Targeted expression of nuclear transgenes in *Chlamydomonas reinhardtii* with a versatile, modular vector toolkit

Author names and affiliations: Kyle J. Lauersen, Olaf Kruse and Jan H. Mussgnug*

Bielefeld University, Faculty of Biology, Center for Biotechnology (CeBiTec), Universitätsstrasse 27, 33615, Bielefeld, Germany.

***Corresponding Author:** Jan H. Mussgnug, jan.mussgnug@uni-bielefeld.de

Present/Permanent address: Bielefeld University, Faculty of Biology, Center for Biotechnology (CeBiTec), Universitätsstrasse 27, 33615 Bielefeld, Germany. Phone: +49 521 106-12260, Fax: +49 521 106-12290

Abstract

We present a versatile vector toolkit for nuclear transgene expression in the model green microalga *Chlamydomonas reinhardtii*. The vector was designed in a modular fashion which allows quick replacement of regulatory elements and genes of interest. The current toolkit comprises two antibiotic resistance markers (paromomycin and hygromycin B), five codon optimized light-emission reporters, including the *Gaussia princeps* luciferase, as well as bright cyan, green, yellow, and red fluorescent protein variants. The system has demonstrated robust functional flexibility with signal options to target the protein of interest to the cytoplasm, the nucleus, cellular microbodies, the chloroplast, mitochondria, or via the endoplasmic reticulum-Golgi apparatus secretory pathway into the culture medium. Successful fluorescent reporter protein fusion to *C. reinhardtii* Rubisco small subunit 1 was accomplished with this system. Localization of the fluorescently tagged protein was observed in the chloroplast pyrenoid via live-cell fluorescence microscopy, the first report of heterologous protein localization to this cellular structure. The functionalities of the vector toolkit, the individual modular elements, as well as several combinations thereof are demonstrated in this manuscript. Due to its strategic design, this vector system can quickly be adapted to individual tasks, and should, therefore, be of great use to address specific scientific questions requiring nuclear recombinant protein expression in *C. reinhardtii*.

Keywords:

Microalga, *Chlamydomonas reinhardtii,* recombinant proteins, fluorescent proteins, luciferase, protein secretion

Introduction

Microalgae are a heterogeneous group of organisms, including photosynthetic pro- and eukaryotes which are sources of biotechnologically relevant products (Hallmann 2007; Wijffels et al. 2013). Growth and inexpensive cultivation with sunlight energy in simple mineral salt solutions make microalgae attractive as sustainable sources of a vast array of natural bio-products with many application potentials. Ease of handling and relatively rapid turnaround time from gene of interest to scale-up compared to other photosynthetic systems are additional advantages of these unicellular production organisms (León-Bañares et al. 2004). Microalgal biotechnology has seen a boom in interest over the last 20 years, especially with regards to the potential of microalgal biomass as third generation biofuel feedstock (Stephens et al. 2010). Currently, microalgae are used as protein rich feed for agri- and aquaculture, sources of pigments and antioxidants for cosmetics and food, sources of oils with variable fatty acid compositions (reviewed in (Hallmann 2007; Draaisma et al. 2013; Wijffels et al. 2013)), and, more recently, as hosts for recombinant proteins (RP) for molecular farming and proposed as hosts for metabolic engineering (Shao and Bock 2008; Eichler-Stahlberg et al. 2009; Neupert et al. 2009; Cordero et al. 2011; Lauersen et al. 2013b; Rasala et al. 2014).

Transgenes can be expressed in microalgae from nuclear, chloroplast, or mitochondrial genomes (Kindle 1990; Bateman and Purton 2000; Remacle et al. 2006) and since the host cells can combine prokaryotic and eukaryotic expression features, they offer some unique possibilities for the production of RPs, especially when their functions depend on specific post-translational modifications or complex protein folding (Rasala and Mayfield 2014). Genetic manipulation has been discussed for several decades as the means to achieve some of the ambitious goals of microalgal technologies, including the competitive production of lipids for sustainable solar-based liquid fuels, or the synthesis of complex chemicals which do not occur naturally, by metabolic engineering (Vieler et al. 2012; Bogen et al. 2013; Liu et al. 2013).

Although several algal species have been transformed, including but not limited to *Nannochloropsis sp.* (Kilian et al. 2011; Radakovits et al. 2012; Vieler et al. 2012), *Phaeodactylum tricornutum* (Apt et al. 1996; Miyahara et al. 2014), and *Dunaliella salina* (Feng et al. 2014a; Feng et al. 2014b), the alga with the most highly developed molecular toolkit is the unicellular Chlorophyte *Chlamydomonas reinhardtii*. This

model microalga has served as a valuable tool for the study of photosynthesis for over 60 years, and has recently developed as a platform to model recombinant algal technologies (Schwarz et al. 2007; Shao and Bock 2008; Eichler-Stahlberg et al. 2009; Neupert et al. 2009; Rasala et al. 2010; Chen and Melis 2013; Lauersen et al. 2013b; Rasala and Mayfield 2014). It is well documented that RP production from *C. reinhardtii* chloroplasts can be economically viable, a feature owing to RP titers of several percent total soluble protein (TSP) and reliable targeted genetic manipulation by homologous recombination (Bateman and Purton 2000; Rasala et al. 2011; Chen and Melis 2013). However, this strategy of gene expression is limited to products which can accumulate as soluble protein within the stroma of the plastid, and which do not require post-translational modifications other than disulphide bridging (Rasala and Mayfield 2014).

Expression of RPs from the nuclear genome, by contrast, results in significantly lower titers of RP, but allows for an array of alternative applications for the recombinant product, including specific subcellular targeting, RP secretion, post-translational modifications including glycosylation, and plasma membrane localization. The capacity for subcellular targeting in the eukaryotic cell provides the possibility of accessing specialized substrates within cellular compartments, a valuable trait for metabolic engineering and *in vivo* bio-product manipulation.

Transformation of the nuclear genome of *C. reinhardtii* was first demonstrated in 1989 (Kindle et al. 1989; Kindle 1990), and for many years was limited to auxotrophy complementation with genomic fragments for phenotype complementation (Rochaix 1995). Various antibiotic selection markers have since been demonstrated as effective for use in nuclear transformation, including the aminoglycoside (3') phosphotransferases *aphVII* and *aphVIII* from *Streptomyces hygroscopicus* and *S. rimosus* (hygromycin B and paromomycin resistance, respectively), as well as the bleomycin-antibiotic-family binding *Sh ble* from *Streptoalloteichus hindustanus* (Lumbreras et al. 1998; Sizova et al. 2001; Berthold et al. 2002). These selection markers were used directly from their bacterial origins without sequence modification, owing to their high GC contents that match the third position GC bias of the *C. reinhardtii* nuclear genome.

Heterologous reporters have also developed considerably in the last ten years, owing in a large part to the rapid decrease of cost and ease of gene synthesis technologies, a requirement to adapt codon usage of target sequences. Luciferases from *Renilla*

reniformis (sea pansy) and *Gaussia princeps* (a mesopelagic copepod) have been codon optimized and successfully expressed from the nuclear genome of *C. reinhardtii* (Fuhrmann et al. 2004; Ruecker et al. 2008; Shao and Bock 2008). In addition to the green fluorescent protein (GFP) (Fuhrmann et al. 1999), a rapid expansion of the spectral palette of available reporters has recently been achieved (Rasala et al. 2013). Robust fluorescent reporter proteins provide many advantages for analysis of subcellular targeting peptides and *in vivo* protein localization through fusion to protein targets. Codon optimized GFP (CrGFP or cGFP) was demonstrated to accumulate in the nucleus as a *Sh ble*-fusion (Fuhrmann et al. 1999; Rasala et al. 2012), and fusion of this reporter with either cNAPL or RAA4 demonstrated chloroplast localization (Glanz et al. 2006; Glanz et al. 2012). In a recent publication, mCerulean and mCherry reporter proteins were successfully targeted to the chloroplast, nucleus, endoplasmic reticulum, or mitochondria (Rasala et al. 2014). In addition, a different report discussed the targeting of the green fluorescent protein to peroxisome-like microbody structures in *C. reinhardtii* (Hayashi and Shinozaki 2012).

A large amount of data now exists surrounding the regulation of nuclear expressed transgenes, for example, the *Sh ble* gene and *Renilla* luciferase were both used to demonstrate the enhancer element effect of introns in heterologous sequences on protein expression levels (Lumbreras et al. 1998; Eichler-Stahlberg et al. 2009). In addition, several endogenous promoters and synthetic fusion promoters have demonstrated reliable gene expression (Sizova et al. 1996; Fischer and Rochaix 2001; Berthold et al. 2002; Neupert et al. 2009). Very recently, successful expression of multiple proteins from a single open reading frame was accomplished via inclusion of viral 2A peptides (Rasala et al. 2012). However, most previous molecular tools were exclusively constructed for specific nuclear target genes and the vector designs did not incorporate options for efficient adaptation, allowing more versatile use.

Here, we describe the development of the novel pOptimized vector system. This system was designed *in silico* and built completely *de novo* to be a modular, standardized, and expandable vector toolkit for general use in microalgal research with *C. reinhardtii*. The system employs a modular genetic element approach for nuclear gene expression, codon optimized variants of four fluorescent proteins, one luciferase, and modified versions of two antibiotic resistance markers are presented as the foundation of this expandable system.

Materials and Methods

Cultivation conditions and *Chlamydomonas reinhardtii* strains

C. reinhardtii UVM4 cultures (graciously provided by Prof. Dr. Ralph Bock) were routinely grown in TAP media with 150 µE light intensity in shake flasks or on TAP agar plates. UVM4 is a ultraviolet light derived mutant of CC-4350 (cw15 arg7-8 mt+ [Matagne 302]) which was co-transformed with the emetine resistance cassette CRY1 and ARG7, and demonstrated expression of nuclear transgenes with high efficiency (Neupert et al. 2009). CC-4350 is available from the Chlamydomonas Resource Center (http:// chlamycollection.org). Transformations were performed with glass bead agitation as previously described (Kindle 1990), with only 6 h recovery of cells in liquid TAP rather than 18 h (Lumbreras et al. 1998). Positive transformants were recovered on TAP agar plates containing respective antibiotics at 10 mg L^{-1} with 150 µE light intensity, and maintained on TAP agar by colony stamping. Routine cultivation of transformants was conducted in liquid TAP media in shake flasks at light intensities indicated above.

Plasmid construction, codon optimization, promoter analysis, and antibiotic resistance genes

The pOptimized vector (pOpt) *Chlamydomonas* expression unit was designed *in silico* and built *de novo* by oligonucleotide annealing gene synthesis into the pBluescript II KS (+) *Escherichia coli* vector (Genscript, USA). The hygromycin B resistance gene (*aphVII*, NCBI accession: CAF31839.1) was modified from that of pHyg4 (Berthold et al. 2002), to remove the redundant restriction sites *Eco*RI (GA(A/G)TTC), *Bsi*WI (C(G/C)TACG), as well as *Aat*II (GACGT(C/G)), and synthesized as part of the first pOptimized vector (pOpt_gLuc_Hyg) between 5' *Hin*dIII and 3' *Xho*I sites. This vector was synthesized containing the synthetic *Gaussia princeps* luciferase (sequence modifications described below) within a second gene expression cassette.

The paromomycin resistance gene (*aphVIII*) region was amplified using vector pJR38 as template (Neupert et al. 2009). First, the region containing the expression cassette from the HSP70A promoter until the TGA of *aphVIII* was amplified from pJR38 and cloned into the pOpt_vector backbone from *Mlu*I to *Xho*I to demonstrate that the downstream region was not important for antibiotic resistance activity. Second, the region directly after the *Msc*I site of pJR38 until the *aphVIII* stop codon was amplified

with primers containing *Hin*dIII-*Xho*I sites and cloned into these sites to replace the hygromycin B resistance gene of pOpt_gLuc_Hyg to build the vector pOpt_gLuc_Paro. These two vectors served as the templates for all variant vectors described in this work.

The amino acid (aa) sequences for synthetic FP variants mCerulean3 (cyan, variant of PDB: 4EN1_A as described in (Markwardt et al. 2011)), mVenus (yellow, (Kremers et al. 2006), NCBI Accession: AAZ65844.1), mRuby2 and Clover (red and green variants, AFR60232.1 and AFR60231.1, respectively (Lam et al. 2012)), and *g*Luc were codon optimized as previously described (Verhaegen and Christopoulos 2002; Ruecker et al. 2008; Lauersen et al. 2013a). All optimized sequences were modified with synonymous codons to remove any redundant restriction sites within the pOptimized vector concept (Table S1). The 329 bp intron 2 (i2) of the Rubisco small subunit 2 (RBCS2) gene of *C. reinhardtii* (NCBI Accession: X04472.1) (Eichler-Stahlberg et al. 2009) was integrated into each reporter to maintain a similar CG/…/GT integration site near the middle of each sequence. Where necessary, synonymous codons were used to modify the sequences to create this motif. Each reporter was synthesized with 5' *Nde*I, *Bgl*II, and *Aat*II restriction sites as well as 3' *Eco*RV and *Eco*RI sites and cloned into either pOpt_gLuc_Hyg or pOpt_gLuc_Paro between *Nde*I and *Eco*RI to create the template vectors.

The four fluorescent reporter sequences above were also codon optimized as single genes for expression in *E. coli* (Genscript, USA, Sequences and NCBI accession numbers provided in Supplemental Data file), and cloned into the pET24a(+) expression vector (Novagen) between *Nde*I and *Xho*I sites, followed by transformation into chemically competent *E. coli* KRX cells (Promega). *E. coli* expressed recombinant proteins were induced overnight with 1 mM IPTG in lysogeny broth (LB media), followed by freeze-thaw lysis, and Strep-Tactin chromatography routinely following manufacturer's protocols (IBA Life Sciences).

Targeting of reporters for secretion from *C. reinhardtii* was achieved by 5' fusion of the *c*CA secretion signal between *Nde*I and *Bgl*II within each pOptimized vector as previously described (Lauersen et al. 2013a; Lauersen et al. 2013b). Confirmation of microbody targeting was achieved by 3' addition of the three PTS1-like sequences, presented in Table S2, between the *Eco*RV and *Eco*RI restriction sites of the pOpt_mVenus_Paro vector. The minimal targeting sequence of the Simian Virus 40 nuclear localization sequence (SV40 NLS) was built by oligonucleotide annealing of

the respective DNA sequence, also presented in Table S2. Construction of both the 36 aa *C. reinhardtii* photosystem I reaction center subunit II (PsaD) and 46 aa mitochondrial ATP synthase subunit A (AtpA) N-terminal targeting peptide DNA sequences was achieved by annealing equimolar ratios of the four primers for each construct, listed in Table S2, in polymerase chain reactions (PCR) followed by restriction digest and ligation between *Nde*I and *Bgl*II sites within the pOpt_mVenus_Paro vector.

Localization of the ribulose-1,5-bisphosphate carboxylase small-subunit 1 (RBCS1) with mVenus fluorescence tagging was achieved by *de novo* synthesis of the RBCS1 mRNA sequence (Genscript, USA) (Goldschmidt-Clermont and Rahire 1986; Genkov et al. 2010), followed by cloning of the sequence between *Nde*I and *Bgl*II in the pOpt_mVenus_Paro vector, resulting in a C-terminal fusion of mVenus to the RBCS1 protein sequence. Truncations of RBCS1 N-terminal region were performed by PCR amplification of the 5' region of the RBCS1 sequence with the primers listed in Table S2.

All cloning in this work was performed with Fermentas Fastdigest restriction enzymes following manufacturer's protocols, alkaline phosphatase and ligation reactions were performed with the Rapid DNA Dephos & Ligation Kit (Roche). All PCRs were performed with Q5 High Fidelity polymerase with GC enhancer solution (New England Biolabs) following manufacturer's protocols and cloning was conducted as described above using primers listed in Table S2. After each cloning step, vector sequences were confirmed by sequencing (Sequencing Core Facility, CeBiTec, Bielefeld University, Germany).

The template vector sequences and feature annotation for pOpt_X_Paro and pOpt_X_Hyg, where X is any of the five reporters described above, have been deposited to NCBI (Access. No. can be found in Table S3). Reporter vectors, including *c*CA secretion variants, for both antibiotic resistance markers, as well as the four *E. coli* reporter expression strains, have been made available at the Chlamydomonas Resource Center (http:// chlamycollection.org), including supporting vector sequence information. All vectors for *C. reinhardtii* transformation created in this work are listed in Table S4, variants not deposited to the Resource Center can be generated with the oligonucleotides listed in Table S2 as described above.

Bioluminescence

Bioluminescence was analyzed as previously described for mutant identification at plate level (Lauersen et al. 2013a). Recombinant Gaussia luciferase was detected in Western or dot-blot analysis with an anti-Gaussia luciferase antibody (New England Biolabs, MA, USA) using secreted *g*Luc standard from *Kluyveromyces lactis* as previously described (Lauersen et al. 2013a).

Fluorescent protein analysis

Positive fluorescent protein (FP) expression was detected from plate level *C. reinhardtii* cultures using a Leica-Binocular MZFLIII system equipped with filters for CFP, GFP, YFP, and DsRED to detect mCerulean3, Clover, mVenus, and mRuby2, respectively. Fluorescence imaging was carried out using a confocal laser-scanning microscope with respective filter sets and excitation and emission wavelengths previously described for each reporter (Table S5) (LSM780, Carl Zeiss GmbH, Germany). mCerulean3 was detected by excitation with 458 nm laser and emission reading between 460-490 nm, Clover was detected by excitation with 488 nm laser and detection between 500-530 nm, mVenus was detected with excitation at 514 nm and emission between 520-550 nm, mRuby2 was detected with excitation at 561 nm and emission between 590-620 nm. Chlorophyll fluorescence was observed independently with stimulation at 488 nm and emission from 650-700 nm. Live cells were allowed to settle in microtiter plates, and 8 µl was pipetted onto a glass slide prior to microscopy analysis.

Supernatants containing fluorescent proteins were harvested from stationary phase cultures by centrifugation for 3 min at 3000xg in 50 mL Falcon tubes, followed by 0.2 µm micro filtration, and concentration using 10 kDa MWCO centrifugal filter units (Millipore) to a final concentration of 10X for each FP. Concentrated media was visualized through Leica-Binocular filters and captured with a digital camera through the auxiliary viewing lens of the microscope. Proteins were analyzed by either Western- or dot-blot analysis using a GFP Rabbit IgG Antibody Fraction HRP Conjugate (Life Technologies, Germany). Green fluorescent protein (GFP) purified from *E. coli* was used as a standard (graciously provided by Prof. Dr. Karl Friehs).

Results

Vector design, codon optimization, and expression of recombinant reporters in *C. reinhardtii*

The general vector design is depicted in Fig. 1. In order to develop a standardized, modular, and easily adjustable gene expression system for *C. reinhardtii*, two separate expression cassettes were included in the vector, one for the expression of an antibiotic resistance marker and the other for the expression of the gene of interest (GOI). The heat shock 70A-Rubisco small subunit 2 fusion promoter with Rubisco small subunit intron 1 (collectively HSP70A-RBCS2-i1) and RBCS2 3' untranslated region (3' UTR) were chosen as regulatory elements for both cassettes, because these elements resulted in efficient transgene expression in previous work (Lauersen et al. 2013a). To achieve an overall modular design, each regulatory element was separated by a unique restriction enzyme recognition sequence (Table S1). In addition, a multiple cloning site for future modifications was added between the two gene expression cassettes.

Two antibiotic resistance genes, *aphVII* and *aphVIII*, were used in this study, and both conferred growth of positive transformants on medium containing the antibiotics hygromycin B (10 mg L^{-1}) or paromomycin (10 mg L^{-1}), respectively (Fig. 2a,c). For paromomycin resistance, we determined that a DNA fragment containing only a 839 bp region up to the TGA stop codon, shorter than the 904 bp *aphVIII* fragment described before (Sizova et al. 2001), was sufficient to confer antibiotic resistance (Fig. 2b,c). The *aphVII* gene for hygromycin B resistance was modified slightly in its nucleotide sequence with synonymous codons to remove unwanted restriction enzyme recognition sequences (Fig. S1), which did not result in a loss of antibiotic resistance activity for selection of positive transformants. Both constructs resulted in transformation efficiencies of ~$1x10^3$ colonies per µg vector for either resistance cassette under optimal transformation conditions.

Since the expression cassette for the GOI was separated from the antibiotic resistance marker, antibiotic resistance does not necessarily guarantee efficient expression of the GOI. Light emitting reporter genes (*g*Luc, mCerulean3, Clover, mVenus, and mRuby2) were therefore cloned into the GOI expression cassette after codon optimization and removal of interfering restriction enzyme cut sites to allow fast evaluation of GOI transgene expression efficiencies. Again, a modular design was

chosen and complementary restriction sites were placed on the 5' and 3' ends so that the GOI reporter can be easily replaced in the expression vector backbone and used in fusion protein concepts (Fig. 2d). Each reporter protein was constructed to contain the second intron of RBCS2 (i2) in the middle of its DNA sequence as it has been previously shown that introns from this gene in their physiological order result in increased transgene expression (Eichler-Stahlberg et al. 2009). As a result, each reporter exhibited expression that was detectable at the plate level (Fig. 3), with single gene products detected at the protein level by Western blot (Fig. S2a-c), indicating no alternative splicing or protein degradation. Rates of false positive transformants were determined in three independent transformations of vectors pOpt_mVenus_Hyg and pOpt_mVenus_Paro. For both vectors it was determined that 30% ± 5% of the antibiotic resistant transformants showed detectable levels of YFP fluorescence on agar plates.

Reporter activities were detectable in colonies at the plate level (Fig. 3a,b,c). In case the 21 aa carbonic anhydrase 1 secretion signal (*c*CA, Fig. 2e) was cloned N-terminal to the reporter proteins, these were efficiently secreted, and were also detectable at the plate level (Fig. 3a,c(vi-ix)) or in culture supernatant (Fig. 3d). Here, secreted reporter proteins accumulated up to ~0.7 mg L^{-1} in standard cultivation conditions (Fig. S3a,b). The unique spectral properties of each reporter protein were visible under fluorescent protein filters in 10X concentrated media samples without further affinity-based purification being necessary (Fig. 3d(i-v)). In case no targeting signal was used, the recombinant reporter proteins were also strongly expressed and accumulated in the cytoplasm to up to 0.25% of TSP (Fig. 3b,c(ii-v), Fig. S3c). For all fluorescence reporter constructs, it was possible to visualize the respective intracellular, cytoplasmic signals by laser scanning confocal microscopy with respective excitation and emission wavelengths (Fig. 4a-d). However, for intracellular expression strains, the mVenus signal proved to be superior to the other markers tested, since this reporter exhibits minimal spectral interference with the chlorophyll pigment background fluorescence as well as a bright signal in confocal microscopy (Fig. 3c(i), Fig. 4b). For these reasons, mVenus was chosen as a reporter for subsequent subcellular targeting studies.

Subcellular targeting and fluorescent protein tagging

Expression of the reporters in the cytoplasm presented a pattern of fluorescence which filled the shape of the cell and was clearly distinguishable from the red chlorophyll background fluorescence signal (Fig. 4a-d). Therefore, cytoplasmic localization provided a signal pattern that was used for comparison of constructs targeted to various cellular compartments. In our setup, the parental strain demonstrated typical red chlorophyll fluorescence background signal but no emission signal in the yellow range (Fig. 5a). The chlorophyll background fluorescence was used to orient the cells (red signal in all confocal microscopy pictures), and the mVenus fluorescence signal was clearly distinguishable in all cases (Fig. 5b-j). Expressed mVenus reporter protein, without targeting signals or fusions, accumulated in the cell cytosol (Fig. 5b). We then used mVenus fusion constructs to successfully target the reporter to the ER/secretion pathway (Fig. 5c), the chloroplast pyrenoid structure (Fig. 5d), intracellular microbodies (with three different signals, Fig. 5e-g), the nucleus (Fig. 5h), the chloroplast (Fig. 5i), and mitochondria (Fig. 5j). The corresponding individual emission channels for each construct are presented in Fig. S4a-j.

For microbody targeting, we tested three separate potential signal peptides with our mVenus containing pOptimized plasmid. These included the C-terminal, 10 aa, proposed type 1 peroxisomal targeting sequence from pumpkin malate synthase (PkPTS1, IHHPRELSRL*, vector Fig. 2f), the C-terminal, 10 aa sequence from *C. reinhardtii* malate synthase (MSPTS1, HIVTKTPSRM*, vector Fig. 2g), as well as a novel, seven aa sequence. The latter is shorter than the malate synthase sequences, but contains similar amino acids in a randomly shuffled sequence order (rPTS1, SWRRSLI*, vector Fig. 2h). Hayashi and Shinozaki (2012) demonstrated that C-terminal addition of these PTS1 sequences to GFP resulted in accumulation of this reporter in small spots within the cell. We were able to confirm these results, since the mVenus reporter was detected in cellular compartments resembling spherical microbodies in confocal laser scanning microscopy analyses (Fig. 5e,f). This localization pattern was also true for the novel seven aa sequence, indicating functional exchangeability of amino acids (Fig. 5g). The signals were located throughout the cell close to the chloroplast and were determined to be not located in mitochondria by co-staining with mitotracker and 3-dimensional modeling of Z-stack confocal microscope imaging (Fig. S5a,b and Movie S1).

The monopartite simian virus 40 large T-antigen nuclear localization signal (SV40 NLS) was first described in 1984 as a sequence responsible for protein import into eukaryotic nuclei (Kalderon et al. 1984). In order to test the greater applicability of reporter localization via peptide transport signals, we cloned nucleotides coding for the 7 aa SV40 NLS (PKKKRKV*, vector Fig. 2i) onto the C-terminus of the mVenus reporter and expressed this construct from the nucleus of *C. reinhardtii*. Colonies exhibiting yellow fluorescence at plate level were analyzed by confocal microscopy to elucidate the subcellular localization of fluorescence signals. SV40 NLS-tagged mVenus accumulated in the nucleus of *C. reinhardtii* cells (Fig. 5h) with fluorescence patterns resembling those recently described from similar mCerulean constructs containing a 2X repeat of this sequence (Rasala et al. 2014), indicating successful nuclear localization of the reporter protein.

To demonstrate the versatility of the pOptimized vector system for full-length protein-FP tagging and *in vivo* localization, we chose the Rubisco small subunit 1 (RBCS1), which had previously been shown to be amenable to cDNA based expression, modification with plant homologues, and intron rearrangements while maintaining proper expression (Genkov et al. 2010). The cDNA sequence of RBCS1 was cloned 5' to the mVenus reporter, so that the fluorescent protein was fused to the C-terminus of the target protein (Fig. 2l). In this orientation, intron 1 (from the fusion promoter) and intron 2 (in the mVenus reporter) of RBCS2 surround the sequence of RBCS1. It was expected that fluorescence signals would accumulate in the chloroplast of transformant cells. Interestingly, the fluorescence signal was localized at the pyrenoid structure of the cells within the chloroplast (Fig. 5d), which potentially indicated that the fusion protein was incorporated into the Rubisco multi-subunit enzyme complex. The full-length fusion protein product was detected at the predicted molecular mass of ~51 kDa by Western blot against the mVenus protein (Fig. S6a).

In order to investigate whether the plastid targeting amino acid sequence of RBCS1 could be used for plastid localization of reporter proteins on its own, constructs containing the 45 aa N-terminal targeting peptide, as well as a series of four 10 aa extensions into the RBCS1 sequence were analyzed for their ability to transport mVenus to the chloroplast (vectors Fig. 2m-q). For construct T1 (truncation 1), with no additional amino acids past the cleavage peptide, mVenus fluorescence could be detected in the chloroplast (Fig. S6b) as was observed with the PsaD targeting peptide (Fig. 5i, vector Fig. 2j). However, T2-5, which contained 10-40 aa downstream of the

RBCS1 cleavage peptide, resulted in highly unstable fluorescence and poor imaging during confocal analyses as well as the detection of only degradation products of the mVenus in Western blotting analyses (Fig. S6b and not shown). Fluorescence signals from chloroplast localized mVenus were difficult to image as the signals were spread throughout the stroma, exhibiting a horseshoe shaped pattern of diffuse fluorescence signal around the darker cytoplasmic space (Fig. 5i, 6d, S6b). It was therefore difficult to focus on one cellular structure to obtain a crisp image in comparison to the strong signal obtained from mitochondrial localized AtpA_mVenus (Fig. 5j, vector Fig. 2k).

The pOptimized system presents the option of multiple construct transformations using two separate proteins in GOI expression cassettes of vectors containing either paromomycin or hygromycin B resistance cassettes. For each of the subcellular localized constructs, we also transformed these into cytosolic mCerulean3 expressing strains generated with hygromycin B selection using the pOpt_mCerulean3_Hyg vector. As shown in Fig. 6, the cytoplasmic signals of the mCerulean3 reporter contrasted clearly with the mVenus signals, which were directed to cytoplasm, microbodies, pyrenoid, or chloroplast, respectively (Fig. 6a-d). For the SV40 NLS tagged mVenus construct, we transformed this into a strain previously transformed with the pOpt_cCA_mCerulean3_Hyg vector, resulting in a faint mCerulean3 signal observed from the secretory pathway radiating outwards from the mVenus signal localized in the nucleus (Fig. 6e). Individual imaging layers and combined channel overlays for these constructs as well as comparisons of control signals derived from the parental strain are presented in Fig. S7a-c, Fig. S8a,b, Fig. S9a,b, and Fig. S10a,b.

To obtain pure recombinant reporter proteins for signal comparisons, each fluorescent reporter was also expressed in *E. coli* and purified using standard Strep-Tactin chromatography. Each reporter demonstrated appropriate spectral properties (Fig. S11a) and migrated in SDS PAGE gels to appropriate positions related to their predicted molecular masses (Fig. S11b).

Discussion

Heterologous gene expression systems for eukaryotic microalgae are important tools for recombinant protein production and are also the prerequisite to adapt cellular pathways for metabolic engineering. However, even for the most established model microalgal species, *C. reinhardtii*, the availability of efficient and versatile vector systems for nuclear transgene expression is still very limited.

Previously, vector pSI103 was constructed by combining the gene expression cassette of the *Sh ble* resistance gene (Lumbreras et al. 1998), containing the RBCS2 promoter, with the HSP70A promoter, RBCS2 intron 1 (HSP70A-RBCS2-i1), and the *aphVIII* gene to confer resistance to paromomycin (Sizova et al. 2001). This antibiotic resistance cassette was used as the basis of subsequent vector pJR38 (Neupert et al. 2009), which expanded expression capabilities by using the *Chlamydomonas* PsaD promoter and 3' UTR to express a codon optimized green fluorescent protein. pJR38 was then used as a backbone to create the vector p*c*CA*g*Luc*Lp*IBP (Lauersen et al. 2013a; Lauersen et al. 2013b) which was used to express a Gaussia luciferase-*Lolium perenne* ice binding protein fusion targeted for secretion from *C. reinhardtii*.

Given the success of the HSP70A-RBCS2-i1 fusion promoter for the reliable expression of the *g*Luc-*Lp*IBP fusion in our laboratory, we chose to expand on the concept of the synthetically derived p*c*CA*g*Luc*Lp*IBP vector (Lauersen et al. 2013a) to establish a platform for standardized vector design and future nuclear gene expression studies from *Chlamydomonas*. The heterogeneity of available vectors for nuclear gene expression currently published for *C. reinhardtii* is the result of independent development of expression systems in multiple research groups over the past decades of research with this species. Our goal with the pOptimized vector system was to create a platform for standardization and expansion of gene expression possibilities, while making the system accessible for general research of this model green alga. The pOptimized vector allows multiple cloning of any element through digestion with strategically designed restriction enzyme recognition sequences (Fig. 1).

In order to achieve a system which entirely lacked redundant restriction sites, the gene of interest and antibiotic expression cassettes were re-designed *in silico*. In addition, a series of reporters, including a luciferase, four fluorescent proteins, and two antibiotic resistance markers were also built *de novo* to fit into this vector concept (Fig. 1, 2).

Each reporter was tested for reliable expression from vectors conferring resistance to either antibiotic and demonstrated multiple subcellular localizations (Fig. 3-6).

Although the work presented here has been conducted with the mutant strain UVM4, which is generally prone to high recombinant gene expression (Neupert et al. 2009), we have observed the reporter constructs to be reliable in other *C. reinhardtii* strains as well, including CC 406, and CC 1883 (not shown).

As this vector system is expandable, we suggest that future synthetic reporter sequences could be designed to fit within the concept, avoiding redundant restriction sites in order to allow the greatest flexibility. We also propose a nomenclature structure for variants of the pOptimized vectors, that each element different from the templates provided be separated by an underscore. For example, the paromomycin resistant cytosolic mVenus and secreted (*c*CA tagged) mVenus vectors are given the names pOpt_mVenus_Paro and pOpt_cCA_mVenus_Paro, respectively, see Table S4 for other examples.

The availability of protein standards is important for quantification of recombinant protein production. A recombinant secreted *g*Luc is available commercially, however, no commercially available standards for the specific fluorescent variants are available. Therefore, we also constructed *E. coli* codon optimized variants of these four reporters, adapted to contain the additional amino acids present from the restriction sites of the pOptimized open reading frame as well as the StrepII tag for ease of isolation and use as reliable standards for future investigations (Fig. S11). It should be noted that the 3' UTR of the GOI expression cassette of the pOptimized vectors contain a sequence for an optional StrepII tag addition, which may function as a Western blotting target when specific antibodies are not available. We chose these fluorescent reporters as bright, monomeric, photostable representatives of each class of FP which can also function as FRET pairs for future investigations using this vector concept (Kremers et al. 2006; Markwardt et al. 2011; Lam et al. 2012).

The bright nature of the reporter proteins allowed practical screening using fluorescence or bioluminescence signals at the plate level (Fig. 3a-c). Although there is a strong variability in transformant reporter expression throughout a mutant population (e.g. Fig. 3a,b), the transformants with the strongest protein expression could be easily identified at plate level, avoiding tedious scale up and blotting. Cytosolic expression of mVenus was robust with accumulation up to ~0.25% TSP (Fig. S3c) indicating the reliability of the vector concept for reporter protein

expression. For constructs with subcellular localization of mVenus, colonies could also be readily screened at plate level (not shown), owing to a lack of background fluorescence from chlorophyll at respective excitation and emission wavelengths (Fig. 3c).

It has been recently described that nuclear, chloroplast, and mitochondrial localization of fluorescent reporters can be reliably accomplished with a 2X repeat of the SV40 NLS, the PsaD N-terminal peptide, and AtpA N-terminal targeting peptide respectively (Rasala et al. 2014). Here, we expand on this concept with demonstration of microbody targeting with two known PTS1-like sequences, confirming recently published results (Hayashi and Shinozaki 2012) as well as presenting a novel, smaller, synthetic PTS1-like signal with similar targeting activity (Fig. 5e-g, and 6b). The reliability of targeting to these structures with multiple signals will likely aid in future proteomic analyses of these microbodies and their constituent components. It is unclear at this time what potential biotechnological relevance these cellular structures have, however, as isolated subcellular compartments, they may present interesting applications for target protein sequestration.

A single copy of the SV40 NLS is able to reliably target mVenus to the nucleus in our analyses (Fig. 5h). Since this sequence must be exposed for its proper recognition (Kalderon et al. 1984), it is understandable why a 2X repeat may be of value for some protein expression targets, as has been recently demonstrated (Rasala et al. 2014).

We demonstrate reliable tagging of the Rubisco small subunit 1 with mVenus and that this protein fusion accumulates in the pyrenoid (Fig. 5d, 6c). To our knowledge, this is the first demonstration of reporter protein localization in the pyrenoid of *C. reinhardtii*. Although the 45 aa N-terminal targeting peptide of RBCS1 could localize mVenus into the chloroplast stroma, extensions of the RBCS1 N-terminal plastid targeting sequence resulted in unstable mVenus, both the fluorescent signal and protein quantity as analyzed by Western blot were unusable for analysis from these transformants (Fig. S6b and not shown). This instability likely indicates that further RBCS1 sequence elements were required for functional assembly with the larger Rubisco complex and subsequent pyrenoid localization. It is likely that these fusions were recognized as misfolded RBCS1 and degraded within the stroma. It was proposed by Genkov et al. (2010) that *C. reinhardtii* Rubisco subunits may contain an as-of-yet unknown pyrenoid localization signal, as hybrid Rubisco small subunits containing plant homologues did not result in pyrenoid formation in complemented

photosynthetic *C. reinhardtii* mutants, these mutants also exhibited reduced photosynthetic growth (Genkov et al. 2010). It is still unclear whether the RBCS1 contains such a localization signal or, rather, its incorporation into the larger Rubisco complex and secondary transport is responsible for this localization. Further truncation experiments and reporter-Rubisco fusions with the pOptimized vector will likely contribute to elucidating this unique feature. Expression of the cDNA sequence alone for RBCS1 was likely assisted by the RBCS2 introns 1 and 2 surrounding its sequence in the pOpt_mVenus_Paro vector, demonstrating the value of the use of RBCS2i2 within the reporter sequences.

Each subcellular localized construct was contrasted with a separately localized reporter protein by a second transformation with pOpt_mCerulean3_Hyg or *c*CA variant vector (Fig. 6). The availability of a secondary antibiotic selection marker allows for a second GOI to be transformed and independently screened for expression from a strain already known to express another reporter. Since reporters expressed from either antibiotic resistance vector, resulted in ~30% of colonies with fluorescence signals detectable at the plate level, we found that secondary transformation, rather than co-transformation, allowed practical pre-screening of strong reporter expression, followed by a subsequent screen for secondary reporter expression. It has been recently published that the incorporation of viral 2A self-cleaving peptides allows multiple gene expression from a single GOI expression cassette (Rasala et al. 2012; Rasala et al. 2014). Recovery on media containing a member of the bleomycin antibiotic family resulted in selection for those transformants which express high titers of target GOI. This strategy is effective for expression of target proteins which may be otherwise difficult to detect, as surviving colonies likely express the target protein. It has also been shown that multiple gene traits can be combined through mating of heterologous marker expressing strains to yield progeny expressing up to four separately targeted reporters (Rasala et al. 2014). The pOptimized vector concept naturally complements these expression strategies and presents multiple new options for combined GOI expression. In this work, multiple gene traits were combined without mating, rather by double transformation and selection mediated by separate antibiotic resistance markers. The ability to use multiple different antibiotic selection methods, as well as separate markers for gene expression, expands the possibilities of complex metabolic engineering from *C. reinhardtii*. Indeed, 2A peptide-bleomycin resistance fusions, devoid of redundant

restriction sites, will likely be among the first new genetic elements designed for the pOptimized system.

In summary, with this work we introduce a new and versatile vector toolkit which can be of great use for projects involving heterologous gene expression in the model green microalga *C. reinhardtii*. All reporter sequences and basic vectors have been deposited at the publicly accessible Chlamydomonas Resource Center (http://chlamycollection.org/) and, therefore, will be freely available for future research projects.

Acknowledgements

The authors would like to acknowledge the CLIB Graduate Cluster Industrial Biotechnology (Federal Ministry of Science & Technology North Rhine Westphalia, Germany (to K.J.L.)), The authors would also like to express thanks to Martina Lummer, Prof. Dr. Thorsten Seidel, Prof. Dr. Karsten Niehaus, Dr. Darius Widera for assistance with fluorescence protein analysis and microscopy, Prof. Dr. Ralph Bock for strain UVM4, as well as to Jan Schwarzhans and Prof. Dr. Karl Friehs for providing the GFP standards. The authors declare that they have no conflict of interest.

Conflict of Interest

The Authors declare that they have no conflict of interest.

References

Apt KE, Kroth-Pancic PG, Grossman AR (1996) Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. Mol Gen Genet 252(5):572–579

Bateman JM, Purton S (2000) Tools for chloroplast transformation in *Chlamydomonas*: expression vectors and a new dominant selectable marker. Mol Genet Genomics 263:404–410

Berthold P, Schmitt R, Mages W (2002) An engineered *Streptomyces hygroscopicus* aph 7″ gene mediates dominant resistance against hygromycin B in *Chlamydomonas reinhardtii*. Protist 153(4):401–412

Bogen C, Klassen V, Wichmann J, La Russa M, Doebbe A, Grundmann M, Uronen P, Kruse O, Mussgnug JH (2013) Identification of *Monoraphidium contortum* as a promising species for liquid biofuel production. Bioresour Technol 133:622–626

Chen H-C, Melis A (2013) Marker-free genetic engineering of the chloroplast in the green microalga *Chlamydomonas reinhardtii*. Plant Biotechnol J 11(7):818–828

Cordero BF, Couso I, León R, Rodríguez H, Vargas MA (2011) Enhancement of carotenoids biosynthesis in *Chlamydomonas reinhardtii* by nuclear transformation using a phytoene synthase gene isolated from *Chlorella zofingiensis*. Appl Microbiol Biotechnol 91:341–51

Draaisma RB, Wijffels RH, Slegers PME, Brentner LB, Roy A, Barbosa MJ (2013) Food commodities from microalgae. Curr Opin Biotechnol 24(2):169–177

Eichler-Stahlberg A, Weisheit W, Ruecker O, Heitzer M (2009) Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. Planta 229(4):873–883

Feng S, Feng W, Zhao L, Gu H, Li Q, Shi K, Guo S, Zhang N (2014a) Preparation of transgenic *Dunaliella salina* for immunization against white spot syndrome virus in crayfish. Arch Virol 159(3):519–525

Feng S, Li X, Xu Z, Qi J (2014b) *Dunaliella salina* as a novel host for the production of recombinant proteins. Appl Microbiol Biotechnol 98(10):4293–300

Fischer N, Rochaix J-D (2001) The flanking regions of PsaD drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. Mol Genet Genomics 265(5):888–894

Fuhrmann M, Hausherr A, Ferbitz L, Schödl T, Heitzer M, Hegemann P (2004) Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. Plant Mol Biol 55(6):869–881

Fuhrmann M, Oertel W, Hegemann P (1999) A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*. Plant J 19(3):353–361

Genkov T, Meyer M, Griffiths H, Spreitzer RJ (2010) Functional hybrid rubisco enzymes with plant small subunits and algal large subunits: engineered rbcS cDNA for expression in *Chlamydomonas*. J Biol Chem 285(26):19833–19841

Glanz S, Bunse A, Wimbert A, Balczun C, Kück U (2006) A nucleosome assembly protein-like polypeptide binds to chloroplast group II intron RNA in *Chlamydomonas reinhardtii*. Nucleic Acids Res 34(18):5337–51

Glanz S, Jacobs J, Kock V, Mishra A, Kück U (2012) Raa4 is a trans-splicing factor that specifically binds chloroplast tscA intron RNA. Plant J 69(3):421–31

Goldschmidt-Clermont M, Rahire M (1986) Sequence, evolution and differential expression of the two genes encoding variant small subunits of ribulose bisphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*. J Mol Biol 191(3):421–432

Hallmann A (2007) Algal transgenics and biotechnology. Transgenic Plant J 1(1):81–98

Hayashi Y, Shinozaki A (2012) Visualization of microbodies in *Chlamydomonas reinhardtii*. J Plant Res 125(4):579–586

Kalderon D, Roberts BL, Richardson WD, Smith AE (1984) A short amino acid sequence able to specify nuclear location. Cell 39(2):499–509

Kilian O, Benemann CSE, Niyogi KK, Vick B (2011) High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. Proc Natl Acad Sci U S A 108:21265–21269

Kindle KL (1990) High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA 87(3):1228–1232

Kindle KL, Schnell RA, Fernández E, Lefebvre PA (1989) Stable nuclear transformation of *Chlamydomonas* using the *Chlamydomonas gene* for nitrate reductase. J Cell Biol 109(6):2589–2601

Kremers G-J, Goedhart J, van Munster EB, Gadella TWJ (2006) Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius. Biochemistry 45(21):6570–6580

Lam AJ, St-Pierre F, Gong Y, Marshall JD, Cranfill PJ, Baird MA, McKeown MR, Wiedenmann J, Davidson MW, Schnitzer MJ, Tsien RY, Lin MZ (2012) Improving FRET dynamic range with bright green and red fluorescent proteins. Nat Methods 9(10):1005–1012

Lauersen KJ, Berger H, Mussgnug JH, Kruse O (2013a) Efficient recombinant protein production and secretion from nuclear transgenes in *Chlamydomonas reinhardtii*. J Biotechnol 167(2):101–110

Lauersen KJ, Vanderveer TL, Berger H, Kaluza I, Mussgnug JH, Walker VK, Kruse O (2013b) Ice recrystallization inhibition mediated by a nuclear-expressed and -secreted recombinant ice-binding protein in the microalga *Chlamydomonas reinhardtii*. Appl Microbiol Biotechnol 97(22):9763–9772

León-Bañares R, González-Ballester D, Galván A, Fernández E (2004) Transgenic microalgae as green cell-factories. Trends Biotechnol 22(1):45–52

Liu B, Vieler A, Li C, Jones a D, Benning C (2013) Triacylglycerol profiling of microalgae *Chlamydomonas reinhardtii* and *Nannochloropsis oceanica*. Bioresour Technol 146:310–316

Lumbreras V, Stevens R. D, Purton S (1998) Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. Plant J 14(4):441–447

Markwardt ML, Kremers G-J, Kraft CA, Ray K, Cranfill PJC, Wilson KA, Day RN, Wachter RM, Davidson MW, Rizzo MA (2011) An improved cerulean fluorescent protein with enhanced brightness and reduced reversible photoswitching. PLoS One 6(3):e17896

Miyahara M, Aoi M, Inoue-Kashino N, Kashino Y, Ifuku K (2014) Highly efficient transformation of the diatom *Phaeodactylum tricornutum* by multi-pulse electroporation. Biosci Biotechnol Biochem 77(4):874–876

Neupert J, Karcher D, Bock R (2009) Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. Plant J 57(6):1140–1150

Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, Posewitz MC (2012) Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*. Nat Commun 3:686

Rasala BA, Barrera DJ, Ng J, Plucinak TM, Rosenberg JN, Weeks DP, Oyler GA, Peterson TC, Haerizadeh F, Mayfield SP (2010) Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. Plant Biotechnol J 8(6):719–733

Rasala BA, Barrera DJ, Ng J, Plucinak TM, Rosenberg JN, Weeks DP, Oyler GA, Peterson TC, Haerizadeh F, Mayfield SP (2013) Expanding the spectral palette of fluorescent proteins for the green microalga *Chlamydomonas reinhardtii*. Plant J 74(4):545–56

Rasala BA, Chao S-S, Pier M, Barrera DJ, Mayfield SP (2014) Enhanced genetic tools for engineering multigene traits into green algae. PLoS One 9(4):e94028

Rasala BA, Lee PA, Shen Z, Briggs SP, Mendez M, Mayfield SP (2012) Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. PLoS One 7(8):e43349

Rasala BA, Mayfield SP (2014) Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. Photosynth Res. doi: 10.1007/s11120-014-9994-7

Rasala BA, Muto M, Sullivan J, Mayfield SP (2011) Improved heterologous protein expression in the chloroplast of *Chlamydomonas reinhardtii* through promoter and 5′ untranslated region optimization. Plant Biotechnol J 9(6):674–683

Remacle C, Cardol P, Coosemans N, Gaisne M, Bonnefoy N (2006) High-efficiency biolistic transformation of *Chlamydomonas* mitochondria can be used to insert mutations in complex I genes. Proc Natl Acad Sci U S A 103(12):4771–4776

Rochaix JD (1995) *Chlamydomonas reinhardtii* as the photosynthetic yeast. Annu Rev Genet 29:209–230

Ruecker O, Zillner K, Groebner-Ferreira R, Heitzer M (2008) Gaussia-luciferase as a sensitive reporter gene for monitoring promoter activity in the nucleus of the green alga *Chlamydomonas reinhardtii*. Mol Genet Genomics 280(2):153–162

Schroda M, Blöcker D, Beck CF (2000) The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. Plant J 21(2):121–131

Schwarz C, Elles I, Kortmann J, Piotrowski M, Nickelsen J (2007) Synthesis of the D2 protein of photosystem II in *Chlamydomonas* is controlled by a high molecular mass complex containing the RNA stabilization factor Nac2 and the translational activator RBP40. Plant Cell 19(11):3627–3639

Shao N, Bock R (2008) A codon-optimized luciferase from *Gaussia princeps* facilitates the in vivo monitoring of gene expression in the model alga *Chlamydomonas reinhardtii*. Curr Genet 53(6):381–388

Sizova I, Fuhrmann M, Hegemann P (2001) A *Streptomyces rimosus aphVIII* gene coding for a new type phosphotransferase provides stable antibiotic resistance to *Chlamydomonas reinhardtii*. Gene 277(1-2):221–229

Sizova IA, Lapina T V., Frolova ON, Alexandrova NN, Akopiants KE, Danilenko VN (1996) Stable nuclear transformation of *Chlamydomonas reinhardtii* with a *Streptomyces rimosus* gene as the selective marker. Gene 181(1-2):13–18

Stephens E, Ross IL, Mussgnug JH, Wagner LD, Borowitzka MA, Posten C, Kruse O, Hankamer B (2010) Future prospects of microalgal biofuel production systems. Trends Plant Sci 15(10):554–564

Verhaegen M, Christopoulos TK (2002) Recombinant Gaussia luciferase. Overexpression, purification, and analytical application of a bioluminescent reporter for DNA hybridization. Anal Chem 74(17):4378–4385

Vieler A, Wu G, Tsai C-H, Bullard B, Cornish AJ, Harvey C, Reca I-B, Thornburg C, Achawanantakun R, Buehl CJ, Campbell MS, Cavalier D, Childs KL, Clark TJ, Deshpande R, Erickson E, Armenia Ferguson A, Handee W, Kong Q, Li X, Liu

B, Lundback S, Peng C, Roston R L, Sanjaya, Simpson JP, Terbush A, Warakanont J, Zäuner S, Farre EM, Hegg EL, Jiang N, Kuo M-H, Lu Y, Niyogi KK, Ohlrogge J, Osteryoung KW, Shachar-Hill Y, Sears BB, Sun Y, Takahashi H, Yandell M, Shiu S-H, Benning C (2012) Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. PLoS Genet 8(11):e1003064

Wijffels RH, Kruse O, Hellingwerf KJ (2013) Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae. Curr Opin Biotechnol 24(3):405–413

Figure Captions

Fig. 1 The pOptimized vector concept, markers, and reporters. Each element of the vector was designed to be surrounded by unique restriction digest recognition sequences, allowing simple removal and opening of the vector backbone for future marker, reporter, or regulatory element investigations. Codon optimized and RBCS2i2 (i2) containing markers were generated to match restriction sites of the vector backbone. Multiple digestion sites on either side of the reporter allow for rapid development of fusion protein targets and targeting peptides. Paromomycin and hygromycin B resistance markers were also developed as single unit modules. H: *C. reinhardtii* heat shock 70A promoter, R: Rubisco small subunit 2 (RBCS2) promoter, i/i2: RBCS2 intron 1 or 2, 3': RBCS2 3' untranslated region, StrepII-TAA: StrepII affinity tag and stop codon (WSHPQFEK*), MCS: multiple cloning site.

Fig. 2 Schematic overview of all vectors used in this study. (a) The marker expression cassette with hygromycin B resistance gene. (b) The initial paromomycin resistance marker cassette, including promoters and 839 bp of *aphVIII* from vector pJR38. The original 67 bp downstream of the TGA stop codon was not required for efficient antibiotic resistance. (c) The pOptimized paromomycin resistance marker was developed by amplification of the upstream region of *aphVIII* (λ : from original *Msc*I restriction insertion site in pJR38 to the start codon) until its TGA stop codon, allowing insertion of this modular genetic element into the pOptimized vector between *Hin*dIII and *Xho*I. (d) Depiction of element orientation of the cytosolic expression reporters. (e) The secretion vectors, maintain the *C. reinhardtii* carbonic anhydrase 1 (*c*CA) secretion signal between *Nde*I and *Bgl*II for efficient secretion of all reporters. (f-h) The microbody targeting vectors, including the 3' addition of the nucleotides encoding for the pumpkin and *C. reinhardtii* malate synthase PTS1 signals (PkPTS1 and MSPTS1), or a smaller, novel PTS1-like sequence (rPTS1), respectively. The signal sequences were cloned between the *Eco*RV and *Eco*RI restriction sites on the marker sequence coding for mVenus. (i) The nuclear localization vector, including one repeat of the SV40 nuclear localization signal. (j,k) N-terminal fusion of the 36 aa targeting peptide of *C. reinhardtii* PsaD and 46 aa targeting peptide of *C. reinhardtii* mitochondrial AtpA to mVenus for chloroplast or mitochondrial targeting, respectively. (l) Fusion construct of RBCS1 cDNA to

mVenus. (m-q) Truncations of various lengths of the RBCS1 N-terminal coding sequence, arrows indicate the cleavage site of the predicted chloroplast targeting peptide. Each construct contains sequences encoding for 10 aa past the previous cleavage peptide (T1-5).

Fig. 3 Plate-level detection of the intracellular and extracellular expression of the five light-emission reporters used in this work. Strains expressing cytoplasmic and secreted reporters (Fig. 2 vector constructs d and e) are shown. (a, b) *g*Luc reporter proteins were detected by bioluminescence assay. Depending on the presence or absence of the *c*CA secretion signal, *g*Luc dependent signals were detected as halos outside of the cells (a) or within the cell colony (b). (c) (i) Fluorescence background signals of parental strain with CFP, YFP, GFP, and DsRed filter settings, respectively. (ii-v) mCerulean3, mVenus, Clover, and mRuby2 expression from respective cytoplasmic pOptimized vectors in transformant colonies under the same filter settings as in (i). (vi-ix) mCerulean3, mVenus, Clover, and mRuby2 expression from respective pOptimized expression and secretion vectors under the same filter settings as in (i). Fusion of with the *c*CA secretion signal resulted in reporter accumulation outside of the cell colonies. (d) Secreted fluorescent protein detection in 10X concentrated culture media from the *c*CA secretion strains (c, vi-ix) without further purification. (i) Transmission light without excitation and emission filters is shown, followed by the same filter settings as in (i) .

Fig. 4 Detection of the intracellular fluorescence of reporter proteins accumulated in the cytoplasm by scanning confocal laser microscopy. Transformants generated from fluorescent reporter constructs described in Fig. 2d are depicted. All reporters, mCerulean3, mVenus, Clover, and mRuby2 (a-d, respectively), demonstrated reliable expression and were clearly visualized with respective excitation and emission wavelengths (see table S5). The pattern of chlorophyll background fluorescence are shown as red signals in the left column, specific fluorescence signals are shown in the middle column as blue, bright yellow, green or orange signals, respectively, and an overlay of the two signals is shown in the right column. Scale bars represent 5 µm.

Fig. 5 Subcellular localization of mVenus reporter with peptide fusions. Chlorophyll background fluorescence is used to orient the signals in the cells. (a) Parental

untransformed strain is imaged with the same laser intensities, demonstrating no background signal in the yellow channel, is presented as a control. (b) Cytosolic mVenus localization (vector construct depicted in Fig. 2d). (c) Remaining mVenus signal from the secretory pathway of *c*CA_mVenus protein secretion strain (vector construct as Fig. 2e). (d) RBCS1_mVenus fusion localized the YFP signal to the chloroplast pyrenoid, (vector construct Fig. 2l). Vectors with pumpkin (e) and *C. reinhardtii* (f) malate synthase PTS1 signals, as well as the novel PTS1-like signal (g) in C-terminal fusion to mVenus exhibited fluorescent signal in spherical microbodies throughout the cell (vector constructs Fig. 2f-h, respectively). (h) The mVenus_SV40 NLS construct (vector Fig. 2i) accumulated in the nucleus of *C. reinhardtii*. (i) 36 aa PsaD N-terminal targeting peptide_mVenus fusion was detected as a diffuse signal within the chloroplast stroma (vector Fig. 2j). (j) The 46 aa N-terminal AtpA targeting peptide_mVenus fusion accumulated in the mitochondrial network and displays reticular, strong fluorescence signals (vector Fig. 2k). Scale bars represent 5 µm.

Fig. 6 Double transformations using separate resistance markers demonstrating expression and clear contrast of separate reporter localizations. Cell lines transformed with pOpt_mCerulean3_Hyg expressing mCerulean3 (cyan) in the cytoplasm were subsequently transformed with mVenus (yellow) reporter constructs. Chlorophyll fluorescence is shown as red background signals. (a) Secondary transformation of mCerulean3 expressing strain with pOpt_mVenus_Paro resulted in simultaneous expression of both reporters in the cytoplasm and overlapping fluorescence signals. (b) Spherical mVenus containing microbodies contrasts against cytosolic mCerulean3 and chlorophyll background signal (mVenus vector construct from Fig. 2h). (c) RBCS1_mVenus pyrenoid localization contrasts against chlorophyll background signal and cytosolic mCerulean3 (mVenus vector construct Fig. 2l). (d) mVenus accumulation in the chloroplast stroma (construct Fig. 2j) contrasts against chlorophyll background fluorescence and mCerulean3 spread throughout the cytoplasm. (e) mVenus_SV40NLS signal (construct Fig. 2i) contrasts against the comparably weak residual pOpt_*c*CA_mCerulean3_Hyg transformed strain and the signal derived from mCerulean3 within the secretory pathway (construct as in Fig. 2e). Scale bars represent 5 µm.

Figure1
Click here to download high resolution image

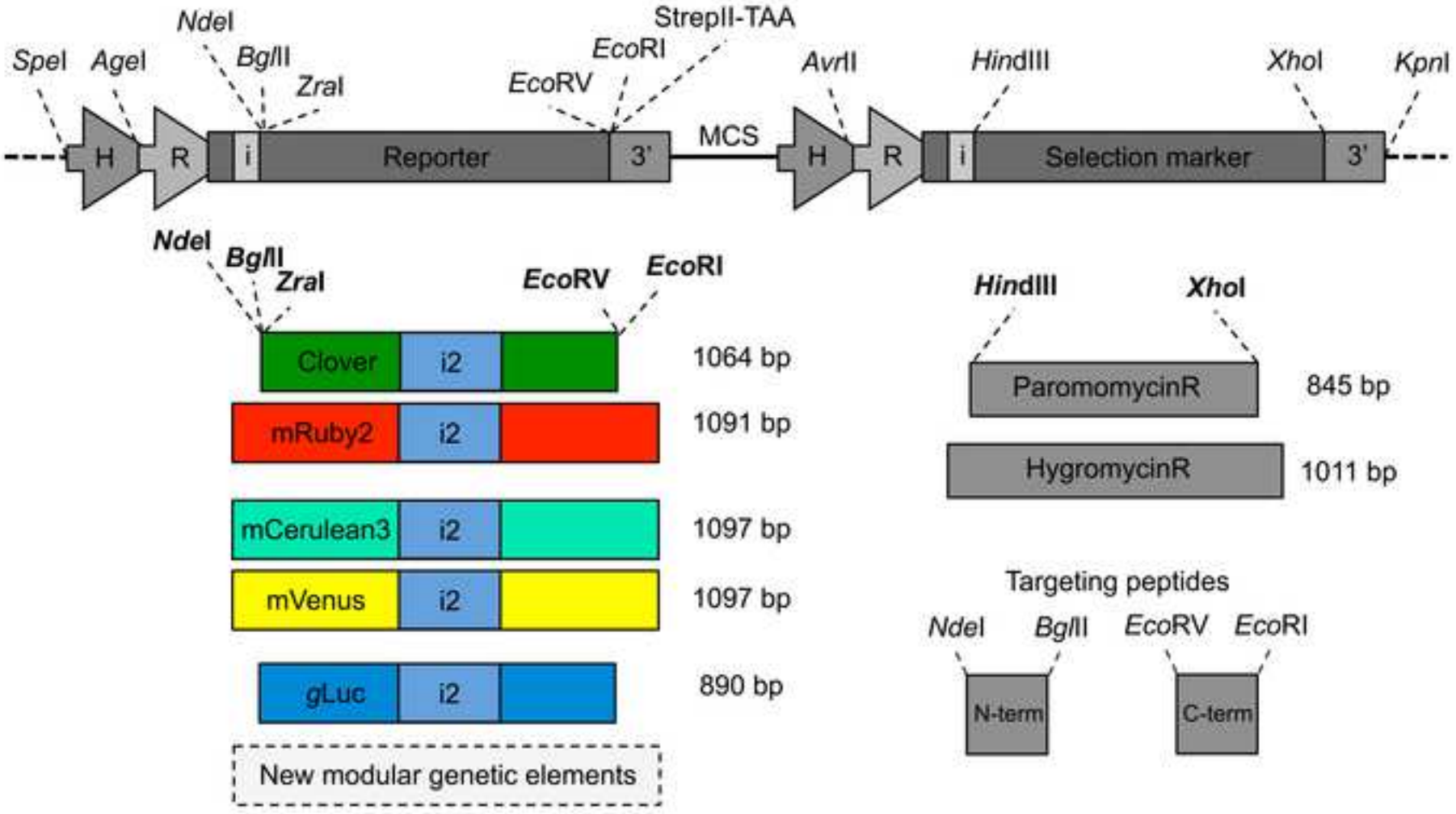


Figure2
Click here to download high resolution image

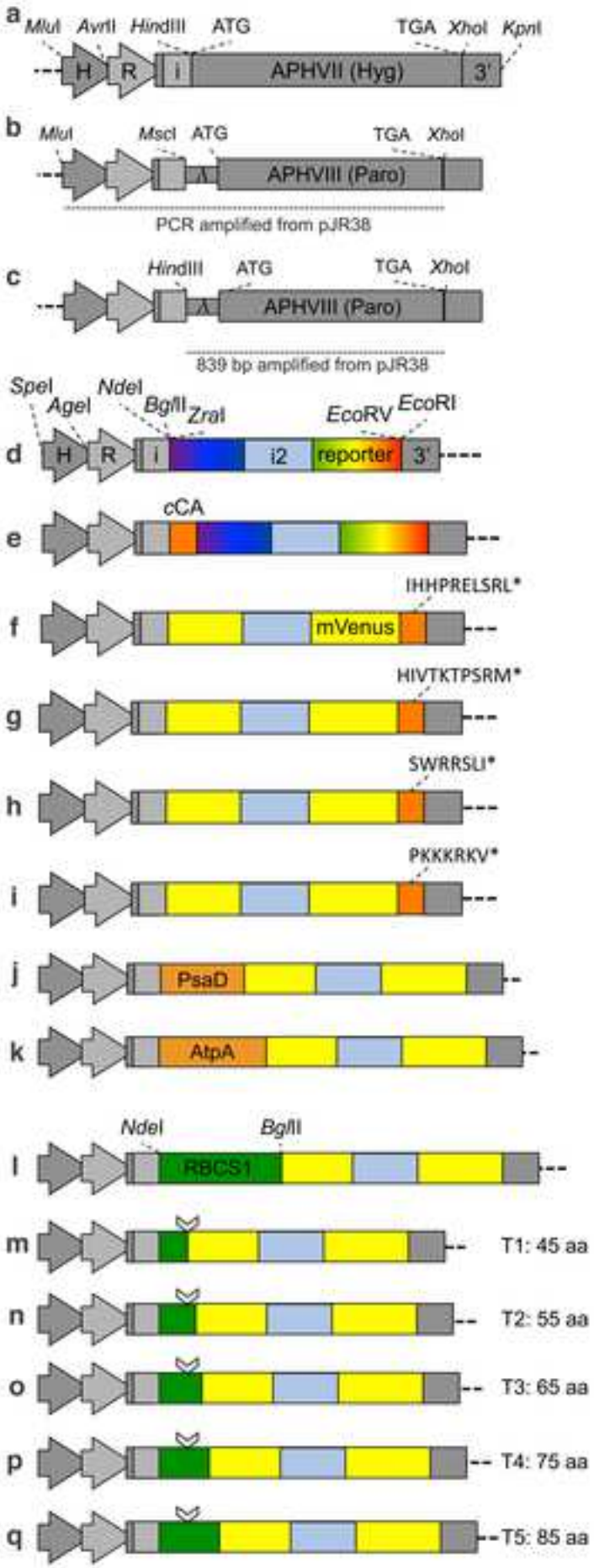


Figure3
Click here to download high resolution image

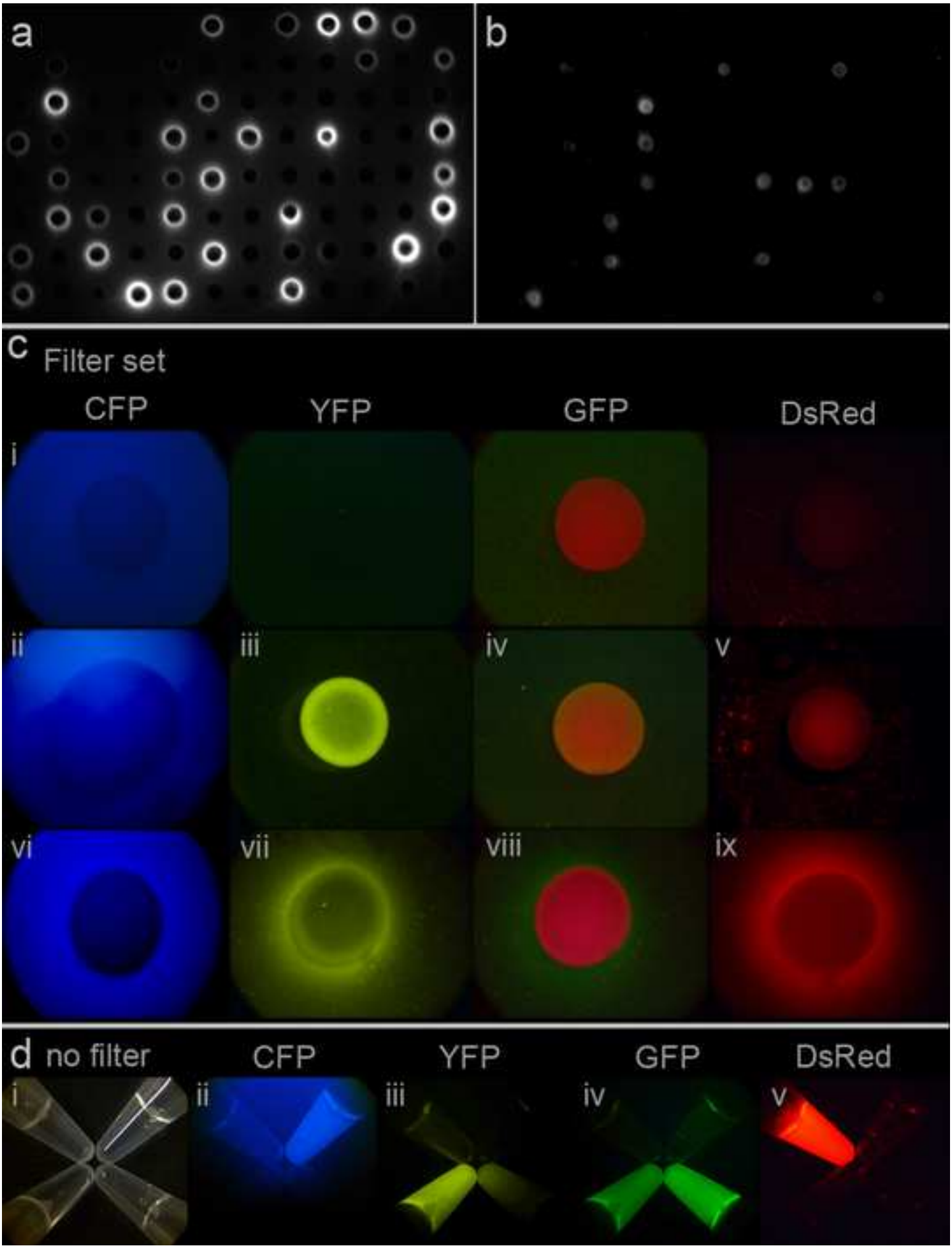


Figure4
Click here to download high resolution image

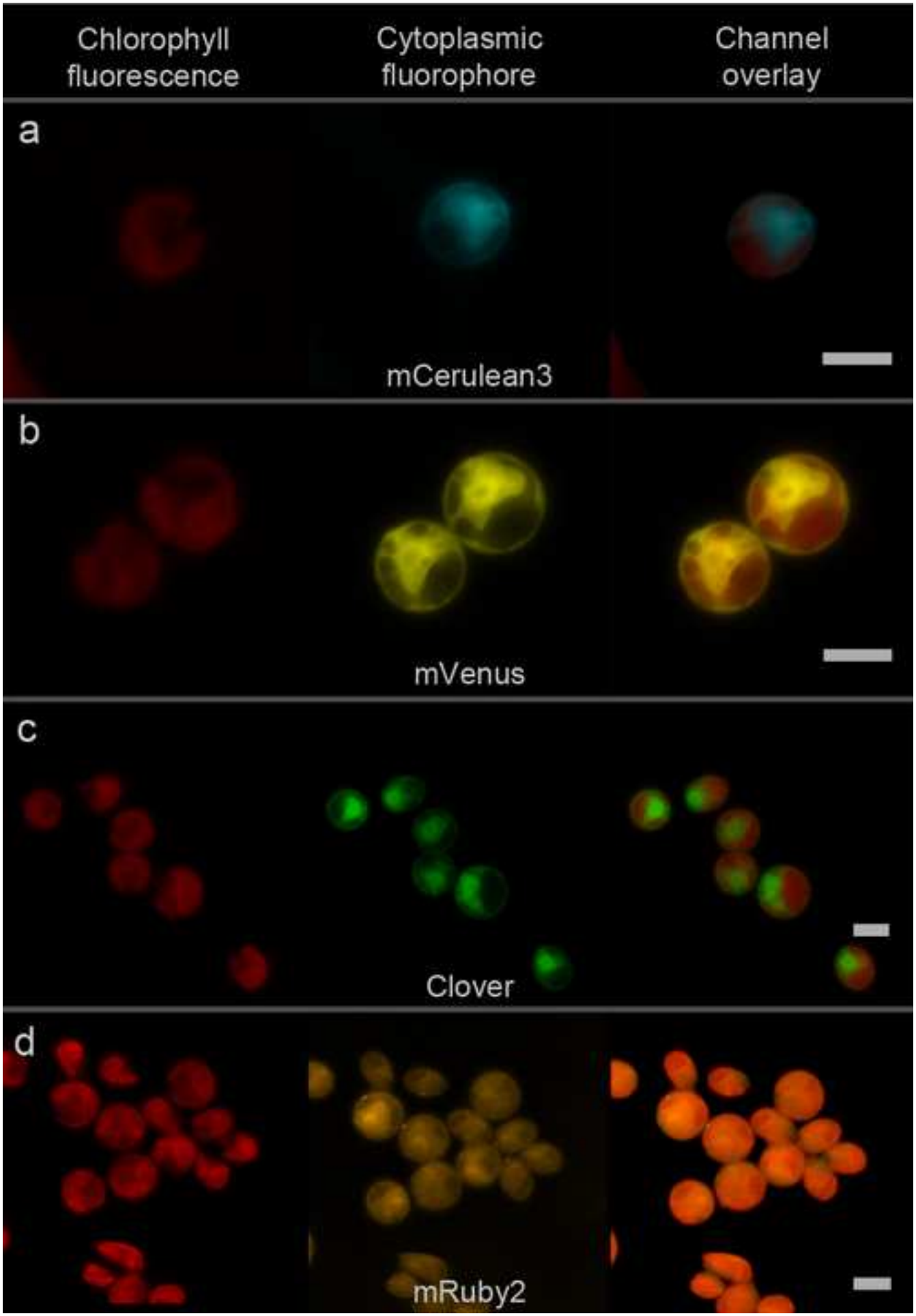


Figure5
Click here to download high resolution image

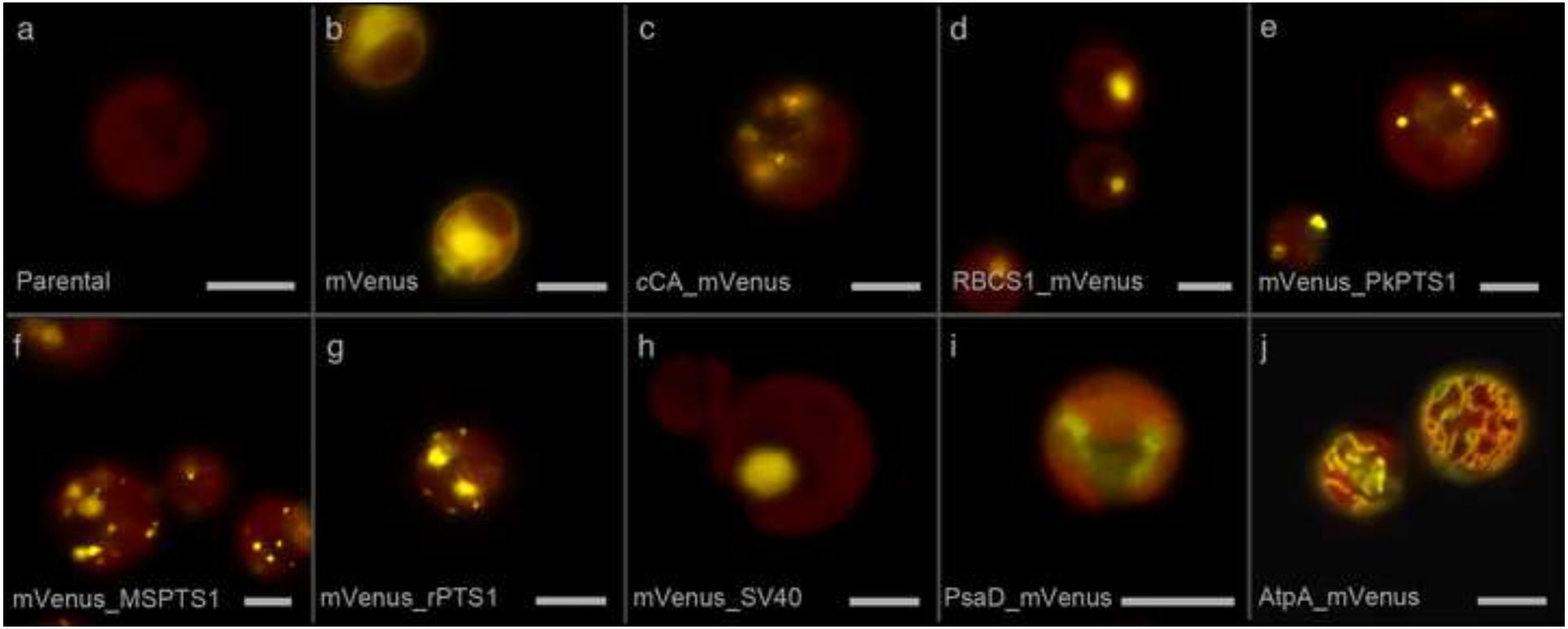
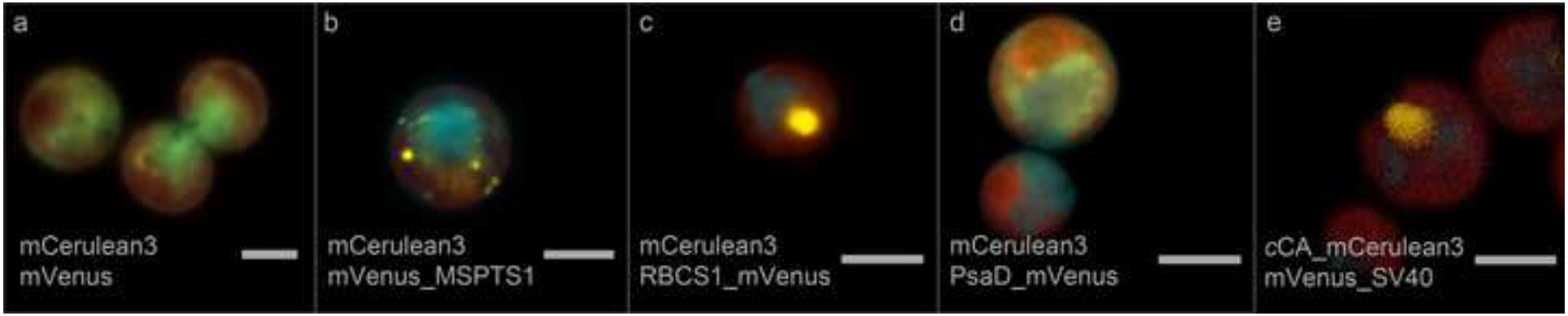


Figure6
Click here to download high resolution image



Supplementary Material

Click here to download Supplementary Material: Online Resource Supporting Information.pdf

Supplementary MovieS1

Click here to download Supplementary Material: Movie S1.mpg