

# Characterization of *Rhodobacter sphaeroides* Gene RSP1977

Arlie Joe Green,<sup>a</sup> & Aaron Setterdahl<sup>a</sup>

*Rhodobacter sphaeroides* is purple, non-sulfur, photosynthetic bacterium whose entire genomic DNA has been sequenced; however, many of the genes in its genome have not been characterized in order to determine their function in the bacterium. Furthermore, many of the genes in *R. sphaeroides* are very similar to gene sequences in other organisms. Therefore, upon determining the functions of *R. sphaeroides* genes, it will help lead to a better understanding of the functions of genes in other organisms. This research is an attempt to isolate, purify and characterize proteins from *R. sphaeroides* using techniques in biochemistry, molecular biology, informatics, and genetics. To date, the gene RSP1977 has been successfully isolated, cloned, and analyzed. It has been attempted to insert the gene into the pSC-B-amp/kan cloning vector, to transform the vector into competent cells, then re-isolate the vector and confirm the insertion of RSP1977. Attempts to insert the gene into the pSC-B-amp/kan cloning vector are currently underway. The next steps, after confirmation, would be to insert the gene into the pET-21b(+) expression vector and express the protein. Once the protein is expressed, its functions will be studied and determined.

**Keywords:** Biochemistry, bacteria, *Rhodobacter sphaeroides*, genetics

## Introduction

*Rhodobacter sphaeroides* is a rod-shaped, gram-negative, purple non-sulfur, photosynthetic, photoheterotrophic bacterium that can be found in soil, in anoxic zones of waters, mud, sludge, and in organic-rich water habitats. It can also be found in all types of aqueous environments including those of extreme conditions such as hot springs, ice caps, acidic and alkaline lakes, and hypersaline environments.<sup>1,2</sup> It is unknown precisely how *R. sphaeroides* can grow in these various conditions, but it is attributed to the organism's uncanny ability to grow by various growth methods. Some of these methods include growing by either aerobic or anaerobic respiration, photosynthesis using bacteriochlorophyll a in the absence of oxygen, fermentation, and diazotrophicity; meaning it can fix nitrogen gas to a usable form if none is present in its environment.<sup>1,2,3,4</sup> *R. sphaeroides* has also been found to be able to detoxify a number of metal oxides and oxyanions as well as a possible tetrapyrrole and carotenoid biosynthesis organism, which have broad applicability.<sup>2,5</sup> The variability of *R. sphaeroides* growth methods and habitats as well as its many interesting abilities has attracted the attention of the scientific community and led to the organism being very well-studied and its entire genomic DNA being sequenced and annotated. This has created the possibility for the discovery that *R. Sphaeroides* has 2 circular chromosomes (CI and CII) and 5 endogenous plasmids (A, B, C, D, E) with some genes being located on one of the two chromosomes, while others are found on both.<sup>1,2,5</sup> Even so, not every gene has been characterized. Characterization of the genes of *R. sphaeroides* could help lead to better understandings of its complex growth mechanisms and internal functions as well as to similar gene sequences in other organisms.

In any living organism there is a set of instructions that dictate the inner workings of that organism. These instructions are called genes, which are found in that organisms DNA. These genes contain the

instructions that allow ribosomes to build the proteins that carry out various tasks in the organism. A protein is a molecule composed of one or more polypeptide chains, which are sequences of amino acids linked together by peptide bonds, and has a specific biological function.<sup>6</sup> The study of the structure, function and expression of proteins is an area of great interest in the field of biochemistry due to its applications called proteomics.<sup>1</sup>

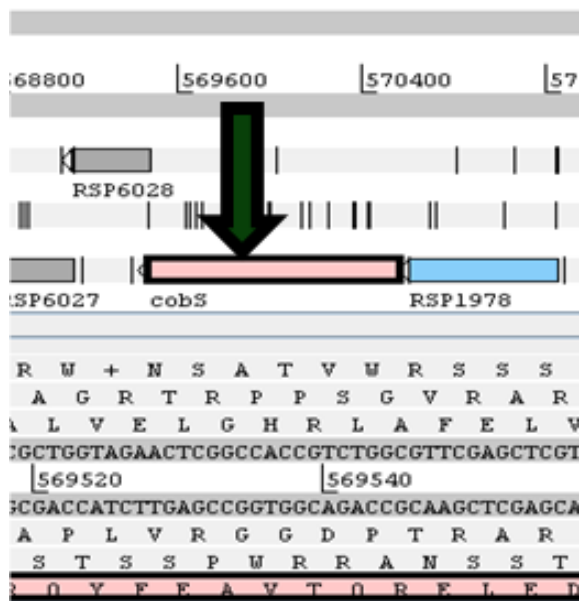
This research project utilizes techniques from biochemistry, molecular biology, genetics, and informatics to select an uncharacterized gene from the genome of *R. sphaeroides*, isolate that gene from its genome, insert the gene into an expression vector, and transform that vector into *Escherichia coli* in order to produce the protein for which that gene encodes. The overall goal of this research is to characterize the protein that is expressed by the gene in order to determine its biological function. This characterization could be done utilizing biochemical techniques such as affinity chromatography and spectroscopy.

## Research

### Gene Selection

The first step in the research process for this project was to select an uncharacterized gene of interest. This was done via the Artemis: Genome Browser and Annotation Tool. Artemis is a Java-based DNA sequence visualization tool and is used for the annotation of bacterial and lower eukaryotic genomes.<sup>7</sup> The chromosomal DNA of *Rhodobacter sphaeroides* has been sequenced and therefore the Artemis viewer can be used to look at the annotated sequence. This allows for visualization of the entire DNA sequence, the individual gene sequences within the DNA, the possible function of those genes, as well as six possible reading frames of amino acids produced from the DNA to form proteins.<sup>7</sup> For this research, individual genes were selected that were uncharacterized but had putative, or possible, functions. The genes selected were

RSP1927, RSP1977 and RSP1762; all of which are located on the CI chromosome of *R. Sphaeroides*. RSP1977 yielded the best preliminary results and was the focus of the research and this paper.



**Figure 1:** Artemis Program showing the gene RSP1977, the DNA sequence and the possible amino acid sequences.

Figure 1 shows a screenshot of the Artemis program and RSP1977 as well as the DNA sequence of the gene and the six possible amino acid sequences for the protein. RSP1977 was found to be 1,095 base pairs long with a proposed function of being a possible cobalt insertion protein (cobS). The cobS protein is a subunit of a larger protein that inserts cobalt into the corrin ring of a tetrapyrrole in order to form cobalamin which is vitamin B-12. The Artemis program is able to determine possible protein functions by comparing the genomic sequence in the viewer to the genomic sequences found in other organisms and determining possible gene encoding sequences and their possible functions based on that comparison.<sup>7</sup>

To further our comparison of RSP1977 with the gene sequences of other organisms we utilized the BLAST program. BLAST stands for Basic Local Alignment Search Tool and is a sequence similarity search program. BLAST compares the imputed sequence to a database containing the DNA sequences of thousands of other organisms and yields a percentage and sequence comparison.<sup>8</sup> Determining the similarities between RSP1977 and the sequences of other organisms gives us a starting point in determining the function of the protein once expressed as well as showing that the research done on *R. sphaeroides* will also apply to other organisms. The BLAST search showed that other bacteria and more complex organisms contained sequences of high similarity to RSP1977 as shown in Table 1.

**Table 1:** Sample BLAST results for RSP1977

Sequences producing significant alignments:							
Accession	Description	Max score	Total score	Query coverage	E-value	Max ident	
CP000143.1	Rhodobacter sphaeroides 2.4.1 chromosome 1, complete sequence	763	763	100%	0.0	100%	
CP000377.1	Rhodobacter sphaeroides ATCC 17029 chromosome 1, complete sequence	739	739	100%	0.0	99%	
CP001130.1	Rhodobacter sphaeroides KD131 chromosome 1, complete sequence	734	734	100%	0.0	99%	
CP0002147.1	Pseudovibrio sp. FO-BEG1, complete genome	524	524	99%	1e-164	70%	
CP000494.1	Bradyrhizobium sp. ITA1, complete genome	509	509	98%	2e-159	67%	
CP000283.1	Rhodopseudomonas palustris Bis15, complete genome	506	506	97%	3e-158	67%	
CP001096.1	Rhodopseudomonas palustris TIE-1, complete genome	504	504	97%	1e-157	67%	
BX572234.1	Rhodopseudomonas palustris CGA009 complete genome; segment 2/16	503	503	97%	4e-157	66%	
CP000282.1	Roseobacter denitrificans OCh 114, complete genome	502	607	97%	0.0	81%	
CP000489.1	Paracoccus denitrificans P01222 chromosome 1, complete sequence	502	641	97%	0.0	83%	
CP000961.1	Rhodobacter sphaeroides ATCC 17025, complete genome	704	704	96%	0.0	99%	
CP002937.1	Paracoccus denitrificans SD1, complete genome	599	640	95%	0.0	85%	
CP002588.1	Polymorphum givum SL003B-26A1, complete genome	493	493	95%	1e-153	67%	
CP002245.1	Agrobacterium sp. H13-3 linear chromosome, complete sequence	493	493	94%	1e-153	69%	
CP000738.1	Ochrobactrum anthropi ATCC 49188 chromosome 1, complete sequence	498	498	93%	1e-155	70%	
CP001212.1	Rhodobacter capsulatus SB 1003, complete genome	602	609	93%	0.0	86%	
CP000031.1	Ruegeria pomeroyi DSS-3, complete genome	619	619	93%	0.0	87%	
CP000781.1	Xanthobacter autotrophicus Py2, complete genome	495	495	93%	2e-154	68%	
CP000290.1	Chelativorans sp. BNC1, complete genome	516	514	92%	3e-161	73%	
CP000282.1	Rhodomicrobium vannielii ATCC 17100, complete genome	493	493	92%	9e-154	68%	
CG179514.1	Uncultured Methylocystis sp. G5C357 putative TrkA family protein (trkA), putative beta-ribof	504	504	92%	4e-158	72%	

**Gene Isolation**

After the comparison searches were ran, the next step was to take the gene sequence for RSP1977 and create specific oligonucleotide primers that would allow us to isolate the gene from the genomic DNA of *R. sphaeroides*. The forward primer was created using the first 20 bases starting from the 5' end of the gene sequence while the reverse primer was created using the complimentary bases for the last 21 bases of the 3' end of the gene sequence. Once the primers were designed they were sent to Integrated DNA Technologies to be synthesized.

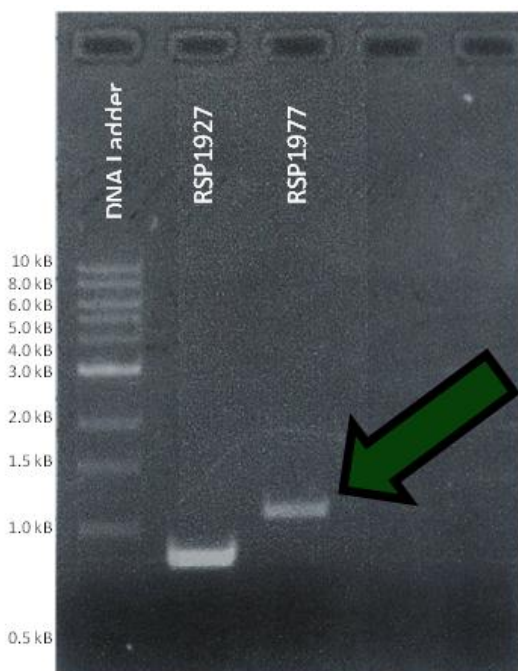
**Table 2: DNA sequences of oligonucleotide primers**

Primer	Sequence	Melting Points (°C)
Forward Primer	5'-ATGGCGACTGCAAGAAGGTC-3'	54
Reverse Primer	5'-TCATTTTCATCGCCGTGGAGGC-3'	56

Upon receiving the synthesized primers, the next step was to isolate RSP1977 using the polymerase chain reaction (PCR) method of isolation and amplification. PCR is a three-step reaction that is repeated multiple times to amplify a specific DNA sequence one million times after twenty five cycles.<sup>6</sup> The reaction involves the use of dNTPs (Finnzymes), template *R. sphaeroides* chromosomal DNA purified using PureLink Genomic DNA Kit (Invitrogen), Phusion DNA Polymerase (Finnzymes), and the designed primers specific to the gene.<sup>9</sup>

The first step of the reaction heats the sample to 98°C in order to denature the chromosomal DNA from double-stranded to two single strands. Then the sample is cooled to 55°C to allow the designed primers to hydrogen bond to their complementary base pairs on the chromosomal DNA strands; the forward primer binds to one strand while the reverse primer binds to the other in order to flank the desired gene and create a double-stranded blunt form of the desired gene. In the final step, the temperature is raised to 75°C, the optimal temperature of activity of the DNA polymerase, so that it extends the primers in the 5' to 3' direction. These three steps are repeated for 25 cycles so that the desired DNA sequence is amplified one million times.<sup>6</sup> The PCR reaction was ran with a total solution containing 50 microliters (µL).

The success of the PCR product was qualitatively assessed by agarose gel electrophoresis which is a widely used separation technique in the analysis of DNA. First, agar is dissolved with heat into the appropriate buffer in an amount to create a 1% agarose solution then cooled in a mold to resolidify forming the gel with an agarose matrix through which the DNA can pass. Next the gel is placed into the electrophoresis box and submerged in TE electrophoresis buffer containing Tris and EDTA to solubilize the DNA while keeping it from degrading. Then 10µL of the PCR product is mixed with blue loading dye containing bromophenol blue and glycerol, to track the DNA as it migrates down the gel, and loaded into the wells of the agarose gel. Next, an electrical current (100V used in our research) is passed through the buffer and the gel resulting in the DNA moving through the gel toward the positive electrode. This occurs because DNA has a uniform negative charge and will therefore travel through an electrical current toward the positive electrode. Since DNA's charge is uniform, the DNA will then travel through the gel matrix based on its size; longer strands of DNA with more mass will travel through the matrix more slowly than smaller DNA fragments resulting in shorter migration distances down the gel. The samples are also run along with a DNA ladder in order to determine the size of the DNA fragments contained in the PCR samples. The ladder is DNA standard that contains fragments of DNA of known lengths so that, based on the separation of these fragments, one can tell the size of a fragment ran in another well of the gel. After running the gel until the blue loading dye can be visually seen to have migrated halfway down the gel, the gel was then placed in a camera that passes ultra violet light through the gel and takes a picture. The ultra violet (UV) light allows for the visualization of the DNA because the agarose gel contained ethidium bromide, which is a dye that fluoresces under UV light and has a high affinity for DNA to which it binds. Figure 2 shows the agarose gel containing the PCR products of the genes RSP1927 and RSP1977. The first lane of the gel contained the DNA ladder, lane 2 contained the PCR product for RSP1927, and lane 3 contained the PCR product for RSP1977. The indicated band for RSP1977 was slightly above the 1.0 kilobase (kB) pair band of the ladder. This was appropriate because RSP1977 is 1095 base pairs in length. This gel therefore indicates successful isolation of the gene RSP1977 from the genomic DNA of *R. sphaeroides*.<sup>9</sup>

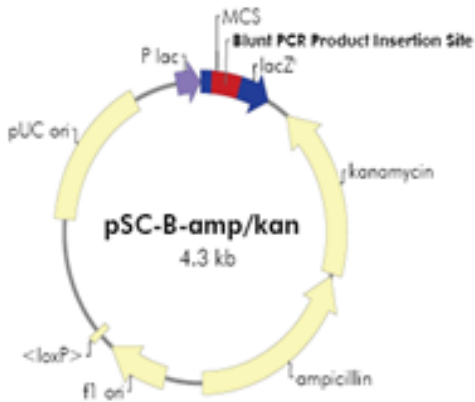


**Figure 2: Electrophoresis Gel of PCR Products**

**Figure 2: Electrophoresis Gel of PCR Products**

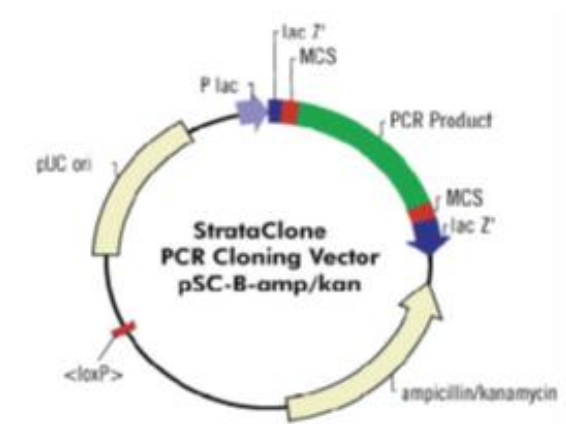
*Insertion of PCR Product into Blunt Vector*

Now that the isolation of the gene RSP1977 was confirmed, the next step in the research was to insert RSP 1977 into the pSC-B-amp/kan blunt vector shown in Figure 3. A plasmid is a circular piece of double stranded DNA which replicates independently of an organisms chromosomal DNA. A vector is a genetically altered or synthetically derived plasmid. Plasmids/vectors are able to replicate separately from the hosts chromosome because they contain *ori* and *rop* replication sites in their sequence which direct the binding of the host cell's replication enzymes.



**Figure 3:** pSC-B-amp/kan Blunt Vector<sup>10</sup>

This allows the independent replication of 5-50 copies of the plasmid per cell. This would allow for further amplification of RSP 1977. In order to ensure the host cell maintains the vector, it contains a selective advantage to the host cell that favors the cells upkeep of the vector. The selective advantage usually allows the host cell to survive under certain growth conditions, such as in the presences of antibiotics; kanamycin and ampicillin as in the StrataClone Blunt vector. This allowed for the experiment to be performed under controlled conditions in which only the host cells that contained the plasmid were able to grow in the presence of these antibiotics.



**Figure 4:** pSC-B-amp/kan Blunt Vector with inserted PCR product<sup>10</sup>

In order to insert the PCR product into the StrataClone Blunt Vector, the protocol from the StrataClone Ultra Blunt PCR Cloning Kit<sup>10</sup> was followed allowing for the insertion of the PCR product into the multiple cloning site (MCS) of the blunt vector. This was accomplished by incubating both the PCR product and blunt vector mixture with Topoisomerase I for five minutes at room temperature. The anticipated product after incubation is shown in Figure 4.

*Transformation and Confirmation*

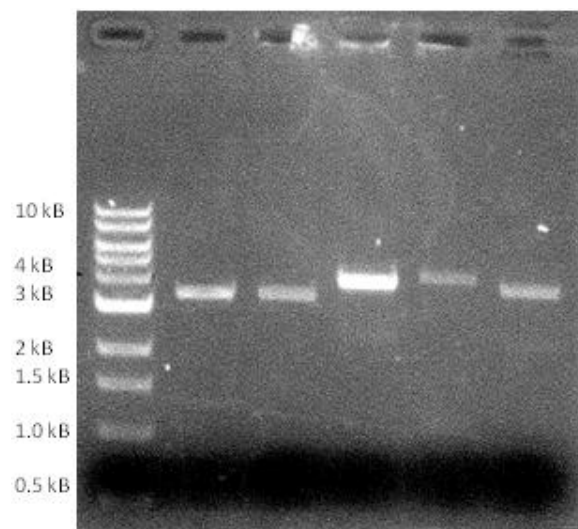
After the blunt vector ligation protocol was completed, the next step was to transform the ligation product into competent *Escherichia coli* cells. We used frozen *E. coli* cells from StrataClone in a tube they call a SoloPack containing CaCl<sub>2</sub>. The SoloPack tube of cells was thawed on ice and 1µL of the ligation reaction mixture was added to the tube and incubated on ice for twenty minutes. The transformation mixture was then heat-shocked at 42°C for forty-five seconds. It is believed that heat shocking causes the release of lipids from the cell membrane causing a decrease in the membrane's fluidity resulting in the formation of pores on the cell's surface. The lowering of the membrane fluidity and pore formation is stabilized by the divalent Ca<sup>2+</sup> cations and aid in the plasmid DNA crossing through the outer cell membrane. Heat shocking almost completely depolarizes the cell's inner membrane and allows the plasmid DNA to cross through the inner membrane of the bacterial cells.<sup>11</sup> After the mixture was heat-shocked, the sample is placed on ice for two minutes to allow the fluidity of the membrane to be restored by the loss of membrane proteins.<sup>11</sup> Next, 250µL of a medium (LB medium) containing the nutrients necessary for cell growth of the *E. coli*, was added to the reaction and the competent cells were allowed to recover at 37°C with agitation for one hour. Then 5µL and 100µL of the competent cells were spread on solid LB-ampicillin plates coated with 40µL of 2% X-gal. The X-gal is a lactose derivative molecule called 5-bromo-4-chloro-3-inodlyl-β-D-galactopyranoside.



**Figure 5:** Blue/White screening Test<sup>9</sup>

X-gal is a substance commonly used in the blue/white screening test of transformed bacteria shown in Figure 5. The X-gal reacts with the enzyme  $\beta$ -galactosidase produced by an operating lacZ gene to form a blue color.  $\beta$ -galactosidase hydrolyzes the glycosidic bond between the galactose and the 5-bromo-4-chloro-3-hydroxyindole. Two molecules of 5-bromo-4-chloro-3-hydroxyindole then spontaneously dimerize to form an insoluble blue product.<sup>12</sup> This blue product is what indicates that the lacZ gene in the vector is functional. This was indicative of the PCR product not being successfully inserted into the blunt vector's multiple cloning site and therefore the gene that codes for  $\beta$ -galactosidase was not interrupted. The white colonies indicated the successful ligation of the gene into the blunt vector and of transformation of the vector into the *E. coli* cells. These white colonies were then chosen from the plate, placed in tubes containing LB medium and overnight liquid cultures were grown.

These overnight cultures were then spun down in a centrifuge, the cells were lysed, and the DNA purified utilizing the MoBio UltraClean Mini Plasmid Prep Kit. The resulting solutions were then tested for purity of the DNA by use of the Nucleic Acid Analysis on the Beckmann DU800 Spectrophotometer. The absorbences of the purified solutions were determined at the wavelengths of 260nm and 280nm. These wavelengths were used because the cyclic nitrogenous bases of DNA will absorb light at these two wavelengths with 260nm being absorbed 1.8-2 times as much as 280nm. Proteins contain amino acids residues that will also absorb light at 280nm. These amino acids are the cyclic tyrosine, tryptophan, and phenylalanine. Therefore, our analysis should have yielded a ratio of 260:280 of 1.8-2 for pure DNA. Ratios less than 1.5 were considered contaminated with protein and were discarded. The DNA concentrations of the pure DNA solutions were then obtained utilizing the Beer-Lambert Law and the known combined molar extinction coefficient of DNA being 0.020.



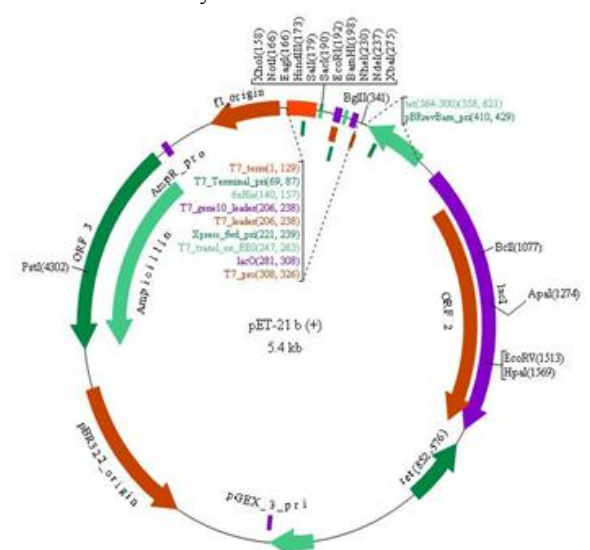
**Figure 6:** Gel Electrophoresis of the Restriction Enzyme Digest Products Using EcoRI Enzyme

To further ensure that the PCR product was successfully inserted into the vector, a restriction enzyme digest was performed to determine if an insert was present in the plasmid and what size the insert was. Restriction enzymes are enzymes that cleave DNA along specific sequences of nucleotides. The blunt vector had two EcoRI sites built into it flanking the position of the insert.

The sample was therefore incubated with the EcoRI enzyme at 37°C for two hours so that the enzyme would cleave the two sites of the vector and would separate the PCR gene product from the vector. This digestion product was then run on a gel using gel electrophoresis as shown in Figure 6. Two separate bands of linear DNA were expected to be seen on the gel, one for the vector at 4300 base pairs and one at 1095 base pairs for the RSP1977 gene insert that had been removed by the restriction digest. As clearly seen, only the band at 4300 base pairs is visible for all of the samples tested. This would indicate that the PCR gene product of RSP1977 was not present in the digested samples. Therefore, at some point in the research, either the gene insert was lost or mutated out of the vector or was simply never successfully inserted and the white colonies on the X-gal containing plates were simply false positives. The steps taken up to this point were repeated several times with the same results.

*Future Plans*

The gene RSP1977 has to date been successfully isolated and cloned, but successful insertion into the pSC-B-amp/kan vector could not be confirmed. Upon confirmation of successful insertion and transformation into competent cells, the next step in the research would be to insert the gene into the pET-21 b(+) expression vector shown in Figure 7. Once this vector is transformed into competent cells, it will produce (express) multiple copies of the protein for which the gene insert encodes. The pET vector contains an MCS that is different from that of the MCS in the blunt vector. The MCS of the pET vector contains several restriction enzyme sites that cleave the DNA with overhanging bases on one strand. These are referred to as “sticky ends.”



**Figure 7:** pET-21 b(+) expression vector

These sites require analogous sites to be on the insert for proper ligation of the insert into the vector. Therefore, new primers for RSP1977 would have to be created. These primers would contain the same bases as the original primers, but the base pair sequences recognized by the two restriction enzymes desired would be added to the primers. One site would be added to the beginning (5' end) of the first primer, the 5' end of the gene, and one would be added to the 5' end of the second primer, the 3' end of the gene. The restriction sites for NdeI and HindIII would be used for these respectively. These new primers would then be synthesized by Integrated DNA Technologies and used to isolate RSP1977 from the blunt vector so that it will then contain the new restriction sites. It would then be digested with NdeI and HindIII, as would the pET vector, and then RSP1977 would be ligated into the expression vector. This ligation product would then be transformed into *E. coli* competent cells which would express the protein for which RSP1977 encodes. The protein would then be collected and purified through affinity chromatography. The purified protein would then be characterized through various biochemical techniques such as spectroscopy, affinity chromatography, and an enzymatic assay. Since the probable function of RSP1977 is a subunit of a cobalt insertion protein, its affinity and uptake of cobalt would be tested as well as its affinity for cobalamin derivatives.

## References

1. Mackenzie, C.; Eraso, J.M.; Choudhary, M.; Roh, J.H.; Zeng, X.; Bruscella, P.; Puskás, Á.; Kaplan, S. Postgenomic Adventures with *Rhodobacter sphaeroides*. *Annu. Rev. Microbiol.* 2007, 61, 283-307.
2. *Rhodobacter sphaeroides*. MicrobeWiki. Kenyon College. [http://microbewiki.kenyon.edu/index.php/Rhodobacter\\_sphaeroides](http://microbewiki.kenyon.edu/index.php/Rhodobacter_sphaeroides). August 20, 2010.
3. Choudhary, M.; Mackenzie, C.; Donohue, T.J.; Kaplan, S. Purple Bacterial Genomics. In *The Purple Phototrophic Bacteria*; Hunter, C.N., Daldal, F., Thurnauer, M.C., Beatty, J.T., Eds.; Advances in Photosynthesis and Respiration; Springer Science + Business Media B.V.: The Netherlands, 2009, 28, 691-706.
4. Madigan, M.R.; Jung D.O. An Overview of Purple Bacteria: Systematics, Physiology, and Habitats. In *The Purple Phototrophic Bacteria*; Hunter, C.N., Daldal, F., Thurnauer, M.C., Beatty, J.T., Eds.; Advances in Photosynthesis and Respiration; Springer Science + Business Media B.V.: The Netherlands, 2009; Vol. 28, pp 1-15.
5. *Rhodobacter sphaeroides* 2.4.1 Home Page. JGI: DOE Joint Genome Institute. U.S. Department of Energy. Office of Science. <http://genome.jgi-psf.org/rhosp/rhosp.home.html>. November 21, 2011.
6. Nelson, D.L.; Cox, M.M. *Lehninger Principles of Biochemistry*, 5<sup>th</sup> ed.; W.H. Freeman and Company: New York, 2008; p 271, 303-338.
7. Rutherford, K.; Parkill, J.; Crook, J.; Horrsell, T.; Rice, P.; Rajandream, M.A.; Barrell, B. Artemis: sequence visualization and annotation. *Bioinformatics* 2000, 16, 944-945.
8. McGinnis, S.; Madden, T.L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 2004, 32, W20-W25.
9. Dr. Setterdahl, A.; Green, A.; Unpublished Research on Characterizing *Rhodobacter sphaeroides* gene 1977.
10. Strata Clone Blunt PCR Cloning Kit Instruction Manual [www.genomics.agilent.com/files/Manual/240218.pdf](http://www.genomics.agilent.com/files/Manual/240218.pdf).
11. Hanahan, D. (1983). "Studies on transformation of *Escherichia coli* with plasmids". *Journal of molecular biology* 166, 4, 557-580.
12. Kiernan, J.A. 2007. Indigogenic substrates for detection and localization of enzymes. *Biotech Histochem.* 2007, 82(2), 73-103.