scale model reconstruction and validation of eukaryotes is certainly more challenging than prokaryotic organisms given the multiple organelles. However such approach applied for complex, multi-tissue plant systems would be of potentially even greater value, since direct measurement of fluxes using pulse chase studies is practically impossible in plants due to signal
dispersion in the large network as well as metabolite and pathway replication in multiple organelles. In this work we attempted to use this systems-based framework to understand the interplay of metabolic network in complex systems, like C4 plant metabolism. This is the first reconstruction used to formulate an \textit{in silico} genome scale model describing the interactions between the two tissue types, M and BS, involved in C4 metabolism. The predicted differences in flux distribution between M and BS cells were in agreement with observed proteomic studies and C4 metabolism features reported in the literature, suggesting that the model can be used to further explore C4 plant metabolism.
Results and Discussion

Characteristics of the Genome-scale Model

The metabolic network reconstruction includes associations between genes, enzymes and reactions to represent C4 plant metabolism based on best available online resources (Table I). Draft models were constructed from all reactions supported in KEGG for the three C4 plants sorghum, maize and sugarcane. Inconsistencies in the KEGG database (e.g., multiple IDs for single metabolites and unbalanced reactions) were resolved as described previously (Quek and Nielsen, 2008; de Oliveira Dal'Molin et al., 2010).

Subcellular compartmentalization plays an important role in plant metabolism. An initial attempt to use WoLF PSORT (http://wolfpsort.org/), a protein subcellular localization predictive tool (Horton et al., 2007), to assign compartments for each gene failed as most calls were ambiguous or inconsistent across the three species. Instead, AraGEM (de Oliveira Dal'Molin et al., 2010) was used to make the compartmentalization calls for each reaction. AraGEM is manually curated for subcellular localisation (cytosol, mitochondrion, plastid, peroxisome or vacuole) and we have not identified any inconsistencies between localisation in Arabidopsis and the limited number of proteins in C4 plants, for which experimental localisation data are available. The compartmentalised draft model covers 3557 genes and 13,114 gene-reaction-association entries for Sorghum, 11,623 genes and 38,829 associations for maize, and 3,881 genes and 13,593 associations for sugarcane (Table II). The large number of gene-reaction associations reflects that – in absence of localisation information – genes associated with a reaction found in multiple compartments must be associated with each compartment.

A functional model of primary metabolism, C4GEM, was derived from the draft model through manual curation. The model describes 1588 unique reactions involving 1755 metabolites (Table
II). Forty seven (47) biomass drain equations describe the accumulation of carbohydrates, amino acids, fatty acid (palmitic acid only), cellulose and hemi-cellulose, representing the major biomass drains for a plant cell, as well as some vitamins and co-factors (Table III). Eighteen (18) inter-cellular exchange reactions (cytoplasm–extracellular) have been included to describe the uptake of light (photons), assimilation/secretion of inorganic compounds (CO$_2$, H$_2$O, O$_2$, NO$_3$, NH$_3$, H$_2$S, SO$_4^{2-}$, PO$_4^{3-}$), translocation of sugars (sucrose, glucose, fructose, and maltose), and amino acids (glutamine, glutamate, aspartate, alanine and serine). Together with biomass drains, the intercellular exchangers define the broad physiological domain of the model, that is, the curated aspects of primary C4 plant metabolism captured by C4GEM. A total of 83 inter-organelle transporters were introduced in the model to achieve the desired functionality. The active, validated scope of C4GEM includes glycolysis (plastidic and cytosolic), pentose phosphate pathway (PPP) (plastidic and cytosolic), tricarboxylic acid cycle (TCA cycle), light and dark reactions (Calvin cycle), NADH/ NADPH redox shuttle between the subcelular compartments, fatty acid synthesis, beta-oxidation and glyoxylate cycle.

Although genome annotation can be used to derive a majority of plant metabolic reactions, there are missing reactions where the genome has not been fully elucidated. During the reconstruction, gaps were found in the primary metabolism for the three C4 models: 131, 135 and 156 gaps for *Sorghum bicolour*, *Zea mays* and *Saccharum officinarum*, respectively. Some of these gaps could be filled by combining all of the C4 plant models. The remaining 100 essential reactions were all found in AraGEM [See list of covered reactions in additional file 1]. The reconstruction allowed a functional annotation of the C4 plant models to ensure that reaction gaps are filled in the context of metabolic functions of nucleotide, carbohydrates, amino acids and nitrogen metabolism, citrate cycle (TCA), carbon fixation and pentose phosphate pathway (PPP).
The final model has 446 singleton or dead-end metabolites (i.e. internal metabolites only used in a single reaction; supplemental data Table S2). A total of 512 reactions are linked to the use or production of dead-end metabolites. These reactions will by definition have zero flux in flux balance analysis. There are several causes of singletons. Most are linked to vitamins, cofactors, and secondary metabolites not currently included in the model (i.e. not described by biomass drains or intercellular transporters). Some result from true gaps in the network, where one or more essential reactions have not yet been assigned. Finally, some result from KEGG’s automatic annotation, in which every reaction known to be catalyzed by a particular EC enzyme in some organism will be included by default, whether or not other reactions required for functionality of the particular reaction are present. Continued curation efforts will focus on resolving these singletons.

In its current form, the model has 183 degrees of freedom. Of these, 47 are associated with the production of biomass components and reduce to a single degree of freedom (growth rate) when a fixed biomass composition is assumed. Another 18 degrees of freedom are associated with intercellular transport. The remaining 118 degrees of freedom represent the maximum cellular scope for using alternative pathways to achieve identical outcomes in terms of growth rate and net transport (e.g. the use of cytosolic or chloroplastic glycolysis). The real scope is further limited by irreversibility constraints as well as regulatory constraints. C4GEM is available in System Biology Markup Language (SBML) format for maize, sorghum and sugarcane (Supplemental Data Files S3, S4, and S5).

**Two-tissue model**

The active scope of C4GEM is similar to AraGEM (de Oliveira Dal'Molin et al., 2010). We have previously demonstrated how AraGEM can be used to explore metabolism in a single tissue, e.g.,
photorespiration in C3 mesophyll. The purpose of this study, however, is to explore the interactions between two distinct tissues, M and BS. Significantly, it represents the first step to represent a plant multi-tissue model.

C4GEM encapsulates the full complement of primary metabolic reactions available to any cell in a C4 plant. Tissue specificity is realised through differential gene expression, i.e., by constraining what part of the network is available to the individual tissues. Thus, M and BS tissue will be described as two instances of C4GEM differing only in terms of the constraints applied to the models (see later). The Kranz anatomy is modelled by connecting these two instances, M and BS, through transporters representing plasmodesmata. The movement of photosynthetic intermediates between mesophyll and sheath cells (Figure 1) is largely or entirely limited to the plasmodesmata (Evert et al., 1977). While plasmodesmata are capable of transporting a broad spectrum of low molecular weight (<800-900 Da) metabolites, significant transport requires high concentration and transporters were only added for major substrate carriers involved in the C4 photosynthetic pathway (subtypes NADP-ME, NAD-ME and PCK), i.e., malate, pyruvate, phosphoglycerate, triose phosphates, aspartate, alanine and phosphates (pyrophosphate and orthophosphate) (Furbank et al., 1989, 1990) (Figure 1).

The final C4 model is described pictorially in Figure 2A. The model based on the annotated genome for sorghum, but uses AraGEM for compartmentalization and the 131 missing, essential reactions in primary metabolism. The model has been extended with 11 intercellular transporters (plasmodesmata), including sucrose translocation and gases efflux.

Mathematically, the large network is broken down into a stoichiometric matrix, S, with a row for each intracellular metabolite and a column for each reaction. Component balances governing a
metabolic network of \( N = \{1, \ldots, n\} \) chemical transformations and \( M = \{1, \ldots, m\} \) chemical entities are

\[
\frac{dc_i}{dt} = \sum_{j=1}^{n} S_{ij} r_j, \forall i \in M
\]  

(1)

where \( C_i \) denotes the concentration of chemical entity \( i \), \( S_i \) is the stoichiometric coefficient of chemical entity \( i \) in transformation \( j \), and \( r_j \) the corresponding flux of transformation \( j \). The time scale of metabolic reactions is short compared to regulation and growth dynamics allowing us to invoke the steady-state assumption

\[
\sum_{j=1}^{n} S_{ij} r_j = 0, \forall i \in M
\]  

(2)

In addition, some reactions are irreversible, so

\[
r_j \geq 0, \forall j \in N_{irreversible}
\]  

(3)

With two tissues communicating, the model is duplicated and the metabolites in plasmodesmata, which are external (unbalanced) for the individual tissues, become internal and balanceable in the two-tissue model. In matrix form, the two-tissue model reads

\[
\begin{bmatrix}
S_M & 0 \\
0 & S_{BS} \\
P_M & P_{BS}
\end{bmatrix}
\begin{bmatrix}
P_M \\
P_{BS}
\end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\]  

(4)

where \( P_M \) and \( P_{BS} \) have a row for each of the 11 metabolites transported via plasmodesmata and a “1” entry in the column for the corresponding transporter. Thus, the third row dictates that the sum of metabolites exported into the plasmodesmata is 0, or in other words, whatever is exported from \( M \) must be taken up by \( BS \) (negative export).

The two-tissue model describes the network topology connecting inputs (e.g., photons, inorganic compounds) to outputs (e.g., biomass). The network is redundant, i.e., there are several paths through the network connecting any given input to any given output. This redundancy is inherent in all biological networks and confers flexibility and robustness to biological systems. The actual
path chosen is defined by enzyme kinetics and regulation and the current model does not consider these.

Equations (3) and (4) define the feasible flux space, i.e., the combinations of fluxes in M and BS that satisfy flux balance and irreversibility constraints. By careful formulation of additional assumptions and – in some instances – an optimality criterion is possible to explore the feasible space and formulate testable biological hypotheses, e.g.,

- Given a certain photon supply, what is the maximum rate of photosynthate production?
- What is the minimum number of photons required to support plant metabolism and what is the optimal flux distribution?
- Is a particular gene deletion expected to affect growth and/or photosynthate production?

It is also possible to incorporate expression data from transcriptomics or proteomics to reduce the feasible solution space by deleting non-expressed reactions. In the following sections, we will illustrate the use of constraint-based analysis.

**Simulations**

In this study, we are interested in exploring, if (a) the observed differences between M and BS metabolism in C4 plants are consistent with an assumption that the plant has evolved to use the network in an optimal manner and (b) if the different C4 subtypes differs in their potential efficiency.

For a photosynthetic system, a logical optimality criterion is photon efficiency, i.e., a plant network will operate with the feasible flux distribution that minimizes photon uptake for fixed rates of biomass synthesis and export of carbon assimilates to other tissues. While no proof exists for the validity of this criterion, we observed for AraGEM that this criterion accurately predicted the classical photorespiration pathway as the most photon efficient way to handle the effect of
the RuBisCo oxygenation reaction and that the predicted cost of photorespiration compared to pure photosynthesis was a 40% increase in photon usage for a 3:1 carboxylation:oxygenation ratio, which is consistent with experimental data (de Oliveira Dal'Molin et al., 2010).

Assumptions and constraints

As a minimum definition of tissue specificity, RuBisCO and phosphoenolpyruvate carboxylase activity is constrained to BS and M tissues, respectively, based on gene expression, enzyme activity and proteome studies (Ghosh et al., 1994; Gutteridge and Gatenby, 1995; Dong et al., 1998; Patel and Berry, 2008) (Table IV). These constraints are common to all the 3 C4 subtypes. Mathematically, the tissue specific activities are captured by equality constraints:

\[
\begin{align*}
 r_{M}^{\text{RuBisCO}} &= 0; \quad r_{BS}^{\text{PEPCase}} &= 0; \quad r_{BS}^{\text{CO}_2} &= 0;
\end{align*}
\]

(5)

In order to represent each C4 subtype, the enzyme used to release CO₂ from the C4 acids is specified (NADPH-ME, NAD-ME or PCK) (Table IV). Similarly, the active C3-C4 transport systems have been specified based on subtypes. Finally, the drain of biomass components, starch accumulation and export of carbon assimilates (sucrose) were fixed based on average measured values.

It should be stressed that this is a minimum definition of C4 subtype metabolism used in this study because we seek to explore questions around optimality. More detailed constraints, e.g., based on transcriptome analysis, may be used when exploring other questions.

**Preferential starch accumulation in BS cannot be explained by photon efficiency**

In C4 plants, starch is principally synthesized in BS, though the amount of mesophyllic starch biosynthesis is species dependent (Black and Mollenha, 1971; Spilatro and Preiss, 1987). In maize, for example, starch biosynthesis occurs only in the BS during normal photosynthesis, but
plants growing in continuous light do accumulate starch in the M, suggesting that photosynthate availability is the limiting factor (Downton and Hawker, 1973).

One explanation for preferential starch biosynthesis in BS could be that it is “less expensive” (i.e., require less photons) to accumulate starch in BS than in M, and hence BS stores are filled before M stores of starch. We tested this hypothesis by comparing the minimum photon requirement predicted with all starch produced in M to the minimum photon requirement for the standard model with all starch produced in BS. We did not observe an energy efficiency advantage for producing starch in BS rather than in M, thus preferential starch production in BS cannot be argued to be the result of an evolutionary pressure for higher energy efficiency. Starch accumulation in BS only was added as an experimentally-based constraint for subsequent analysis.

**Sucrose translocation**

In order to sustain other plant tissues, fixed carbon must move out of the photosynthetic cells and into phloem (Braun and Slewinski, 2009). For most plants, this occur by loading sucrose into the phloem and transporting it from source tissues (net exporters) to sink tissues (net importers), where sucrose in unloaded. Sucrose drain from BS to vascular parenchyma has been added to represent the translocation of carbon for the sustenance of the sink tissues. It unclear if sucrose produced in mesophyll or BS is transported through plasmosdesmata during C4 photosynthesis or if malate and pyruvate (for NADP-ME subtypes, like maize) are the main carbon flow during the carbon assimilation. We tested this hypothesis *in silico* by allowing sucrose exchange between both cells during C4 photosynthesis. According to the flux variability analysis (see Methods) sucrose can be exchanged during this cooperative process but malate and pyruvate
show a much greater flux in magnitude (about 100 folds). Sucrose exchange between both cells did not affect the main contrasts in metabolic activity by comparing BS and M.

**In silico flux profile correlates with maize chloroplast proteome studies**

The optimal flux distribution predicted by the two tissue model shows a number of reactions through which there is a significant difference in the flux observed in BS and M (Figure 3). While the magnitude of flux is not directly linked to the amount of enzyme present, one might reasonably expect that a significant increase in flux in a tissue in most cases is matched by a significant increase in the expression of the corresponding enzyme. We compared metabolic flux predictions to a large dataset comparing BS and M chloroplast proteomes in maize (Majeran et al., 2005; Majeran et al., 2008; Friso et al., 2010). Since the optimality criterion does not necessarily yield a unique flux distribution, flux variability analysis (see Methods) was used to establish the range of values for each flux for which optimal photon usage could be achieved. All but four fluxes – ATP/ADP and OAA/Malate transporters, Glyceraldehyde 3-phosphate dehydrogenase and triose phosphate isomerise – were uniquely defined by the optimality criterion.

For 50 of the 66 proteins or protein complexes annotated in the proteome assigned to enzymatic reactions, differential expression (or absence hereof) was consistent with predicted flux differences (Table V). Of these 50, two predictions were a direct consequence of the model assumptions for RuBisCO and starch synthesis (Table IV). Twenty of the reactions were assigned to fatty acid biosynthesis and the average and median BS-M ratios for proteins involved in fatty acids synthesis were 0.91 and 0.94, respectively, supporting the assumption that the demand for fatty acids is similar in M and BS.
The simulations show that flux through enzymes involved in carbon fixation, PPP, and starch synthesis are upregulated in BS, while flux through enzymes involved in purine, pyruvate and nitrogen metabolism and nitrogen assimilation are upregulated in M. The model also highlights contrasts in flux distribution through isoenzymes distributed in multiple organelles, such as higher flux through malate dehydrogenase (MDH) in M plastids, and higher flux through malic enzyme in BS plastids. Interestingly, in silico analysis shows no difference in flux through chloroplastic pyruvate dehydrogenase during photosynthesis, indicating an essential need to maintain malate and pyruvate pools required for operation of the C4 cycle. These flux predictions agree with the proteome studies.

Of the 16 reactions not accurately predicted, six are involved in chlorophyll and isoprenoid synthesis. While C4GEM has been validated for its ability to synthesize chlorophyll, the biomass content of chlorophyll was not specified in the model. Since BS chloroplasts have lower chlorophyll b levels, most of the predictions would agree with the proteome data if the content was accurately specified in the biomass equation. The reason for higher protoporphyrinogen oxidase activity in BS is unclear given that enzyme catalysis an early step in chlorophyll biosynthesis. Another seven of the reactions not predicted in the simulated fluxes relate to the expected overexpression of oxidative stress response activities in M tissues. The model currently does not describe the generation of reactive oxygen species or their downstream products, and hence it does not predict the need for these activities. Similarly, the overexpression of three enzymes involved in degradation pathways was not predicted. The photon optimal solution for making biomass clearly will not predict degradation, which is a wasteful process. The flux through maintenance and turn-over processes need to be expressed explicitly to be accounted for in the model.
The model predicts cyclic electron flow in bundle chloroplast for NADPH-ME subtype

Whereas linear electron transport (LET) in photosynthesis produces both ATP and NADPH, the cyclic electron transport (CET) around photosystem I has been shown to produce only ATP (Shikanai, 2007; Endo et al., 2008). Two alternative routes have been shown for CET; NAD(P)H dehydrogenase (NDH)-and ferredoxin:plastoquinone oxidoreductase (FQR)-dependent flows, but their physiological relevance has not been elucidated in detail. Meanwhile, because C4 photosynthesis requires more ATP than does C3 photosynthesis to concentrate CO₂, it has not been clear how the extra ATP is produced. LET alone cannot produce enough ATP to satisfy the stoichiometry required for CO₂ fixation. The shortage of ATP may be compensated for by the function of CET. Interestingly, two subtypes of C4 photosynthesis, NAD-ME and NADP-ME, have been shown to have different cell-specific ATP requirements. The ATP/NADPH ratio required in NAD-ME species is higher in M cells than in BS cells, but the opposite is true for NADP-ME species (Moore, 1984; Takabayashi et al., 2005). This cell-type-specific ATP requirement suggests that the activity of CET should be higher in M cells of NAD-ME species and in BS cells of NADP-ME species, if CET plays a critical role in supplying the additional ATP needed in C4 photosynthesis.

C4GEM predicts conventional linear electron transfer (LET) in M chloroplasts and cyclic electron transfer (CET) in BS chloroplast for NADP-ME species (Figure 4). One explanation for preferential CET in BS could be that it requires less photons to get the extra ATP in BS chloroplast than in M chloroplast. We tested this hypothesis by comparing the minimum photon requirement predicted with CET in M chloroplast to the minimum photon requirement for the model with the CET in BS chloroplast. C4GEM highlights that CET in BS is energetically more efficient (about 25% less photons required) in NADP-malic enzyme type metabolism than if
CET was active in M. Our simulations are consistent with the proteome data showing high NDH activity in BS (Table V), which is backed by previous studies showing that ndh genes expression are upregulated in BS for subtypes NADP-ME (Dariel et al., 2006; Shikanai, 2007).

**The model predicts little difference in theoretical quanta requirement for the C4 subtypes**

The costs of concentrating CO₂ in BS in NADP-ME and NAD-ME type plants consist of 2 ATP per CO₂ fixed for regeneration of PEP, plus the amount of ATP required to pump extra CO₂ (overcycle CO₂) to compensate for the CO₂ that leaks out of the BS. In PEP-CK type, one extra ATP and 0.5 extra NADPH are required in addition to the amount required for overcycling.

Our flux simulations show little difference in theoretical quantum yield of photosynthesis among the three C4 decarboxylation types (Figure 4), also discussed in the literature [60]. Such result indicates that these distinct biochemical subtypes amongst C4 plants represent different, equally optimal biochemical solutions to the same problem.

Although the theoretical yield is similar for all the C4 subtypes, in practice differences in quantum yield is observed (Hatch et al., 1995). The reasons for the differences in quantum yield between decarboxylation types have proved elusive to identify. Our results suggest that these differences are not due the distinct C4 decarboxylation pathways. It has been suggested that if CO₂ leakage (from BS back into M) occurs, this would change quantum yield in between C4 plants (Furbank et al., 1989). However, estimates of leak rates show little correlation between C4 subtype and quantum yield. Our flux simulations show little change in total quanta requirement under the assumption that leakage occurs (Figure 4), but the simulation show considerable differences in activity of linear and cyclic electron transport around PSI and PSII for each of the
C4 species. These results are consistent with the differences in photochemical activity in C4 subtypes observed in the literature (Moore, 1984; Takabayashi et al., 2005; Romanowska and Drozak, 2006). As discussed previously, photosynthetic electron transport can involve either a linear flow from water to NADP, via photosystem II (PSII) and photosystem I (PSI) or a cyclic flow just involving PSI. Because C4 subtypes have different cell-specific ATP requirements [3, 63], there are differences in activity of PSI and PSII (Romanowska and Drozak, 2006). It is found for example that NAD-ME and PCK species have substantial PSII activity in the BS tissues, while NADP-ME species have little PSII activity.

**Conclusion**

C4GEM is the first genome scale metabolic reconstruction of C4 plants. The two tissue model based on C4GEM is the first attempt of constructing and integrating context-specific metabolic networks in multi-cellular interactions. The use of this model for *in silico* flux analysis illustrates the potential of using genome scale models to explore complex and compartmentalized networks to derive non-trivial hypotheses. The *in silico* predicted differences in the metabolic activity in BS and M during C4 photosynthesis agreed with observed differences in proteome data for many plastidic enzymatic reactions. The fact that *in silico* fluxes predicted assuming optimal photon use agree with observed proteome differences and C4 metabolic features reported in the literature lend support to the assumption that gene expression in BS and M has evolved to realise photon optimality.
Material and Methods

Metabolic reconstruction

The reconstruction process employed here is the same as the ones used for *Mus musculus* (Quek and Nielsen, 2008) and Arabidopsis GEMs (de Oliveira Dal'Molin et al., 2010). The workflow for the development and refinement of the C4 genome-scale model (C4GEM) is illustrated in Figure 5. The process of the metabolic reconstruction naturally lends itself to an iterative approach and subsequent rounds of model refinement facilitate the testing of hypothesis.

The network reconstruction followed the steps:

1) *Extraction of the metabolic information*: A gene-centric organization of metabolic information was adopted, in which each known metabolic gene is mapped to one or several reactions.

The metabolic elements contained in C4GEM (genomic metabolic reactions) were collected from Kyoto Encyclopedia of Genes and Genomes (KEGG, Release 49.0, August 3, 2009) (Kanehisa et al., 2008) for three C4 plant models: maize, sorghum and sugarcane (stored in excel spreadsheet).

2) *Manual curation*: The metabolic information was merged into a single generic plant cell model, and subsequently curated using literature information. AraGEM was used as a core model to cover gaps found in primary metabolism of the C4 plant models. Curation was performed in Excel, accounting for cellular enzyme localization.

3) *Enzyme localization*: enzymes were assigned to different compartments according to literature evidence or enzyme localization databases (e.g., PPDB - a Plant Proteome DataBase for *Arabidopsis thaliana* and *Zea mays*);

4) *Gene association*: in many cases there are many genes for a particular enzyme (i.e., isoenzymes) found in multiple compartments. Attempts to assign localisation based on *in silico*
predictions failed (see Results & Discussion) and generally all genes for a particular reaction had to be assigned to all organelles to which the reaction had been assigned in 3).

5) **Stoichiometric matrix**: The set of unique reactions ID were extracted and stored as a stoichiometric matrix (Java application). In this step, multiple entries for a reaction in a particular compartment appearing in the Excel gene-enzyme-reaction table are collapsed to a single reaction entry. The stoichiometric matrix is used for flux balance analysis.

**Inter-cellular and extracellular transport elements**

A multi-tissue model capturing the interactions between BS and M was formulated by combining two instances of the C4GEM and adding the movement of photosynthetic intermediates between mesophyll and sheath cells (Figure 1), which are restricted largely or entirely to the plasmodesmata (Evert et al., 1977). Transport through the bridge channel plasmodesmata was restricted to the main metabolites that give the characteristic of the C4 photosynthetic pathway, inferred from literature (Furbank et al., 1989, 1990), i.e., malate, pyruvate, phosphoglycerate, triose phosphates, and phosphates (pyrophosphate and orthophosphate) (Figure 1). The extra cellular transporters were included for water and gases exchange flux through the cell wall (Evert et al., 1977), biomass drains for M and BS and Sucrose drain from BS to vascular parenchyma to represent the translocation of carbon for the sustenance of the sink tissues (Braun and Slewinski, 2009). We have also considered gases efflux (CO₂ and O₂) to be exchanged through plasmodesmata to test the hypothesis, for example, that some percentage of CO₂ that is released from the C4 acids in BS leaks back to M, as reported in the literature (Farguhar and Hatch, 1983; Furbank et al., 1990; Hatch et al., 1995) (Please, see Results and Discussion session).
**Optimality criterion**

For the C4 metabolic model, the corresponding photon minimisation solution can be expressed as

\[
\min_{r_{M}, r_{BS}} \frac{r_{M}}{r_{BS}} \text{ photon uptake}
\]

Subject to

\[
\begin{bmatrix}
S_M & 0 \\
0 & S_{BS}
\end{bmatrix}
\begin{bmatrix}
0 \\
P_M
\end{bmatrix}
\begin{bmatrix}
r_M \\
r_{BS}
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0
\end{bmatrix}
\]

\[1 \leq \begin{bmatrix}
r_M \\
r_{BS}
\end{bmatrix} \leq u\]

The last constraint can be use to define the previously mentioned irreversibility constraints as well as tissue specificity (e.g., 0 flux for RuBisCO in M). It is also used to define the fixed output of the system, in the form of biomass drains. In order to simplify comparisons, biomass synthesis rates were assumed to be the same for the two tissue types (except for starch, see below) and estimated based on literature values (Poorter and Bergkotte, 1992, 1992). Inorganic carbon was assumed to be limited in BS (no CO₂ uptake) and free in M, where CO₂ is initially fixed by (PEPC) to form C₄ dicarboxylic acids. It was further assumed that the two tissues can freely exchange other inorganic compounds with the environment while the exchange rate for organic compounds is assumed zero except for exchange via plasmodesmata.

**Flux simulations**

Following manual curation, flux balance analysis was applied to confirm that each biomass component could be synthesized and in the subsequent model validation. C4GEM was compiled and curated as described for the compartmentalised C3 plant model (de Oliveira Dal'Molin et al., 2010). From the data collection, a 2D reaction centric SBML (System Biology Markup
Language, [www.sbml.org](http://www.sbml.org) database was generated. The stoichiometric matrix, $S$, and reversibility constraints (defining $v_{\text{min}}$), were extracted from the SBML database in MATLAB (Version 9, The MathWorks) and the linear programming problems were solved using the MOSEK Optimization Toolbox for MATLAB (Version 6). Constraints were applied in Matlab ($C4Flux$ tool box) to represent cell type interactions (Figure 2B). Reactions were activated or deactivated based on gene and enzyme activity studies (Westhoff et al., 1991; Bassi et al., 1995) for the specified cells (e.g., RuBisCO is not present is Mesophyll, but present in BS chloroplasts). Biomass drains (see the following section for more details) were also specified based on literature values (Poorter and Bergkotte, 1992, 1992). Optimum flux distribution were simulated by linear programming and visualized on a metabolic flux map (which represents the central metabolism of a compartmentalised plant cell) drawn in Excel (Figure 3).

**Biomass synthesis and network functionality**

In the final step of the manual curation process, network gaps were identified based on the models ability to produce biomass components from substrates. Biomass drain reactions are incorporated into C4GEM as the accumulation terms of the biomass precursors (e.g., “Starch = Starch\_biomass”), in order to simplify the task of uncovering the pathway gaps in the each of the biosynthetic routes separately. The list of biomass components considered are shown in Table 2 and includes major structural and storage components as well as trace elements such as vitamins. For each biomass component, the following linear programming problem was formulated and solved

$$\begin{align*}
\text{maximise} & \quad v_i \\
\text{subject to} & \quad Sv = 0 \\
& \quad v_{\text{min}} \leq v \leq v_{\text{max}}
\end{align*}$$
where $v_i$ is the corresponding biomass drain reaction. In this study, the problem was solved for leaf tissue (photons as energy source, CO$_2$ as carbon source and nitrate as nitrogen source) to represent biomass synthesis during C4 photosynthesis. The model can be also be used for non-photosynthetic tissues, when the minimization criterion is applied for carbon source usage (e.g., sucrose).

**Flux variability analysis**

Due to the lack of experimental constraints, flux solutions generated by linear programming for large-scale models are often not unique. Despite having identical network topology and maintaining the same set of constraints and objective function, it is possible to have different flux distributions that give rise to the same objective function value, with each flux distribution representing a different metabolic state (e.g., different reaction utilization and/or directionality) (Lee et al., 2000; Papin et al., 2002). Instead of a single flux distribution, it is therefore important to characterize the flux space encompassing all alternate optimal solutions before making reasonable interpretation of whether a metabolic phenotype is sufficiently different between two systems (e.g., BS versus M). An efficient linear programming-based strategy was used to quantify the range and variability of flux values, which involves minimizing and subsequently maximizing the flux value of all reactions in the network while keeping the same optimum objective function value (Mahadevan and Schilling, 2003). This method identifies the range of fluxes that is possible within all the possible alternate optimal solutions. The following linear programming problem was formulated and solved

\[
\begin{align*}
\text{maximise} & \quad v_i \\
\text{subject to} & \quad Sv = 0 \\
& \quad c^Tv = Z_{\text{obj}} \\
& \quad v_{\text{min}} \leq v \leq v_{\text{max}}
\end{align*}
\]
and

\[
\begin{align*}
\text{minimize} & \quad v_i \\
\text{subject to} & \quad S v = 0 \\
& \quad c^T v = Z_{\text{obj}} \\
& \quad v_{\text{min}} \leq v \leq v_{\text{max}}
\end{align*}
\]

where \( v_i \) represent each reaction in the network, \( c \) is the objective function vector and \( Z_{\text{obj}} \) is the objective function value calculated from the first optimization instance.

This method was used to identify and interpret the main contrasts between the two systems, M and BS, as shown in Table V (See Results and Discussion session).

**Supplemental Material**

Table S1: Reactions with no ORFs in all C4 models, covered by AraGEM.
Table S2: List of singletons.
Supplemental data S3: Maize C4GEM_vs1.0; SBML format.
Supplemental data S4: Sorghum C4GEM_vs1.0; SBML format.
Supplemental data S5: Sugarcane C4GEM_vs1.0; SBML format.
References


Black CC, Mollenha H (1971) Structure and Distribution of Chloroplasts and Other Organelles in Leaves with Various Rates of Photosynthesis. Plant Physiology 47: 15-


Bull TA (1969) Photosynthetic Efficiencies and Photorespiration in Calvin Cycle and C4-Dicarboxylic Acid Plants. Crop Science 9: 726-


Hatch MD, Kagawa T (1973) Enzymes and Functional Capacities of Mesophyll Chloroplasts from Plants with C4-Pathway Photosynthesis. Archives of Biochemistry and Biophysics 159: 842-853


Maize Genome Sequencing Project (2009). In: http://www.maizegenome.org


Oberhardt MA, Palsson BO, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Molecular Systems Biology 5: -


in the Mesophyll and Bundle-Sheath Cells of Monocotyledonous Nadp-Malic Enzyme-Type-C4 Plants. Planta 184: 377-388


Figure legends

Figure 1. Schematic representation of the photosynthetic metabolism of three C4 subtypes distinguished according to the decarboxylating enzyme. NADP-ME, NADP requiring malic enzyme; NAD-ME, NAD requiring malic enzyme; PCK, PEP carboxykinase. Numbers refer to enzymes: (1) PEP carboxylase, (2) NADP-malate dehydrogenase, (3) NADP-malic enzyme, (4) Pyruvate-Pi dikinase, (5) Rubisco, (6) PEP carboxykinase, (7) Alanine amino trasnferase, (8) Aspartate aminotrasferase, (9) NAD-malate dehydrogenase, (10) NAD-malic enzyme. Some steps were occulted for the sake of simplicity.

Figure 2. Elements of the reconstructed C4GEM network and the two tissue model.

Figure 3. Flux map of the central metabolism.

Relative flux distributions (BS/M) between bundle sheath (BS) and Mesophyll (M) during C4 photosynthesis. Green and red lines represent increased or decreased flux in BS, respectively. Gray lines represent flux values that do not differ significantly between cells.

Figure 4. Flux simulation for each C4 subtype with no leakage of CO2 and then considering 50% of CO2 leakage out of bundle sheath. A. Effect on linear electron flow (LEF) and cyclic electron flow (CEF). Flux ratios:(i) NADP-ME mode:NAD-ME mode;(ii) NAD-ME mode:NADP-mode; (iii) NAD-ME mode:NADP-mode (iv) CO2 leakage mode:NADP-ME mode. B. Effect on total absorbed quanta; Flux ratio: related to CO2 leakage mode. Photosysthem I (PSI); photosystem II (PSII); Mesophyll (M); Bundle sheath (BS). NADP-malic
enzyme subtype (NADP-ME); NAD-malic enzyme subtype (NAD-ME); PEP carboxykinase subtype (PCK).

**Figure 5.** C4 genome scale-model reconstruction process.
## Tables

**Table I.** Online resources for the reconstruction of the metabolic network of C4 plants.

<table>
<thead>
<tr>
<th>Database</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genome Database</strong></td>
<td></td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> Genome (SorghumDB)</td>
<td><a href="http://www.plantgdb.org/SbGDB/">http://www.plantgdb.org/SbGDB/</a></td>
</tr>
<tr>
<td><em>Zea mays</em> Genome (MaizeGDB)</td>
<td><a href="http://www.maizegdb.org/">http://www.maizegdb.org/</a></td>
</tr>
<tr>
<td>Resource for Comparative Grass Genomics (GrameneDB)</td>
<td><a href="http://www.gramene.org/">http://www.gramene.org/</a></td>
</tr>
<tr>
<td>Grass Regulatory Information Server (GrassiusDB)</td>
<td><a href="http://grassius.org/">http://grassius.org/</a></td>
</tr>
<tr>
<td><strong>Pathway Databases</strong></td>
<td></td>
</tr>
<tr>
<td>ExPASy Biochemical Pathways</td>
<td><a href="http://www.expasy.ch/cgi-bin/search-biochem-index">http://www.expasy.ch/cgi-bin/search-biochem-index</a></td>
</tr>
<tr>
<td><strong>Enzymes Databases</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Enzyme/Protein Localization Databases</strong></td>
<td></td>
</tr>
<tr>
<td>AraPerox (Arabidopsis Protein from Plant Peroxisomes)</td>
<td><a href="http://www.araperox.uni-goettingen.de/">http://www.araperox.uni-goettingen.de/</a></td>
</tr>
<tr>
<td>PPDB (Plant proteome database)</td>
<td><a href="http://ppdb.tc.cornell.edu/default.aspx">http://ppdb.tc.cornell.edu/default.aspx</a></td>
</tr>
</tbody>
</table>
**Table II.** Characteristics of C4GEM based on the genomic information* of Sorghum, Sugarcane and Maize.

<table>
<thead>
<tr>
<th>Elements</th>
<th>C4 plant model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Sorghum bicolor</strong></td>
</tr>
<tr>
<td>Gene-reaction-association entries</td>
<td>13114</td>
</tr>
<tr>
<td>Genes (unique)</td>
<td>3557</td>
</tr>
<tr>
<td>Metabolites</td>
<td>1755</td>
</tr>
<tr>
<td>Unique reactions</td>
<td>1588</td>
</tr>
<tr>
<td>Extra cellular transporters</td>
<td>18</td>
</tr>
<tr>
<td>Transporters (Intercellular-plasmodesmata)</td>
<td>11</td>
</tr>
<tr>
<td>Transporters (Inter-organelle)</td>
<td>83</td>
</tr>
<tr>
<td>Biomass drains</td>
<td>47</td>
</tr>
<tr>
<td>Metabolic Reactions from primary metabolism not assigned to any particular gene (covered by AraGEM)</td>
<td>131</td>
</tr>
</tbody>
</table>

* Extracted from Kyoto Encyclopedia of Genes and Genomes.
<table>
<thead>
<tr>
<th></th>
<th>Starch, sucrose, fructose, glucose, maltose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbohydrates and Sugars</td>
<td></td>
</tr>
<tr>
<td>Cell wall</td>
<td>Lignin (4-coumaryl alcohol, coniferyl alcohol, sinapyl alcohol), cellulose, hemi-cellulose (xylose)</td>
</tr>
<tr>
<td>Amino acids</td>
<td>Alanine, arginine, aspartate, asparagine, cystein, lysine, leucine, isoleucine, glutamate, glutamine, histidine, methionine, phenylalanine, proline, serine, tyrosine, tryptophan, valine</td>
</tr>
<tr>
<td>Nucleotides</td>
<td>ATP, GTP, CTP, UTP, dATP, dGTP, dCTP, dTTP</td>
</tr>
<tr>
<td>Fatty acids</td>
<td>C16:0 (Palmitic acid)</td>
</tr>
<tr>
<td>Vitamins and cofactors</td>
<td>Biotin, coenzyme A, riboflavin, folate, chlorophyll, nicotinamide, thiamine, ubiquinone,</td>
</tr>
</tbody>
</table>

*Coefficients were estimated based on literature data (Guinn, 1966; Poorter and Bergkotte, 1992). Although individual drains for vitamins and cofactors were tested, it was not added in the stoichiometric biomass equation as it accounts for minor compounds.

Coefficients for biomass drain: 0.001 Maltose + 0.096 Sucrose + 0.053 Glucose + 0.053 beta-D-Fructose + 0.265 Starch + 0.305 Cellulose + 0.282 D-Xylose + 0.04 4-Coumaryl alcohol + 0.048 Coniferyl alcohol + 0.056 Sinapyl alcohol + 0.026 Hexadecanoic acid + 0.107 L-Alanine + 0.39 L-Asparagine + 0.162 L-Aspartate + 0.024 L-Cysteine + 0.39 L-Glutamate + 0.35 L-Glutamine + 0.14 Glycine + 0.06 L-Isoleucine + 0.14 L-Leucine + 0.002 L-Methionine + 0.08 L-Phenylalanine + 0.01 L-Proline + 0.021 L-Threonine + 0.086 L-Valine + 0.04 L-Tryptophan + 0.03 L-Tyrosine + 0.1 L-Valine + 0.001 L-Lysine + 0.0027 dATP + 0.0017 dGTP + 0.0019 dCTP + 0.0028 dTTP + 0.0029 ATP + 0.0039 GTP + 0.0033 CTP + 0.0029 UTP + 30 ATP = Biomass + 30 ADP + 30 Orthophosphate.
Table IV. Set of constraints used for C4 subtypes based on literature.

<table>
<thead>
<tr>
<th>Constraints</th>
<th>NADP-ME subtype</th>
<th>NAD-ME subtype</th>
<th>PCK subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>BS</td>
<td>M</td>
</tr>
<tr>
<td>RuBisCO (plastid)</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>PEPCase (cytosol)</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>CO₂ uptake</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>NADP-ME (plastid)</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>NAD-ME (mitochondrion)</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>PCK (cytosol)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Malate transport (plasmodesmata)</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pyruvate transport (plasmodesmata)</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Aspartate transport (plasmodesmata)</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Alanine transport (plasmodesmata)</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Biomass *</td>
<td></td>
<td>Fixed growth</td>
<td>Fixed growth</td>
</tr>
</tbody>
</table>

M: Mesophyll; BS: bundle sheath; RuBisCO: Ribulose-1,5-bisphosphate carboxylase oxygenase; PEPcase: Phosphoenolpyruvate carboxylase; PCK: PEP carboxykinase; NADP-ME: NADP malic enzyme; NAD-ME: NAD malic enzyme. (-) flux limited: constraint to zero. (+) free flux: estimated through optimization. *Starch synthesis (biomass drain) was limited in M, and the remaining biomass components were assumed to be evenly distributed in both cells.
Table V. Regulation of key enzymes/transporters during C4 photosynthesis in mesophyll (M) and bundle sheath (BS) predicted by C4GEM and correlated/no correlated to Proteomics studies.

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>EC</th>
<th>Metabolism</th>
<th>P</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferredoxin-NADP(+) reductase 1 (FNR-1)</td>
<td>1.18.1.2</td>
<td>Photosynthesis/LET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAD(P)H dehydrogenase (NDH)</td>
<td>1.6.5.2</td>
<td>Photosynthesis/CET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RuBisCO</td>
<td>4.1.1.39</td>
<td>Carbon fixation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triosephosphate isomerase</td>
<td>5.3.1.1</td>
<td>Glycolysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphoglycolate phosphatase</td>
<td>3.1.3.18</td>
<td>Photorespiration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fructose-bisphosphate aldolase</td>
<td>4.1.2.13</td>
<td>Glycolysis/PPP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyruvate Dehydrogenase (PDH E1)</td>
<td>1.2.4.1</td>
<td>Glycolysis/TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutamine synthetase</td>
<td>6.3.1.2</td>
<td>Nitrogen metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspartate aminotransferase</td>
<td>2.6.1.1</td>
<td>Carbon fixation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphate/Triose-phosphate (IEP30)</td>
<td>translocator</td>
<td>Chloroplast membrane</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphate/Phosphoenolpyruvate (IEP33)</td>
<td>translocator</td>
<td>Chloroplast membrane</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP/ADP translocator</td>
<td>translocator</td>
<td>Chloroplast membrane</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-oxoglutarate/malate (DiT1)</td>
<td>translocator</td>
<td>Chloroplast membrane</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-oxoglutarate/malate (DiT2)</td>
<td>translocator</td>
<td>Chloroplast membrane</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAPDH (GAP-A)</td>
<td>1.2.1.12; 1.2.1.13</td>
<td>Glycolysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphoglycerate kinase</td>
<td>2.7.2.3</td>
<td>Glycolysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphoribulokinase</td>
<td>2.7.1.19</td>
<td>Carbon fixation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribulose-5-phosphate-3-epimerase</td>
<td>5.1.3.1</td>
<td>PPP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transketolase</td>
<td>2.2.1.1</td>
<td>PPP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribose-5-phosphate isomerase</td>
<td>5.3.1.6</td>
<td>PPP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADP-glucose pyrophosphorylase</td>
<td>2.7.7.27</td>
<td>Starch synthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Starch synthase</td>
<td>2.4.1.21</td>
<td>Starch synthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carbonic anhydrase</td>
<td>4.2.1.1</td>
<td>Carbon fixation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatty acid biosynthesis (20 enzyme average)</td>
<td></td>
<td>Fatty Acids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malic enzyme</td>
<td>1.1.1.40</td>
<td>Carbon fixation/TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malate dehydrogenase (MDH)</td>
<td>1.1.1.37</td>
<td>Carbon fixation/TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyruvate orthophosphate dikinase</td>
<td>2.7.9.1</td>
<td>Pyruvate metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferredoxin-dependent Glu synthase</td>
<td>1.4.7.1</td>
<td>Nitrogen metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferredoxin-nitrite reductase</td>
<td>1.7.7.1</td>
<td>Nitrogen metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenylyl kinase</td>
<td>2.7.4.3</td>
<td>Purine metabolism</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P: Proteomics studies (BS/M): Maize chloroplast BS:M protein accumulation (Majeran et al., 2005; Majeran et al., 2008; Friso et al., 2010). F: Flux analyses (BS/M): Relative flux distribution (in this study). Green: Increased protein accumulation (P) or increased flux (F); Red: Decreased protein accumulation (P) or decreased flux (F); Grey: no

**Table V.** Continue. Regulation of key enzymes/transporters during C4 photosynthesis in mesophyll (M) and bundle sheath (BS) predicted by C4GEM and correlated/no correlated to Proteomics studies.

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>EC</th>
<th>Metabolism</th>
<th>P</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geranylgeranyl reductase</td>
<td>1.3.1.83</td>
<td>Chlorophyll and Isoprenoids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zeta-carotene desaturase</td>
<td>1.14.99.30</td>
<td>Chlorophyll and Isoprenoids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phytoene dehydrogenase</td>
<td>1.14.99.-</td>
<td>Chlorophyll and Isoprenoids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protochlorophyllide oxidoreductase</td>
<td>1.3.1.33</td>
<td>Chlorophyll and Isoprenoids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorophyll synthase</td>
<td>2.5.1.62</td>
<td>Chlorophyll and Isoprenoids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protoporphyrinogen oxidase</td>
<td>1.3.3.4</td>
<td>Chlorophyll and Isoprenoids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flavin reductase</td>
<td>1.5.1.30</td>
<td>Oxidative stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutathione peroxidase</td>
<td>1.11.1.9</td>
<td>Oxidative stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ascorbate peroxidase</td>
<td>1.11.1.11</td>
<td>Oxidative stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldo/keto reductase</td>
<td>1.1.1.21; 1.1.1.2</td>
<td>Oxidative stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 1 DAD1-like acylhydrolase</td>
<td>4.2.1.60</td>
<td>Oxidative stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lysophospholipase</td>
<td>3.1.1.5</td>
<td>Oxidative stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipoxygenase 2</td>
<td>1.13.11.12</td>
<td>Oxidative stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldose-1-epimerase</td>
<td>5.1.3.3</td>
<td>Degradation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>beta-D-glucosidase</td>
<td>3.2.1.21</td>
<td>Degradation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pheophorbide a oxygenase</td>
<td>1.14.12.20</td>
<td>Degradation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P Proteomics studies (BS/M): Maize chloroplast BS:M protein accumulation (Majeran et al., 2005; Majeran et al., 2008; Friso et al., 2010). † Flux analyses (BS/M): Relative flux distribution (in this study). Green: Increased protein accumulation (P) or increased flux (F); Red: Decreased protein accumulation (P) or decreased flux (F); Grey: no change, White: not predicted. PPP: Pentose phosphate pathway. GAPDH: Glyceraldehyde 3-phosphate dehydrogenase. CET: cyclic electron flow. LET: linear electron flow. PEP: phosphoenolpyruvate.
Figure 1: Schematic representation of the photosynthetic metabolism of three C4 subtypes distinguished according to the carboxylating enzyme. NADP-ME, NADP requiring malic enzyme; NAD-ME, NAD requiring malic enzyme; PCK, PEP carboxykinase. Numbers refer to enzymes: (1) PEP carboxykinase, (2) NADP-malate dehydrogenase, (3) NADP-malic enzyme, (4) Pyruvate-Pi dikinase, (5) Rubisco activity, (6) Malate:NADP+ transhydrogenase, (7) Aspartate aminotransferase, (8) Aspartate aminotransferase, (9) NAD-malate dehydrogenase, (10) NAD-malic enzyme.
A. Main elements of the reconstructed genome-scale models.

B. Matlab

1) Specify cell types for interaction: Eg.: Mesophyll (M), Bundle sheath (BS) and plasmodesmata exchange reactions (C4 photosynthesis).

2) Duplicate the generic single cell model to represent each cell:

\[ BS = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \]

\[ M = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \]

3) Deactivate reactions based on gene expression/proteome studies for the specified cells:

Eg.: Constrain ribulose bisphosphate carboxylase (Rubisco) flux to zero in Mesophyll. Rubisco is not present in Mesophyll but present in Bundle (chloroplast).

4) Specify rates for external transporters and biomass drains;

5) Simulate optimum flux distributions by linear programming (FBA);

(Eg. optimize for minimum photons uptake);

6) Visualization/Fluxmap

Relative flux distribution (BS/M).
Figure 3. Flux map of the central metabolism. Relative flux distributions (BS/M) between bundle sheath (BS) and Mesophyll (M) during C4 photosynthesis. Green and red lines represent increased or decreased flux in BS, respectively. Gray lines represent flux values that do not differ significantly between cells.
Figure 4. Flux simulation for each C4 subtype with no leakage of CO2 and then considering 50% of CO2 leakage out of bundle sheath. Effect on linear electron flow (LET), carbonic acid electron flow (CEF), Mesophyll (M); Bundle sheath (BS). NADP-malic enzyme subtype (NADP-ME); NAD-malic enzyme subtype (NAD-ME); PEP carboxykinase subtype (PCK).
Figure 5. C4 genome-scale model reconstruction process.