Edinburgh Research Explorer

Progress and challenges of engineering a biophysical carbon dioxide concentrating mechanism into higher plants

Citation for published version:

Digital Object Identifier (DOI):
10.1093/jxb/erx133

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Experimental Botany

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Progress and challenges of engineering a biophysical carbon dioxide concentrating mechanism into higher plants

Authors:  
Benjamin D. Rae1,2  
Benedict M. Long2  
Britta Förster2  
Nghiem D. Nguyen1,2  
Christos N. Velanis3  
Nicky Atkinson3  
Wei Yih Hee2  
Bratati Mukherjee1,2  
G. Dean Price1,2  
Alistair J. McCormick3,†

ORCID:  
0000-0003-2199-1205  
0000-0002-4616-2967  
0000-0002-5231-0954  
0000-0002-4322-4755  
0000-0003-4212-1209  
0000-0002-6279-8164  
0000-0001-5906-4912  
0000-0002-7255-872X

ResearcherID:  
C-3045-2011  
B-3799-2008  
O-7908-2015  
B-1656-2017  
O-7908-2015  
C-9505-2009  
B-1558-2008

Address:  
1 Australian Research Council Centre of Excellence for Translational Photosynthesis  
2 Research School of Biology  
2 The Australian National University  
2 134 Linnaeus Way  
2 Acton ACT 2601  
2 AUSTRALIA

3 SynthSys & Institute of Molecular Plant Sciences  
3 School of Biological Sciences  
3 University of Edinburgh  
3 Edinburgh  
3 EH9 3BF  
3 UK

† Corresponding Author

Running title: Engineering biophysical CCMs into higher plants

Abstract: 133

Length: 11088 (main text)

Figures: 4

Tables: 1

Key words: algae, carboxysome, cyanobacteria, photosynthesis, RuBisCO, pyrenoid, transporter
Abstract

Growth and productivity in important crop plants is limited by the inefficiencies of the C3 photosynthetic pathway. Introducing CO2-concentrating mechanisms (CCMs) into C3 plants could overcome these limitations and lead to increased yields. Many unicellular micro-autotrophs, such as cyanobacteria and green algae, possess highly efficient biophysical CCMs that increase CO2 concentrations around the primary carboxylase enzyme, RuBisCO, to enhance CO2 assimilation rates. Algal and cyanobacterial CCMs utilise distinct molecular components, but share several functional commonalities. Here we outline the recent progress and current challenges of engineering biophysical CCMs into C3 plants. We review the predicted requirements for a functional biophysical CCM based on current knowledge of cyanobacterial and algal CCMs, the molecular engineering tools and research pipelines required to translate our theoretical knowledge into practice, and the current challenges to achieving these goals.

Introduction

Generating new C3 crop varieties with increased yields is critical for safeguarding future food security (Long et al., 2015). To achieve this goal, much research has focused on improving the efficiency of photosynthetic CO2 uptake by the primary carboxylase, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). RuBisCO might be the most important limiting component that restricts photosynthetic productivity (Parry et al., 2013). Compared to other enzymes, RuBisCO has a slow turnover rate ($k_{cat}$) and a relatively low affinity for CO2. Nonetheless, RuBisCO in higher plants has a $K_m$ for CO2 at ambient O2 ($K_{c}^{air}$) close to the CO2 concentration in a C3 leaf mesophyll cell (Galmés et al., 2014), indicating functional adaptation to its operating environment. RuBisCO also catalyses a competitive reaction with O2 (approximately 30% of activity is apportioned to oxygenation in C3 plants), leading to substantial metabolic costs associated with photorespiration (Sharkey, 2001). Thus, increasing the carboxylation efficiency of RuBisCO and reducing photorespiration are important strategies for improving yields in C3 crops (Betti et al., 2016; Sharwood, 2017).

Photosynthetic organisms have evolved to compensate for the slow and promiscuous nature of RuBisCO. Different strategies have emerged through evolutionary history as organisms adapted both to changes in atmospheric gas composition and growth environment (Shih, 2015). For example, C3, C4 and CAM modes of photosynthesis have evolved to occupy specialist niches in terrestrial environments. C3 plants predominate in temperate ecosystems (Woodward et al., 2004) and have compensated for the limitations of RuBisCO by maintaining high stromal RuBisCO concentrations. Thus, C3 plants maximise CO2 capture by increasing the likelihood of carboxylation events within the leaf at the cost of higher rates of photorespiration and water use. C4 plants dominate tropical/subtropical grasslands (Edwards et al., 2010; Woodward et al., 2004) and have adapted a biochemical CO2 concentrating mechanism (CCM), typically shared between two adjacent leaf cell...
types (mesophyll and bundle sheath). C₄ plants utilise cytosolic phosphoenolpyruvate carboxylase in
the mesophyll to perform biochemical fixation of inorganic carbon (Cᵅ) as the bicarbonate ion (HCO₃⁻)
into four-carbon organic acids, which are transported to bundle sheath chloroplasts where RuBisCO
is primarily localised for subsequent decarboxylation and CO₂ fixation. The biochemical CCMs found
in CAM species utilise temporal separation to capture CO₂ as malic acid within the mesophyll vacuole
during the night followed by decarboxylation and CO₂ fixation by RuBisCO during the day. As an
adaptation to drought stress, CAM species can facilitate CO₂ fixation while stomata are closed,
ensuring minimal water loss (Keeley and Rundel, 2003).

In aquatic environments, additional limitations on photosynthesis are imposed by the relatively low
solubility and diffusion of CO₂ through water. Furthermore, the effects of temperature and pH/ionic
strength on CO₂ solubility and the equilibrium ratios of carbonate species (H₂CO₃, HCO₃⁻, CO₃²⁻) are
such that supply of inorganic carbon (Cᵅ) is both limited and sometimes highly variable (Mangan et
al., 2016). Capture of Cᵅ in the form of the less membrane-permeable, yet usually more predominant
bicarbonate ion, is a major driving force in CCMs found in aquatic photosynthetic organisms. Algae
and cyanobacteria have evolved highly efficient biophysical CCMs that can enhance CO₂ levels
around RuBisCO by up to 1000-fold, making these the most effective biological Cᵅ uptake systems yet
identified in nature (Wang et al., 2015). Their success is evident in that no less than 80% of
photosynthetic Cᵅ uptake in aquatic environments is driven by biophysical CCMs, which accounts for
roughly half of global primary productivity (Raven and Beardall, 2015).

**Increased ambient CO₂ improves crop yield**

Several decades of experimental data clearly show that C₃ crops grown under elevated CO₂
concentrations consistently have higher yields, demonstrating that productivity and growth
 efficiencies are improved by increasing CO₂ levels at the RuBisCO active site (Leakey et al., 2009).
These findings are complemented by the numerous examples in nature where CCMs have evolved to
enhance the operating efficiency of RuBisCO. Thus, introduction of CCM components into C₃ crops
using new synthetic biology based approaches is considered a promising strategy to increase
 photosynthetic productivity and environmental resilience (Furbank, 2016; Long et al., 2016; Meyer et
al., 2016; Schuler et al., 2016; Wang et al., 2017). The biophysical CCMs found in algae and
cyanobacteria are particularly attractive due to their high levels of efficiency and establishment in a
single cell.

**Common characteristics of biophysical CCMs**

Biophysical CCMs have four common characteristics: i) energy-dependent uptake of Cᵅ, ii)
dehydration of bicarbonate to CO₂ via a carbonic anhydrase (CA) located nearby RuBisCO, iii)
assembly of RuBisCO into subcellular microcompartments (cyanobacterial carboxysomes) or regions (algal pyrenoids), and iv) systems to minimise CO₂ diffusion from the site of carboxylation (Figure 1). Carboxysomes are essential CCM components in photosynthetic cyanobacteria, whereas pyrenoids are found in most but not all eukaryotic microalgae (Badger et al., 1998; Giordano et al., 2005). Furthermore, carboxysomes are permanent subcellular features, whereas pyrenoid assembly/disassembly is circadian and C₆ dependent (Mitchell et al., 2014). Accordingly, key components of the cyanobacterial CCM are well characterised, and detailed models exist for predicting the best routes for introduction into C₃ chloroplasts (McGrath and Long, 2014; Price et al., 2011a; Price et al., 2013). However, the factors that regulate activity and assembly are still unclear and the focus of current investigation. In contrast, eukaryotic algal CCMs are relatively less well understood overall and likely more complex due to the nested-compartmentalised structure of the eukaryotic cell (Wang et al., 2015).

In this review, we highlight recent progress and current challenges of engineering biophysical CCMs into C₃ plants. Specifically, we outline progress towards the incorporation of inorganic carbon transporters into the chloroplastic inner membrane, utilising both algal and cyanobacterial genes as sources, and emphasising the need for further knowledge of their activation and energisation. We also draw attention to current knowledge of the RuBisCO microbodies (pyrenoids) and microcompartments (carboxysomes) and how this informs us to their potential reconstruction within C₃ plant chloroplast hosts. The review also touches on the costs of microcompartment construction within the chloroplast and synthetic approaches to generating a chloroplastic CCM.

Model-guided strategies for improvement of C₃ photosynthesis

Mathematical modelling reveals that modest photosynthetic benefits might be achievable by incorporating bicarbonate transporters and their ancillary systems into the chloroplast envelope of C₃ plants (McGrath and Long, 2014; Price et al., 2011a; Price et al., 2013). Bicarbonate transporters could be harnessed independently or in concert from different species (e.g. algae and cyanobacteria) to boost C₅ levels in the chloroplast. However, native chloroplastic CA activities, which are highly abundant in C₃ plants (Meyer et al., 2016), would need to be removed or significantly re-regulated to avoid dissipation of any C₅ gradient generated by active transport (McGrath and Long, 2014; Price et al., 2013). These features are a necessary precondition for downstream introduction of RuBisCO aggregating factors, such as carboxysome and pyrenoid-like structures. Additional aggregation of RuBisCO into carboxysomes is predicted to enhance the photosynthetic potential of higher plants by ca. 60% (McGrath and Long, 2014). Earlier models have also indicated that the presence of a pyrenoid significantly improves the maintenance of elevated CO₂ concentrations around RuBisCO (Badger et al., 1998), suggesting that pyrenoid-like structures would likewise be beneficial in C₃ plants in the presence of a functional chloroplastic C₅ uptake system.
McGrath and Long (2014) predicted that the further addition of complex thylakoid membrane localised systems to reduce CO₂ leakage (multi-subunit NDH-1 complexes, NDH-1₃ and NDH-1₄) would only marginally improve photosynthesis. Thus far, the technical challenges associated with localising multi-gene complexes on the thylakoid membrane have precluded introduction of NDH-1₃ and NDH-1₄ as a major engineering focus. The increasing availability of component libraries for gene expression and development of rapid assembly approaches for multigene constructs may simplify this goal (Patron, 2016). An alternative CO₂ recapture system is found in algae, which is conceivably simpler, and consists of a stromal soluble complex of two proteins (LCIB and LCIC in Chlamydomonas) (Jin et al., 2016). Expression of LCIB/LCIC may represent the simplest means of achieving the CO₂ recapture step of the CCM process outlined by McGrath and Long (2014).

**Introducing bicarbonate transporters into the chloroplast envelope**

* Cyanobacterial bicarbonate transporters

To date, five Cᵢ uptake systems have been identified in cyanobacteria that differ in expression patterns (constitutive vs inducible) as well as substrate affinity and uptake rates (Figure 1) (Price et al., 2011a; Price et al., 2008). Many cyanobacterial genomes possess only a subset (Rae et al., 2011), suggesting the full complement of Cᵢ uptake systems is not required to accumulate sufficient intracellular Cᵢ in their natural environments. Common to all systems is that Cᵢ uptake is an active process relying on various forms of energisation.

While models suggest the greatest improvement in C₃ photosynthetic efficiency would require the introduction of a complete CCM into higher plant chloroplasts, they also indicate that the introduction of one or two active bicarbonate transporters alone can increase photosynthetic efficiency significantly (McGrath and Long, 2014; Price et al., 2011a). Light-saturated CO₂ assimilation rates could be increased by 9% by using a single cyanobacterial bicarbonate transporter [Bicarbonate Transporter A (BicA)], while utilising all cyanobacterial bicarbonate transporters together could improve photosynthetic rates by as much as 16% (McGrath and Long, 2014). Notably, McGrath and Long (2014) found that introducing bicarbonate transporters into the plasma membrane would result in no additional benefits in photosynthetic efficiency. Observations in field-based CO₂-enrichment studies for C₃ crops show, on average, a 17% increase in biomass accumulation based on improvements in photosynthetic efficiency (Long et al., 2006). Modelled photosynthetic improvements using cyanobacterial bicarbonate transporters alone (i.e. a partial biophysical CCM) indicate that active bicarbonate transporters could translate into a 5-10% increase in biomass on a daily basis. However, the compounding interest effect could easily lead to a larger increase in biomass accumulated at harvest time. This effect has already been observed with recent strategies for crop plant photosynthetic improvement (Kromdijk et al., 2016; Simkin et al., 2015), where relatively small
changes in photosynthetic efficiency led to significant increases in endpoint biomass production. Furthermore, downstream addition of carboxysomes or pyrenoids, and pluralist approaches drawing from all photosynthetic improvement strategies will provide additive improvement.

Three different transporters located in the cyanobacterial plasma membrane actively take up bicarbonate. The Bicarbonate Transporter 1 (BCT1) complex, encoded by the *cmpABCD* operon, is an ATP-binding cassette (ABC) type transporter [traffic ATPase family (Omata *et al.*, 1999)]. Even though fewer subunits constitute a functional BCT1, the same multi-gene engineering challenges exist as for NDH-1 complexes. Therefore, the two single-gene bicarbonate transporters, BicA and Sodium-dependent Bicarbonate Transporter A (SbtA), encoded by the *bicA* and *sbtA* genes, respectively, are considered the most suitable candidates for expression in higher plant chloroplast envelopes (Price *et al.*, 2004; Shibata *et al.*, 2002). BicA belongs to the large SLC26A/SulP type protein family of anion carriers and channels, present in many prokaryotic and eukaryotic organisms. SbtA is widely distributed amongst cyanobacteria and found in some bacteria, but is absent from eukaryotes. Structurally, SbtA has no resemblance to other known transport proteins (Price and Howitt, 2014). Activity of both transporters depends on a cell-inward directed sodium ion (Na\(^+\)) gradient, and most likely both function as HCO\(_3^-\)/Na\(^+\) symporters. However, direct evidence is lacking and stoichiometry of transported ions needs to be determined. Based on partial homology to low resolution structures of bacterial SulP proteins (Compton *et al.*, 2011), it is likely that BicA forms homodimers, while there is some proteomic evidence for tetramers of SbtA (Zhang *et al.*, 2004). Nevertheless, the active configuration and assembly mechanisms of BicA and SbtA have not yet been elucidated. Important structural information about the expression of the mature protein in the plasma membrane has been gained from topology mapping and superimposition of known structures of similar proteins. The N- and C-termini of BicA have been shown to be cytoplasmic exposed (Shelden *et al.*, 2010), whereas both N- and C-termini of SbtA are unusual in that they protrude into the periplasmic space (Price *et al.*, 2011b).

**Expression of cyanobacterial membrane transporters in planta**

Both BicA and SbtA have been successfully introduced into C\(_3\) plants using chloroplast and nuclear transformation approaches (Pengelly *et al.*, 2014; Rolland *et al.*, 2016; Uehara *et al.*, 2016). Introduction of transgenes directly into the chloroplast genome may negate the need for a chloroplast targeting signal for insertion into the chloroplast inner envelope membrane (IEM) (Figure 2). On the other hand, the vast majority of IEM targeted proteins are encoded in the nuclear rather than the chloroplast genome, with only a few known exceptions - Ycf1/Tic214 identified in *Arabidopsis thaliana* (Arabidopsis hereafter) (Kikuchi *et al.*, 2013), and CemA in pea (Sasaki *et al.*, 1993) and its orthologue Ycf10 in Chlamydomonas (Rolland *et al.*, 1997).
Insertion of chloroplast- and nuclear-encoded proteins into the thylakoid membrane is thought to occur via four pathways. The first three are analogous to bacterial systems, and require additional factors [i.e. the signal recognition particle (SRP), secretory (Sec), and twin-arginine translocase (Tat)/ΔpH-dependent pathways (Aldridge et al., 2009)]. The fourth pathway is less common and relies on spontaneous insertion. For example, several membrane-spanning proteins are able to insert unassisted into the thylakoid membrane [e.g. photosystem subunits PsbY, PsaG, PsaK (Zygadlo et al., 2006)]. It is important to note though that those proteins contain at most four membrane spans, and therefore require significantly less threading of protein domains across the lipid bilayer than insertion of BicA and SbtA, which contain 14 and 10 transmembrane/membrane integral domains, respectively. Detailed analysis of insertion requirements of PsaG suggests the positively charged stromal loop region between the membrane spans is essential. This is consistent with the “positive inside rule” proposed for spontaneous insertion of membrane integral proteins in bacteria (Dalbey et al., 2011).

It is not known how membrane integral proteins such as BicA and SbtA are sorted into the plasma membrane in cyanobacteria, or in the special cases of Ycf1/Tic214 and CemA from the stroma to the IEM in chloroplasts. Based on topology mapping, BicA and SbtA may follow the positive inside rule, as positive charges are present on their cytoplasm-exposed domains (Price et al., 2011b; Shelden et al., 2010). Hence, the chloroplast transformation approach relies on either spontaneous insertion of BicA or SbtA into the IEM, or sufficient compatibility of these proteins with other chloroplastic insertion pathways. Given the prokaryotic origin of chloroplasts it is conceivable that the membrane protein insertion pathways have not completely diverged from the ancestral cyanobacterial system. Mature BicA protein of the expected size was detected throughout the life of the plant both in the thylakoid and the IEM fraction in Nicotiana tabacum (tobacco hereafter) chloroplast transformants (Pengelly et al., 2014), whereas this approach has been unsuccessful for SbtA, as protein did not accumulate (Förster unpublished). In both cases, the original cyanobacterial coding sequences were transferred to the same region of the inverted repeat in the chloroplast genome and expressed from the same promoter. Evidently, there are unknown factors that differentially affect synthesis, stability and targeting of BicA and SbtA. Physiological characterisation of mature BicA-expressing plants suggested the BicA protein was inactive (Pengelly et al., 2014). It remains unclear whether BicA was inserted into the IEM incorrectly or whether other BicA-specific regulatory factors (e.g. post-translational modifications and/or interacting proteins or metabolites) are missing in chloroplasts. The latter seems more likely, as BicA inserts correctly into the plasma membrane when expressed in Escherichia coli, evidenced by successful topology mapping (Shelden et al., 2010). Active uptake of radioactively-labelled bicarbonate mediated by BicA could not be detected using E. coli as a heterologous expression system (Du et al., 2014). Likewise, BicA showed no bicarbonate uptake activity when expressed in Xenopus laevis oocytes (Förster, Parker unpublished). Taken together, this...
emphasizes the absolute necessity to understand post-translational regulation/activation mechanisms of BicA to make it a useful candidate for C\textsubscript{i} uptake.

Nuclear expression and chloroplast targeting of cyanobacterial bicarbonate transporters

The alternative to transgene expression in the chloroplast is nuclear transformation. This approach requires synthesis of precursors of BicA and SbtA that have peptide additions that facilitate correct targeting and insertion into the chloroplast IEM, via a stop-transfer or post-import mechanism (Viana et al., 2010). Even though our understanding of these components and processes is growing, we are far from being able to predict the right signal peptide for each protein. Using a trial-and-error approach, a cleavable transit peptide (cTP) followed by a potentially uncleaved membrane protein leader sequence (MPL) of known chloroplast-IEM targeted proteins, was fused to BicA or SbtA, resulting in successful insertions in the IEM of \textit{Nicotiana benthamiana} and Arabidopsis (Rolland et al., 2016; Uehara et al., 2016). Rolland et al. (2016) identified two Arabidopsis cTP+MPL sequences, \textit{AtPLGG1} and \textit{AtHP59}, which effectively direct the GFP-tagged bicarbonate transporters to the IEM in \textit{N. benthamiana}. The \textit{AtPLGG1} leader can correctly target both, SbtA and BicA, while the \textit{AtHP59} leader is only effective in directing SbtA. This demonstrates that the effectivity of a leader sequence is highly dependent on the protein to be delivered to the IEM, and must be determined for each foreign membrane protein separately.

There is some evidence that information for protein insertion into the IEM may be contained elsewhere in the mature protein sequence (Endow et al., 2016; Oh and Hwang, 2015; Okawa et al., 2014; Theg and Scott, 1993). Uehara et al. (2016) were able to direct \textit{Staphylococcus} protein A-tagged BicA and SbtA chimera to the IEM in Arabidopsis using the cTP and the mature IEM located Arabidopsis Cor413im1 protein. One possibility is an MPL may be contained in the Cor413im1 sequence leading up the first transmembrane domain, suggesting a similar targeting structure (cTP+MPL) was generated as in Rolland et al. (2016). Interestingly, BicA was also targeted to the IEM when Cor413im1 was fused to the C-terminus of BicA. All IEM-inserted chimeras tested by Rolland et al. (2016) had an N-terminus cTP+MPL fusion, and domain swapping experiments showed that MPL sequences had to be in the correct orientation and matched with a compatible cTP.

If we assume that the IEM targeting signal is not, or only partially, removed from the transporter upon IEM insertion, the positioning of the IEM signal may have important implications with respect to the functionality of the transporters, particularly for SbtA. In \textit{E. coli} (Du et al., 2014) and in cyanobacteria (Förster, Price unpublished), we have shown previously that SbtA activity is abolished by adding epitope tags, green or yellow fluorescent proteins (GFP, YFP). Furthermore, bicarbonate transporters can only function effectively on the chloroplast envelope if they are inserted in the correct orientation. From the work of Uehara et al. (2016) using detergent and protease treatments, and
Rolland et al. (2016) using visualisation of BicA::GFP and SbtA::GFP fusions by confocal microscopy, we know both transporters were integrated into the IEM but not their orientation or exact N-terminus amino acid sequence.

Characterisation of cyanobacterial bicarbonate transporter activity in transgenic plants

Achieving expression of BicA and SbtA in the IEM was a significant milestone. Future research now faces two key challenges. Firstly, bicarbonate transporter activity on the IEM cannot be readily measured directly. Researchers currently rely on measurements of leaf photosynthetic traits (e.g. the CO$_2$ compensation point and carbon isotope discrimination ($\Delta^{13}$C)) and plant biomass accumulation as indicators of a change in C$_i$ uptake capacity (Pengelly et al., 2014). These measurements are largely indirect and assume that bicarbonate uptake will lead immediately to an improvement in photosynthetic efficiency, as suggested by modelling (McGrath and Long, 2014; Price et al., 2011a). However, there are several confounding factors that may mask or counteract phenotypic effects of bicarbonate uptake, including i) the expression levels required for an active bicarbonate transporter to elevate C$_i$ in a plant chloroplast are not yet clear, ii) photosynthetic improvement may not translate into an increase in growth due to other limiting factors (e.g. light, temperature, source-sink dynamics) under the given growth conditions and iii) transporter activity may have detrimental side effects on native metabolism. The latter is a potential issue for BicA and SbtA, as both transporters are HCO$_3^{-}$/Na$^+$ symporters. An increased influx of Na$^+$ into the chloroplast needs to be matched with a Na$^+$/H$^+$ antiporter activity on the IEM to maintain ion homeostasis and prevent a build-up of stromal Na$^+$ from the stroma. Higher plants have endogenous HCO$_3^{-}$/Na$^+$ symporters in the IEM, such as the well-characterized NHD1 transporter in Arabidopsis (Müller et al., 2014), but whether native transporters provide sufficient Na$^+$ extrusion or additional capacity is required in the presence of BicA and SbtA remains unclear (Price, 2011). Direct measurements of bicarbonate/CO$_2$ uptake can be performed on isolated chloroplasts, either by monitoring changes in photosynthetic gas exchange using membrane-inlet mass spectrometry (MIMS), or by measuring uptake of radioactively-labelled [14C]-bicarbonate into chloroplast using the silicon oil centrifugation method. Neither technique is simple and requires isolation of intact and functional chloroplasts, which is difficult for many plant species. One solution may be to develop improved MIMS techniques to detect C$_i$ uptake activity at the leaf level.

Transport regulation of cyanobacterial bicarbonate transporters ex situ

A further key challenge is to overcome the apparent inactivity of the bicarbonate transporters ex situ and, potentially, to regulate activity when expressed in plants. The observed lack of activity in plants, particularly of BicA, is not yet understood. Requirements for a sufficient Na$^+$ gradient to drive activity are met - cytosolic levels of Na$^+$ are estimated at 1-3 mM, which is in the range of Na$^+$ concentration required for half-maximal activity of bicarbonate uptake by BicA and SbtA (Price et al., 2013).
Elevated Na⁺ in the chloroplast stroma is unlikely to be problematic due to the presence of Na⁺/H⁺ antiporters, and Na⁺-coupled transporters, in the chloroplast membrane (Price et al., 2013). Furthermore, no serious effect of increased cytoplasmic bicarbonate is expected (Price et al., 2011a; Price et al., 2013). One possible cause for a lack of function is incorrect insertion in the IEM. A bimolecular fluorescence complementation (BiFC) system could be suitable for determining the orientation of BicA and SbtA in planta (Waadt et al., 2008). Previously, BiFC-based approaches have been successfully used to determine the orientation of TOC75 in the outer chloroplast envelope membrane (OEM) in Arabidopsis (Chen et al., 2016).

In cyanobacteria, BicA and SbtA activity are regulated by environmental cues. One important feature is the rapid inactivation of bicarbonate transport in the dark (Förster, unpublished), which is likely to avoid futile Cᵢ uptake and energy wastage when photosynthetic CO₂ fixation does not occur. The underlying molecular mechanisms that activate/deactivate BicA and SbtA are unknown. Heterologous expression of SbtA in E. coli has shown that SbtA is constitutively active in the absence of photosynthetic CO₂ fixation (Du et al., 2014) However, when co-expressed with a companion protein of unknown function (SbtB), SbtA is inactive. The genome of every sequenced cyanobacterial species containing sbtA inextricably has a sbtB gene in the same operon or in close proximity. The SbtB protein shows 21% identity to the cyanobacterial PII/GlnB proteins that are regulators of nitrogen and carbon metabolism in bacteria. Structurally, SbtB trimers (unpublished crystal structure: ncbi.nlm.nih.gov structure 3DFE) fold in a similar way to trimeric GlnB and GlnK (a PII homolog). In E. coli, the ammonium channel AmtB is inactivated upon binding of GlnK (Conroy et al., 2007). Therefore, we postulate a similar “lock-down” function of SbtB when bound to SbtA, which is further corroborated by in vitro interaction of SbtA with SbtB in E. coli protein extracts (Du et al., 2014).

SbtA proteins from at least two different cyanobacterial species bind efficiently to their cognate, immobilized SbtB orthologs during immobilized metal affinity chromatography (IMAC). However, analysis of cyanobacterial mutants lacking SbtB suggests additional post-translational mechanisms are in place to regulate SbtA activity (Förster, Price unpublished). Bioinformatics revealed highly conserved, putative phosphorylation sites are present in BicA, SbtA and SbtB, which suggests phosphorylation could be involved in regulation. Furthermore, high molecular weight protein complexes containing SbtA or BicA can be detected in native protein extractions from cyanobacteria, raising the question whether interactions with other proteins and or metabolic intermediates determine BicA or SbtA activity (Förster, Price unpublished). Clearly, we need to improve our understanding of post-translational regulation of Cᵢ uptake in cyanobacteria to order to engineer active forms of BicA and SbtA into higher plant chloroplasts.
Algal bicarbonate transporters

The algal CCM is most well understood in the model green alga Chlamydomonas. The Chlamydomonas CCM is dynamic and only induced in the light and under limiting CO₂, which results in the expression of many hundreds of CCM-associated genes to sustain photosynthetic growth (Brueggeman et al., 2012; Fang et al., 2012; Miura et al., 2004). Several of these genes have been confirmed as essential or contributing to the functionality of the algal CCM through molecular, genetic and physiological studies of CCM mutants (Wang et al., 2015).

Two putative Cᵢ transporters have been identified as primary targets for engineering Cᵢ uptake into higher plants using the algal CCM (Figure 1). The currently accepted CCM model based on Chlamydomonas indicates that active Cᵢ uptake under limiting CO₂ is driven by High Light-Activated 3 (HLA3) and Low Carbon-inducible A (LCIA), located on the plasma membrane and chloroplast envelope, respectively. The model predicts a third Cᵢ transporter may be located on the thylakoid membrane to facilitate transport Cᵢ into the pyrenoid, but a candidate has yet to be identified. HLA3 is a putative ATP-dependent ABC-type transporter belonging to the multidrug resistance related protein family that pumps bicarbonate from the periplasm to the cytosol (Im and Grossman, 2002). LCIA is a putative anion transporter and a homolog of the bacterial formate transporter FocA, a symmetric pentamer that closely resembles an aquaporin channel (Wang et al., 2009). Similar to FocA, LCIA forms a protein complex under native conditions (Yamano et al., 2015). Although a Cᵢ uptake mechanism has yet to be confirmed, LCIA is believed to shuttle cytosolic bicarbonate into the stroma (Wang and Spalding, 2014). Both HLA3 and LCIA have been localised in vivo (Yamano et al., 2015), and capacity for Cᵢ uptake has been observed in vitro in Xenopus oocytes (Atkinson et al., 2016; Mariscal et al., 2006).

An additional component of interest is Low Carbon-inducible 1 (LCI1), a small protein localised primarily at the plasma membrane (Ohnishi et al., 2010). LCI1 is highly induced under low CO₂ and LCI1 over-expression studies have shown enhanced rates of Cᵢ uptake in Chlamydomonas. Nevertheless, topology predictions have indicated that the hydrophobic domains of LCI1 are too short to support a transmembrane channel, at least not in the monomeric form (Meyer and Griffiths, 2013). In addition, Cᵢ uptake was not observed in vitro in Xenopus oocytes (McCormick, unpublished data). These data suggest that LCI1 is not directly involved in Cᵢ uptake, but may act as an ancillary component to enhance the activity of other transporters, such as HLA3 (Figure 1). Notably, no close homologs of LCI1 are present in any other available algae proteome.

Engineering approaches for the introduction of algal CCM components have thus far focused on nuclear transformation, as the key CCM components identified in Chlamydomonas are encoded for in the nuclear genome (Atkinson et al., 2016). Highlighting one of the potential advantages of a
eukaryotic CCM, the majority of algal CCM components appear to localise to the appropriate locations in higher plants without modification of the coding sequence, indicating that higher plants are able to recognise cTPs from *Chlamydomonas*. In particular, LCIA and LCIB/LCIC were able to localise to the chloroplast envelope and stroma, respectively. Atkinson *et al.*, (2016) also demonstrated that LCII could be re-localised to the chloroplast stroma using a higher plant cTP, but not the envelope, suggesting again that LCII may not be capable of traversing membranes. Although initial attempts to re-localise HLA3 to the chloroplast envelope were ineffective, recent attempts have yielded more success (McCormick, *unpublished*). This again highlights efforts by Rolland *et al.* (2016), where selection of an appropriate transit signal sequence is required for targeting particular proteins to the chloroplast envelope (Figure 2).

Stable expression of either HLA3 or LCIA has not yet resulted in a significant growth advantage or a measurable change in photosynthetic capacity in higher plants (Atkinson *et al.*, 2016). It is possible that algal transporters may require additional regulation for activity *in planta*, similar to BicA and SbtA (Du *et al.*, 2014; Pengelly *et al.*, 2014). Yamano *et al.* (2015) have recently shown that HLA3 and LCIA act co-operatively when the CCM is active in *Chlamydomonas*, suggesting that both transporters may be required for efficient C\textsubscript{i} uptake. If LCIA is a passive channel for bicarbonate anions rather than an active transporter, C\textsubscript{i} uptake into the stroma would need to proceed against an electrical gradient, as the stromal side of the chloroplast envelope carries a negative charge (Wang and Spalding, 2014). Active C\textsubscript{i} transport at the plasma membrane would be required (e.g. by HLA3) to facilitate a sufficiently high bicarbonate concentration gradient in the cytosol to overcome this limitation. Likewise, HLA3 would be unable to deliver bicarbonate to the chloroplast effectively in the absence of LCIA. In fact, double insertion mutants for LCIA and HLA3 have a lower affinity for C\textsubscript{i} and slower growth rates compared with HLA3 or LCIA single-insertion mutants in *Chlamydomonas* (Duanmu *et al.*, 2009a; Yamano *et al.*, 2015).

**Ancillary requirements to algal bicarbonate transporter function**

It is apparent that additional features are required for a biophysical CCM to function efficiently, in both algae and cyanobacteria, including structures for RuBisCO aggregation (discussed below), components to reduce C\textsubscript{i} leakage, and removal of inhibitory CA activities (Meyer *et al.*, 2016). For example, expression of a heterologous CA in the cytosolic compartment of cyanobacteria or mis-targeting of the native carboxysomal CA is sufficient to “short-circuit” the CCM, resulting in loss of the ability to accumulate internal C\textsubscript{i} (Price and Badger, 1989a, b). Likewise, in *Chlamydomonas*, removal of putative CA enzymes LCIB and/or CAH3 [a \(\alpha\)-form CA located in the thylakoid lumen and inside the pyrenoid when the CCM is active (Blanco-Rivero *et al.*, 2012; Sinetova *et al.*, 2012)] disrupts the capacity of the algal CCM to function under sub-ambient CO\textsubscript{2} conditions (Duanmu *et al.*, 2009b; Jin *et al.*, 2016; Wang and Spalding, 2006). Notably, disruption of LCIA compromises C\textsubscript{i}
uptake and growth in Chlamydomonas under very limiting CO₂ levels, where photosynthesis is thought to be driven primarily by active HCO₃⁻ uptake, but disruption of both LCIA and LCIB is lethal under the same conditions (Wang and Spalding, 2014; Yamano et al., 2015), suggesting Cᵢ transporters may require specific stromal CA activities for a functional CCM.

Engineering and localisation strategies for algal Cᵢ transporters and other CCM components have been largely guided by models for introducing the cyanobacterial CCM into plants (McGrath and Long, 2014; Price et al., 2013). The model of McGrath and Long (2014) predicts that improvements in photosynthesis from localisation of cyanobacterial Cᵢ transporters in the chloroplast envelope are not enhanced by additional Cᵢ transporter activity on the plasma membrane. However, growing knowledge of the apparent co-dependencies and redundancies in the eukaryote biophysical CCM illustrate the requirement for predictive models specifically focused on the multi-layered algal CCM system (Wang et al., 2015). The algal CCM (at least in Chlamydomonas) appears to employ different CCM components at different limiting Cᵢ conditions. Improved understanding of the molecular mechanisms and catalytic characteristics of algal CCM components, such as LCIA and HLA3, will aid the development of models to test whether retention of algal CCM components in the plasma membrane and/or chloroplast envelope is a tractable strategy in plants. Such models should also include the critical role of pH that has been demonstrated for the cyanobacterial CCM (Mangan et al., 2016).

The generation of an elevated bicarbonate pool within the chloroplast by way of active bicarbonate transporters is the first step toward introducing a complete biophysical CCM into higher plant chloroplasts. While these transporters alone theoretically lead to enhanced photosynthetic performance (Price et al., 2011a), the addition of further CCM components, such as a RuBisCO microcompartment (i.e. a carboxysome) or microbody (i.e. a pyrenoid), will allow for progression towards the maximum theoretical performance (McGrath and Long, 2014).

Carboxysome expression in higher plants: challenges and current status

Two convergently-evolved carboxysome sub-types with different biogenesis

Carboxysomes are protein-bound prokaryotic organelles that function to separate the CO₂ fixation step of photosynthesis from the cytoplasmic milieu (Rae et al., 2013a; Rae et al., 2013b). It is thought that enhancement of CO₂-fixation is conferred by a selectively-permeable shell which provides some level of diffusional resistance to CO₂, while allowing free transit of other RuBisCO substrates and products (Mangan et al., 2016; Menon et al., 2010). These microbodies have evolved at least twice, once in diverse proteobacterial species, and incorporating Form-1A RuBisCO, and again in cyanobacteria, incorporating Form-1B RuBisCO - the higher-plant RuBisCO lineage (Badger and
Bek, 2008; Tabita, 1999). These convergent forms of carboxysome are named for the phyletic affinity of their RuBisCO enzymes: α-carboxysomes are the proteobacterial type, containing the proteobacterial RuBisCO Form-1A, and β-carboxysomes are the cyanobacterial type, and encapsulate RuBisCO Form-1B (Badger and Bek, 2008). To complicate matters somewhat, the marine/oligotrophic cyanobacterial clade encompassing the genera Cyanobium, Synechococcus, and Prochlorococcus gained the α-carboxysome by horizontal gene transfer (Rae et al., 2011). Accordingly, this lineage of cyanobacteria is referred to as α-cyanobacteria, and the remainder are β-cyanobacteria (Badger et al., 2002).

Bioinformatic and structural comparison of the proteins of α- and β-carboxysomes has shown that both contain phyletically related, albeit divergent, protein shell structures (Heinhorst et al., 2014). However, the internal enzymes of each carboxysome form are quite different. Each has a unique type of CA, and the principal structural proteins of each form are not evolutionarily related. The fine details of carboxysome structure is not our focus, and we divert the reader’s attention to recent comprehensive reviews of carboxysome structure, function, and evolution (Heinhorst et al., 2014; Kimber, 2014; Rae et al., 2013a; Rae et al., 2013b). To briefly summarise: α-carboxysomes are assembled from proteins with the Cso nomenclature (Carboxysome), where the CsoS2 protein is responsible for packaging the α-carboxysomal CA (CsoSCA) and Form-1A RuBisCO within shells composed of multiple paralogs of CsoS1, and CsoS4 structural proteins (Cai et al., 2015). β-Carboxysomes are assembled from proteins with Ccm nomenclature (CO2 concentrating mechanism) whereby at least two proteins derived from the ccmM gene assemble the β-carboxysomal CA (CcmM or CcaA or both, depending on the species) along with Form-1B RuBisCO (Cot et al., 2008; Long et al., 2007), and these arrayed enzymes are bridged by CcmN to shell structures formed by paralogs of CcmK, CcmO, CcmP, and CcmL structural proteins (Cai et al., 2013; Kinney et al., 2012; Rae et al., 2012). These represent two fundamentally different biogenesis pathways whereby β-carboxysomes primarily rely on the formation of a luminal RuBisCO complex (Cameron et al., 2013; Chen et al., 2013) while empty α-carboxysome structures can form in the absence of its RuBisCO cargo (Menon et al., 2008; Rae et al., 2013a).

**Carboxysome-mediated metabolism**

In theoretical structure-function models, the cyanobacterial carboxysome operates as a discrete physico-metabolic unit, and the two types of carboxysome operate in essentially the same way (Figure 3) (Whitehead et al., 2014). It is helpful to view the steady-state kinetics of carboxysomes as sequential metabolic steps, whereby intracellular C\textsubscript{i}, which exists primarily as the bicarbonate ion in model cyanobacteria, transits the outer protein shell, is interconverted with CO\textsubscript{2} by the carboxysomal CA(s) (CsoSCA in α-carboxysomes, and CcmM and/or CcaA in β-carboxysomes) and fixed into 3-phosphoglycerate (3-PGA) by RuBisCO before release into the cytosol. By virtue of the outer shell
functioning as a CO₂ barrier, the carboxysome lumen is comparatively CO₂-rich, and this environment favours RuBisCO carboxylation to the extent that photorespiration is negligible in some cyanobacterial species (Birmingham et al., 1982; Cheng and Colman, 1974).

Adaptive evolution of carboxysomal RuBisCOs has resulted in these enzymes occupying different catalytic optima than ‘free’ RuBisCO. Carboxysomal RuBisCOs are comparatively fast enzymes, but have lower CO₂/O₂ specificity (S<sub>CO₂</sub>) (Tcherkez et al., 2006). These characteristics allow the cyanobacterial cell to make a comparatively small investment in RuBisCO biosynthesis, especially in comparison to C₃ plants. In theoretical models, carboxysomal RuBisCO operates more effectively because of its localisation with CA at the top of the CO₂ concentration gradient. However, recent work reiterates that merely co-habiting RuBisCO with CA is not enough to enhance photosynthesis appreciably (Gonzalez-Esquer et al., 2015; Mangan and Brenner, 2014; Mangan et al., 2016). Rather, the important function is the resistance to CO₂ diffusion, within an elevated pool of HCO₃⁻, which allows carboxysomes to operate effectively. This contrasts with pyrenoids, whose structure superficially resembles carboxysomes, but apparently lack this clearly-defined barrier (Figure 3).

Folding, assembly, and activation of cyanobacterial RuBisCO in higher plants

The two subunits of Form-1 RuBisCO enzymes are translated and assembled via a complex system of post-translational modifications, and folding and assembly chaperone proteins to form hexadecameric holoenzymes of eight large and eight small subunits (Houtz et al., 2008; Whitney et al., 2011; Wilson and Whitney, 2015). Many of the steps in this pathway remain obscure and are poorly understood. However, it is clear that phyletically distant RuBisCOs have co-evolved with distinct types of folding and activation chaperone proteins (Whitney et al., 2015). For example, the two RuBisCO forms found in cyanobacteria [Form-1A RuBisCO in α-cyanobacteria and Form-1B RuBisCO in β-cyanobacteria (Rae et al., 2013a; Rae et al., 2013b)] have complex and differing chaperone requirements. In some species with Form-1A RuBisCO, the α-carboxysome RuBisCO assembly factor (acRAF) is required for efficient RuBisCO folding (Wheatley et al., 2014), while RuBisCO activation proceeds by CbbQ-CbbO and CbbX (the Cbb nomenclature stands for the Calvin-Benson-Basham cycle) (Sutter et al., 2015; Tsai et al., 2015). Species with Form-1B RuBisCO require chaperonin proteins, RuBisCO protein X (RbcX) and RuBisCO accumulation factor 1 (RAF1), which interact with the large subunit (LSU) of RuBisCO to assist in LSU folding (Emlyn-Jones et al., 2006; Hauser et al., 2015; Kolesinski et al., 2014), while RuBisCO activase (Rca) is required for RuBisCO activation (Li et al., 1999).

The roles of RuBisCO activation chaperones within carboxysomes remains somewhat unclear. Evidence suggests that carboxysomes are persistent through cell-division (Chen et al., 2013) highlighting that, once incorporated into the carboxysome, any requirement for RuBisCO activation would probably need to occur within the structure. The activase module CbbQ appears to be a minor
component of the α-carboxysome shell in *Halothiobacillus neapolitanus* carboxysomes (Sutter et al., 2015). How it might act on inhibited RuBisCO active sites within the carboxysome remains unclear. Indeed, the requirement for re-activation may be low within β-carboxysomes due to the relatively high rate of inhibitor release from Form-1B RuBisCO active sites when catalytic mis-fire occurs (Pearce, 2006). This detail is currently unknown for Form-1A RuBisCOs. The β-cyanobacterial activase, Rca, was not associated with carboxysomes in cells of *Synechococcus elongatus* PCC7942 (Synechococcus hereafter) (Friedberg et al., 1993), while the α-cyanobacterial activase, CbbX, has not been reported in α-carboxysomes. However, inactivation of the *rca* gene in *Anabaena* led to light-dependent effects on RuBisCO activation and cell growth (Li et al., 1999), highlighting its importance. Notably, a proportion of RuBisCO is always found exterior to the carboxysome in both α- and β-cyanobacteria (McKay et al., 1993), which would be more accessible to activases. Within green algae, activase enzymes and RuBisCO co-habit the pyrenoid (McKay et al., 1991) and, while Rca is required for optimal photosynthesis in Chlamydomonas, its absence is partially compensated for by an active CCM (Pollock et al., 2003).

Understanding the importance of chaperones for engineering heterologous RuBisCO enzymes into higher plants is currently an active area of research. Whitney et al. (2015) showed that the content and assembly rate of Arabidopsis RuBisCO (Form-1B) in transgenic tobacco was improved 2-3 fold when co-transformed with Arabidopsis RAF1 (AtRAF1). This work demonstrated that cognate chaperones can significantly improve the biogenesis of heterologous RuBisCOs, even in closely related higher plant species (Whitney et al., 2015). In contrast, RuBisCO from the β-cyanobacterium *Synechococcus* introduced into tobacco did not require or greatly benefit from the presence of RbcX (Lin et al., 2014b; Occhialini et al., 2016). Although the cyanobacterial RuBisCO assembled and was active, the RuBisCO content of these transgenic lines was very low, comparable to the RuBisCO content achieved by expression of archaeal RuBisCO enzymes (Wilson et al., 2016). Furthermore, the catalytic characteristics of the cyanobacterial RuBisCO were not sufficient to support plant growth under ambient CO₂ levels, highlighting the expectation that carboxysomes alone, or their cognate RuBisCOs, are not expected to improve C₃ photosynthetic performance (McGrath and Long, 2014). Lines co-transformed with the β-carboxysomal RuBisCO-interlinking protein, CcmM-35, had a higher RuBisCO content, suggesting that CcmM-35 might act to stabilise the cyanobacterial RuBisCO in planta (Occhialini et al., 2016). Whether co-transformation with *Synechococcus* RAF1 would result in further improvements remains to be tested.

**Appropriate carboxysome operons for chloroplast engineering**

Current efforts to express β-carboxysomes in higher plant chloroplasts have utilised genes from the well-characterised carboxysome of *Synechococcus* (Lin et al., 2014a; Lin et al., 2014b; Occhialini et al., 2016). Lin, Occhialini and co-workers have been successful in transiently expressing shell
(CcmK, L and O) and lumen (CcmM and N) proteins, which have been imported into the chloroplasts of N. benthamiana to form carboxysome-related structures (Lin et al., 2014a). They have also generated crucial luminal RuBisCO-CcmM35 complexes within tobacco chloroplasts (Lin et al., 2014b; Occhialini et al., 2016). While these breakthrough results are highly informative, the carboxysome gene set of Synechococcus is complicated by the existence of additional genes and translational controls that are crucial for the formation of functional carboxysomes (Long et al., 2016). For example, there are several critical isoforms of CcmK (Rae et al., 2012), required in unknown quantities within the shell. The RuBisCO-binding protein CcmM is complicated by expression of two protein isoforms (Long et al., 2010), with a shorter form arising from a ribosomal-entry site within the reading frame of ccmM (Long et al., 2007; Long et al., 2005; Ludwig et al., 2000). The carboxysomes of Synechococcus and related species also contain the additional carbonic anhydrase CcaA (Long et al., 2011; So and Espie, 1998), whereas other β-carboxysomes make use of an active γ-CA at the N-terminal of the CcmM protein (de Araujo et al., 2014; Peña et al., 2010).

There is also the potential requirement of an additional shell protein, CcmP, which may play a role in the passage of large metabolites across the shell (Cai et al., 2013). The requirement for RuBisCO ancillary proteins (mentioned above) is still not completely clear and their presence within the β-carboxysome is completely unknown. The complexity of the β-carboxysome is underlined by the arrangement of carboxysome-related genes into clusters, compared with more tightly regulated operons for α-carboxysomes (Cai et al., 2008; Heinhorst et al., 2014).

Even if expression of all the components required for carboxysome assembly is achieved, we see carboxysomes with smaller genetic footprints as the most logically appealing for C₃ chloroplast transformation (Long et al., 2016). Not only is the genetic load lower, and the potential for recombination between homologous genes significantly reduced, but self-assembly of comparatively simpler carboxysomes might be more efficient in the chloroplast. Cai et al. (2016) were successful in engineering relatively simple synthetic carboxysome shells and cargo in E. coli using genes from the β-cyanobacterium Halothece sp. PCC7418, demonstrating that both simple gene units and synthetic approaches to the challenge of chloroplastic carboxysome expression have significant merit. Self-assembly of the carboxysome is a topic of ongoing investigation, but current evidence suggests that carboxysomes can assemble from their component parts with remarkable ease (Cameron et al., 2013; Chen et al., 2013).

Several simple candidate operons exist for both α- and β-carboxysomes. Gloeobacter violaceus PCC7421 has the smallest known β-carboxysome gene content (Nakamura et al., 2003), and is also likely to encode a CA-competent form of CcmM, obviating the need for a separate carboxysomal CA gene. Moreover, G. violaceus carboxysomes appear to require only five outer shell proteins (CcmK1, CcmK2, CcmO, CcmL and CcmP). However, G. violaceus likely still expresses two isoforms of...
CcmM. The high-light adapted *Prochlorococcus* clade have the simplest known α-carboxysomes in terms of gene content with as few as nine structural genes (Roberts *et al.*, 2012), and with some expressing a single form of CsoS2 (the α-carboxysomal equivalent of CcmM) (Chaijarasphong *et al.*, 2016; Roberts *et al.*, 2012). Thus, based on gene number and apparent ease of expression, α-carboxysome operons are currently the most attractive candidates for chloroplastic transformation.

RuBisCO fixation rates are assumed to approach their theoretical maximum for a fully functional biophysical CCM in a C_3* chloroplast (McGrath and Long, 2014). Thus, the catalytic properties of extant carboxysomal RuBisCOs must be considered when choosing appropriate sources for engineering. Based on measurements thus far, RuBisCOs from α-carboxysomes appear catalytically inferior to those from β-carboxysomes. *Prochlorococcus* and *Cyanobium* Form-1A RuBisCOs have $k_{cat}$ values of 6.6 and 9.0 s⁻¹, respectively (Shih *et al.*, 2016; Whitehead *et al.*, 2014), compared to reports of up to 14.4 s⁻¹ for Form-1B RuBisCO from *Synechococcus* (Occhialini *et al.*, 2016; Whitehead *et al.*, 2014). The current lack of catalytic data for α-carboxysomal Form-1A RuBisCOs warrants greater analysis of this diverse family of enzymes to identify potential high-performance candidates. However, some α-carboxysomes are amenable to sequestration of heterologous RuBisCOs (Menon *et al.*, 2008), which opens up the possibility of engineering structurally simple carboxysomes containing kinetically superior RuBisCOs into C_3 chloroplasts.

The generation of a carboxysome-based biophysical CCM in the chloroplast will be a multi-step process. In the absence of functional transporters, much work has focused on optimising carboxysomes expression in the chloroplast (Long *et al.*, 2016; Price *et al.*, 2013). The latter is challenging, as the functional performance of expressed carboxysomes is difficult to determine *in planta*. The loss of single protein components from the carboxysome (either due to genetic lesion or stoichiometric perturbations) can lead to a loss or decrease of function, even though morphology the carboxysome may not appear altered (Bonacci *et al.*, 2012; Cai *et al.*, 2009; Rae *et al.*, 2013a). Thus, the ability to isolate carboxysome structures expressed in the chloroplast, and to determine their function, will be essential. While synthetic β-carboxysomes have been expressed and purified from *E. coli* (Cai *et al.*, 2016), successful purification of intact, functional β-carboxysomes from wild-type (WT) cyanobacteria has been elusive (Long *et al.*, 2005; Long *et al.*, 2011). Conversely, α-carboxysomes can be purified intact (Gonzales *et al.*, 2005; Roberts *et al.*, 2012; So *et al.*, 2004), and some findings indicating function have been reported (Cai *et al.*, 2009; Menon *et al.*, 2010). The development of further assays for carboxysome functionality *in vitro* will be important to guide future engineering strategies.
Costs of carboxysome expression in the chloroplast - how many do we need?

Conservative predictions estimate between eight and 13 separate polypeptides will be required to generate a functional α- or β-carboxysome in a C₃ chloroplast (Long et al., 2016). While there are still unknowns regarding the absolute protein structure of both α- and β-carboxysomes, existing knowledge of protein stoichiometries for each type (Rae et al., 2013a; Roberts et al., 2012) makes it possible to estimate the nutritional and spatial costs of biosynthesis in plants. In Table 1 we have calculated the costs of carboxysome synthesis in a C₃ leaf under two possible scenarios (in a fully functional CCM where Cᵢ accumulation allows for the operation of RuBisCOs at their catalytic maximum). First, the predicted costs of both α- and β-carboxysomes are estimated assuming the replacement of the WT C₃ enzyme with an equivalent number of carboxysomal RuBisCOs (Table 1, upper rows). Under this scenario, the additional structural and functional components of carboxysomes (i.e. shell, CA and RuBisCO binding proteins) mean that there is additional nitrogen, carbon and energy investment for the construction of carboxysomes. However, these costs are offset by the expected enhancement in carbon fixation rates within a fully-functional chloroplastic CCM when we model a scenario in which the theoretical in vitro maximum carboxylation rates (Vₖ,max) are matched with those expected in WT leaves (Table 1, bottom rows). Given that the kₑ values for carboxysomal RuBisCOs are three to four-fold higher than C₃ counterparts, and assuming a chloroplastic CCM can deliver CO₂ concentrations to carboxysomal RuBisCOs in excess of their relatively high Kₑ (and minimizing the adverse effects of low Sₐ/O), our calculations actually predict significant overall savings in carbon, nitrogen, total protein, and energy investments under this scenario. One caveat here is that the costs of the bicarbonate transporters required to elevate the stromal Cᵢ concentration are not included. Nonetheless, a 44-63% saving in leaf protein and 40-65% saving in stromal volume dedicated to RuBisCO could be expected in the scenario presented, and is likely to significantly offset the costs of transporter synthesis. Notably, predicted savings in the energy cost of peptide synthesis, under conditions which support the same theoretical Vₖ,max as WT plants (Table 1), may also contribute to the cost of energy-driven bicarbonate transport uptake. These predicted gains are facilitated also in part by the extremely high concentration of RuBisCO active sites within both carboxysome types [10.9-14.1 mM (Whitehead et al., 2014)], which approach or exceed the maximum concentrations predicted in C₃ stromae (Pickersgill, 1986). Differential costs between the two carboxysome types are primarily due to their intrinsic RuBisCO kinetic maxima.

In order to achieve the equivalent number of RuBisCO active sites in a chloroplastic carboxysome as in a WT plant, there are significant costs in both protein requirement and stromal volume (Table 1). However, this represents a substantial over-investment and is likely not required with a fully-functional chloroplastic CCM. The estimated carboxysome numbers required to achieve the same Vₖ,max as WT plants suggest leaf carboxysomal RuBisCO active site concentrations need only be 2.8 to 4.5-fold lower than that of WT plants. While this estimate is heavily dependent on the relative
improvement in \( k_{\text{cat}} \) between the WT RuBisCO and its replacement carboxysomal counterpart, our calculations indicate that even low levels of carboxysomal RuBisCO expression may provide a benefit within the context of a chloroplastic CCM. For example, the level of Synechococcus RuBisCO expression reported by Occhialini \textit{et al.} (2016) was low, but those values do approach the relative levels required in a fully-functional chloroplastic CCM. While the data presented in Table 1 present a simplified approach to estimating carboxysome requirements, this analysis highlights the value in modelling plant performance to understand if there is an over- or under-investment in carboxysome synthesis.

\textbf{Structure and dynamics of the algal pyrenoid}

Pyrenoids are subcellular organelles located in the plastid stroma of almost all eukaryotic microalgae, as well as diatoms, dinoflagellates, coccolithophores and a single group of land plants, the hornworts (Hanson \textit{et al.}, 2014; Hopkinson \textit{et al.}, 2016; Schmid, 2001; Vaughn \textit{et al.}, 1990). Species may contain a single pyrenoid (as in Chlamydomonas) or several. Like the carboxysome, the fine structure of the pyrenoid is well known to be formed of a dense protein aggregate (Bertagnolli and Nadakavukaren, 1970; Kowallik, 1969; Leadbeater and Manton, 1971; Markowitz and Hoffman, 1974), which we now interpret to be primarily RuBisCO (Engel \textit{et al.}, 2015). Although pyrenoids are absent in some chlorophytes that appear to retain a capacity for active \( C_3 \) uptake (Morita \textit{et al.}, 1998), pyrenoids likely play a key role in the efficiency of biophysical CCMs in eukaryotes where they are present, and are associated with species with high internal \( C_3 \) levels (Figure 3) (Badger \textit{et al.}, 1998; Morita \textit{et al.}, 1999). While the catalytic properties of microalgal RuBisCOs enable photosynthesis at ambient \( CO_2 \) (Jordan and Ogren, 1981), disruption of the pyrenoid inhibits induction of the algal CCM in species where pyrenoids are present (Genkov \textit{et al.}, 2010; Ma \textit{et al.}, 2011; Mackinder \textit{et al.}, 2016; Meyer \textit{et al.}, 2012). Thus, pyrenoids are thought to be of similar importance to the carboxysomes (Giordano \textit{et al.}, 2005), and their structures do share several key features: a dense aggregation of spatially arranged RuBisCOs and an intrinsic internal CA activity (CAH3 in Chlamydomonas) (Figure 4). However, there are several marked structural distinctions. For example, pyrenoids do not have a defined shell structure, but are typically (but not always) surrounded by a sheath of starch plates. It remains unclear if the starch layer inhibits \( CO_2 \) leakage in a manner akin to the carboxysome shell. In Chlamydomonas under \( C_3 \)-limiting conditions, the enigmatic LCIB/LCIC complex is concentrated in a discrete region around the starch layer, but can also be found in the surrounding stroma (Wang and Spalding, 2014). LCIB/LCIC may be involved in unidirectional re-capture of \( CO_2 \) released from the pyrenoid and thylakoid membranes (Jin \textit{et al.}, 2016). Pyrenoids are also traversed by a dense network of specialised thylakoid membranes that fenestrate the starch sheath and fuse to form internal minitubules contiguous with the chloroplast stroma (Engel \textit{et al.}, 2015). The internal minitubules are thought to provide a channel for the substrates and products of RuBisCO, while the surrounding larger tubules facilitate bicarbonate diffusion to CAH3 localised within the
trans-pyrenoidal thylakoids (Meyer et al., 2016). As the thylakoid membranes traversing the pyrenoid are PSI-enriched, Smith and Griffiths (Smith and Griffiths, 2000) have suggested that the pyrenoid could also provide a means of spatially separating the oxygenic PSII complex from RuBisCO to reduce the occurrence of competitive photorespiratory reactions. Nevertheless, pyrenoidal RuBisCOs do have greater $S_{CO}$ values than those from carboxysomes (Badger et al., 1998).

### Protein determinants of pyrenoid assembly and ultrastructure

Identifying factors that regulate assembly of the pyrenoid is currently an active area of research. Pyrenoids appear considerably more dynamic than carboxysomes, and undergo changes in size, composition and density depending on $C_i$ availability and light (Mitchell et al., 2014; Ramazanov et al., 1994). Under photoautotrophic growth at high $CO_2$ levels (1% or higher), less than half of RuBisCO is located in the pyrenoid, and starch accumulates in the stroma rather than around the pyrenoid (Borkhsonious et al., 1998). Following CCM induction at air levels of $CO_2$ (0.04%) or below, >90% of RuBisCO aggregates in the pyrenoid, and the pyrenoid starch sheath is formed consequently. Similarly, pyrenoid RuBisCO content increases following diel transitions from dark to light, which coincides with light-dependent (or dawn) induction of the CCM (Mitchell et al., 2014).

The structure of the RuBisCO small subunit (SSU) plays a key role in pyrenoid assembly. Hybrid RuBisCOs that contain native Chlamydomonas LSUs and a higher plant SSU cannot form pyrenoids (Genkov et al., 2010). Meyer et al. (2012) elegantly demonstrated that two α-helices located on the SSU periphery are more hydrophobic in species with pyrenoids, and are a critical feature for assembly in Chlamydomonas. In contrast, the LSU does not appear to be important. Pyrenoid morphology remained unaffected by mutations in the LSU, even those resulting in an inactive RuBisCO (Meyer, 2010; Rawat et al., 1996). The pyrenoid is known to contain two additional non-thylakoid-specific proteins, Rca and the recently characterised Essential Pyrenoid Component 1 [EPYC1, previously LCI5 (Mackinder et al., 2016; McKay et al., 1991)]. Null mutants for Rca show an expected decrease in the amount of active RuBisCO, but no effect on pyrenoid morphology or $C_i$ affinity (Pollock et al., 2003). In contrast, EPYC1 is required for pyrenoid assembly and CCM function, and is believed to be a linker chaperone that facilitates aggregation of RuBisCO within the pyrenoid structure during CCM induction (Figure 4). EPYC1 is an intrinsically disordered protein with four nearly identical repeat regions, and each repeat is suggested as a site for RuBisCO binding, possibly by interactions with the SSU. Thus, EPYC1 may play a similar role to the α-carboxysomal RuBisCO-binding protein CsoS2 which is also highly disordered and binds SSUs (Cai et al., 2015).

Successful engineering of the pyrenoid into higher plants will require a better understanding of pyrenoid biogenesis, regulation and integration with other CCM components. Fortunately, recent pioneering research has already started to provide important details. For example, CAH3 and EPYC1...
undergo multiple phosphorylations when CO₂ is limiting (Blanco-Rivero et al., 2012; Mackinder et al., 2016). Similarly, models have implicated phosphorylation in activity regulation of the LCIB/LCIC complex (Jin et al., 2016). Identifying components involved in post-translational modification and signalling is ongoing. Recently, the thylakoid localised Ca²⁺-binding protein (CAS) was shown to re-localise to minitubules in the pyrenoid during CCM induction in Chlamydomonas (Wang et al., 2016). CAS may operate within a Ca²⁺- and CO₂-mediated chloroplast retrograde signalling pathway to regulate expression of nuclear encoded CCM components, such as HLA3 and LCIA. Although it remains unclear how thylakoids come to traverse pyrenoids, newly identified CCM components may shed light on the relationship between RuBisCO, thylakoids, and starch surrounding the pyrenoid (Powell, 2016), and known key regulators involved in pyrenoid assembly and CCM function, such as CIA5 and CIA6 (Fukuzawa et al., 2001; Ma et al., 2011; Xiang et al., 2001).

Towards building a pyrenoid in higher plants

As with the cyanobacterial CCM, developing a chloroplastic pyrenoid-based CCM in C₃ plants will likely require a step-wise incorporation of pyrenoidal components, and modification of native RuBisCO (Meyer et al., 2012). In this respect, the algal CCM may have a tractable advantage over the cyanobacterial CCM. Although cyanobacterial RuBisCOs have high $k_{\text{cat}}$ rates, their catalytic efficiencies in higher plants is hampered by low values for $K_c$ and $S_{\text{CO}}$ (Lin et al., 2014b; Occhialini et al., 2016). This reflects their co-evolution within the carboxysome structure, and adaptation to higher CO₂ concentrations compared to RuBisCOs prevailing in either angiosperm chloroplasts or algal pyrenoids (Meyer and Griffiths, 2013). In contrast, the catalytic characteristics of algal RuBisCOs are close to those of RuBisCOs in C₃ plants (Galmés et al., 2014; Meyer et al., 2012; Prins et al., 2016). Thus, in a development pipeline for engineering the pyrenoid-based CCM, initial introduction of an algal RuBisCO is likely neutral compared with the expected drop in photosynthetic performance and growth during the development of a cyanobacterial CCM (McGrath and Long, 2014).

If the algal LSU is not required for pyrenoid assembly, as evidence suggests [see the previous section (Meyer, 2010; Rawat et al., 1996)], engineering a plant with RuBisCO capable of pyrenoid assembly may require replacement only of the higher plant SSU. Thus, requirements for co-transformation of additional RuBisCO assembly chaperones, as seen for the LSU (Bracher et al., 2015; Whitney et al., 2015), may be limited to those needed for the SSU. Currently, there is little evidence to suggest that the SSU requires assembly chaperones in the chloroplast (Wilson and Whitney, 2015). Recent work in Arabidopsis has shown that modified SSUs competent for pyrenoid assembly can be incorporated into C₃ plant RuBisCO with minimal effects on RuBisCO catalysis, content and plant growth at ambient CO₂ (Atkinson et al., 2017). These Arabidopsis lines are proving to be useful backgrounds in which to i) examine pyrenoid assembly chaperones, such as EPYC1, and other candidates as they emerge, ii)
clarify the nature of SSU-associated interactions, and requirements for RuBisCO aggregation in planta, and iii) integrate other essential CCM components involved in active C\textsubscript{i} uptake.

**Synthetic approaches to constructing a biophysical CCM in higher plants**

*Algal CCMs*

Due to initiatives such as the 1KP project (onekp.com), the increasing availability of sequencing data has made it possible to begin exploring how ubiquitous the CCM components of the Chlamydomonas CCM are in other eukaryotes with biophysical CCMs. For example, some components, such as LCIA and HLA3, appear to be common, while others such as LCI1, are not (Meyer and Griffiths, 2013). Currently the majority of sequenced genomes remain poorly annotated, but this growing resource could offer a useful tool to mine alternative components for building a synthetic biophysical CCM.

Although the Chlamydomonas pyrenoid is clearly the best understood, it is also arguably one of the most structurally complex. Pyrenoids do vary significantly in terms of shape, size and apparent complexity between different species. Although some algae and bryophytes possess highly complex pyrenoids (i.e. with traversing minitubules and/or thylakoid invaginations, sometimes with and without a starch sheath present), the simplest pyrenoids, as viewed in transmission electron micrographs, appear to be homogenous bodies with no starch sheaths and often without any clear localisation near the thylakoids, as seen in *Laminaria* spp. (Chi, 1971), and many Haptophytes (Bendif *et al.*, 2011; Griffiths, 1970). These types of pyrenoid might be the simplest in terms of structural protein components, and thus may be the most feasible models for assembly in higher plant chloroplasts. Recent work has shown that a wide diversity of Haptophyte algae with ‘simple’ pyrenoids, or in some cases no pyrenoid at all, are capable of high photosynthetic rates (Rae and Rickaby, *unpublished*). However, it is clear we still know very little about the molecular make up of pyrenoids, and assembly components, in other eukaryotes. Growing evidence of the contribution of single cell C\textsubscript{4} like CCMs in some algae and diatoms suggests that additional complexities also may exist in certain species (Derelle *et al.*, 2006; Hopkinson *et al.*, 2016). Thus, improving our current molecular and physiological understanding of biophysical CCMs in other globally significant eukaryotes will be important for optimising the chances of successful transfer to higher plants. The hornworts are particularly attractive candidates for further study, as they are phyletically the closest group to angiosperms that retain pyrenoids and biophysical CCMs (Li *et al.*, 2017). The pyrenoid-bearing hornwort, *Anthoceros agrestis*, was recently established as a tractable model system and may prove useful as a resource for comparative studies (Szövényi *et al.*, 2015).
Cyanobacterial CCMs

Efforts to increase the existing genomic data for the cyanobacteria phylum have demonstrated the huge variety of potential CCM component homologs available for exploitation (Shih et al., 2013). Gaudana et al. (2015) showed genes with homology to the key transporters BicA, SbtA (and putative regulator SbtB) were widely conserved in 128 different sequenced genomes. Further research may reveal the malleability of the cyanobacterial CCM and minimal requirements for functionality. Nonetheless, several good examples of minimal CCMs are evident in nature. While the minimal carboxysome gene sets of G. violaceus and Prochlorococcus spp. were discussed earlier, the endosymbiotic cyanobacteria Richelia intracellularis, Anabaena azollae, Synechococcus spongariurn, and the endosymbiont of Paulinella chromatophora, converge on a minimal gene set comprising one identifiable bicarbonate transporter, one CO₂-uptake/scavenging system, and minimal carboxysome operons. The free-living β-cyanobacterium Trichodesmium erythraeum and α-cyanobacterium Prochlorococcus marinus also have a small set of identifiable CCM components (Hopkinson et al., 2014; Kranz et al., 2011).

Rational designs for a chimeric biophysical CCM

Recent progress highlights the possibility of introducing CCM components from either the nuclear genome, the plastid genome, or both (Atkinson et al., 2016; Atkinson et al., 2017; Lin et al., 2014a; Lin et al., 2014b). These findings are exciting as they open up several opportunities, including parallel development of transformant lines and their subsequent conjunction through crossing, or transformation of crop species currently recalcitrant to chloroplastic transformation. However, nuclear expression and import into the chloroplast will require a better understanding of the effect of any transit peptide scars on functional CCM components in the chloroplast. Previous successes in appropriate localisation of algal CCM components in the stroma and chloroplast envelope, and formation of carboxysome structural intermediates from small proteins when fused with native transit peptides are promising (Atkinson et al., 2016; Lin et al., 2014a).

With such a wide availability of tools, the structural and functional concepts of biophysical CCM components from eukaryotes and cyanobacteria could be potentially recapitulated via synthetic biology approaches. For example, a feasible solution for maximising Cᵢ uptake may be a chimeric approach, such as combining Cᵢ uptake systems from cyanobacteria and algae. In addition, orthogonal regulators could be used to control transporter activity. Efforts are underway to engineer rhodopsin-driven bicarbonate transporters, which would enable activity to be regulated by light and photosynthetic activity (Gaudana et al., 2015). The co-encapsulation of a RuBisCO and CA into a simplified (single protein) capsid shell was recently demonstrated by (Frey et al., 2016). While this ‘primordial carboxysome’ system showed marginal changes in RuBisCO catalysis over the naked enzyme, it established a logical foundation for further development. For example, such systems could
utilise directed evolution to advance the functionality of synthetic carboxysomes. While the functional role of the carboxysome shell is still under research scrutiny, designing a simplified microcompartment that approaches the functionality of evolved carboxysomes with a minimal set of genetic components (e.g., Cai et al., 2016; Gonzalez-Esquer et al., 2015) may be a useful strategy. Alternatively, a combinatorial approach may be more beneficial (e.g. Fuentes et al., 2016) where multiple genes could be expressed in the chloroplast, covering every possible requirement of an active carboxysome. Given recent progress in using mutant libraries and high-throughput screening approaches in Chlamydomonas to detect novel putative eukaryotic CCM components (Li et al., 2016; Powell, 2016), the latter approach may also be useful for studying pyrenoid assembly. Our understanding of pyrenoid biogenesis and regulation still lingers far behind that of carboxysomes, but recent advances in high fidelity image analyses of pyrenoid architecture (Engel et al., 2015) and high-throughput genetics may help to provide a rapid catch up.

Future work

Much progress has been made in defining the requirements for introducing biophysical CCMs, and engineering different CCM components into C_3 plants. Future work requires focus in key areas. In particular, characterisation of the processes involved in regulating bicarbonate transporter activities is critical to facilitate elevated C_i levels in C_3 chloroplasts. Price et al. (2011a) noted that an important research focus toward a C_3 chloroplasmatic CCM is an understanding of the activation and energization processes for bicarbonate transporters. Initial modelling indicates the cost of C_i import is minimal and is offset by reductions in photorespiration (Price et al., 2011a). Nonetheless, further accurate modelling of the energy costs is needed to confirm that a C_3 chloroplast can meet the challenge. A related issue that has received little focus but will need to be addressed is the burden and potential limitations of protein expression on the C_3 chloroplast envelope (Breuers et al., 2012). This is an exciting time for algal CCM research, as many new details of the structure and physiology of the eukaryotic pyrenoid are currently emerging. These should be used to inform algal-specific models for engineering C_3 plants. While our understanding of the components required for carboxysomes is significantly more mature than that of pyrenoids, improved techniques to study carboxysome assembly and functionality in planta (including the chaperone requirements for carboxysomal RuBisCOs) will greatly assist efforts to engineer a functional biophysical CCM into higher plants.

Acknowledgements

Work highlighted in this review was funded by the UK Biotechnology and Biological Sciences Research Council (AJM, NA and CNV supported through grants BB/I024453/1 and BB/M006468/1). BML, BF, WYH and GDP are supported through a sub-award from the University of Illinois as part of the Realizing Increased Photosynthetic Efficiency (RIPE) project, funded by the Bill & Melinda Gates Foundation. BM, BDR, NDN and GDP are supported by Australian Research Council, Centre
of Excellence grant for “Translational Photosynthesis” (CE140100015). The authors thank Niall Mangan (University of Washington, USA) for comments on Figure 3, and Murray Badger and Rob Sharwood (Australian National University, Australia) for comments on Table 1.
References


but are apparently unable to generate CO₂ within the carboxysome. *Plant Physiology* **91**, 514-525.


Rolland V, Badger MR, Price GD. 2016. Redirecting the cyanobacterial bicarbonate transporters BicA and SbtA to the chloroplast envelope: soluble and membrane cargos need different
chloroplast targeting signals in plants. *Frontiers in Plant Science* 7


### Tables

Table I. Estimated carboxysome requirements in tobacco leaves in a fully-functional chloroplastic CCM.

<table>
<thead>
<tr>
<th>Rubisco source</th>
<th>$k_{cat}$ ($s^{-1}$)</th>
<th>RuBisCO active sites ($\mu mol$ m$^{-2}$ leaf)</th>
<th>RuBisCO holoenzymes or carboxysomes (m$^{-2}$ leaf)</th>
<th>Energy cost of peptide synthesis (kJ m$^{-2}$)</th>
<th>% Stromal volume</th>
<th>Nitrogen cost (mg N m$^{-2}$ leaf)</th>
<th>Carbon cost (mg m$^{-2}$ leaf)</th>
<th>Protein mass (mg m$^{-2}$ leaf)</th>
<th>Protein cost/saving (mg m$^{-2}$ leaf)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT <em>N. tabacum</em></td>
<td>3.2</td>
<td>25</td>
<td>$1.88 \times 10^{18}$</td>
<td>2.62</td>
<td>13</td>
<td>288</td>
<td>915</td>
<td>1691</td>
<td>-90</td>
</tr>
<tr>
<td>$\alpha$-carboxysome</td>
<td>9.0</td>
<td>25</td>
<td>$1.24 \times 10^{16}$</td>
<td>3.97</td>
<td>22</td>
<td>468</td>
<td>1440</td>
<td>2681</td>
<td>-1160</td>
</tr>
<tr>
<td>$\beta$-carboxysome</td>
<td>14.4</td>
<td>25</td>
<td>$1.68 \times 10^{15}$</td>
<td>4.07</td>
<td>21</td>
<td>492</td>
<td>1497</td>
<td>2851</td>
<td>-1160</td>
</tr>
</tbody>
</table>

To achieve the same RuBisCO active site density as WT tobacco (25 $\mu mol$ m$^{-2}$)

<table>
<thead>
<tr>
<th>Rubisco source</th>
<th>$k_{cat}$ ($s^{-1}$)</th>
<th>RuBisCO active sites ($\mu mol$ m$^{-2}$ leaf)</th>
<th>RuBisCO holoenzymes or carboxysomes (m$^{-2}$ leaf)</th>
<th>Energy cost of peptide synthesis (kJ m$^{-2}$)</th>
<th>% Stromal volume</th>
<th>Nitrogen cost (mg N m$^{-2}$ leaf)</th>
<th>Carbon cost (mg m$^{-2}$ leaf)</th>
<th>Protein mass (mg m$^{-2}$ leaf)</th>
<th>Protein cost/saving (mg m$^{-2}$ leaf)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT <em>N. tabacum</em></td>
<td>3.2</td>
<td>25</td>
<td>$1.88 \times 10^{18}$</td>
<td>2.62</td>
<td>13</td>
<td>288</td>
<td>915</td>
<td>1691</td>
<td>-90</td>
</tr>
<tr>
<td>$\alpha$-carboxysome</td>
<td>9.0</td>
<td>8.9</td>
<td>$4.40 \times 10^{15}$</td>
<td>1.41</td>
<td>7.8</td>
<td>166</td>
<td>512</td>
<td>953</td>
<td>738</td>
</tr>
<tr>
<td>$\beta$-carboxysome</td>
<td>14.4</td>
<td>5.6</td>
<td>$3.73 \times 10^{14}$</td>
<td>0.91</td>
<td>4.6</td>
<td>109</td>
<td>333</td>
<td>634</td>
<td>1057</td>
</tr>
</tbody>
</table>

To achieve the same theoretical *in vitro* maximum carboxylation rates as WT tobacco ($V_{c,max}$, 80 $\mu mol$ m$^{-2}$ s$^{-1}$)

Tobacco (*Nicotiana tabacum*) RuBisCO $k_{cat}$ (3.2 s$^{-1}$) and active site density (25 $\mu mol$ m$^{-2}$) values are taken from (Sharwood, 2017) and are used to calculate a theoretical *in vitro* maximum carboxylation rate ($V_{c,max}$) of 80 $\mu mol$ m$^{-2}$ s$^{-1}$. Values for $\alpha$- and $\beta$-carboxysome RuBisCO $k_{cat}$ are those of *Cyanobium* sp. PCC7001 and *Synechococcus elongatus* PCC7942, respectively (Whitehead et al., 2014). Carboxysome numbers are calculated in terms of RuBisCO active site numbers per m$^2$ leaf area for carboxysomes of 90 nm diameter [representing *Prochlorococcus marinus* MED4 $\alpha$-carboxysomes for which protein stoichiometries have been estimated (Roberts et al., 2012)] and 172 nm diameter [representing *S. elongatus* PCC7942 $\beta$-carboxysomes for which protein stoichiometries have been estimated (Long et al., 2011)]. Carboxysomes are assumed to be idealised icosahedrons containing 152 and 1120 RuBisCO holoenzymes in a Kepler packing arrangement for both $\alpha$- and $\beta$-carboxysomes respectively, where the internal and total volumes are calculated allowing for shell thicknesses of 4 nm for $\alpha$-carboxysomes (Iancu et al., 2007) and 5.5 nm for $\beta$-carboxysomes (Kaneko et al., 2006). Protein, carbon and nitrogen masses of each carboxysome type were determined using published stoichiometries and amino acid sequences from both *P. marinus* MED4 and *S. elongatus* PCC7942. Protein masses for $\alpha$- and $\beta$-carboxysomes were determined to be 130.39 MDa and 1021.91 MDa, respectively. The contributions of proteins CcmP and CcmN to $\beta$-carboxysome masses, along with any RuBisCO activation and folding chaperones, have not been included in the calculations due to their unknown contribution, which is likely to be relatively small. The amino acid sequence of *N. tabacum* RbcL and RbcS were used to calculate carbon and nitrogen masses and a molecular mass of 541 kDa. Numbers of amino acids required for the synthesis of tobacco RuBisCO and the entire carboxysome of each type were used to estimate the energy cost of peptide synthesis at a rate of 160 kJ mol$^{-1}$ for each peptide bond. The cost of the tobacco RbcS chloroplast targeting sequence is included in its energy cost. RuBisCO holoenzymes from each species were assumed to occupy a volume of 905 nm$^3$ (based on a spherical diameter of 12 nm). Stromal volumes were calculated from mesophyll surface area per unit leaf area (15 m$^2$ m$^{-2}$) and stromal volume per mesophyll surface area (0.87 x 10$^{-6}$ m$^3$ m$^{-2}$) as used by (Tholen and Zhu, 2011). Protein cost/savings are calculated from total protein masses for carboxysome requirements compared to wild-type (WT) tobacco plants with free RuBisCO and assumes the total replacement of free RuBisCO with carboxysomes. A negative value represents a net cost and a positive value represents a net saving.
Figure Legends

Figure 1. Introduction of key algal and cyanobacterial CCM components into C₃ plant chloroplasts. Algal (A, dark grey) and cyanobacterial (B, light grey) components are involved in three unique features: inorganic carbon (Cᵢ) uptake (HLA3, LCIA, BCT1, BicA and SbtA), RuBisCO aggregation into sub-cellular organelles (shown as the pyrenoid and carboxysome from algae and cyanobacteria, respectively) and anti-leakage systems (LCIB/LCIC, NDH-13 and NDH-14) to minimise loss of CO₂. Putative regulators of HLA3 (LCI1 and SbtA (SbtB) activity are indicated. Together, components of the cyanobacterial and algal CCMs could constitute a functional synthetic biophysical CCM. We intentionally present components of the algal and cyanobacterial CCMs together as constituents of both systems could be used in concert to achieve the intended goal of enhanced CO₂ concentration at the site of carbon fixation. For detailed recent reviews of these CCMs see Wang et al. (2015), Long et al. (2016), Meyer et al. (2016), Rae et al. (2013a), and Rae et al. (2013b).

Figure 2. Transformation strategies for engineering a biophysical CCM into higher plant chloroplasts. Stable nuclear- and chloroplast-genome transformation are both feasible strategies for introducing CCM components. Chloroplast transformation (right) is advantageous for expression of stromal CCM components and gene operons (e.g. Lin et al., 2014b), but is less predictable for delivery to the envelope (dashed line) and thylakoid membrane (dotted line) (Pengelly et al., 2014). Robust chloroplast transformation techniques are currently established in only a very limited number of plant species (Bock, 2014). Nuclear transformation (left) requires that CCM components be modified to carry an appropriate chloroplast transit signal sequence. The latter approach has successfully delivered CCM components to the stroma and inner envelope membrane (IEM) (Atkinson et al., 2016; Atkinson et al., 2017; Rolland et al., 2016; Uehara et al., 2016). Multigene assembly tools offer the opportunity to deliver several CCM components to different locations following a single transformation step (Patron, 2016). A disadvantage of nuclear transformation is the current limited understanding of the transit signal sequences required for correct orientation of transmembrane CCM components (e.g. bicarbonate transporters) in the IEM.

Figure 3. Concentration of inorganic carbon in the carboxysome and pyrenoid RuBisCO-microcompartments. Carbon fixation proceeds by “outside-in” and “inside-out” concentrating mechanisms in carboxysomes and pyrenoids, respectively. Bicarbonate (HCO₃⁻) diffuses into the cyanobacterial carboxysome and conversion to CO₂ proceeds by the activity of an internal carbonic anhydrase (CA). High internal CO₂ concentrations are maintained by a selectively-permeable protein shell (dotted line, left). In the algal pyrenoid, bicarbonate diffuses down a concentration gradient within the lumen of thylakoid membranes traversing the pyrenoid. Lumenal CA activity is localised within the pyrenoid. Here, bicarbonate is rapidly converted to CO₂, which diffuses into the
surrounding RuBisCO aggregate. A starch sheath localised around the pyrenoid (dotted line, right) may function as a barrier to maintain high internal CO₂ concentrations.

**Figure 4.** Sub-structural organisation of the cyanobacterial α- and β-carboxysomes, and the pyrenoid from Chlamydomonas. In both α- and β-carboxysomes (left), pore structures on the outer shell barrier facilitate bicarbonate (HCO₃⁻) uptake. Bicarbonate is converted to CO₂ by one of several potential carbonic anhydrases [CsoSCA (CA) in the case of α-carboxysomes and CcaA (A) or CcmM (M) in the case of β-carboxysomes], probably located at the subsurface of the each carboxysome shell. In α-carboxysomes Form-1A RuBisCO (filled grey circles) is anchored internally via the highly disordered CsoS2 protein (black line), whereas in β-carboxysomes Form-1B RuBisCOs are interlinked via full-length form of CcmM to interact with the internal surface of the shell, and via a short isoform of CcmM to generate a regular, ordered array of enzymes (only a portion of this internal structure is shown). Carboxysome figures adapted from Long *et al.* (2011) and Rae *et al.* (2013a; 2013b). The Chlamydomonas pyrenoid (right) consists of a dense semi-organised protein aggregation, primarily of RuBisCO, traversed by a network of specialised thylakoid membranes (minitubules not shown here), and surrounded by a starch sheath. Bicarbonate is taken up into the thylakoid lumen and converted into CO₂ by a lumenal α-type carbonic anhydrase (CAH3) localised within the pyrenoid. EPYC1 is abundant in the pyrenoid when the CCM is active, and is shown as a linker chaperone that stabilises RuBisCO within the pyrenoid structure. In Chlamydomonas, EPYC1 has four nearly identical repeat regions (shown in dark grey and light grey), and each repeat is suggested as a site for RuBisCO binding. Within each repeat, a less-disordered domain containing a predicted α-helix may interact with the two outward facing α-helices of the small subunit of RuBisCO (SSU).