Identification and Characterization of a Potential D-Lactate Oxidizing Enzyme from *Rhodobacter sphaeroides*

Research Thesis

Presented in partial fulfillment of the requirements for graduation *with honors research distinction* in Microbiology in the College of Arts and Sciences of The Ohio State University

By

Jacob J. DeVos

The Ohio State University

April 2020

Project Advisor:

Dr. Birgit E. Alber, Department of Microbiology

Examination Committee:

Dr. Charles Daniels, Department of Microbiology

Dr. Patrice Hamel, Department of Molecular Genetics

Abstract:

A cell like *Rhodobacter sphaeroides* can convert carbon sources into intermediates that are able to enter central carbon metabolism, which is the process of producing metabolic precursors that are used to generate cellular biomass. D-Lactate is one of these potential carbon sources and the purpose of the study is to better understand how it is metabolized by *R. sphaeroides* and the genes involved in these processes. Wild-type *R. sphaeroides* underwent a transposon mutagenesis procedure in which a random portion of the DNA of the organism was interrupted, which inactivated the normal function of that gene. From this library, a mutant was isolated that was unable to grow using D-lactate but was able to utilize all other carbon substrates tested including L-lactate. The transposon insertion site was mapped to the gene *rsp_1018* which is annotated to encode an iron-sulfur subunit of a glycolate oxidase. Two other nearby genes are likely to encode two further subunits of this protein complex and based on nucleotide spacing it is likely that these three genes are co-transcribed along with another downstream gene of unknown function. In light of the fact that glycolate and D-lactate are structurally similar, we propose that *rsp_1018, rsp_1019,* and *rsp_1020* encode for subunits of an enzyme that oxidizes D-lactate to pyruvate; pyruvate is an intermediate of central carbon metabolism. This conclusion is consistent with the D-lactate-negative phenotype of the mutant. Results indicate that growth of the transposon mutant on D-lactate can be restored by introducing the genes *rsp_1018* through *rsp_1020* on a plasmid. There is also some evidence to indicate that the overexpression of the *rsp_1018* gene product is detrimental to the cell. Our results suggest that *rsp_1018* encodes a subunit of a D-lactate oxidizing enzyme that is required by *R. sphaeroides* to metabolize D-lactate.

Table of Contents:

Introduction:

Lactate is often thought of as the end product of many metabolic processes such as lactic acid fermentation in bacteria which require certain cellular adaptations to counteract its acidic nature. Recently, lactate has even been thought of as the true end product of glycolysis in eukaryotic cells. However, many bacteria are able to utilize lactate by oxidizing it into pyruvate, a central carbon intermediate, where it is then able to enter carbon metabolism of the cell to form all of the carbon-based molecules required for cellular function. Bacteria are also able to oxidize lactate into carbon dioxide which provides cellular energy during respiration. Some microorganisms are only able to utilize lactate as either a carbon source or an energy source, but most bacteria such as *R. sphaeroides* and *E. coli* are able to utilize it for both processes (Sheng *et al.,* 2015).

Rhodobacter sphaeroides is a metabolically diverse organism that has the ability to grow with a wide variety of carbon sources including both enantiomeric isomers of lactate. When the cells are grown phototrophically, the carbon substrate provided, such as D-lactate is used exclusively as a carbon source and not as an energy source, which better allows for the study of carbon assimilation.

Lactate, a three carbon acid molecule, can exist in nature as one of two different stereoisomer forms, as L- or D-lactate. L- and D-Lactate oxidizing enzymes, which are responsible for the conversion of lactate to pyruvate or *vice versa*, belong to many evolutionarily unrelated enzyme families (Engqvist *et al,* 2009). Lactate has been found to be processed in bacteria by both NAD-dependent lactate dehydrogenases (nLDH) and NAD-independent lactate dehydrogenases (iLDH) (Wang *et al*, 2014) (Jiang *et al,* 2014). In both of these categories of enzymes, there exists those that exhibit chiral specificity meaning that L- and D-lactate require different enzymes

to be processed. This means that even though the two forms of lactate differ only in chirality around a single carbon atom, this is enough of a distinguishing factor that they require separate enzymes to metabolize. The exact mechanism and genes involved in the processing of D-lactate to pyruvate in *Rhodobacter sphaeroides* was previously unknown, but it is believed that a gene encoding for a subunit of a D-lactate oxidizing enzyme has been identified in this study.

The goal of this study was to better understand the mechanism by which *R. sphaeroides* is able to metabolize D-lactate so that it can enter into central carbon metabolism and be processed into cell carbon. In order to accomplish this, a forward genetics approach was taken via random transposon mutagenesis that identified a gene in the genome of *R. sphaeroides* that appeared to be necessary for its ability to grow with D-lactate. In order to confirm this function, an additional functional copy of the gene was reintroduced to the mutant and this appeared to restore full cellular function. This growth was closely monitored via growth studies to be able to determine any phenotypic difference in the isolates of *R. sphaeroides* that were used. In this way, the specificity of the gene product was also examined by growing the cells with either D- or L-lactate as the sole carbon source provided. The apparent functionality and necessity of *rsp_1018* points to it encoding a subunit of a D-lactate dehydrogenase enzyme. This study hopes to act as the first step in order to accurately determine the cellular function of *rsp_1018* and surrounding genes in *R. sphaeroides.*

Materials and Methods:

Bacterial strains, plasmids, and primers. All bacterial strains, plasmid vectors, and genetic primers used are given in Table S1.

Growth conditions. *Rhodobacter sphaeroides* 2.4.1(DSM 158) cells were grown aerobically and in the dark at 30°C as well as anaerobically with a nitrogen atmosphere in the light at 30°C. *E. coli* cell strains were grown aerobically at 37°C. Minimal medium for purple non-sulfur bacteria and Luria broth (LB) were used. For minimal media, the sodium salt of the carbon source was added to a final concentration of 10 mM in 1000 mL along with 1.2 g NH_4Cl (22 mM), 0.2 g $MgSO_4 \cdot 7H_2O$ (0.8 mM), 0.07g $CaCl_2 \cdot 2H_2O$, and the components in the concentrations as described in Table S2. Luria broth was made with 10 g tryptone, 5 g yeast extract, 5 g NaCl, at pH 7.0. When necessary, the following antibiotics at the concentrations indicated were added to the minimal medium for the growth of *R. sphaeroides*: kanamycin (20 µg/mL) and spectinomycin (25 µg/mL). When *E. coli* cells were grown in the presence of an antibiotic on LB media the following concentrations were used: ampicillin (100 µg/mL), spectinomycin (50 µg/mL), and kanamycin (50 µg/mL).

Isolation of *R. sphaeroides* transposon mutants. Equal masses of *R. sphaeroides* wild-type 2.4.1 and *E. coli* BW20767, which carries the pRL27 plasmid, harvested during exponential growth phase were combined and placed as a single drop on an LB plate. After incubation of the plate aerobically for 15 hours, the cells were scraped off the plate, resuspended in minimal media, diluted, and plated on minimal L-malate/kan media. Well isolated colonies were screened for growth aerobically on minimal media with kanamycin and the following carbon sources: D-malate, acetate, L-lactate, D-lactate, and L-malate (control). Strains that were observed to not grow or grow in a detectably different manner on a particular carbon source were

chosen and these were then rescreened on all carbon sources. Strains that demonstrated severely hampered growth were chosen and again screened for both aerobic and photoheterotrophic growth including the additional carbon sources: succinate, butyrate, and 3-hydroxypropionate.

Isolation of genomic region with transposon insert. Genomic DNA of selected strains were isolated using GeneJet DNA purification kit. The sample was digested with *Nco*I restriction enzyme and ligated using T4 DNA ligase, yielding circular DNA. *E. coli* DH5αλpir was transformed with the ligation mixture and plasmids were isolated after overnight growth using the GeneJet plasmid isolation kit. The length of the plasmid was analyzed via 0.8 % agarose gel electrophoresis following restriction digest using the diagnostic *Nco*I enzyme. The concentration of the plasmid was determined using a NanoDrop spectrophotometer. Plasmid was sent for Sanger sequencing (Genewiz) using primer tnpRL_17_1 and primer tnpRL_13_2, which bind to the sequence at either end of the transposon insert. The gene interrupted by the transposon is present in the sequenced DNA region and was identified using the BLAST algorithm. The sequencing results using both the upstream and the downstream primers were analyzed and were found to map to the same gene.

Construction of in-frame deletion and complementation plasmids. Using wild-type *R. sphaeroides* genomic DNA as a template, a precise in-frame deletion construct was constructed using a two-step method reliant on homologous recombination which allowed for precise alteration of the chromosome. Also, complementation constructs were constructed in order attempt to restore function that was lost by the interrupted gene.

For the generation of the inactivation construct for *rsp_1018*, a pair of PCR reactions were performed using 5x GXL buffer, dNTPs, and Prime STAR GXL DNA polymerase with wild-type *R. sphaeroides* DNA along with a pair of primers. The upstream PCR reaction used

delta RSP_1018 upF and delta RSP_1018 upR. The downstream reaction used delta RSP_1018 dnF and delta RSP_1018 dnR. The PCR products were analyzed using a 0.8 % agarose gel electrophoresis and the bands of the proper length were isolated using the GeneJet Gel Extraction Kit. An assembly PCR was performed using Prime Star GXL DNA polymerase, the upstream and downstream PCR reaction products, and the primers delta RSP_1018 upR and delta RSP_1018 dnF. The product was again excised from a gel and purified, yielding PCR delta RSP_1018.

The complementation constructs for *rsp_1018* as well as *rsp_1018, rsp_1019,* and *rsp_1020* follow the same procedures as above except for the assembly steps as the inserts are already one piece. The PCR product PCR RSP_1018_comp was generated using primers RSP_1018_com_upF and RSP_1018_comp_downR. The triple gene complementation construct utilized RSP_1018_com_upF and RSP_1018_comp_downR2, yielding PCR RSP_1018 to 1020_region_comp.

All three PCR products, PCR delta RSP_1018, PCR RSP_1018_comp, and PCR RSP_1018 to 1020_region_comp were cloned into pUC19 by digesting each PCR product and an aliquot of pUC19 with appropriate restriction enzymes based on their sequence. The appropriately cut PCR products were ligated with the corresponding cut pUC19 using T4 DNA ligase.

E. coli DH5α chemically competent cells were transformed with each ligation mixture and transformants were selected on LB medium with ampicillin and X-Gal/IPTG, providing a blue/white screen. White colonies were picked to be inoculated in LB media, grown, and underwent plasmid extraction. A test digestion was performed on the plasmids using a restriction enzyme that would produce resolvable bands when analyzed by gel electrophoresis.

E. coli DH5α competent cells were again transformed with the plasmids whose identities were confirmed via a test digest. These cells were then grown in a larger culture to isolate pure samples of the plasmids. The identity of the plasmids were confirmed using Sanger sequencing (Genewiz) using the appropriate sequencing primers. Plasmid pJD7 has PCR delta_RSP_1018 inserted into pUC19. Plasmid pJD4 has PCR RSP_1018_comp inserted into pUC19. Plasmid pJD6 has PCR RSP_1018 to 1020_region_comp inserted into pUC19.

The region of interest of each of these plasmids then had to be inserted into another vector that would allow the plasmid to be introduced into *R. sphaeroides*. Plasmid pJD7, the plasmid containing the sequence for inactivation of *rsp_1018* via a clean deletion, was ligated into pK18mobsacB (Schäfer *et al.*, 1994), a suicide vector, to yield pJD2. This was done by digesting pJD7 and pK18mobsacB using corresponding restriction enzymes and ligating the pieces together; *E. coli* DH5α cells were then transformed using the ligation. Isolated colonies were tested on LB+10 % sucrose for sensitivity to sucrose as the cells possessing the correct target plasmid exhibit an inability to grow on media containing sucrose but are able to grow in the presence of kanamycin. Cells exhibiting appropriate growth had their plasmid isolated and *E. coli* DH5α cells were retransformed with the plasmids. Plasmid was then reisolated and confirmed utilizing a test digestion. This resulted in the production of pJD2, which is a suicide vector that contains an in-frame deletion of *rsp_1018* of 2415 bp, as well as1.1 kb upstream and 1.3 kb downstream regions to allow for homologous recombination. The remaining open reading frame encoded by the truncated gene is a 56-amino acid peptide.

For the generation of the plasmid to be used for complementation studies of the *rsp_1018* gene, the pJD3 plasmid was digested and the gene-containing region was ligated into pSC75, which contains the 283-bp *rrnB* constitutive promoter and a consensus ribosome binding site.

After transforming the competent cells with the ligation product, the identity of the plasmid was confirmed via a test digest. This resulted in the production of pJD4, to be used for the complementation of *rsp_1018,* which contains the coding region of the gene attached to the *rrnB* constitutive promoter in the pBBR-derived vector pSC75 (Carlson *et al.,* 2019). Also, the production of this gene on the plasmid changed the start codon sequence from TTG, as it is found natively, to ATG which acts as a stronger start codon; the insert also includes 57-bp of the region downstream of *rsp_1018.*

For the generation of the plasmid to be used for complementation studies of the *rsp_1018, rsp_1019,* and *rsp_1020* genes, the pJD5 plasmid was digested and the gene-containing region was ligated into pBBRsm2MCS5(MC) (Schneider *et al.* 2012). After transforming the competent cells with the ligation product, the identity of the plasmid was confirmed via a test digest. This resulted in the production of pJD6, to be used for the three-gene complementation of *rsp_1018, rsp_1019,* and *rsp_1020* while utilizing the native promoter that is most likely present directly upstream of the coding region of *rsp_1020.* Included on the plasmid insert is the 57-bp region downstream of *rsp_1018* and the 84-bp region upstream of *rsp_1020.*

Conjugation of *R. sphaeroides* with plasmid constructs. *E. coli* SM10 competent cells were transformed with the plasmid construct to be used for complementation, pJD4 or pJD6, and *E. coli* S17-1 competent cells were transformed with the plasmid construct for inactivation, pJD2. Both *E. coli* cells and *R. sphaeroides* were grown with LB media until exponential growth phase was reached. The cells were harvested by centrifugation and equal masses of both cells were placed together as a single drop and grown aerobically at 30 ℃ for 15 hours, scrapped of, diluted, and plated on minimal L-malate/spec media or L-malate/kan, depending on the

plasmid's selection factor. Two isolated colonies of each were then grown, and a stock was made for further classification and study. The presence of pJD4 and pJD6 was confirmed by the isolation of plasmid and test digestion. The success of the single cross over of pJD2 into the *R. sphaeroides* chromosome was selected for by plating on MM L-malate/kan as well as for sucrose sensitivity.

Colony PCR. In order to determine if a crossover event had occurred, a master mix was made by combining GXL buffer (1x), GXL dNTPs (0.2 mM each), forward and reverse primers (0.5 µM), ddH_2O and GXL 0.5 µL per 20 µL reaction; the concentrations given are the final concentrations. A single colony was inoculated in each reaction tube by touching a colony with a sterile toothpick, washing in sterile water, and inserting into PCR tube. The PCR reaction was performed at 98 ℃ for 10 seconds, 60 ℃ for 15 seconds, and 68 ℃ for one minute per kB for 35 cycles. The PCR products were then resolved on a 8 % agarose gel via electrophoresis.

Double-crossover strain isolation. One downstream and one potential upstream single-crossover *R. sphaeroides* isolate was chosen and 0.1 mL of liquid culture of the cells was inoculated in 1 mL MM L-malate without any antibiotic overnight at 30 ℃ while shaking. 100 µL of a 10x dilution of each isolate was spread plated on MM L-malate + 10% sucrose and incubated at 30 ℃ until the colonies appeared. Twenty-five colonies of both the potential upstream and the downstream isolates were screened by patching on MM L-malate, MM L-malate/kan, and MM L-malate +10% sucrose and incubated at 30 ℃ aerobically until grown. Colony PCR was performed on the isolates that demonstrated the appropriate phenotypic growth of the target double-crossover; this was growth on the sucrose and no antibiotic plates, but no growth on the kanamycin plate. The colony PCR was analyzed for bands of the correct length for the target double-crossover *R. sphaeroides* isolated with the clean deletion.

Transformation. Freshly thawed 200µL of *E. coli* competent cells were combined with an appropriate amount of genetic material (1 µL of a plasmid or 30 – 50 µL of a ligation mixture). The mixture was incubated on ice for 20 minutes, heat-shocked at 42℃ for 2 minutes, and then incubated on ice for 10 minutes. An aliquot of 800µL of LB was added to the cells and incubated at 37℃ for one hour. The cells were then spread plated on an LB plate with the appropriate antibiotic using 25 - 75 µL for the plasmid transformation and all the cells for a ligation transformation after concentration of the cells via centrifugation. The cells were incubated overnight at 37℃.

Ligation. Combined 5 µL 10x T4 DNA ligase buffer, 2 µL T4 DNA ligase, double-digested vector plasmid, and double-digested DNA insert, and ddH_2O to a total volume of 50 µL. Used four times more insert than vector based on molar ration found by DNA size and concentration. Incubated at room temperature overnight.

Growth Experiments. Growth was monitored via measurement of optical density (OD) at 578nm. The appropriate *R. sphaeroides* culture was pregrown phototrophically to stationary phase in minimal L-malate media with antibiotic when appropriate in the concentrations described above. An aliquot of 0.15 mL of the preculture was inoculated in 4.5 mL of minimal media with the chosen carbon source and antibiotic, when appropriate, to an initial OD of about 0.02-0.05. The anaerobic cell cultures were grown at 30 ℃ and in light (3,000 lx). The aerobic cultures were grown at 30 ℃ in the dark while shaking. The carbon sources used were L-lactate and D-lactate. OD readings were taken every few hours, especially when the cells have entered exponential growth phase. Readings were taken until growth has appeared to have ceased for multiple readings. A final OD was taken a full day after the experiment has concluded to determine the final growth yield. If the optical density exceeded a reading of 1, cells were

removed and diluted 1:10 in ddH_2O and the OD of the diluted sample was determined. Doubling times for the cell cultures were determined dividing the difference in time between two time points by the difference in the logarithm base 2 of the optical densities of those time points. At least three data points taken during exponential growth are necessary and the doubling time is found between all sets of points given and averaged out. The standard deviation of these values are found to quantify the precision of the doubling time calculated.

Computational Methods. For protein sequence analysis, the NCBI blastp program was used (Altschul *et al*, 1990). Domains were defined and analyzed utilizing the CDD database found on NCBI. The multiple sequence alignment used for the formation of the consensus logo was formed with Clustal Omega (Medeira *et al*, 2019; https://www.ebi.ac.uk/Tools/msa/clustalo/). Sequences for this alignment were chosen by taking the protein sequences of different organisms that aligned with the query. The bacteria protein sequences used were broken down as follows: 5 were from the genus *Rhodobacter*, 5 were from the order *Rhodobacterales*, excluding the genus *Rhodobacter*, and the remaining 10 were from the phylum *Proteobacteria*, excluding the order *Rhodobacterales.* The consensus logo was formed using the Berkeley Consensus Logo Generator (Crooks *et al*, 2004; weblogo.berkeley.edu/logo.cgi).

Results:

Transposon mutagenesis. A transposon mutant library of *Rhodobacter sphaeroides* was constructed and individual mutants were screened for their ability to use specific carbon substrates as a carbon source and as an energy source. From an initial screen with the carbon substrates of D-malate, L-malate, acetate, L-lactate, D-lactate five mutants were chosen for further analysis based on their perceived phenotype of restricted growth with one or more carbon sources. A second screen was performed with the additional carbon substrates of 3-hydroxypropionate, butyrate, and succinate. For four of the five mutants, the initial phenotype observed was not confirmed. For all five mutants, the transposon site was mapped onto the genome which identified the gene that was interrupted.

Table 1. Phenotypic and genotypic characterization of characterized *R. sphaeroides* transposon mutants

Carbon Sources:	AU18JD44		AU18JD78		AU18JD90		AU18JD143		AU18JD199	
	O_2	light	O_2	light	O_2	light	O_2	light	O_2	light
D-malate	+	+	+	+	+	+	+	+	+	+
acetate	(+)	+	+	+	+	+	+	+	+	+
L-lactate	+	+	+	+	+	+	+	+	+	+
D-lactate	+	+	—	—	+	+	+	+	+	+
L-malate	+	+	+	+	+	+	+	+	+	+
3-hydroxypropionate	+	+	+	+	+	+	+	+	+	+
butyrate	+	+	+	+	+	+	+	+	+	+
succinate	+	+	+	+	+	+	+	+	+	+
Gene Interrupted:	Intergenic Region		*rsp_1018*		*rsp_4131*		*rsp_1011*		*rsp_0552*	

The transposon of mutant AU18JD44 was initially selected as an aerobic acetate-negative mutant. The transposon was mapped to an intergenic region (position 35687 of CP047041.1). The transposon of mutant AU18JD90 was selected as an aerobic acetate-negative and D-lactate-negative mutant. The transposon was mapped to *rsp_4131* and the gene product belongs to the CitB DNA-binding response regulator protein family (COG2197, pfam00072, E-value: 2.31e-44). The transposon of mutant AU18JD143 was selected as an aerobic acetate-negative mutant. The transposon was mapped to *rsp_1011* and the gene product is a hypothetical protein that does not belong to any known protein family. The transposon of mutant AU18JD199 was selected as an aerobic acetate-negative mutant. The transposon was mapped to *rsp_0552* and the gene product belongs to the EmrE multidrug efflux SMR transporter protein family (COG2076, pfam00893, E-value: 2.98e-29).

Only one of these mutants demonstrated a strong and repeatable phenotype of not being able to utilize a carbon source for metabolism and this mutant was selected for further study, that being AU18JD78. This mutant had a strong, repeatable phenotype of not being able to grow utilizing D-lactate as a carbon source. The transposon insertion site of AU18JD78 was mapped to gene *rsp_1018*, which was confirmed by both sequencing primers.

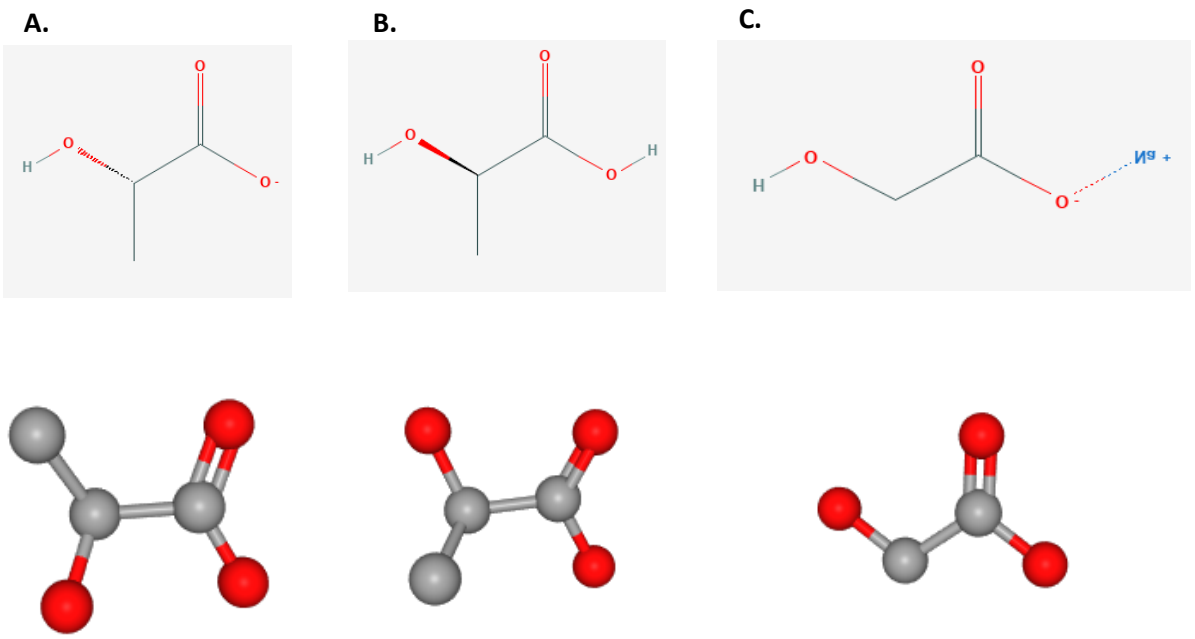

Figure 1. 2D Skeletal structures and 3D ball and stick models of L-lactate (A), D-lactate (B), and glycolate (C). Chirality around carbon #2 is different between the two isomers of lactate. Glycolate has a similar structure to the lactates except for the absence of the methyl group on carbon #2; this eliminates chirality. Even though the structures of L- and D- lactate are alike except for their chirality, they appear to be processed by divergent metabolic pathways. D-lactate appears to share a metabolic enzyme with glycolate, which converts the alcohol group to a carbonyl group.

Genomic context of selected mutant. Mutant AU18JD78 showed no growth on D-lactate both aerobically and phototrophically but was able to grow on all other carbon sources tested including L-lactate, differing from D-lactate only by the chirality of a carbon atom (Fig. 1). According to the sequence, the transposon interrupted the genome by inserting itself into the genome on chromosome 1, at position 2778106 (CP030271). The interrupted gene *rsp_1018* is surrounded closely by the genes *rsp_1017*, *rsp_1019*, and *rsp_1020* in close enough proximity to suggest that the genes are co-transcribed (Fig. 2).

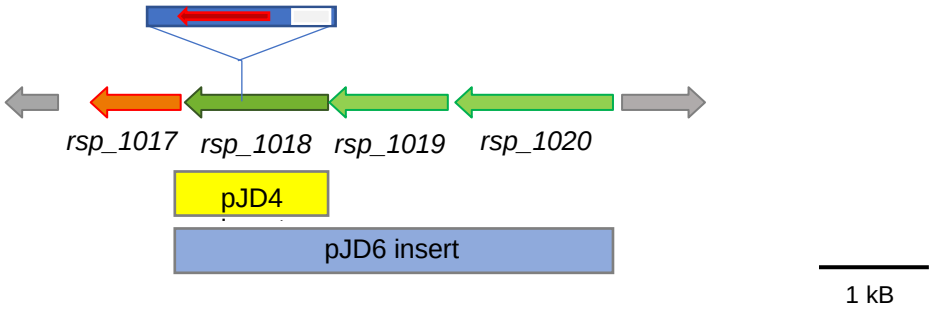

Figure 2. Genomic context of the transposon insertion site of mutant AU18JD78. The transposon insertion occurred in a gene annotated to encode a glycolate oxidase iron-sulfur subunit, GlcF (*rsp_1018*). The transposon carries a kanamycin resistance gene (red) with the same direction of transcription of the mutated gene and surrounding genes. The surrounding genes of interest are annotated as a "putative glycolate oxidase subunit protein" and as a "FAD-binding protein" (*rsp_1019*), "putative glycolate oxidase subunit protein, GlcD" (*rsp_1020)*, and "V8-like Glu-specific endopeptidase" (*rsp_1017*). The genomic regions that are included on the plasmid vectors to form pJD4 and pJD6 to be used for complementation studies are also shown.

Attempted Single-Gene Complementation of *rsp_1018*. In order to test if the interruption of *rsp_1018* is indeed responsible for the D-lactate-negative phenotype, a plasmid containing the intact *rsp_1018* was attempted to be introduced into the mutant. The plasmid pJD4 that contained the full-length coding region for the gene *rsp_1018* behind a constitutive promoter was attempted to be introduced into AU18JD78 via a conjugation procedure. This resulted in transconjugants being recovered at a frequency several orders of magnitude less than what was expected. The size of the plasmid obtained from these isolates differed from the size of the original plasmid pJD4. Sequencing of this plasmid will have to be attempted to determine its identity. A single-gene complementation construct inserted into AU18JD78 was unable to be obtained. Wild-type *R. sphaeroides* cells with the correct pJD4 plasmid were also not successfully isolated. This points to the plasmid not being able to properly enter, replicate, and be passed to cell progeny. The mechanism and reasoning for this occurrence will have to be examined further. One potential explanation is that over production of the *rsp_1018* gene

product, due to the change of the start codon sequence, TTG to ATG, and the addition of a constitutive promoter, is detrimental to normal cellular function.

Three-Gene Complementation Construct. To avoid over-expression of *rsp_1018*, a plasmid pJD6 was utilized that contained a genomic region that included the coding region for all three genes, *rsp_1018, rsp_1019,* and *rsp_1020* along with a 84-bp upstream region that likely contains the native gene promoter was introduced into AU18JD78 via a conjugation procedure. This resulted in a number of isolates that demonstrated the proper antibiotic resistance and the presence of the plasmid pJD6 in the *R. sphaeroides* cells was confirmed via a plasmid isolation and a test digestion. The lengths of the bands were consistent with those of pJD6. A three-gene complementation construct inserted into AU18JD78 was successfully obtained, and the resulting strain is called AU18JD78 (pJD6). The plasmid pBBRsm2MCS5 without an insert, a so-called empty vector, was also introduced into the mutant to produce AU18JD78 (pBBR). Both plasmids were also successfully introduced into wild-type *R. sphaeroides* to produce wt (pJD6) and wt (pBBR).

Cellular Growth. The growth of *R. sphaeroides* wild-type and mutants was studied so as to better understand how the edited genome effects the cell's ability to utilize different carbon substrates. Growth experiments were conducted on wild-type *R. sphaeroides* cells and the mutant AU18JD78 that has a transposon interrupting gene *rsp_1018.* The growth was tested for both of these strains containing no plasmid, the empty vector, and the three gene complementation construct plasmid with its native promoter. Cells were grown with either L- or D-lactate as the carbon source in the light or aerobically in the dark with L- or D-lactate as the source of carbon and energy.

Since the cells grow via two distinct mechanisms, respiratory and phototrophy, the differences between these two were analyzed first. Across all isolates that grew with the given carbon source, the growth yield was much less for those cells growing aerobically, with an OD halting at around 0.7, as compared to those growing anaerobically, whose final OD approached 2. This result was expected and is consistent with the growth of *R. sphaeroides* because under aerobic conditions the carbon substrate has to act as both a carbon source and an energy source while for growth in the light all carbon can be utilized as cell carbon. Also, the aerobically-grown cells showed a faster growth rate than compared to the same isolates with the same carbon substrate. For cells that grow on either L- or D-lactate, there does not appear to be a difference in growth rate between the two carbon substrates while using the same growth mode (Table 2).

Although wild-type *R. sphaeroides* grew well under all the conditions tested, the mutant AU18JD78 was unable to grow in the presence of D-lactate as the sole carbon substrate either aerobically or in the light (Fig. 3, C-D). In order to test if the D-lactate-negative growth phenotype was due to the insertion of the transposon in the *rsp_1018* gene, a functional copy of the gene, together with *rsp_1019* and *rsp_1020,* was introduced to the mutant and its growth was characterized. The inclusion of the complementation plasmid pJD6 to AU18JD78 successfully restored growth to *R. sphaeroides* with D-lactate (Fig. 3, C-D). The complementation construct brought both the growth rate and the growth yield of AU18JD78 back up to that of the wild-type, although more trials will have to be conducted to confirm. This evidence strongly suggests that the genes encoded on pJD6 are involved in the metabolism of D-lactate and that these gene products are required for these cellular processes. Particularly, it suggests the importance of RSP_1018 for this cellular function because that is what is being encoded for on the plasmid that is not functional on the chromosome of AU18JD78, however the transcription of *rsp_1019* and

rsp_1020 may also be affecting growth. The presence of the empty vector did appear to decrease the growth rate of *R. sphaeroides* when it was introduced to both the wild-type and AU18JD78. The inclusion of pJD6 in the wild-type *R. sphaeroides* did not appear to generally have an effect on the growth rate of wild-type cells, but more trials will have to be performed to confirm this result.

	Table 2: Doubling Times of *R. sphaeroides* Isolates Under Differing Growth Conditions (hours) +/- Standard Deviation					
Growth conditions	wt	AU18JD78	wt (pJD6)	AU18JD78 (pJD6)	wt (pBBR)	AU18JD78 (pBBR)
L-Lactate (light)	[4.3+/-0.1]	[4.5+/-0.2]	(5.2+/-0.1)	[5.4 +/-0.6]	[5.3+/-0.4]	[7.3+/-0.5]
L-Lactate (aerobic)	[3.8+/-0.3]	[3.3+/-0.1]	[3.8+/-0.3]	[3.6+/-0.3]	4.1+/-0.2	3.8+/-0.3
D-Lactate (light)	4.2+/-0.3	n.g.	4.0+/-0.1	4.2+/-0.3	5.0+/-0.4	n.g.
D-Lactate (aerobic)	[3.9+/-0.2]	n.g.	[3.7+/-0.1]	[3.9+/-0.1]	4.5+/-0.2	n.g.

*Doubling times (in hours) provided in brackets are the average of two growth experiments and those provided in parentheses are the results of a single experiment. Those values without either are the average of three trials, making them the most likely to be precise and trustworthy. n.g. no growth

Representative Growth Curves for *R. sphaeroides* Wild-Type and Mutant Strains Grown in Minimal Media Containing D- or L-Lactate as the Carbon Substrate

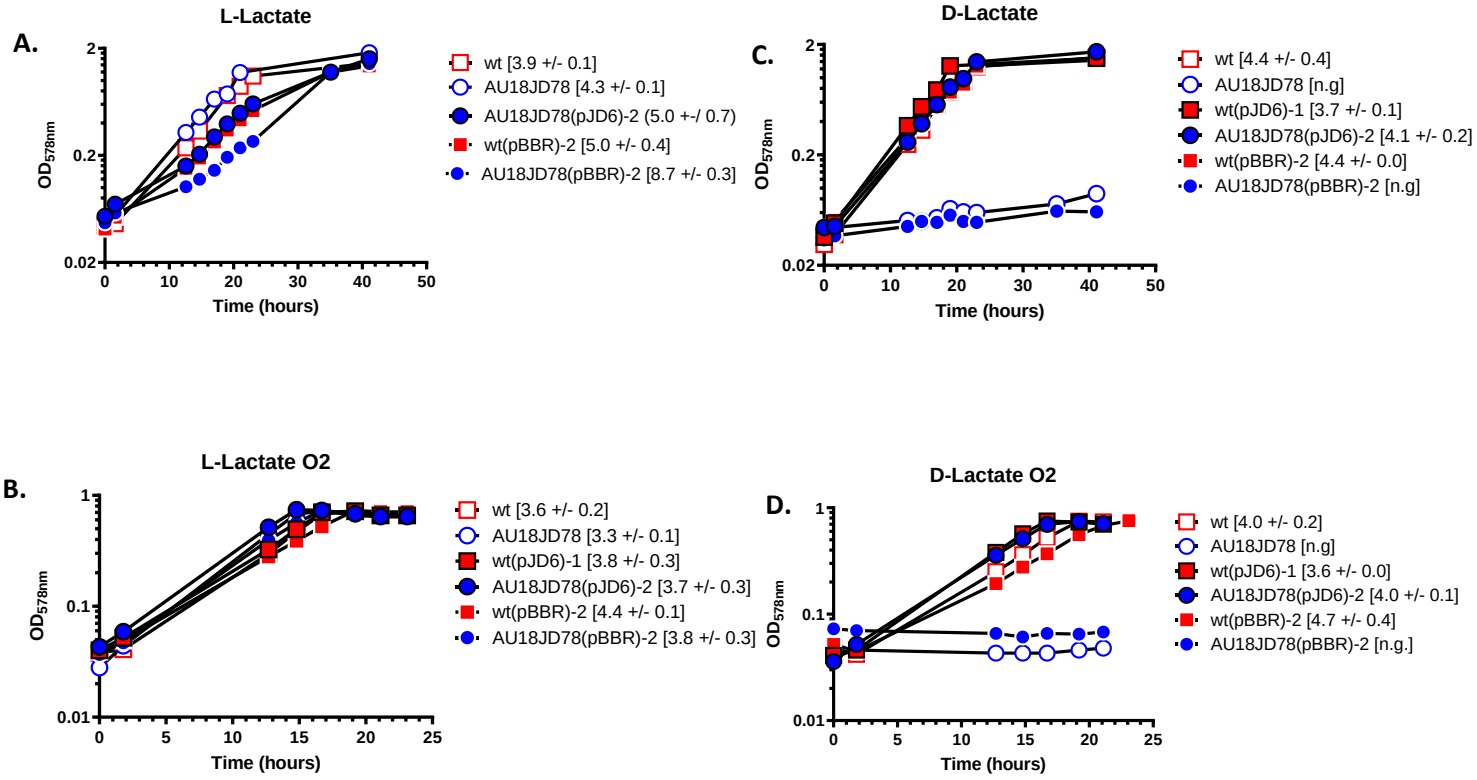

Figure 3. Phototrophic and aerobic growth of wild-type *R. sphaeroides* and AU18JD78 mutant and their derivatives with L- or D-lactate as the sole carbon substrate. The graphical results for each growth condition are provided on a logarithmic scale which allows for the representation of exponential growth. Panel A. Phototrophic growth with L-lactate (left) or D-lactate (right) as the carbon source. Panel B. Aerobic growth with L-lactate (left) or D-lactate (right) as the carbon substrate. The doubling times, shown next to each graph, were determined from data points derived during exponential growth. Strain AU18JD18 contains a transposon insertion in the *rsp_1018* gene. Plasmid pJD6 contains the genes *rsp_1018-1020* together with the native promoter. Plasmid pBBR represents the empty plasmid control.

Bioinformatic Analysis. It has been shown that some function provided by the genes *rsp_1018 – 1020* are required for D-lactate dependent growth. These gene and their protein products are by no means unique in *R. sphaeroides* as there are regions of high sequence identity across many different species of bacteria in which the gene products' function are predicted to be homologous. The most studied of these genes are present in *Escherichia coli*. The protein sequence encoded by *rsp_1018* aligns to GlcF in *E. coli* (AYG18054.1) as it possesses a 44% protein sequence identity (E-value: 4e-116). Both gene products belong to the GlcF super family of proteins (PRK11274, E-value: 2e-165) and possesses two [4Fe-4S] motifs (pfam13534, E-value: 5.5e-9. Also notable is that directly upstream of *glcF* in *E. coli* are *glcE* (AYG18053.1) and *glcD* (AYG18052.1) whose protein products align to the protein products of *rsp_1019* (ABA80202.1) and *rsp_1020* (ABA80203.1) with a 38% (E-value: 3e-58) and 49% (E-value: 1e-151) sequence identity, respectively. Based on this high sequence identity across all three genes in both *E. coli* and *R. sphaeroides*, it is predicted that they share a homologous cellular function.

A notable feature of *rsp_1018* and its homologs found across many different bacteria is the presence of two [4Fe-4S] binding motifs (CXXCXXCXXXCP) as these iron sulfur clusters are a feature of many oxidoreductase enzymes. Sequences that are highly conserved between organisms are considered essential as they have not changed as different species have evolved. The iron sulfur cluster sequence found in RSP_1018 is very highly conserved in gene homologs across a wide variety of bacteria species (Fig. 4, D.). This suggests the importance of this protein domain on the function performed by the gene products.

A.

glycolate oxidase iron-sulfur subunit [Escherichia coli str. K-12 substr. MG1655]

Sequence ID: AYG18054.1 **Length:** 407 **Number of Matches:** 1

Range 1: 1 to 403 GenPept Graphics ▼ Next Match ▲ Pre

Score	Expect	Method	Identities	Positives	Gaps
334 bits(856)	4e-116	Compositional matrix adjust.	185/419(44%)	255/419(60%)	18/419(4%)

```
Query  1    MQTTFTDEQLQDPGVARSNEILRACVHCGFCTATCPTYQVLGDELDSPRGRIYLIKDMLE  60
            MQT  T+E  Q+     ++ ILRACVHCGFCTATCPTYQ+LGDELD PRGRIYLIK +LE
Sbjct  1    MQTQLTEEMRQNARALEADSILRACVHCGFCTATCPTYQLLGDELDGPRGRIYLIKQVLE  60

Query  61   AGRPADARTVKHIDRCLGCLACMTTCPSGVHYMHLVDHARDHIEKTYRRPLFERLLRWTV  120
             G     +T +H+DRCL C  C TTCPSGV Y +L+D  RD +E+  +RPL ER+LR  +
Sbjct  61   -GNEVTLKTQEHLDRCLTCRNCETTCPSGVRYHNLLDIGRDIVEQKVKRPLPERILREGL  119

Query  121  AQLLPYPARFRAALRLAQLGRPFAGLVPDARLKAMLALAPHHVPKPSPNDRPQVFEATGP  180
             Q++P PA FRA  ++  + RPF      A+L A    A     KP P  R
Sbjct  120  RQVVPRPAVFRALTQVGLVLRPFLPEQVRAKLPAETVKA-----KPRPPLR---------  165

Query  181  KRMRVALLTGCAQRALDTDINDATIRLLRRMGCEVVIPKGMGCCGALTHHMGRAEQSHGL  240
             + RV +L GCAQ  L  + N AT R+L R+G  V+     GCCGA+ +H+   E+
Sbjct  166  HKRRVLMLEGCAQPTLSPNTNAATARVLDRLGISVMPANEAGCCGAVDYHLNAQEKGLAR  225

Query  241  AATNIRAWMAEKRARGLDAIVINTSGCGTTIKDYGHIFRNDPLAAEAA-EVSALALDVTE  299
            A  NI AW     A G +AI+   SGCG  +K+YG + +ND L A+ A +VS LA+D+ E
Sbjct  226  ARNNIDAWWPAIEA-GAEAILQTASGCGAFVKEYGQMLKNDALYADKARQVSELAVDLVE  284

Query  300  LLERLGLPEGAPKG-LRVAYHSACSLQHGQRIRQAPKELLARAGFEVTEPADPHLCCGSA  358
            LL    L + A +G  ++A+H  C+LQH Q++    +++L R GF +T+  D HLCCGSA
Sbjct  285  LLREEPLEKLAIRGDKKLAFHCPCTLQHAQKLNGEVEKVLLRLGFTLTDVPDSHLCCGSA  344

Query  359  GTYNLLHPEISAELKARKVATLEARKPDVVSAGNIGCIMQIGSGMGVPVVHTVELLDWA  417
            GTY L HP+++ +L+  K+  LE+ KP+++   NIGC   + S     V H +E+++ A
Sbjct  345  GTYALTHPDLARQLRDNKMNALESGKPEMIVTANIGCQTHLASAGRTSVRHWIEIVEQA  403
```

B.

glycolate oxidase subunit GlcE [Escherichia coli str. K-12 substr. MG1655]

Sequence ID: AYG18053.1 **Length:** 350 **Number of Matches:** 1

Range 1: 24 to 346 GenPept Graphics ▼ Next Match ▲ Pr

Score	Expect	Method	Identities	Positives	Gaps
180 bits(457)	3e-58	Compositional matrix adjust.	132/343(38%)	187/343(54%)	24/343(6%)

```
Query  19   PLRIRGGGTRA-IGR-TDGAPLETAGLAGVRLYEPGALTLVAGAGSPLGEIEAMLAAEGQ  76
            PL I+G  ++A +GR   G  L+     G+  Y+P  L + A  G+PL  IEA L + GQ
Sbjct  24   PLVIQGSNSKAFLGRPVTGQTLDVRCHRGIVNYDPTELVITARVGTPLVTIEAALESAGQ  83

Query  77   RLPFEAPDLRGLLGREGVSTLGGVVAANGSGPRRIQTGACRDSLIGVRFVDGVGQVVKNG  136
             LP E P      G E  +T GG+VA   +GPRR  +G+ RD ++G R + G G+ ++ G
Sbjct  84   MLPCEPPHY----GEE--ATWGGMVACGLAGPRRPWSGSVRDFVLGTRIITGAGKHLRFG  137

Query  137  GRVMKNVTGYDLVKLMAGSHGTLGVLSEVAFKLLPLPETELTLALPDLAPERAVEAMAAA  196
            G VMKNV GYDL +LM GS+G LGVL+E++ K+LP P   L+L   +++ + A+  +A
Sbjct  138  GEVMKNVAGYDLSRLMVGSYGCLGVLTEISMKVLPRPRASLSLRR-EISLQEAMSEIAEW  196

Query  197  LGSPFDVSGAAHWPGRGTFLRIEGFEASVRYRAERLRRLLTGFGEAAEVESPWAAIRDVA  256
               P  +SG  ++     ++R+EG E SV+      R LL G   A +    W  +R+
Sbjct  197  QLQPLPISGLCYF-DNALWIRLEGGEGSVK----AARELLGGEEVAGQF---WQQLREQQ  248

Query  257  -PFHGREGDVWRLSLKPSEAPAVAARAGGAALFDWGGGLVWILTEPGTDLRARIAPFA-G  314
             PF    G +WR+SL PS+AP +     G  L DWGG L W+ +    +   RIA  A G
Sbjct  249  LPFFSLPGTLWRISL-PSDAPMMD--LPGEQLIDWGGALRWLKSTAEDNQIHRIARNAGG  305

Query  315  HATLVRGTAPVAAFHPEPLPLAALAAGLRARFDPRGLFNPGLM  357
            HAT  R +A    F P   PL      L+ + DP G+FNPG M
Sbjct  306  HAT--RFSAGDGGFAPLSAPLFRYHQQLKQQLDPCGVFNPGRM  346
```

C.

glycolate oxidase [Escherichia coli str. K-12 substr. MG1655]

Sequence ID: AYG18052.1 **Length:** 499 **Number of Matches:** 1

Range 1: 29 to 473 GenPept Graphics ▼ Next Match ▲ Pre

Score	Expect	Method	Identities	Positives	Gaps
429 bits(1104)	1e-151	Compositional matrix adjust.	220/445(49%)	292/445(65%)	2/445(0%)

```
Query  25   LPADAVIHDEAETRAYECDALTAYRCPPLAAVLPRSTAEVSAVLRICHEERVPVVPRGSG  84
            +P   ++H + E    YECD L+AYR  PL  VLP+   +V+A+L +CH  RVPVV RG+G
Sbjct  29   VPGLEILHTDEEIIPYECDGLSAYRTRPLLVVLPKQMEQVTAILAVCHRLRVPVVTRGAG  88

Query  85   TSLAGGALPTADCVILGVARMNRVLETDYANRFIRVETGRTNLSVTGAVEVDGFFYAPDP  144
            T L+GGALP    V+L +AR   +L+ +   R  RV+ G  NL+++ AV      +YAPDP
Sbjct  89   TGLSGGALPLEKGVLLVMARFKEILDINPVGRRARVQPGVRNLAISQAVAPHNLYYAPDP  148

Query  145  SSQLACAIAGNIAMNSGGAHCLKYGVTTNNLLGVTMVLMDGTVVEIGGAHLDAPGLDLLG  204
            SSQ+AC+I GN+A N+GG HCLKYG+T +NLL + +  +DG  + +G   LD+PG DLL
Sbjct  149  SSQIACSIGGNVAENAGGVHCLKYGLTVHNLLKIEVQTLDGEALTLGSDALDSPGFDLLA  208

Query  205  VICGSEGQLGVVTEATLRILPKPEGARPVLMGFGSNEVAGACVSDIIRAGILPVAIEFMD  264
            +  GSEG LGV TE T+++LPKP  AR +L  F S E AG  V DII  GI+P  +E MD
Sbjct  209  LFTGSEGMLGVTTEVTVKLLPKPPVARVLLASFDSVEKAGLAVGDIIANGIIPGGLEMMD  268

Query  265  RPCIRATEAFAKAGYP-DCEALLIVEVEGSPAEIDDQLARILEIARRHDPVELRESRSEE  323
               IRA E F  AGYP D EA+L+ E++G  +++ +   R+ +I  +    ++R ++ E
Sbjct  269  NLSIRAAEDFIHAGYPVDAEAILLCELDGVESDVQEDCERVNDILLKAGATDVRLAQDEA  328

Query  324  ESRRIWLGRKSAFGAMGQIN-DYMCLDGTIPVTELPRVLRRIGELAAEAGLEVANVFHAG  382
            E  R W GRK+AF A+G+I+ DY C+DGTIP    LP VL  I  L+ +  L VANVFHAG
Sbjct  329  ERVRFWAGRKNAFPAVGRISPDYYCMDGTIPRRALPGVLEGIARLSQQYDLRVANVFHAG  388

Query  383  DGNMHPLILFDANRPGDLERCERLGAEILKLCVEVGGCLTGEHGVGVEKRDLMGVQFAPA  442
            DGNMHPLILFDAN PG+  R E LG +IL+LCVEVGG ++GEHG+G EK + M  QF
Sbjct  389  DGNMHPLILFDANEPGEFARAEELGGKILELCVEVGGSISGEHGIGREKINQMCAQFNSD  448

Query  443  DLEAQMRVKDVFDPRWLLNPAKVFP  467
            ++     VK  FDP  LLNP K  P
Sbjct  449  EITTFHAVKAAFDPDGLLNPGKNIP  473
```

D.

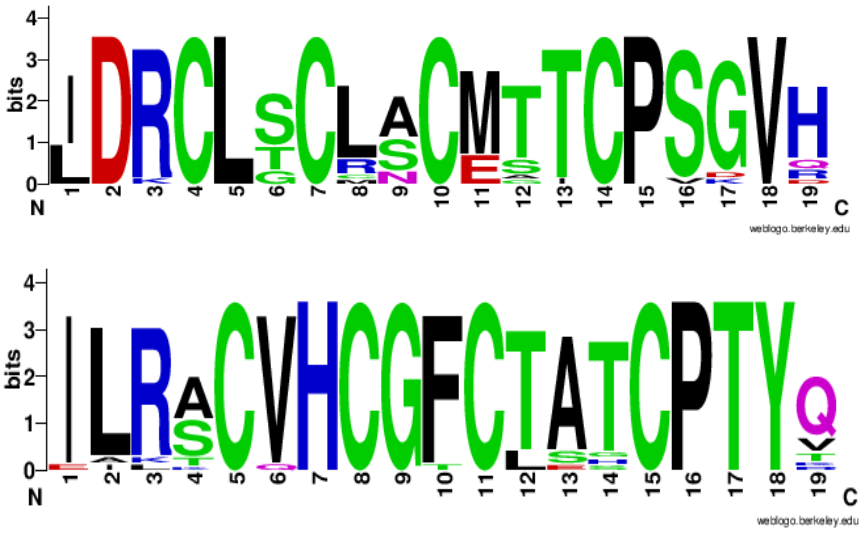

Figure 4. Sequence alignments for homologous proteins from between *R. sphaeroides* (query) and *E. coli* (subject). Alignments show high sequence identity and similarity across the entire protein sequence. This combined with very low E-values strongly suggest that these genes are homologous and perform related functions between the two organisms. A) Alignment between RSP_1018 from *R. sphaeroides* (ABA80201.1) and GlcF from *E. coli* (AYG18054.1). B) Alignment between RSP_1019 from *R. sphaeroides* (ABA80202.1) and GlcE from *E. coli* (AYG18053.1). C) Alignment between RSP_1020 from *R. sphaeroides* (ABA80203.1) and GlcD from *E. coli.*(AYG18052.1) D) Consensus logos constructed after multiple sequence alignment for proteins that aligned to RSP_1018. Protein sequences from an array of 20 diverse species of bacteria were used to make the logo. The proteins exhibited between a 99% and a 20% sequence identity to RSP_1018. This shows very highly-conserved regions of the protein and all proteins exhibited two nearby characteristic spatial arrangements of cysteine residues that indicate a [4Fe-4S] cluster (Cys-X-X-Cys-X-X-Cys-X-X-X-Cys-Pro). The highly conserved nature of this portion of the aligned proteins suggests its cellular importance for the function of the gene product. Multiple sequence alignment results can be found here: https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalo-I20200415-125625-0664-20150998-p2m

Attempted In-Frame Deletion of *rsp_1018*. In order to confirm the results obtained from the initial transposon mutagenesis, a clean deletion of the gene would have to be obtained. In order to obtain a clean *rsp_1018* deletion mutant, a plasmid was constructed that contained a truncated *rsp_1018* gene as well as 1.1 Kb upstream and 1.3 Kb downstream regions. The pJD2 suicide vector was introduced into *R. sphaeroides* by conjugation and multiple single-crossovers were selected and isolated. These isolates contained both the wild-type copy of the target region as well as the insertion of the truncated version of *rsp_1018* which was able to be introduced to the chromosome via homologous recombination on the chromosome. The double-crossover outgrowth procedure was completed, and the patching of these colonies revealed that many demonstrated the correct phenotype indicative of the retention of the truncated gene and the elimination of the wild-type copy. This had to be confirmed via colony PCR and all of these results for the numerous isolates tested were inconclusive. A successful *R. sphaeroides* isolate containing a clean, in-frame deletion of *rsp_1018* was unable to be obtained and the experiment needs to be continued.

Discussion:

Identification of a potential D-lactate oxidizing enzyme. The interruption of the gene that encodes for RSP_1018 in *R. sphaeroides* resulted in an inability of the cell to grow utilizing D-lactate as a carbon source and the reintroduction of this gene on a plasmid restored growth. Lactate has to be oxidized to pyruvate in order for it to enter into central carbon metabolism of the cell. Due to the phenotype of the AU18JD78 mutant, it would appear that *R. sphaeroides* is lacking this ability with D-lactate, but not with L-lactate. Growth is then able to be restored by the introduction of a copy of the *rsp_1018* gene, confirming its necessity for the cell to utilize D-lactate. Due to this gene's genomic context it is believed that it encodes for a subunit of an enzyme along with its surrounding genes. The results of this study point to RSP_1018 acting as a subunit of a D-lactate oxidizing enzyme.

Gene overexpression preventing normal cellular function. The plasmid pJD4 was used to attempt to introduce a functional copy of *rsp_1018* back to the mutant. Attempts to obtain *R. sphaeroides* cells that were successfully able to obtain and replicate this plasmid were unsuccessful even as a parallel procedure to introduce pJD6 showed great success. The differences between these two plasmids would be the first place to look for a possible explanation to account for these differences in success.

Both plasmids were constructed from a similar pBBR-based vector; the difference between the two is that the vector used for the construction of pJD4 contained the promoter *rrnB* directly upstream of the multiple cloning site. This *rrnB* promoter found in *R. sphaeroides* is a strong constitutive promoter found upstream of the 16SrRNA gene which increases the frequency of transcription initiation. In order for this promoter to act on a gene that codes for mRNA, a strong ribosome binding site was added. (Dryden & Kaplan, 1993). The use of this

promoter has been demonstrated to allow for high gene expression under all growth conditions. Because in pJD4 the *rrnB* promoter was fused to a functional copy of *rsp_1018,* this means that in *R. sphaeroides* the gene would be highly expressed leading to the production of large amounts of gene product. This differs from the promoter present in pJD6 because there is no promoter fusion present on the vector. The genes *rsp_1018, rsp_1019,* and *rsp_1020* are all present adjacently on chromosome 1 of *R. sphaeroides* and the short distances between each of these genes indicate that they are most likely co-transcribed. Because *rsp_1020* is the first of the genes, the native promoter used to activate the expression of these genes should be found directly upstream. For this reason, a 84-bp upstream region was included as part of the insert to form pJD6 and since the plasmid appears to be able to successfully restore cell function, the native promoter is likely present. The use of the constitutive *rrnB* promoter versus the native promoter would result in the production of much higher amounts of the gene product.

A change in the sequence of the start codon also accounts for an increase in protein synthesis. The wild-type copy of *rsp_1018* begins with the rare start codon sequence of TTG. The presence of an alternative start codon other than ATG has been found to result in at least a three-fold decrease in protein expression (Panicker *et al,* 2015). The use of TTG as the start codon has such a detrimental impact on protein production that it has been found to be negatively selected against by prokaryotes when it is changed from ATG; when TTG is used it requires other sequence mutations to compensate (Belinki *et al.,* 2017). The fact that *R. sphaeroides* possesses the TTG start codon on the gene *rsp_1018* means that relatively little gene product is produced than would be expected with another start codon, but this level of gene expression is enough for *R. sphaeroides* to function normally. The change in pJD4 from TTG to ATG as the

start codon of *rsp_1018* would indicate that a lot more gene product is being produced than would natively.

The addition of the constitutive promoter as well as the alteration of the start codon is believed to result in the severe overexpression of *rsp_1018* which prevents normally cellular function and growth. The exact mechanism by which this occurs is unknown, but it could be the result of a number of factors. Simply, the production of one protein could prevent the production of other necessary proteins or the high concentrations of RSP_1018 may have some detrimental effect on the cell. Whatever the explanation, it is likely that too much gene expression and protein production was a reason in which a mutant of *R. sphaeroides* could not be obtained that properly expressed the pJD4 complementation plasmid.

Comparison of *rsp_1018* and surrounding genes to well-studied homologs. The gene product RSP_1018 (ABA80201.1) of *R. sphaeroides* has been identified as a subunit of a potential D-lactate oxidizing enzyme. The function of this protein has not been yet experimentally tested in *R. sphaeroides* as this would require the isolation of protein to perform catalytic tests on difference substrates. However, there exists protein orthologs in a diverse array of organisms that have been shown to this D-lactate oxidizing activity which will be further discussed. In *Escherichia coli*, the protein GlcF (AYG18054.1) has a 44% sequence identity to RSP_1018 across the entire length of the protein including the presence of two highly conserved, cysteine-rich sequence indicative of two [4Fe-4S] clusters (pfam13534) as well as a number of other conserved cysteine residues which could be involved in association with itself or other proteins. In *E. coli* there also exists two additional genes directly upstream of *glcF*, being *glcE* (AYG18053.1) and *glcD* (AYG18052.1). There also exists two genes upstream from *rsp_1018* in *R. sphaeroides*, those being *rsp_1019* (ABA80202.1), and *rsp_1020* (ABA80203.1). The gene

products of *rsp_1019* and *glcE* have an amino acid sequence identity of 38%, whereas GlcD and RSP_1018 are 49% identical on the protein level, including five conserved cysteine residues. The overall rarity of cysteine in protein sequences makes their conservation likely to hold significance as this is not likely to occur randomly. In both organisms, the genes in question are in close enough proximity that they are likely co-transcribed. From the high sequence identities, it is believed that these genes proteins in both organisms are homologous.

The sequence conservation between GlcEFD and RSP 1018 - 1020 is significant because previous studies have been conducted on these proteins in *E. coli* and they represent a glycolate dehydrogenase based on their studied metabolic activity; it has also been shown that the protein products are 14% more effective at metabolizing D-lactate than glycolate and it has been shown that these enzymes show a severely hampered ability to metabolize L-lactate, at only 14% the rate of D-lactate (Lord, 1972). The necessity of each of the genes described in *E. coli* have been documented as it has been shown that the expression of each gene is required for the metabolism of glycolate, thus indicating an active three subunit enzyme, something that was unknown in previous experiments of this enzyme (Pellicer *et al,* 1996). The ability of the enzyme to selectively metabolize D-lactate over L-lactate and the necessity of each of the subunits of the enzyme aligns with the phenotype of the *rsp_1018* mutant in *R. sphaeroides.*

The activity of another enzyme homologous for those previously described has been studied in *Chlamydomonas reinhardtii* as well as its specificity for a given carbon substrate (Aboelmy and Peterhansel, 2014). The product of gene *gyd1* (XP_001695381.1) in this organism contains regions of high homology for both RSP_1018 and RSP_1020, with regions of high sequence identity for RSP_1019 being found elsewhere in the genome. This enzyme annotated as glycolate dehydrogenase demonstrated considerably higher enzyme activity with D-lactate versus

glycolate and very low activity detected with L-lactate. These results support the findings of this study of the ability of RSP_1018 and its associated subunits to act as a D-lactate oxidizing enzyme.

Potential identification of the electron acceptor for RSP_1018. The molecule that acts as the electron acceptor for this enzyme is worthy of consideration especially since it has not been be previously identified. Both the isolated enzymes of *E. coli* and *C. reinhardtii* require the addition of an artificial electron acceptor such as 2,6-dichloroindophenol (DCIP) or phenazine methosulfate (PMS) to show metabolic activity, but what is fulfilling this function *in vivo* is still unknown.

The standard reduction potential for pyruvate/lactate is – 0.19 V which means that in order for this reaction to be exergonic under standard conditions, the standard reduction potential of the reduction of the electron acceptor has to be more positive than – 0.19 V. This eliminates the possibility of NAD^+ as the electron acceptor as it has a standard reduction potential of – 0.32 V under standard conditions (Alberty, 2004). Flavins, such as FAD and FMN, have reduction potentials that are determined by their protein environment, however since flavins are prosthetic groups and are not free in the cell, they cannot act as the electron acceptor. While the use of the heme-bound respiratory cytochrome *c* would be thermodynamically favored with a E_0 of + 0.25 V, it has been determined to be localized to the periplasm of Gram-negative bacteria cells (Hamel *et al.*, 2009). Analysis of all three protein subunits, RSP_1018, RSP_1019, and RSP_1020, by the transmembrane topology and signal peptide predictor Phobius does not indicate the presence of a signal peptide meaning that they are not likely to be exported to the periplasm where the cytochrome *c* is located (Käll *et al.*, 2004). Of course, the reduction of O_2 to H_2O_2 has a very high reduction potential, however this is unlikely to act as the electron acceptor

because the enzyme appears to have activity under both aerobic and anaerobic conditions as indicated by the growth experiments. It is worth nothing that the electron acceptor could differ between the two growth modes. A flavoprotein, an electron carrying protein with a flavin like FAD or FMN as the prosthetic group, would be an acceptable electron acceptor energetically. However, a flavoprotein has been documented to require the interaction with membrane-bound proteins to facilitate electron transfer, which it does not appear RSP_1018, or its subunits, is (Vogt *et al.,* 2019). A quinone would also work, but it is also associated with the membrane and would require another protein that brings the enzyme close to the membrane. From this analysis, a prediction of the electron acceptor cannot be made, however certain molecules can be eliminated. A key to determining the electron acceptor may lie in certain domains present in the protein sequences of the enzyme subunits. There has been evidence to suggest that similar proteins are associated electrostatically with the membrane, but it is unknown if this protein shares any of those required characteristics (Dym *et al,* 2000).

The homologs of RSP_1019 in *R. sphaeroides* and GlcE in *E. coli* both contain a conserved FAD binding domain (pfam01565, COG0227, E-value: 1.46e-32). What is also notable is that *C. reinhardtii*, which lacks a homolog for *glcE*, did not show any activity of the enzyme in question when provided with flavin adenine dinucleotide (FAD). This aligns with previous findings that electron-transferring flavoproteins are utilized for the oxidation of D-lactate (Brockman and Wood, 1975). NAD^+, O_2, and cytochrome *c* were also independently tested in the previous study and they also demonstrated no activity with the enzyme. The presence of the [4Fe-4S] cluster domain of RSP_1018 would potentially be able to take only one electron at a time from the flavin. This is significant because the oxidation of lactate requires the transfer of two electrons at a time, but the iron-sulfur cluster can only accept one electron which

could again be transferred to an electron acceptor (Brzóska *et al.*, 2006). While it is still unknown what the electron acceptor is for this D-lactate oxidizing enzyme in *R. sphaeroides*, knowledge of the protein sequences as well as the standard reduction potentials required for an exergonic reaction to occur will allow for the possibilities to be narrowed down.

References:

1. Aboelmy MH, Peterhansel C. 2014. Enzymatic characterization of *Chlamydomonas reinhardtii* glycolate dehydrogenase and its nearest proteobacterial homologue. *Plant Physiol Biochem* 79:25–30.
2. Alberty RA. 2004. Standard apparent reduction potentials of biochemical half reactions and thermodynamic data on the species involved. *Biophys Chem* 111:115–122.
3. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410.
4. Belinky F, Rogozin IB, Koonin E V. 2017. Selection on start codons in prokaryotes and potential compensatory nucleotide substitutions. *Sci Rep* 7:1–10.
5. Brockman BHL, Wood WA. 1975. d-Lactate dehydrogenase of *Peptostreptococcus elsdenii* 41:309–312.
6. Brzóska K, Męczyńska S, Kruszewski M. 2006. Iron-sulfur cluster proteins: Electron transfer and beyond. *Acta Biochim Pol* 53:685–691.
7. Carlson SJ, Fleig A, Kelsey Baron M, Berg IA, Alber BE. 2019. Barriers to 3-Hydroxypropionate-dependent growth of *Rhodobacter sphaeroides* by distinct disruptions of the ethylmalonyl coenzyme A pathway. *J Bacteriol* 201:1–13.
8. Chen J, Li Y, Zhang K, Wang H. 2018. Whole-genome sequence of phage-resistant strain Escherichia coli DH5α. *Genome Announc* 6:5–6.
9. Crooks GE, Hon G, Chandonia JM, Brenner SE WebLogo: A sequence logo generator, *Genome Research*, 14:1188-1190, (2004)
10. Dryden SC, Kaplan S. 1993. Identification of cis-acting regulatory regions upstream of the rRNA operons of *Rhodobacter sphaeroides. J Bacteriol* 175:6392–6402.
11. Dym O., Pratt EA., Ho C., Eisenberg D. 2000. The crystal structure of D-lactate dehydrogenase, a peripheral membrane respiratory enzyme. *Proc Natl Acad Sci USA.* 97:9413-9418.

12. Engqvist M, Drincovich MF, Flügge UI, Maurino VG. 2009. Two D-2-hydroxy-acid dehydrogenases in *Arabidopsis thaliana* with catalytic capacities to participate in the last reactions of the methylglyoxal and β-oxidation pathways. *J Biol Chem* 284:25026–25037.

13. Finn RD, Basran J, Roitel O, Wolf CR, Munro AW, Paine MJI, Scrutton NS. 2003. Determination of the redox potentials and electron transfer properties of the FAD- and FMN-binding domains of the human oxidoreductase NR1. *Eur J Biochem* 270:1164–1175.

14. Hamel P, Corvest V, Giegé P, Bonnard G. 2009. Biochemical requirements for the maturation of mitochondrial c-type cytochromes. *Biochim Biophys Acta - Mol Cell Res* 1793:125–138.

15. Jiang T, Gao C, Ma C, Xu P. 2014. Microbial lactate utilization: Enzymes, pathogenesis, and regulation. *Trends Microbiol* 22:589–599.

16. Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027–1036.

17. Lord JM. 1972. Glycolate oxidoreductase in *Escherichia coli*. *BBA - Bioenerg* 267:227–237.

18. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019 Jul;47(W1) W636-W641. doi:10.1093/nar/gkz268.

19. Metcalf W., Larsen R., Wilson M., AMG. 2002. Genetic analysis of pigment biosynthesis in *Xanthobacter autotrophicus* Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria. *Arch Microbiol* 178:193–201.

20. Metcalf W, Jiang W, Daniels LL, Kim SK, Haldimann A, Wanner BL. 1996. Conditionally replicative and conjugative plasmids carrying lacZα for cloning, mutagenesis, and allele replacement in bacteria. *Plasmid* 35:1–13.

21. Norrander J, Kempe T, Messing J. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26:101–106.

22. Panicker IS, Browning GF, Markham PF. 2015. The effect of an alternate start codon on heterologous expression of a PhoA fusion protein in *Mycoplasma gallisepticum*. *PLoS* One 10:1–10.

23. Pellicer MT, Badía J, Aguilar J, Baldomà L. 1996. glc Locus of *Escherichia coli*: Characterization of genes encoding the subunits of glycolate oxidase and the glc regulator protein. *J Bacteriol* 178:2051–2059.

24. Puhler R Simon; UPA. 1983. A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Nat Biotechnol* 1:784–791

25. Rogatzki MJ, Ferguson BS, Goodwin ML, Gladden LB. 2015. Lactate is always the end product of glycolysis. *Front Neurosci* 9:1–7.

26. Schäfer A, Tauch A, Jsger W, Kalinowski J, Thierbachb G, Piihler A. 1994. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutumicum. Gene* 145:49–5201.

27. Schneider K, Asao M, Carter MS, Alber BE. 2012. *Rhodobacter sphaeroides* uses a reductive route via propionyl coenzyme A to assimilate 3-hydroxypropionate. *J Bacteriol* 194:225–232.

28. Sheng B, Xu J, Zhang Y, Jiang T, Deng S, Kong J, Gao C, Ma C, Xu P. 2015. Utilization of D-lactate as an energy source supports the growth of *Gluconobacter oxydans*. *Appl Environ Microbiol* 81:4098–4110.

29. Vogt MS, Schühle K, Kölzer S, Peschke P, Chowdhury NP, Kleinsorge D, Buckel W, Essen LO, Heider J. 2019. Structural and Functional Characterization of an Electron Transfer Flavoprotein Involved in Toluene Degradation in Strictly Anaerobic Bacteria. *J Bacteriol* 201:1–17.

30. Wang L, Cai Y, Zhu L, Guo H, Yu B. 2014. Major role of NAD-dependent lactate dehydrogenases in the production of L-lactic acid with high optical purity by the thermophile *Bacillus coagulans. Appl Environ Microbiol* 80:7134–7141.

Acknowledgement:

Funding for this research was supported by the National Science Foundation, grant number MCB60047897, and a research scholarship was provided by The Ohio State University Honors College of Arts and Sciences. The largest thank you to Dr. Birgit Alber for not only her advisement and support for this research project, but also for challenging my thinking and for influencing the educational path that I wish to pursue. Thank you to Dr. Charles Daniels for your advice and guidance as well as for serving on my examination committee. Thank you to Dr. Patrice Hamel for serving as a member of my examination committee. Thank you to Dr. Steven Carlson for your teaching of lab concepts and techniques. Thank you to Dr. Tabita for the use of the NanoDrop spectrometer and to Dr. Rappleye for the gel imaging system. I really appreciate all of my fellow undergraduate researchers in the lab including Suzy, Supriya, Maxwell, Olivia, and most of all, my lab partner in crime Kerrigan.

Appendix:

TABLE S1. Bacterial strains, plasmids, and primers used in this study

Strain, plasmid, or primer	Description	Source
Strains and Isolates		
R. sphaeroides 2.4.1	DSM 158	DSMZ, Alber Lab
E. coli BW20767	*E. coli* containing pRL27 for transposon mutagenesis	Larsen *et al.*, 2002
E. coli DH5α	Competent *E. coli* cells used for transformation and replication of a plasmid	Chen *et al.*, 2018
E. coli DH5αλpir	Competent *E. coli* cells used for transformation and replication of a plasmid	Alber Lab
E. coli S17-1	Kanamycin resistant donor strain used in conjugation	Simon *et al*, 1983
E. coli SM10	Spectinomycin resistant donor strain used in conjugation	Simon *et al*, 1983
AU18JD78	*R. sphaeroides* transposon mutant interrupting *rsp_1018* (glycolate oxidase iron-sulfur subunit)	This study
AU18JD44	*R. sphaeroides* transposon mutant interrupting *rsp_4084* (acetyltransferase)	This study
AU18JD90	*R. sphaeroides* transposon mutant interrupting *rsp_4131* (DNA-binding response regulator)	This study
AU18JD143	*R. sphaeroides* transposon mutant interrupting *rsp_1011* (hypothetical protein)	This study
AU18JD199	*R. sphaeroides* transposon mutant interrupting *rsp_0552* (QacE family quaternary ammonium compound efflux SMR transporter)	This study
wt::pJD2, 1-5	Wild-type *R. sphaeroides* with pJD2 single crossover, downstream	This study
wt:pJD2, 6	Wild-type *R. sphaeroides* with pJD2 single crossover, unknown insertion location	This study
wt(pBBR)	Wild-type *R. sphaeroides* containing the pBBRsm2MCS5 empty vector plasmid	This study
wt(pJD6)	Wild-type *R. sphaeroides* containing the complementation plasmid pJD6	This study
AU18JD78(pBBR)	Mutant *R. sphaeroides* AU18JD78 containing the pBBRsm2MCS5 empty vector plasmid	This study
AU18JD78(pJD6)	Mutant *R. sphaeroides AU18JD78* containing the complementation plasmid pJD6	This study
Plasmids		
pRL27	Suicide plasmid vector used for transposon mutagenesis that contains transposase and kanamycin resistant transposon insertion sequence	Metcalf et al., 1996
pUC19	High replication plasmid with multiple cloning site, ampicillin resistance, and *lacZα* gene used for replication of constructed gene sequence	Norrander *et al.*, 1983
pK18mobsacB	Suicide vector used for the insertion of the truncated gene deletion sequence into *R. sphaeroides*	Schäfer *et al*., 1994
pSC75	pBBR-derived plasmid that contains the *rrnB* promoter	Carlson *et al.*, 2019
pBBRsm2MCS5	Plasmid vector with multiple cloning site and spectinomycin resistance able to be replicated by *R. sphaeroides*	Schneider *et al.*, 2012
pJD2	Ligation of pJD7 insert into pK18mobsacB with *Xba*I and *Eco*RI	This study
pJD3	Ligation of PCR RSP_1018_comp into pUC19 with *Kpn*I and *Nde*I, sequenced	This study
pJD4	Ligation of pJD3 insert into pSC75 with *Nde*I and *Kpn*I	This study
pJD5	Ligation of PCR RSP_1018 to 1020_region_comp into pUC19 with *Xba*I and *Kpn*I, sequenced	This study
pJD6	Ligation of pJD5 insert into pBBRsm2MCS5 with *Xba*I and *Kpn*I	This study
pJD7	Ligation of PCR deltaRSP_1018 into pUC19 with *Eco*RI and *Xba*I, sequenced	This study
Primers		
tnpRL17_1	Sequencing primer for forward upstream direction for pRL27 plasmid used for transposon mutagenesis (AACAAGCCAGGGATGTAACG)	Alber Lab
tnpRL13_2	Sequencing primer for downstream reverse direction for pRL27 plasmid used for transposon mutagenesis (CAGCAACACCTTCTTCACGA)	Alber Lab
puc_For	Sequencing primer for forward upstream direction pUC19 plasmid vectors (CGCCAGGGTTTTCCCAGTCACAC)	Alber Lab
puc_Rev	Sequencing primer for reverse downstream direction pUC19 plasmid vectors (AGCGGATAACAATTTCACAGG)	Alber Lab
puc_Nde	Sequencing primer for pUC 19 plasmid vectors using NdeI site (GGAGCAGACAAGCCCGTCAG)	Alber Lab
delta RSP_1018 upF	PCR primer for upstream forward region of *rsp_1018* (CGTGCAGAAGCCGCAATGGACACAG)	This study
delta RSP_1018 upR	PCR primer with *Xba*I site for upstream reverse region of *rsp_1018* and assembly PCR (PCR deltaRSP_1018 full) (TCCGGCCAAGG**TCTAGA**CGTTGGATG)	This study
delta RSP_1018 dnF	PCR primer with *Eco*RI site for downstream forward region of *rsp_1018* and assembly PCR (PCR deltaRSP_1018 full) (CGGC**GAATTC**AAGAGCCTCCCAAG)	This study
delta RSP_1018 dnR	PCR primer for downstream reverse region of *rsp_1018* (GTCCATTGCGGCTTCTGCACGACGGTCGAGCTTCTGGACTG)	This study

delta_pJD1_downR2seq	Sequencing primer for pJD7 (CGAGGTGGCCTTCAAGCTTC)	This study
delta_pJD1_upF2seq	Sequencing primer for pJD1 and pJD7 (AGGGCCAGATCGAAC)	This study
		This study
RSP_1018_com_upF	PCR primer with *Kpn*I site for upstream forward region for gene isolation of *rsp_1018* (PCR RSP_1018_comp) and *rsp_1018*, *rsp_1019*, and *rsp_1020* construct (PCR RSP_1018 to 1020_region_comp) (TGA**GGTACC**GCATGGATGCTCCGGAAG)	
RSP_1018_comp_downR	PCR primer with *Nde*I site for downstream reverse region for gene isolation of *rsp_1018* (PCR RSP_1018 to 1020_region_comp) (GGAGAGC**CATATG**CAGACGACCTTC)	This study
delta_pJD5_upF5	Sequencing primer for pJD5 (ACCTCGACGCAGAGCTTCAG)	This study
delta_pJD5_upF4	Sequencing primer for pJD5 (CGTAACCGGTGACATTCTTC)	This study
delta_pJD5_upF3	Sequencing primer for pJD5 (TCGTCGGTGAAGGTCGTCTG)	This study
delta_pJD5_upF2	Sequencing primer for pJD5 (CAGCCGCTCGTGTTGATGAC)	This study
RSP_1018_comp_downR2	PCR primer with *Xba*I site for downstream reverse region for gene isolation of *rsp_1018* and *rsp_1018*, *rsp_1019*, and *rsp_1020* construct (PCR RSP_1018 to 1020_region_comp) (TGGCTCTAGATGCGGGCACGATAATGCAAG)	This study
rsp_1018_dn_I	PCR primer for identification of location of inactivation plasmid after single cross over; anneals to downstream region of *rsp_1018* (ATCAGCCCGAGGATCGTGAGGAAG)	This study
rsp_1018_up_I_comp	PCR primer for identification of location of inactivation plasmid after single cross over; anneals to upstream region of *rsp_1018* on the complimentary strand (TGACGGGAGAGCCTCTTGCAGAC)	This study
78region_dn	PCR primer for identification of location of inactivation plasmid after single cross over; anneals to downstream region outside of *rsp_1018* (TGTTGTAGGGCGGATAGGTG)	This study
78region_up_comp	PCR primer for identification of location of inactivation plasmid after single cross over; anneals to upstream region outside of *rsp_1018* (TCGAGGTGGCCAATGTCTTC)	This study

Table S2. Additional components and concentrations for minimal media.

Solution	Composition for 1 Liter	Volume (mL/L)	Concentration in Medium
1 M Phosphate Buffer pH 6.7	1 M K_2HPO_4 1 M KH_2PO_4	15	15 mM
100 x Trace Element Solution	500 mg Disodium EDTA 300 mg $FeSO_4 \cdot 7H_2O$ 3 mg $MnCl_2 \cdot 4H_2O$ 50 mg $CoCl_2 \cdot 6H_2O$ 1 mg $CuCl_2 \cdot 2H_2O$ 2 mg $NiCl_2 \cdot 6H_2O$ 3 mg $Na_2MoO_4 \cdot 2H_2O$ 5 mg $ZnSO_4 \cdot 7H_2O$ 2 mg H_3BO_3 (boric acid)	10	1 % each
667 x Vitamin Solution	100 mg Cyanocobalamin (Vitamin B_{12}) 300 mg Pyridoxamine-2 HCl (Vitamin B_6) 100 mg Calcium-D(+)-Pantothenate 200 mg Thiamine chloride (Vitamin B_1) 200 mg Nicotinic acid 80 mg 4-Aminobenzoic acid (Vitamin H_1) 20 mg D(+)-Biotin	1.5	0.15 % each