CHAPTER TWENTY-THREE

MODIFICATION OF THE GENOME OF RHODOBACTER SPHAEROIDES AND CONSTRUCTION OF SYNTHETIC OPERONS

Paul R. Jaschke,* Rafael G. Saer,* Stephan Noll,† and J. Thomas Beatty*

Contents
1. Introduction 520
2. Gene Disruption and Deletion 522
   2.1. General scheme 522
   2.2. Construction of the ΔRCLH mutant 526
3. Construction of Synthetic Operons 527
   3.1. Operon objectives and composition 527
   3.2. Host strain 527
   3.3. Expression vector 528
   3.4. Regulation of synthetic operon expression 529
4. Future Directions 532
   4.1. Genome modification 532
   4.2. Synthetic operons 533
References 533

Abstract
The α-proteobacterium Rhodobacter sphaeroides is an exemplary model organism for the creation and study of novel protein expression systems, especially membrane protein complexes that harvest light energy to yield electrical energy. Advantages of this organism include a sequenced genome, tools for genetic engineering, a well-characterized metabolism, and a large membrane surface area when grown under hypoxic or anoxic conditions. This chapter provides a framework for the utilization of R. sphaeroides as a model organism for membrane protein expression, highlighting key advantages and shortcomings. Procedures covered in this chapter include the creation of chromosomal gene deletions, disruptions, and replacements, as well as the construction of a

* Department of Microbiology and Immunology, University of British Columbia, Life Sciences Centre, Vancouver, British Columbia, Canada
† Gene Bridges GmbH, Im Neuenheimer Feld 584, Heidelberg, Germany

ISSN 0076-6879, DOI: 10.1016/B978-0-12-385075-1.00023-8 All rights reserved.
synthetic operon using a model promoter to induce expression of modified photosynthetic reaction center proteins for structural and functional analysis.

1. Introduction

The defining trait of purple photosynthetic bacteria, such as in the genus *Rhodobacter*, is the ability to perform anoxygenic photosynthesis, an ancient form of photosynthesis that does not generate oxygen (Xiong and Bauer, 2002). Both *Rhodobacter sphaeroides* (formerly *Rhodopseudomonas sphaeroides*) and *Rhodobacter capsulatus* have been studied since the 1940s, but herein we focus on *R. sphaeroides*. *R. sphaeroides* is a member of the α-proteobacteria, and descended from bacteria that eventually became mitochondria in eukaryotic cells (Yang et al., 1985). Interestingly, most genome-sequenced strains of *R. sphaeroides* contain two circular chromosomes, as well as a variable number of plasmids (NCBI accession numbers CP000143-CP000147; DQ232586-DQ232587). Other characteristics of *R. sphaeroides* include an aerobic metabolism that functions in O₂ concentrations from atmospheric down to microaerophilic, in addition to anaerobic respiration using substances such as dimethyl sulfoxide (DMSO) as a terminal electron acceptor (Zannoni et al., 2009). *R. sphaeroides* is able to fix N₂ as well as CO₂ (Calvin–Benson–Bassham cycle), although organic compounds are preferred over CO₂ (Madigan, 1995; Romagnoli and Tabita, 2009).

The key environmental signal that controls the bioenergetic properties of *R. sphaeroides* is the concentration of O₂, and cultures shift from aerobic to photosynthetic metabolism in response to a reduction in oxygen tension. During adaptation, the cytoplasmic membrane expands and folds inward to create a highly invaginated intracytoplasmic membrane system, which houses the photosynthetic apparatus (Chory et al., 1984). Photosynthetic membrane protein complexes execute cyclic electron transfer, and pump protons from the cytoplasm to the periplasmic space. This electrochemical gradient is utilized by the cell to generate ATP. In autotrophic growth, electrons may be bled out of the electron transport chain for use in synthesis of NAD(P)H, while electrons enter by oxidation of H₂ (Herter et al., 1997) or reduced sulfur compounds such as H₂S (Brune, 1995).

The core of the photosynthetic apparatus is a dimer of the reaction center (RC) complexes surrounded by the light-harvesting 1 (LH1) complex and the PufX protein (Qian et al., 2005; Scheuring et al., 2004, 2005). The LH1 complex consists of α/β heterodimer subunits that bind two coupled bacteriochlorophyll (BChl) pigments that absorb light (Kohler, 2006). About 24–28 LH1 heterodimers form an S-shaped structure surrounding two RCs (Fig. 23.1A). Each RC contains three proteins called L, M, and H; the structurally similar (33% sequence identity) RC L and M proteins consist largely of five transmembrane helices with pseudo-twofold symmetry, whereas the RC H protein has only one transmembrane helix.
and a large cytoplasmic domain (Fig. 23.1B; Yeates et al., 1988). The overlapping *pufL* and *pufM* genes encoding the RC L and M proteins are located within a 66.7 kb region of chromosome 1 called the photosynthesis gene cluster, or PSGC (Fig. 23.2). The *pufLM* genes are flanked by the *pufBA* genes (encoding LH1 α/β proteins) upstream and *pufX* downstream, all of which are transcribed as a polycistronic mRNA. In contrast, the RC H gene *puhA* is transcribed from an operon 38 kb distant from the *puf* genes, but still within the PSGC (Chen et al., 1998; Donohue et al., 1986). The *pucBAC* operon, a further 22 kb separated from *puhA* (Fig. 23.2), codes for the LH2 complex which acts as a variable-sized antenna, funneling photons toward the RC–LH1 core complex (Gabrielsen et al., 2009).

A large catalog of research has accumulated on *R. sphaeroides*, and many of the basic parameters of metabolism have been examined and quantified, allowing the construction of models of the electron transport chain (Klamt et al., 2008). Further, transcriptomic and proteomic characterization of cells in various growth modes have been published (Arai et al., 2008; Callister et al., 2006; Zeng et al., 2007), along with an understanding of oxygen and redox signal transduction pathways and the mechanism of regulation of some promoters (Bauer et al., 2009; Eraso and Kaplan, 2009; Eraso et al., 2008; Moskvin et al., 2007; Oh and Kaplan, 2000; Roh et al., 2004).

*R. sphaeroides* has garnered considerable interest for biotechnology applications. Recently, a large multidisciplinary project has been initiated to look into biological hydrogen production of this organism (Curtis et al., 2010). Additionally, *R. sphaeroides* has been targeted in a structural genomics initiative.

Figure 23.1 Representation of the *Rhodobacter sphaeroides* photosynthetic reaction center and light-harvesting 1 complex. (A) the RC/LH1/PufX supercomplex dimer: RC, reaction center complex; LH1, light-harvesting 1 complex subunit, an α/β protein heterodimer; PufX, the PufX protein (needed for quinone exchange). (B) The RC artificially in isolation from the LH1 to show the organization of the RC proteins and cofactors: M, the RC M protein; L, the RC L protein; H, the RC H protein; P, the “special pair” or “primary donor” dimer of BChls; B_B and B_A, accessory BChls; H_B and H_A, bacteriopheophytins; Q_B and Q_A, quinones.
Laible et al., 2004, 2009) for use as a high-throughput membrane protein expression system, because of its inducible promoters and extensive and well-characterized membrane system. Additional potential applications include: metal nanoparticle synthesis (Narayanan and Sakthivel, 2010), reduction of odors in large-scale farming waste (Kobayashi, 1995; Schweizer, 2003), heavy metal bioremediation (Italiano et al., 2009; Van Fleet-Stalder et al., 2000), production of plant hormones (Rajasekhar et al., 1999a,b), and photovoltaics (Lebedev et al., 2008; Takshi et al., 2009). Thus, R. sphaeroides has much to offer synthetic biologists who are willing to leave the more familiar model organisms behind. In fact, several teams (Utah State and Washington University) in the 2009 International Genetically Engineered Machine (iGEM) competition used R. sphaeroides as a chassis for their projects.

This chapter is intended as an introduction to a bacterial chassis that has potential to create devices and study phenomena outside of the realm of possibility of the dominant model organisms. We will outline methods to: (1) delete or disrupt R. sphaeroides genes and (2) construct synthetic operons expressed in trans from broad-host range plasmids.

2. Gene Disruption and Deletion

2.1. General scheme

This section outlines how to create a null mutation (knockout) in a R. sphaeroides gene, using strain 2.4.1 as an example. The two circular chromosomes of 3.2 Mb (RefSeq NC_007493) and 0.9 Mb (NC_007494)
in length average 69% G+C. There are also five plasmids ranging from 37 to 114 kb in length. Essential functions are shared by both chromosomes, with the majority of the genes that encode the photosynthetic apparatus found grouped in the PSGC on chromosome 1 (Fig. 23.2; Choudhary and Kaplan, 2000).

Unlike *Escherichia coli*, there is not an efficient transformation or electroporation method for introduction of DNA into *R. sphaeroides*, perhaps because of endogenous restriction enzymes (D. Jun and J. T. Beatty, unpublished). Therefore, the directed genetic manipulation of the genome of *R. sphaeroides* requires the construction of circular gene replacement suicide vectors that encode for the desired gene modification(s).

As outlined in Fig. 23.3, the general scheme for generation of a directed *R. sphaeroides* gene knockout consists of several phases: (1) cloning the gene

---

**Figure 23.3** Generation of knockout in the *R. sphaeroides* genome. (A) Cloning of the *puf*QBALMX operon and flanking sequences in plasmid pUC19. (B) Insertion of an antibiotic-resistance gene (Ω) in place of the *puf*QBALMX genes. (C) Transfer of Ω and *puf* operon flanking sequences (Δ*puf*), to suicide plasmid pNHG1 (*tet*, tetracycline resistance; *sacB*, levansucrase gene). (D) Conjugation of recombinant suicide plasmid and integration into a *puf*QBALMX flanking region by homologous recombination. (E) Resolution of the cointegrate leading to a mixed wild type (I) and mutant (II) cell population.
of interest in *E. coli*, (2) disruption of the plasmid-borne gene sequence by deletion and/or insertion of an antibiotic-resistance marker, (3) conjugation of a disrupted copy of the gene into *R. sphaeroides*, (4) selection and screening for an initial single-recombination, followed by counter-selection for a double-recombination event.

The gene of interest is first amplified from the chromosome of *R. sphaeroides*, using PCR protocols modified for high GC-content DNA. We have found that Platinum Pfx (Invitrogen) and Vent (NEB) work well. DMSO is routinely used at 3–10% (v/v) final concentration in the PCR mixes, to lower melting temperature and reduce secondary structure of the chromosomal DNA template. In our hands, enhancing buffers provided in PCR enzyme kits that are designed for use with high-GC templates do not work as well as DMSO added to the kit’s standard buffer.

The PCR product is subcloned by classical methods into a high copy number *E. coli*-compatible vector (Fig. 23.3A), commonly pUC19 (Messing, 1983). Alternatively the commercially available TA (Zhou *et al.*, 1995) or TOPO cloning systems (Shuman, 1994) may be used. Upon plasmid purification from *E. coli*, the gene of interest is then cut with appropriate restriction enzyme(s) and either ligated to yield an unmarked deletion or, as shown in Fig. 23.3B, ligated with an antibiotic-resistance marker such as the spectinomycin resistance gene on the Ω cartridge (Prentki and Krisch, 1984), or the neo gene on the KIXX cartridge (Barany, 1985). Both resistance markers are functional in *E. coli* and *R. sphaeroides*, which simplifies subsequent selection and screening steps. The disrupted gene should have >0.4 kb of continuous flanking sequence identity to the chromosomal target locus on each end, to allow for efficient homologous recombination into the genome and subsequent recovery of the desired mutant.

Tangentially, we note that disruption of a gene 5′ of other genes in an operon may interfere with transcription of 3′ genes (a polar effect). The chance of a polar effect cannot be eliminated, but it can be reduced best by use of a translationally inframe deletion, or to a lesser likelihood by use of the KIXX cartridge in the same transcriptional orientation as the disrupted gene. However, to rule out polar effect(s), complementation in *trans* by a plasmid borne wild-type gene should be performed in the final *R. sphaeroides* mutant. The Ω cartridge was designed to halt translation and transcription (Prentki and Krisch, 1984), and is almost guaranteed to have a polar effect when inserted 3′ of a single promoter that drives transcription of multiple genes in an operon. The different phenotypic effects of KIXX (usually nonpolar) and Ω cartridge (polar) disruption of genes may be used to experimentally define and dissect operons of uncertain composition, as described for *R. capsulatus* (Aklujkar *et al.*, 2000) and *R. sphaeroides* (Chen *et al.*, 1998).

As shown in Fig. 23.3C, the mutant gene is transferred to an appropriate suicide plasmid that encodes (i) an origin of replication (usually from ColE1
or p15A) that allows maintenance in *E. coli* but not in *R. sphaeroides*, and (ii) an origin of transfer sequence (*oriT*) needed for conjugation from the appropriate *E. coli* strain to *R. sphaeroides*. Examples of suicide vectors in common use are the pLO-series, the pSUP-series, and pNHG1 (Jeffke et al., 1999; Lenz et al., 1994; Simon et al., 1983).

The preferred diparental conjugation method utilizes *E. coli* donor strain S17-1, which contains key genes that facilitate the transfer of *oriT*-containing plasmids (such as the suicide plasmid) into *R. sphaeroides* cells (Simon et al., 1983). Triparental mating is also efficient, utilizing a DH10B or other auxotrophic donor strain harboring the plasmid of interest along with an HB101(pRK2013) helper strain (Ditta et al., 1985).

After transfer to *R. sphaeroides*, the suicide plasmid cannot replicate, and so selection for the antibiotic–resistance marker on the plasmid ensures that cells in colonies that arise on an agar medium containing the appropriate antibiotic have the plasmid integrated into the chromosome by homologous recombination (Fig. 23.3D). Using the methodology described above, the frequency of RecA-dependent homologous recombination is on the order of $10^{-4}$ per potential plasmid recipient for a single event (a crossover on one or the other side of the disrupted gene).

After isolation of strains where a single crossover has occurred, growth in liquid culture *without selection* well into stationary phase (5–10 generations) allows time for a second homologous recombination to occur. As shown in Fig. 23.3E, there are two possibilities for this event: one is that the suicide vector will reform and leave the genome, thereby restoring the state prior to the first recombination (Fig. 23.3E(I)); alternatively, the disrupted copy of the gene of interest may be left in the chromosome, while the suicide plasmid backbone leaves the chromosome with the wild-type copy of the gene (Fig. 23.3E(II)).

In addition to a selectable marker, the backbone of a suicide plasmid may contain a counter-selectable marker, which under appropriate growth conditions, allows for selection of colonies of cells that have undergone plasmid loss. A frequently used system is the *sacB* gene from the Gram-positive *Bacillus subtilis*, which allows for counter-selection by growing cells on an agar medium containing a high concentration of sucrose (Gay et al., 1985). In the presence of sucrose, the *sacB*-encoded levansucrase polymerizes fructose from the degradation of sucrose that, in Gram-negative species, inhibits colony formation (Gay et al., 1983; Steinmetz et al., 1983). Apparently, the native promoter of *sacB* functions in *R. sphaeroides*, but *sacB* has also been put under the control of the *R. sphaeroides* puc promoter in pJE2864 (Eraso and Kaplan, 2002), to improve the efficiency of the selective process.

Thus, by using a *sacB*-containing plasmid, plating recipient cells on an agar medium that contains sucrose (10–15%) results in a great enrichment of cells that have lost the plasmid. In the case where an antibiotic–resistance
marker has been inserted into the gene of interest, the relevant antibiotic is included in the agar medium to inhibit the growth of cells that have retained the native gene. Colonies are screened for the presence of the disrupted or deleted gene, as indicated by a change in the size of PCR product, using primers that flank the gene of interest.

2.2. Construction of the ΔRCLH mutant

The ΔRCLH mutant (Tehrani and Beatty, 2004) serves as a good example of the gene disruption and deletion techniques described above. This mutant contains deletions of the puhA gene (encodes the RC H protein), the pucBA genes (encode the LH2 proteins), and the pufBALMX operon (coding for LH1, RCL, RC M, and PufX proteins). These modifications resulted in a mutant that does not contain any structural protein of the photosynthetic complexes, which was created to serve as a null background in which modified photosynthetic complexes could be expressed (see Section 3).

The pathway to the ΔRCLH mutant began with the creation of a translationally inframe deletion of the puhA gene (Chen et al., 1998). The deletion was obtained using a plasmid-borne copy of puhA in E. coli and “loop-out” oligonucleotide mutagenesis, to replace a 561 bp segment of the coding region with an EcoRV site. This technique removes a section of DNA by using oligonucleotide primers to bridge two separate parts of the gene, causing the intervening sequence to loop-out and be lost upon amplification. The modified puhA gene was then inserted into the suicide vector pSUP203 and conjugated into the R. sphaeroides strain PUH1 (Chen et al., 1998). After selection for tetracycline resistance resulting from a single homologous recombination event, the resultant strain was grown in liquid medium and plated onto solid medium in the absence of selection. Colonies were replica-plated to identify colonies that had lost the tetracycline resistance marker on the suicide plasmid, because the pSUP203 vector lacks the counter-selection marker sacB. Tetracycline sensitive colonies were screened for a decrease in size of the puhA sequence by Southern blot hybridization and a clone was named ΔPUHA (Chen et al., 1998).

The method outlined in Fig. 23.3 was used to delete the pufBALMX operon from the ΔPUHA strain, using the pNHG1::PUFDEL suicide plasmid. This suicide plasmid was constructed in several steps, starting with a modified pUC19 plasmid (pAli2) at the subcloning stage (Tehrani and Beatty, 2004). A 4.6 kb chromosomal DNA fragment containing the puf operon was cloned into pAli2 and modified by replacing the BspEI to BclI sequence (from pufB to pufX, inclusive) with a linker (Tehrani and Beatty, 2004). This markerless deletion was transferred as an EcoRI fragment into the suicide plasmid pNHG1 (Jeffke et al., 1999) to generate pNHG1::
PUFDEL, which was conjugated into *R. sphaeroides* ΔPUHA, followed by selection for tetracycline resistance (integration by homologous recombination), and followed by counter-selection on a sucrose-containing medium, and screening for the desired double-crossover event. To delete the *pucBA* genes in the resultant strain, the construction and deployment of the pNHG1::DELPUC suicide plasmid utilized similar principles (Tehrani and Beatty, 2004).

With the creation of the mutant ΔRCLH, which lacks all photosynthetic complexes, we had a blank slate that allows expression of a wide variety of engineered photosynthetic complexes. We describe below how the ΔRCLH strain was used as a key ingredient in the expression of plasmid-borne synthetic operons of RC genes.

### 3. Construction of Synthetic Operons

The aim of this section is to describe how we initially created and expressed synthetic operons in *R. sphaeroides*. The main principles of design and implementation are similar to principles guiding work on *E. coli*, but several differences between these systems are highlighted.

We first turned to the design of synthetic operons to aid in the study of mutant RC proteins within the native host. Several general considerations that must be kept in mind when designing synthetic expression systems will be explored within the context of the *R. sphaeroides* host system: (1) operon objectives and composition, (2) utilization of a suitable background strain, (3) use or design of an appropriate expression vector, (4) choice of genetic control elements.

#### 3.1. Operon objectives and composition

This example focuses on expressing site-directed mutants of endogenous genes within a synthetic operon in *R. sphaeroides*; see Laible *et al.* (2009) for a review of foreign gene expression in this host. Our general goal was to create a system for expression of variants of the RC genes to further our work on fundamental and applied aspects of RC structure and function (Lin *et al.*, 2009; Takshi *et al.*, 2009).

#### 3.2. Host strain

We chose the ΔRCLH strain (Tehrani and Beatty, 2004) for this purpose, because it contains precise deletions of the genes encoding the photosynthetic complexes, as outlined in Section 2.2. Additionally, because this strain cannot grow photosynthetically without a functional RC complex, the
photosynthetic growth phenotype served as a simple test of RC electron transfer efficiency. Although a deletion was made within the chromosomal *puhA* operon, the expression of downstream chromosomal genes was needed for maximal production of photosynthetic complexes (Aklujkar *et al*., 2005; Chen *et al*., 1998). The determination of the capability for photosynthetic growth and measurement of photosynthetic culture growth rate, coupled with absorption spectroscopy of cells, are rapid and simple ways to evaluate the functional properties of RC variants.

Using the methods described in Section 2, it should be feasible to create many different types of *R. sphaeroides* host strains, depending on the process that is to be engineered. This methodology may be used to deliver novel genes and operons to the genome, as well as creating knockouts.

### 3.3. Expression vector

A well-designed vector backbone can simplify synthetic operon design and facilitate the genetic manipulations necessary for its construction. The approach is to modify a preexisting vector by tailoring of the backbone to the nature of their work. This may include adding or removing restriction sites, and other key sequences.

Plasmids are introduced into *R. sphaeroides* from *E. coli* by conjugation using shuttle vectors that are stably replicated within both organisms. Some examples of broad host-range plasmids currently used in *R. sphaeroides* research include: pRK415, pBBR1, pJRD215, and pATP19P (Davison *et al*., 1987; Keen *et al*., 1988; Kovach *et al*., 1994; Tehrani and Beatty, 2004).

Typically, a synthetic operon would first be created in a small, high-copy *E. coli* vector such as pUC19 (Messing, 1983), and subsequently transferred to a broad host-range plasmid as a cluster of genes on a single DNA fragment. This is because most broad host-range plasmids are large, low copy number, and lack a wide variety of unique restriction sites—therefore, there are practical reasons for why it is easier to create a synthetic operon in an *E. coli* cloning vector before moving the operon into a plasmid capable of replication in *R. sphaeroides*.

To create a plasmid backbone for *R. sphaeroides* mutant RC expression, the hypoxia-inducible *puc* promoter (Lee and Kaplan, 1995) was inserted into the broad host–range plasmid pRK415 as a 0.75 kb HindIII fragment, along with part of the multiple cloning site of pUC19, such that the resultant plasmid pATP19P (Tehrani and Beatty, 2004) now had seven unique restriction sites for insertion of genes downstream of the promoter. A copy of the *puhA* gene was inserted downstream of the *puc* promoter as a 1.3 kb BamHI fragment yielding plasmid pATSHR. The two additional (native or mutant) RC genes, *puL* and *puM*, were added by inserting a 4.5 kb EcoRI fragment that contains the *puFQBALMX* cluster (Tehrani and Beatty, 2004). Another derivative of pATP19P was created by adding a
*puhA* gene modified by the addition of six histidine codons on the 3'-end of the gene (Abresch *et al.*, 2005). This plasmid, p6His-C, was found to yield amounts of the His-tagged RC H protein sufficient for purification of the RC or RC/LH1/PufX complex using Ni/NTA affinity chromatography (Abresch *et al.*, 2005; Jaschke and Beatty, 2007; Lin *et al.*, 2009). The C-terminal 6× His tag was used because an N-terminal 6× His tag disrupted RC formation (unpublished).

The plasmid pATSHR described above was used to express deletions of RC genes in *R. sphaeroides* strain ΔRCLH, to investigate protein–protein interactions and membrane-insertion (Tehrani and Beatty, 2004). The method we have typically used to create the desired RC gene mutants is to first create the desired modification to RC genes in an *E. coli* high copy number plasmid, then to transfer the mutant gene as either a BamHI to SacI fragment (for *puhA* mutants), or a SacI to EcoRI fragment (*puf* mutants).

The synthetic RC expression operon was created using native coding sequences, but several specific modifications were necessary to achieve the desired results. Our first design of the synthetic operon was found to express RC in insufficient quantities to enable photosynthetic growth of the host strain (Fig. 23.4). A search for potential mRNA stem-loop structures indicated a sequence shortly after the 3' end of the *puhA* gene that might attenuate transcription into the downstream *pufQBALMX* genes. Replacement of 83 bp, starting 3 bp downstream of the His-tagged *puhA* stop codon, with a SacI restriction site yielded plasmid pRS1. It was found that *R. sphaeroides* strain ΔRCLH(pRS1) was capable of photosynthetic growth (Fig. 23.4). This synthetic RC expression system using the ΔRCLH host strain and plasmid pRS1 has been used by our group to rapidly create, express, and purify a large number of RCs with modifications of all three subunits.

We also considered several factors prior to the creation of this synthetic operon, including (1) the characteristics of the novel mRNA made from this synthetic operon, and whether it would be resistant to nucleases and allow an appropriate level of translation; (2) whether the RC genes could be expressed in *trans* or whether the assembly of the RC complex required *cis*-active factors not present on our plasmid. The successful generation of fully functional RC/LH1/PufX core complexes from our synthetic operon indicates that no essential information was encoded in the relative genome locations of the RC genes within the PSGC.

### 3.4. Regulation of synthetic operon expression

Regulation is one of the most important features of a synthetic operon, and regardless of whether genes are expressed in the native or heterologous host, cryptic regulatory sequences may be present in the coding or intergenic regions as was seen in the first iteration of our synthetic RC operon (Fig. 23.4A).
Figure 23.4  Design and construction of the synthetic RC/LH1 gene cluster. (A) The puc promoter and puhA gene, including puhA downstream sequences that later were found to attenuate transcription, were inserted into a plasmid pRK415 backbone; the pufQBALMX genes were inserted downstream of puhA sequences, but the expression of RC genes was found to be insufficient for photosynthetic growth of strain ΔRCLH. Sequences immediately following the puhA stop codon were removed, and it was found that RC gene expression was sufficient for photosynthetic growth of ΔRCLH. (B) Map of plasmid pRS1, which contains the final synthetic operon; bent arrow indicates the puc promoter; key unique restriction sites are indicated; Te', plasmid encodes resistance to tetracycline.
Genes within the operon should be coupled to appropriate transcriptional promoters and attenuators to obtain desired levels of mRNA synthesis, and to appropriate ribosome binding sites (RBSs) for desired levels of mRNA translation. The stability of mRNA also affects the level of gene expression, and it is interesting that the *Rhodobacter puf* operon was an early model system in this area (Chen et al., 1988), although *E. coli* has emerged as the prokaryotic model system (Schuck et al., 2009). Codon sequence composition affects translation efficiency, and codon usage varies significantly between species, approximately as a function of genome GC-content (Kane, 1995; Lee et al., 2009).

If construction of a synthetic operon requires isolation of genes of interest from the native host, as opposed to gene synthesis, it may or may not be desirable to include the native regulatory sequences. This decision is colored by how well defined these elements are in the host strain.

Little is known about fundamental properties of *Rhodobacter* transcription promoters, except that they often differ from well-understood *E. coli* promoters in −10 and −35 sequence composition (Leung, 2010; Swem et al., 2001), and hence may not be recognized in a heterologous host. We routinely use the *R. sphaeroides puc* promoter because it is thought to be a strong promoter, and can be regulated by control of culture aeration. A fructose-inducible promoter from *R. capsulatus* was reported to have a high-dynamic range (Duport et al., 1994), and presumably the *R. sphaeroides* homologue would function similarly. Recently, a description was published of the *puc* promoter fused to *lacO* under the control of *lacIq*, so that the promoter is induced under low concentrations of O₂ only when IPTG is present (Hu et al., 2010).

*Rhodobacter rho*-independent transcription terminators appear to be similar to *E. coli* terminators (Chen et al., 1988). Several algorithms to detect transcriptional terminators, such as TransTermHP (Kingsford et al., 2007) (http://transterm.cbc.cbcb.umd.edu/) or RNAfold (Hofacker et al., 1994; McCaskill, 1990; Zuker and Stiegler, 1981) (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) can be used to scan the 3′ end of ORFs to find potential terminators that may be present.

The RBS provides another level of regulation (Nakamoto, 2009). If a gene lacks an obvious RBS, it is possible to introduce an RBS sequence to ensure that translation initiation does not pose a bottleneck in protein expression. An analysis of RBS effects on protein expression in *E. coli* was published (Salis et al., 2009), and an online RBS calculation tool for genome-sequenced prokaryotes may be found at https://salis.psu.edu/software/.

Genes may be synthesized for expression in another species, and the coding sequences modified to match the codon usage frequency of the heterologous host (Villalobos et al., 2006). Issues of translation efficiency and relationships to tRNA abundance in *R. sphaeroides* are somewhat unclear, despite recent advances (Cannarozzi et al., 2010; Tuller et al., 2010).
4. Future Directions

4.1. Genome modification

Although the genetics of *R. sphaeroides* allows modification of the genome in a reasonable time-frame, work on this organism is still far from the speed and ease of *E. coli* genetics. Some of the difficulty with working with this organism is due to the high GC-content of the genome, and the longer doubling-time of *R. sphaeroides* (3–5 h) compared to *E. coli* (20–30 min). Neither of these factors can be remedied in the near future, but additional tools could aid the *in vitro* and *in vivo* manipulation of the construct prior to the conjugation step.

Red/ET cloning (“recombineering”) has come to be recognized as superior technology for the size- and sequence-independent manipulation of DNA in *E. coli* and related *Enterobacteriacea*. Cells that express λ phage-derived red genes, or their functional prophage rac equivalents, promote base-precise exchange of linear single- or double-stranded donor DNA into the bacterial chromosome. Therefore, only short flanking homology arms are required; see Sawitzke *et al.* (2007) and references therein.

Recombineering methods for direct genome targeting in non-*Enterobacteriaceae* have been recently developed for *Mycobacterium tuberculosis* (van Kessel and Hatfull, 2007) and *Pseudomonas syringae* (Swingle *et al.*, 2010), but no such system exists for *R. sphaeroides*. However, methods for plasmid recombineering in *E. coli* (Noll *et al.*, 2009; Thomason *et al.*, 2007) facilitate the engineering of gene replacement vectors for *Rhodobacter* and other species as outlined below.

A drug marker suitable for *E. coli* and *Rhodobacter* has to be flanked by ~50-base tails homologous to the subcloned target region. This can easily be achieved by PCR. Thereby, the primer design determines exactly where the cassette recombines into the plasmid because no specific recombination sites are required. The recombination step takes place *in vivo*. To minimize unwanted side effects of plasmid recombineering, that is, multimer formation and mixtures of mutated and parental plasmids, low amounts of substrate plasmid (~10 ng) and linear marker (~100 ng) should be coelectroporated into Red/ET proficient *E. coli* cells. Nevertheless, the isolation of monomeric recombinant plasmids requires careful monitoring of the plasmid topology and a retransformation step (Noll *et al.*, 2009; Thomason *et al.*, 2007).

Recombineering approaches allow freedom from the need for restriction cleavage sites, and an antibiotic-resistance marker can be recombined into an appropriate suicide plasmid to obtain a base-precise disruption or deletion of the subclone target gene. However, as outlined in Section 2.1, a modification of a cloned gene does not necessarily need to be marked by a
resistance cassette. Interestingly, all kinds of markerless plasmid modifications (deletion, insertion, replacement) can be introduced in a two-step “hit and fix” approach (Noll et al., 2009).

Therefore, the PCR-amplified marker used in the first Red/ET step has to introduce a unique restriction site into the target plasmid. This can easily be achieved by oligo design. Following drug selection and isolation of recombinant plasmids, nonselectable DNA coding for all kinds of modification(s) is used to replaces the cassette and the unique restriction site in the second step. Upon selective digestion of parental plasmids (i.e., unique restriction site elimination) and retransformation, recombined plasmids are obtained with reasonable efficiency. Given its flexibility, plasmid recombination should prove to be a welcome alternative for the construction of gene replacement vectors in \textit{R. sphaeroides}.

Additionally, use of the Flp/FRT system from \textit{Saccharomyces cerevisiae} (Sadowski, 1995; Schweizer, 2003) and the Cre/LoxP system from bacteriophage P1 (Abremski et al., 1986; Sternberg et al., 1986), which are commonly used in \textit{E. coli} for marker removal, would facilitate the generation of markerless \textit{R. sphaeroides} mutants in a fraction of the time of the traditional methods. To our knowledge, no group is actively working on adapting these systems for use in \textit{R. sphaeroides}.

### 4.2. Synthetic operons

It is often desirable when generating synthetic operons to utilize proteins stemming from a wide variety of different species. This “mix and match” approach may prove useful in the creation of novel protein systems with function not found in nature. In such cases, there may be difficulty in finding a good host strain, because no one strain may be able to express at appropriate levels all the heterologous genes in a gene system. In order to overcome such a barrier, it may be necessary to refactor coding sequences such that they are better suited for a particular organism, as well as change the regulatory elements to match the heterologous host. For example, Widmaier et al. (2009) changed both the codons and regulatory sequences of spider silk genes to obtain high-level synthesis and secretion of spider silk in \textit{Salmonella} SPI-1 T3SS.

### REFERENCES


