

**Molecular Genetic Studies on the Light Harvesting
– Reaction Centre (LH-RC) Genes in the Purple
Non-Sulphur Bacterium *Blastochloris viridis***

Adhie Dwi Putranto Reksodipuro

Submitted in part of fulfilment for the Degree of M.Sc

University of Glasgow

Division of Cell and Molecular Biology

February 2009

© Adhie Dwi Putranto Reksodipuro 2009

Declaration

I declare that this thesis has been written in accordance with University regulations and is less than 50,000 words in length. All work contained herein was perform by author unless otherwise stated.

Adhie Dwi Putranto Reksodipuro

February 2009

Abstract

The 3D structure of *Blastochloris (Bl.) viridis* photosynthetic reaction centre (RC) was the first membrane protein structure determined by X-ray crystallization (Deisenhofer, et al., 1984; Deisenhofer, et al., 1985). At the same time, Michel *et al* determined the amino acid sequence of the reaction centre polypeptides (Michel, 1982; Michel, et al., 1985; Michel, et al., 1986; Wiessner, et al., 1990). However, recent investigations on a high resolution protein structure of this RC (Roszak, et al., unpublished observations) have found that several amino acids positions in the sequence did not fit the newer, higher resolution structure. The aim of this project, therefore, was to re-sequence the genes of the β and α subunits of the light-harvesting complex (LHC) and the L, M, H and Cytochrome *c* subunits of photosynthetic reaction centre (RC) from *Blastochloris (Bl.) viridis* to see if new sequence data would produce a better fit to the high resolution x-ray structure.

All these genes were re-sequenced and the following changes were found; 5 changes for *pufM* (M29, M63, M70, M74 and M164), 8 for *pufC* (C43, C77, C84, C252, C277, C287, C288 and C323), and 2 for *puhA* (H216 and H256). The new amino acids that replaced the previous sequence at each position in PufM were as follows; Ile for Val (M29), Ala for Ser (M63), Gly for Leu (M70), Leu for Ala (M74) and Ala for Thr (M164), for PufC: Pro for Ala (C43), Met for Ile (C77), Glu for Gln (C84), Ser for Thr (C252), Met for Leu (C277), Thr for Ala (C287), Val for Ser (C288) and Gln for Lys (C323), and finally, for PuhA: Asp for Glu (H216) and Ala for Ser (H256).

The new sequences were then fitted into the electron density maps of the higher resolution crystal structure. The changes in the primary structure of the polypeptides resulted in a much better fit. In particular the new sequence data resulted in a significant change in the structure of the reaction centre carotenoid so that it is now much more consistent with data from other reaction centre structures.

Acknowledgements

Firstly, I would like to thank Allah for all the air that I breathe, the food that I eat, all unexpected experiences and all guidance in life that keeps me reminded that there's limit for everything.

I would also like to take this opportunity to thank my supervisor, Professor Richard J. Cogdell, for all his assistance, encouragement, guidance and patience. He accepted me as his student in the first place, even my English ability was limited. I will not have been able to set my foot in Glasgow without his recommendation. His energy and sense of humour helped ease the pressure of my one year of study. He was always available and kept me sane and on track with all of his great advices. On a personal level, he is a friend as well as a mentor, and I will always be grateful for his many words of wisdom.

I owe thanks to all of the current group members for useful discussions, encouragement, and friendship. I thank Vladimira for all her guidance in the lab and being the most patience person when explaining every single case about the project and also introduce me to several Czech words (I already forgot all of it now, sorry Vladi). I thank Alastair for all his help of recognising that I am destroying his language by writing my thesis in English (haha!). I really appreciate for his willingness for 'reviewing' my thesis and for teaching me how to write in English properly. From now on I will support his football club, The Arsenal FC (maybe to become the EPL runner up! ManU will be the Champions, of course). I thank Aleks for all his time in producing the electron density pictures and explaining everything very precisely. I also thank June for all her help in getting me familiar with the lab. and to show me how to make my first 'Bose' media. I also thank Mads who gave me a chance to be introduced me 'a wee bit' to his project, even if I did not have time finish it.

I also thank Tatas who always supported me when I was 'lost' in the lab. and to his wife, Shinta Limantara, who trained me in all the things required for my study and convinced me to always aim high. A lot of thanks to Niall, Gillian, Susan and Alette who were always friendly to me and gave me the opportunity to improve my English as well as try to understand 'Scottish', language and culture every time I got confused. Finally, I would also like to thank Gillian Murray for her kindness of teaching me how to do Ceilidh dancing, even though I stepped onto her feet lots of time. Again, thank you all for the greatest experience in my life.

I acknowledge the financial support from the National Education Department of Indonesia (DEPDIKNAS) for my study at the University of Glasgow. I also thank Satya Wacana Christian University for their link with the University of Glasgow that allowed me to have the chance to go to Glasgow.

I also need to thank my mother and my brothers (Heru and Rory) for always supporting me no matter what path I chose and for always providing guidance and encouragement. I really hope my father can see me and be proud of me from his final resting place.

Big thanks to Tante Wiwiet and Pa' de Larmo, who treated me like their son when I was lost, back then. I must also thank Om Aryo and Tante Ninuk, Eyang Tien, Logan and Claire for all their kind support, Om Budi and Tante Nanan and also MaUd and Mba Tanti for their financial help during my stay in Glasgow, Om Astu for his greatness to introduce me with the leader of this DEPDIKNAS scholarships project, Om Eko and Tante Reny for giving me two 'precious' Donington's GP tickets (I discovered the love of my life there). Also I thank the rest of my family; Tante Hanny, Om Pri, Tante Titiek, Om Mahakam, Mba Nie, Iwin, Harry, Yanti, Diptok, Tomi, Soni, Fitri and Hendro

Thanks too to Agustina, Novita and Ratih, for believing in all important matters that related to the preparation of our study since the first day we were candidates for the scholarship.

I also thank all members of the Scotland's Indonesian Student Association (PPI-Scotland) for all their help in gave from a place to stay during my first days in Glasgow right up to now. For all other friends who I cannot mention, thanks for all your help. Lastly, special thank's to my soon to be wife (Insya Allah; by the will of Allah, Amin), Retno, for all her love and support. You're the best and I love you.

Abbreviations

Å	Angstrom
α	Alpha
AppA	Activation of photopigment and <i>puc</i> expression
ATP	Adenosine triphosphate
B800	Bacteriochlorophylls that absorb at 800 nm
B850	Bacteriochlorophylls that absorb at 850 nm
B875	Bacteriochlorophylls that absorb at 875 nm
Bchl <i>a</i> / Bchl <i>b</i>	Bacteriochlorophyll <i>a</i> / <i>b</i>
Bpheo/ A/ B	Bacteriopheophytin in the A branch/ in the B branch
β	Beta
<i>Bl</i>	<i>Blastochloris</i>
bp	Base pair
C20/40	Carbon at position 20/40
Car/ ^T	Carotenoid/triplet state
cDNA	Coding deoxyribonucleic acid
Chl <i>a</i> /Chl <i>b</i>	Chlorophyll <i>a</i> / Chlorophyll <i>b</i>
°C	Degree Celcius
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acids
dNTP	Deoxyribonucleotide triphosphate
dTTP	Deoxythymidine triphosphate
EDTA	Ethylene-diamine tetra acetic acid
EM	Electron micrograph
fs	Femtosecond
Fl	Fluorescence
FnrL	Fumarate nitrate regulator
HOMO	Highest unoccupied molecular orbital
ICM	Intracytoplasmic membrane
IPTG	Isopropyl β -D-thiogalactosidase
ISC	Intersystem crossing
kb	Kilobase (10^3 base)
kD	Kilodalton (10^3 dalton)
L	Litre
LH1	Light-harvesting complex 1
LH2	Light-harvesting complex 2
LHC	Light-harvesting complex
LUMO	Lowest unoccupied molecular orbital
μ l	Microlitre (10^{-6} litre)
M	Molar
NADPH	Nicotinamide adenine dinucleotide phosphate
ng	Nanogram (10^{-9} gram)

nM	Nanomolar (10^{-9} molar)
ns	Nanosecond (10^{-9} second)
ORF	Open reading frame
P/ P*	Special pair/ excited state
PCR	Polymerase chain reaction
PET	Photoinduced electron transfer
PGC	Photosynthetic gene cluster
Ph	Phosphorescence
PpsR	Photopigment suppression
Prr/B/A	Photosynthetic response regulator/ B/A
ps	Picosecond (10^{-12} second)
PS1	Photosystem 1
PS2	Photosystem 2
PSU	Photosynthetic unit
Puf	Photosynthetic unit-fixed
Q _A	Primary electron acceptor ubiquinone
Q _B	Secondary electron acceptor ubiquinone
<i>Rb</i>	<i>Rhodobacter</i>
RC	Reaction centre
rev/min	Revolutions per minute
RNA	Ribonucleic acid
RNAse	Ribonucleases
<i>Rps</i>	<i>Rhodopseudomonas</i>
<i>Rsp</i>	<i>Rhodospirillum</i>
RT-PCR	Reverse transcription PCR
S ₀	Ground state
S ₁	First singlet excited state
S ₂	Second singlet excited state
SDS	Sodium dodecyl sulphate
SpB	<i>sphaeroides puf</i> binding factor
T ¹	First triplet excited state
TspO	Tryptophan rich sensory protein
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Table of Content

Abstract	iii
Acknowledgements.....	iv
Abbreviations	vii
Table of Contents	viii
List of Figures.....	xi
List of Tables	xv
 Chapter 1 Introduction.....	 1
1.1. Photosynthesis	1
1.2. Photosynthetic bacteria.....	2
1.2.1. Purple bacteria	2
1.3. The photosynthetic pigments involved in the light reactions of purple bacterial photosynthesis	5
1.4. The energy transfer pathway in purple photosynthetic bacteria.....	10
1.5. Bacterial light harvesting complexes.....	11
1.5.1. The structure of the light harvesting complexes.....	12
1.5.2. How the light harvesting complexes function	16
1.6. The reaction centre:	20
1.6.1. The structure of the reaction centre	20
1.6.2. How the reaction centres function	23
1.7. Genetics	23
1.7.1. Regulation of photosynthetic unit expression mainly in the level of gene-expression	25
1.8. Polymerase chain reaction	27
1.8.1. Basic principle of PCR	27
1.8.2. DNA polymerase	30
1.8.3. Real time PCR	31
1.8.4. Reverse transcription	33
1.9. DNA Sequencing.....	34
1.9.1. The Sanger dideoxy method	34
1.10. Aim of this project.....	37

Chapter 2 Materials and Methods.....	39
2.1. General culturing methods and cell storage	39
2.2. Growth of photoheterotrophic cultures	39
2.3. Cell harvesting	40
2.4. Genomic DNA isolation	41
2.5. Quantification of the DNA by gel electrophoresis	42
2.6. Polymerase chain reaction	42
2.7. DNA cloning	43
2.8. DNA sequencing	45
2.9. Experimental method overview	45
 Chapter 3 Re-sequencing the genes of light harvesting complex and reaction centre from <i>Bl. viridis</i>	 47
3.1. Introduction	
3.2. Isolation of genomic DNA from <i>Bl. viridis</i>	
3.3. Primers and fragment Amplification	
3.3.1. Amplification of the gene fragments	59
3.3.2. Testing the primers to be used for sequencing	64
3.3.3. PCR products that required cloning before they could be sequenced	66
3.4. Fragment cloning	
3.5. Nucleotide sequences of <i>pufB</i> , <i>pufA</i> , <i>pufL</i> , <i>pufM</i> , <i>pufC</i> and <i>puhA</i>	76
 Chapter 4 Fitting the new sequence data with the high resolution structure of <i>Bl. viridis</i> Reaction Centres	 86
4.1. M subunit	
4.2. C subunit	
4.3. H subunit	96
4.4. Conclusions	96
 References	 107
Appendix 1 – Growh Media	121
Appendix 2 – Standard Solutions	124
Appendix 3 – PCR product (insert): vector Molar Ratios calculation.....	126
Appendix 4 – Abbreviations and Masses of common amino acids.....	127

Appendix 5 – Primer Design Positions in Previous DNA sequence.....	128
Appendix 6 – Raw Sequencing Data	131
Appendix 7 – Amino Acid Translation into The New Sequences	136

List of Figures

Figure 1. Schematic classification of purple photosynthetic bacteria.....	3
Figure 2. Schematic figure of purple photosynthetic bacteria position in their natural habitat.	4
Figure 3. Schematic illustration of different structural types of intracytoplasmic membranes (ICM) where the photosynthetic unit is found.	5
Figure 4. Structures of Chl <i>a</i> and Bchl <i>a</i>	6
Figure 5. Absorption spectra of Chlorophyll <i>a</i> , Chlorophyll <i>c</i> and Bacteriochlorophyll <i>a</i> in diethyl ether.	7
Figure 6. Structure of Bacteriopheophytin <i>a</i> , a metal-free analogue of bacteriochlorophyll.....	7
Figure 7. Chemical structure of several carotenoids found in purple bacteria.....	9
Figure 8. Chemical structure of quinone.....	9
Figure 9. Schematic illustration of the photosynthetic membrane in purple bacteria.....	10
Figure 10. A sucrose gradient is the first step to isolate and purify of LH2 and RC-LH1 complexes from <i>Rps. acidophila</i> strain 10050.....	12
Figure 11. Illustration of the LH2 complex of <i>Rps. acidophila</i>	14
Figure 12. A schematic model of LH1 that interrupted by PufX protein.....	15
Figure 13. Electronic ground and excited state levels of a typical molecule	16
Figure 14. A Jablonski diagram showing the electronic states of a molecule.	17
Figure 15. A schematic model of energy and electron transfer.	18
Figure 16. Illustration of energy transfer times in the light harvesting complex.	19
Figure 17. The photoprotective reactions of carotenoids in quenching singlet oxygen.....	20
Figure 18. Structure of the <i>Bl. viridis</i> reaction centre	21
Figure 19. RC cofactor arrangement (Deisenhofer, et al., 1989), Special pair (P), Bacterioclorophyll (B_A , B_B), Bacteriopheophytin (B_{pHeo_A} , B_{pHeo_B}), Quinone (Q_A , Q_B).	22
Figure 20. Physical and genetic map of the photosynthetic gene cluster of <i>Rb. sphaeroides</i> 2.4.1..	24
Figure 21. DNA amplification with PCR.....	28
Figure 22. Schematic drawing of the PCR cycle.	29
Figure 23. The figure shows real-time PCR response figure between fluorescence and number of amplification cycle.....	32
Figure 24. Exonuclease probe: A single-stranded DNA probe is labelled with a fluorescent dye and a fluorescent quencher..	33
Figure 25. The structure of dideoxy analog; hydrogen (shown in red) replaced hydroxyl so that further elongations will be prevented.....	35
Figure 26. Strategy for chain-termination method for DNA sequencing.....	35

Figure 27. Automated fluorescence detection of oligonucleotide fragments produced by the dideoxy method.....	37
Figure 28. Bacterial cultures in bottles, an anaerobic jar and on a plate.....	40
Figure 29. Gel electrophoresis results of ladder used. Ladders were named based on size difference between each fragment, (A) 1.5kb ladder and (B) 10kb ladder.	42
Figure 30. The promoter and multiple cloning sequences of the pGEM®-T Easy vectors	44
Figure 31. pGEM®-T Easy vector circle map and sequence reference points	44
Figure 32. Overview of the experimental methodology.	46
Figure 33. The genetic map of the <i>puf</i> operon DNA region between <i>pufB</i> , <i>pufA</i> , <i>pufL</i> , <i>pufM</i> and <i>pufC</i>	48
Figure 34. The <i>2Fo-Fc</i> electron density at 1σ level (in blue) and the <i>Fo-Fc</i> density at 3σ level	49
Figure 35. Four examples of putative mistakes in the published amino acid sequence identified based on the electron density maps.....	50
Figure 36. An example of gel electrophoresis result of DNA isolation that compares the use of TE buffer and water.	54
Figure 37. Gel electrophoresis analysis of each step analysis of the DNA isolation.	55
Figure 38. Gel electrophoresis of total genomic DNA obtained when the chloroform wash step was omitted..	56
Figure 39. Illustration of the primers position in the genes map..	58
Figure 40. Gel electrophoresis analysis of the amplification of the <i>pufC</i> fragment (1,561 bp).	62
Figure 41. Gel electrophoresis analysis of the amplification of the <i>pufC</i> fragment.....	63
Figure 42. Gel electrophoresis analysis of the large scale amplification of the <i>pufC</i> fragment (1,564 bp).	63
Figure 43. Gel electrophoresis analysis of the result of testing <i>pufC390virF</i> and <i>pufCvirR</i> primers	65
Figure 44. Examples of sequencing results that obtained by Sanger's dideoxy method.	66
Figure 45. Previous published <i>puhA</i> DNA sequence obtained from GeneBank under accession number; <i>Bl. viridis</i> , X02659.....	67
Figure 46. DNA Fragment resulting from the amplification using the <i>pufC138virF</i> and <i>pufCvirR</i> primers.	67
Figure 47. Gel electrophoresis results of <i>puhA</i> (823 bp) protocol optimisation.	69
Figure 48. Gel electrophoresis result of large scale <i>puhA</i> amplification.	69
Figure 49. Pictures of the plates of the blue-white colour screening to test for successful transformation of <i>puhA</i> in the pGEM© T-Easy plasmid vector.	72
Figure 50. A gel showing the results of the Ruscony test on the white colonies, to test for the presence of the <i>puhA</i> insert.	73

Figure 51. A Gel showing the analysis of the results of the PCR amplification of the inserts of the plasmids that gave the white colonies shown in Figure 49.....	75
Figure 52. A gel showing the result of a larger scale isolation of the plasmid that contained the <i>pufA</i> insert.	76
Figure 53. Sequencing results of <i>pufB</i> and <i>pufA</i> genes using both the <i>pufBvirF</i> and <i>pufL104virR</i> primers.	78
Figure 54. Comparisons between the newly obtained and the previously published amino acid sequences of the PufB and PufA.....	79
Figure 55. Sequencing results of the <i>pufL</i> gene using <i>pufL504virR</i> (upper) and <i>pufLvirF</i> (bottom).	79
Figure 56. Sequencing results of <i>pufM</i> gene using <i>pufL429virF</i> (upper) and <i>pufM47virF</i> (bottom).	80
Figure 57. Sequencing results of the <i>pufC</i> gene using the following primers <i>pufC541virR</i> (upper), SP6 (middle) and T7 (bottom).	81
Figure 58. Sequencing results of the <i>pufA</i> gene using the following primers T7 (upper) and SP6 (bottom).	82
Figure 59. Comparison between the newly obtained and the previously published amino acid sequence of PufL.....	83
Figure 60. Comparison between the newly obtained and the previously published amino acid sequence of PufM..	83
Figure 61. Comparison between the newly obtained and the previously published amino acid sequence of PufC.	84
Figure 62. Comparison between the newly obtained and the previously published amino acid sequence of PufA.	84
Figure 63. The $2Fo-Fc$ electron density at 1σ level (in blue) and the $Fo-Fc$ density at 3σ level (in green-red) at the carotenoid binding site.....	88
Figure 64. Ball-and-stick models of the 1,2-dihydroneurosporene (newly refined conformation as found in the reaction centre from <i>B. viridis</i>	89
Figure 65. Multiple sequence alignment of M-subunit RCs of several species.....	90
Figure 66. Electron density fitted with the old and the new sequence data at position M74.....	91
Figure 67. Electron density fitted with the old and the new sequence data at position M63.....	92
Figure 68. Electron density fitted with the old and the new sequence data at position M164.....	93
Figure 69. Electron density fitted with the old and the new sequence data at position M29.....	94
Figure 70. Electron density fitted with the old and the new sequence data at position C77.....	98
Figure 71. Electron density fitted with the old and the new sequence data at position C252.....	99
Figure 72. Electron density fitted with the old and the new sequence data at position C277.....	100
Figure 73. Electron density fitted with the old and the new sequence data at position C287.....	101

Figure 74. Electron density fitted with the old and the new sequence data at position C288.....	102
Figure 75. Electron density fitted with the old and the new sequence data at position C323.....	103
Figure 76. Electron density fitted with the old and the new sequence data at position C84.....	104
Figure 77. Electron density fitted with the old and the new sequence data at position H216.....	105
Figure 78. Electron density fitted with the old and the new sequence data at position H256.....	106

List of Tables

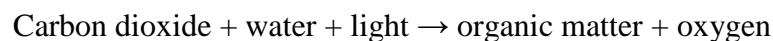
Table 1. List of <i>Rb. sphaeroides</i> genes known or proposed to play a role in regulation of PSU gene expression	26
Table 2. The various volumes used for gel electrophoresis shown in Figure 36.	53
Table 3. List of all primers involved in PCR amplification and sequencing.	58
Table 4. List of primers (in pairs) used in the amplifications.	59
Table 5. PCR protocol for the isolation of the <i>pufC</i> fragment.	61
Table 6. List of primers involved in the DNA sequencing.	65
Table 7. PCR protocol for <i>puhA</i> fragment isolation.	68
Table 8. Ligation protocol used for <i>puhA</i> cloning.	71
Table 9. The PCR protocol for amplification of the <i>puhA</i> insert.	74
Table 10. Summary of all amino acid differences obtained from the new sequence data.	85

Chapter 1 Introduction

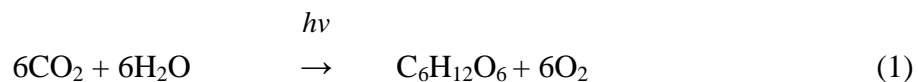
1.1. Photosynthesis

Photosynthesis is the process that uses energy from the sun and ultimately provides all our food and has provided all of the fossil fuels.

The time between the years 1771 and 1804 was an exciting period in the history of photosynthesis (Blankenship, 2002). The role of oxygen in photosynthesis was established by Joseph Priestly in the period 1733-1804 (Jaffe, 1976) and that of light was demonstrated by Jan Ingenhousz between 1730-1799. The involvement of carbon dioxide in photosynthesis was shown by Jean Senebier between 1742-1809 and Nicolas de Saussure discovered the importance of water in the period 1767-1845. As a result of these studies the basic chemical equation of photosynthesis was established, which could be written as:



It was another sixty years before the careful measurements of the photosynthetic quotient, which is the ratio between the carbon dioxide assimilated and the oxygen produced, were made by a Frenchman, T. B. Boussingault in 1864 (Blankenship, 2002). Photosynthesis can be thought of as the photochemical reduction of CO_2 into organic substances (Berg, et al., 2002; Blankenship, 2002). The minimal balanced equation for photosynthesis can now be written as:



Many years after the overall reaction was established, attention has turned to elucidating the details of the mechanism of the process. The photosynthetic process can be conveniently divided into two parts, the light reactions and the subsequent dark reactions. The light reactions are there where solar energy is used to power the synthesis of the ATP and NADPH. The dark reactions, commonly called the Calvin cycle (named after the Nobel Prize winner Melvin Calvin in 1961), are where carbohydrate (carbon) is synthesized from CO_2 in an enzymatic process (Berg, et al., 2002; Blankenship, 2002).

1.2. Photosynthetic bacteria

Plants are not the only types of organism that are able to photosynthesise. Most species of bacteria merely survive by respiration, however, there are others that are able to use photosynthesis to grow (Berg, et al., 2002; Blankenship, 2002). In 1941, Cornelis van Neil carried out a series of experiments on bacterial photosynthesis in which no oxygen was involved. According to van Neil's findings, the bacteria must have access to a reducing substrate in order to provide hydrogen for the further reduction of CO₂ (van Neil, 1941). The general equation he proposed is:



This reaction is a general form of the reaction (1). H₂A may be H₂, H₂S, H₂O or a variety of simple organic molecules. When H₂A is H₂O, A₂ is oxygen.

There are five major types of photosynthetic bacteria are now recognised; purple bacteria, green sulfur bacteria, green non sulfur bacteria, heliobacteria and cyanobacteria (Balows, et al., 1992; Madigan, et al., 1997). The first four are anoxygenic bacteria, meaning that they do not evolve oxygen.

Purple bacteria are one of main groups of photosynthetic organisms that have had an important role in revealing how the process works. These bacteria have a very simple form of photosynthesis and so have been the subject of detailed structural and spectroscopic studies, making them the best understood of all photosynthetic organisms in terms of energy collection and primary electron transfer.

1.2.1. Purple bacteria

Purple bacteria can be classified into sulphur and non-sulphur branches (Imhoff, 1995) dependent on their ability or not to use sulphur as the electron donor in reaction (2). There are two families of purple sulphur bacteria, namely, *Chromaticeae* and *Ectothiorhodospiraceae*. The *Chromaticeae* deposit globules of elemental sulphur inside the cell, whereas *Ectothiorhodospiraceae* produce deposits outside the cell (Imhoff, 1984A).

The *Rhodospirillaceae* family of the non-sulphur purple bacteria is the most diverse family among the phototropic purple bacteria (Imhoff, et al., 1989). It can be divided into 7

genera, namely *Rhodopseudomonas*, *Rhodospirillum*, *Rhodobacter*, *Rhodocyclus*, *Rhodomicrobium*, *Rubrivax* and *Rhodophila* (Willems, et al., 1991; Imhoff, 1995). A schematic classification of purple photosynthetic bacteria can be seen in Figure 1.

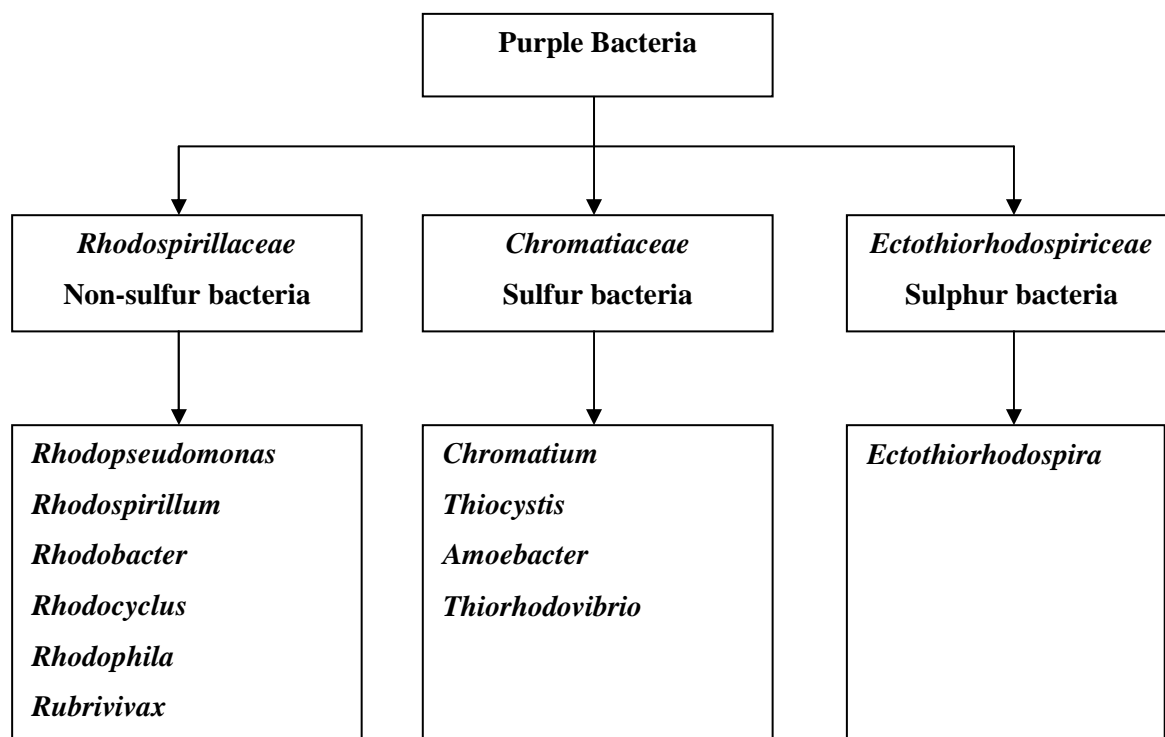


Figure 1. Schematic classification of purple photosynthetic bacteria (Imhoff, et al., 1989; Willems, et al., 1991; Imhoff, 1995)

Most of the Rhodospirillaceae are able to survive by respiration under aerobic and dark conditions. In this growth condition the cells are largely un-pigmented as they survive by respiration. However, when the growth conditions become anaerobic and light is available the cells are very quickly able to change from respiration to photosynthesis (Law, et al., 2008).

The natural habitat of these organisms is in the anaerobic layer in ponds and lakes below the chlorophyll-containing oxygenic phototrophs, see Figure 2 (Cogdell, et al., 2006). In this situation the lack of oxygen causes the cytoplasmic membranes, where photosynthesis takes place, to become invaginated and extended in area. As the result, intracytoplasmic membranes (ICM) are created, into which the cell inserts the necessary complexes to harvest light energy (Drews, et al., 1995). These complexes contain all the necessary pigments and proteins required for light harvesting (Law, et al., 2008).

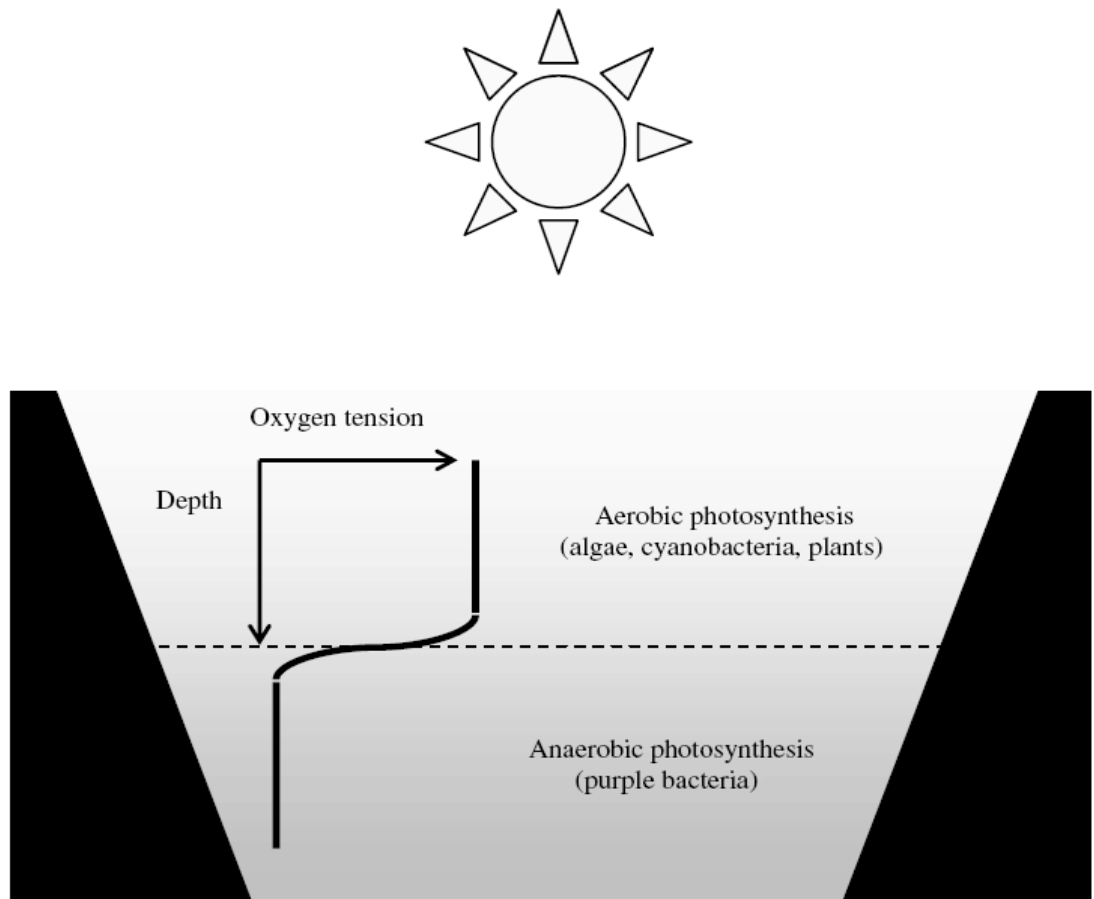


Figure 2. Schematic figure of purple photosynthetic bacteria position in their natural habitat (Cogdell, et al., 2006).

The key point is that the spectrum of the solar energy that reaches the purple bacteria has been filtered by the chlorophyll (Chl)-containing oxygenic phototrophs. These remove the blue and red light (up to 750 nm) and allow only green and near infra red light (above 750 nm) to pass through. It is only these wavelengths of light that are available for purple bacteria. Carotenoid pigments contained in the purple bacteria capture green light while bacteriochlorophyll (Bchl) captures the red light (Cogdell, et al., 2006; Frank, et al., 1991). Most purple bacterial species contain Bchl *a*, whereas a few contain Bchl *b*.

By themselves pigments, such as Bchls *a*, are not able to harvest energy in purple bacteria. It is only when they are bound to specific proteins in the correct orientation that they are able to perform their specific functions in the photosynthetic process. For example when pigments are correctly bound to the reaction centre (RC) polypeptides, they participate in electron transfer, whereas if they are bound to the Light-Harvesting (LH) apoproteins, they can participate in energy transfer (Cogdell, et al., 2006). More details about properties of these pigments will be discussed in section 1.3.

Purple bacteria develop ICM membranes when oxygen is removed from the culture, however, the architecture of the ICM varies from species to species. The shape of these structures is very stable and, therefore, has a great taxonomical value (Oelze, et al., 1972). Several structural types of ICM from purple non-sulphur bacteria are shown in Figure 3.

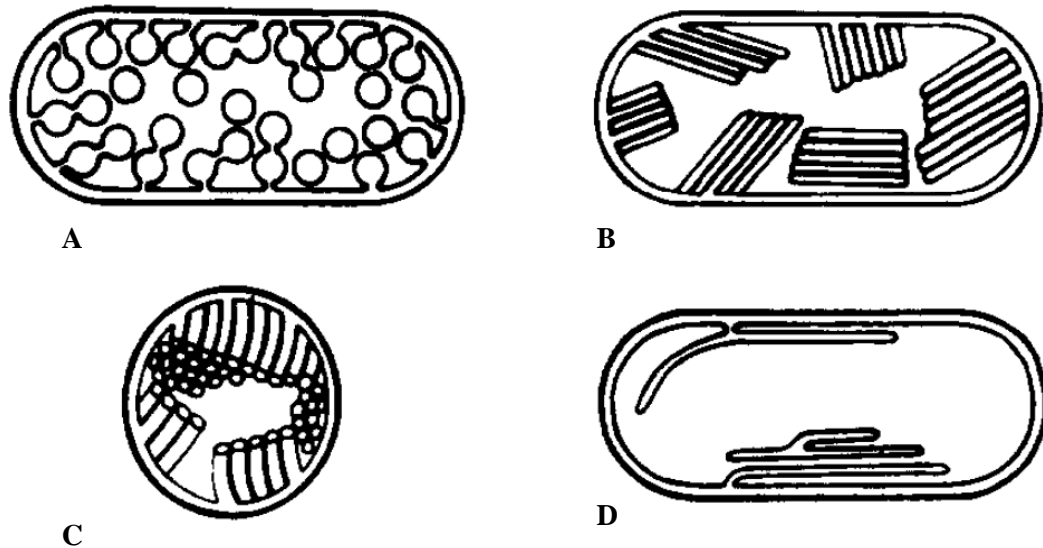


Figure 3. Schematic illustration of different structural types of intracytoplasmic membranes (ICM) where the photosynthetic unit is found (Oelze, et al., 1972). (A) Vesicle type (found in *Rsp. rubrum*). (B) Thylakoid-like, arrange in regular stacks (found *Rsp. molischianum*). (C) Tubuli type (found in *Thiocapsa pfennigi*). (D) Large lamellae membranes (found in *Rps. acidophila*).

1.3. The photosynthetic pigments involved in the light reactions of purple bacterial photosynthesis

In purple bacteria, Bchl *a* is the main light harvesting pigment (Cogdell, et al., 1996). Bacteriochlorophyll is an analogue of the Chl molecules found in green plants. It consists of the bacteriochlorin ring and a phytol chain attached to C-17. The bacteriochlorin ring is planar with a Mg^{2+} ion in the centre, coordinated to four pyrrole nitrogen atoms. A comparison of the chemical structures of Chl *a* and Bchl *a* is presented in Figure 4.

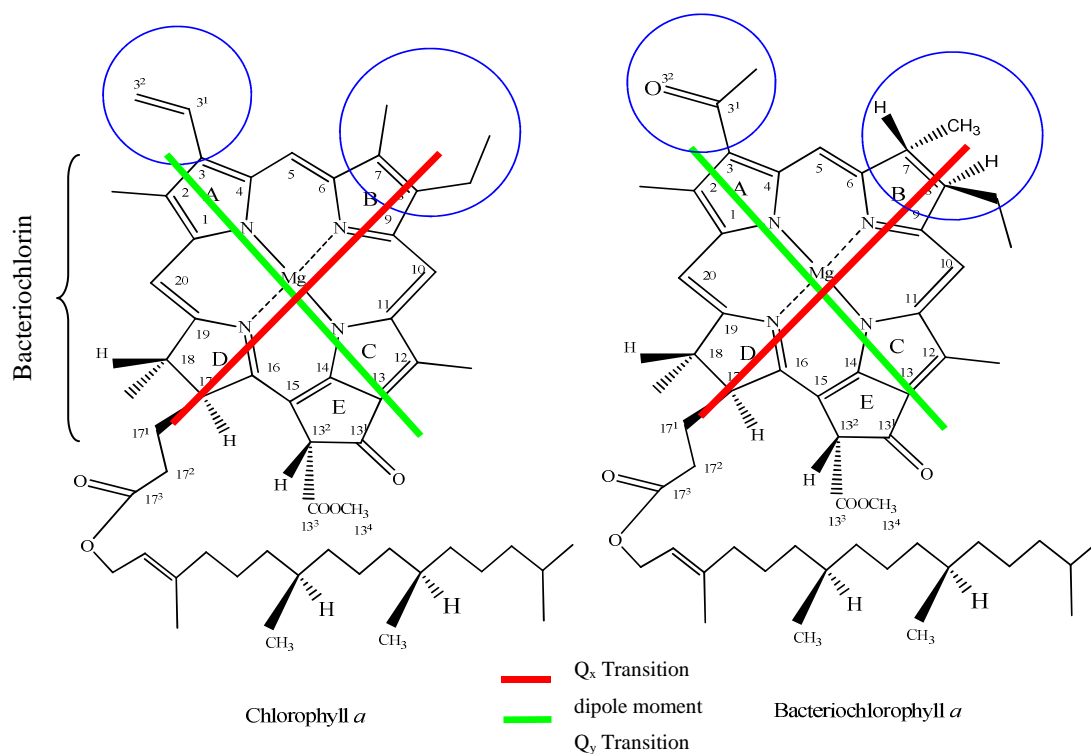


Figure 4. Structures of Chl *a* and Bchl *a* (Kobayashi, et al., 2007). The numbering system is according to the IUPAC standard system. The acetyl group at C-3 and single bond in ring B between C-7 and C-8 are the main differences found in Bchl *a* (marked in blue circles as compared to Chl *a*). The directions of Q_x and Q_y transition dipole moments are shown in red and green respectively.

The structural differences between Bchl *a* and Chl *a* are the acetyl group at the C-3 position and the single bond in ring B between C-7 and C-8. Also in the Chl molecule there is a double bond between C-7 and C-8 in ring B.

Because of their asymmetrically conjugated π system, the electronic transitions of Bchl have two main transition dipole moments, Q_x and Q_y , shown in Figure 4 leading to a characteristic absorption spectrum. The Q_x absorption band can be found in the visible region (700-420 nm) at ~ 570 nm, whereas the Q_y absorption band is in the near infra red region (NIR; >770 nm). A third absorption band, called the Soret band, is located in the blue region of the spectrum (~ 380 nm) and is illustrated in Figure 5.

Bacteriopheophytin (Bpheo) is similar to Bchl but differs only in the absence of the Mg^{2+} ion at the centre of the bacteriochlorin ring (Figure 6).

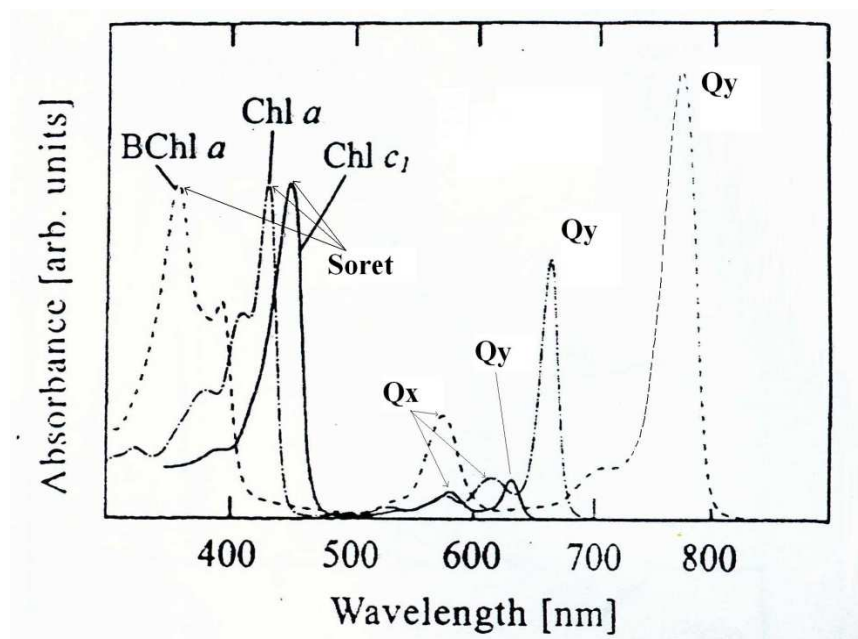
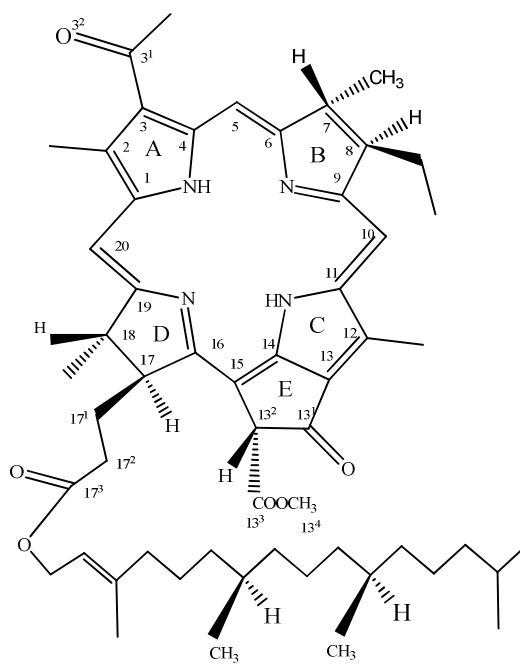


Figure 5. Absorption spectra of Chlorophyll *a*, Chlorophyll *c* and Bacteriochlorophyll *a* in diethyl ether (Kobayashi, et al., 2007).



Bacteriopheophytin *a*

Figure 6. Structure of Bacteriopheophytin *a* (Blankenship, 2002), a metal-free analogue of bacteriochlorophyll.

Carotenoids are long conjugated molecules, which are synthesized by the condensation of molecules of isoprene. They are joined head-to-tail, forming C_{20} , and then tail-to-tail between two C_{20} 's, to form C_{40} . This C_{40} molecule is the precursor of all carotenoids, and is called phytoene (Britton, 1998; Gross, 2006). All other carotenoids are formed from this precursor by reactions such as hydroxylation, desaturation and methylation (Takaichi, 1999; Rodriguez-Amaya, 2001). Carotenoids are found in all known wild-type photosynthetic organisms and in many non photosynthetic organisms (Frank, et al., 2000). They have several important functions in photosynthesis, such as accessory light harvesting, photoprotection via the quenching of Chl (Bchl) triplet states, singlet oxygen scavenging, excess energy dissipation and structure stabilisation (Frank, et al., 1996). The chemical structure of several carotenoids found in purple bacteria can be seen in Figure 7.

Some quinone molecules are also involved in the light-reactions of purple bacteria. Ubiquinones are found in all purple bacteria, whereas menaquinones are found in some purple but in all green bacteria (Imhoff, 1984B). These molecules are involved in the redox reactions in the electron transfer chain within the RC. The chemical structure of ubiquinone is illustrated in Figure 8.

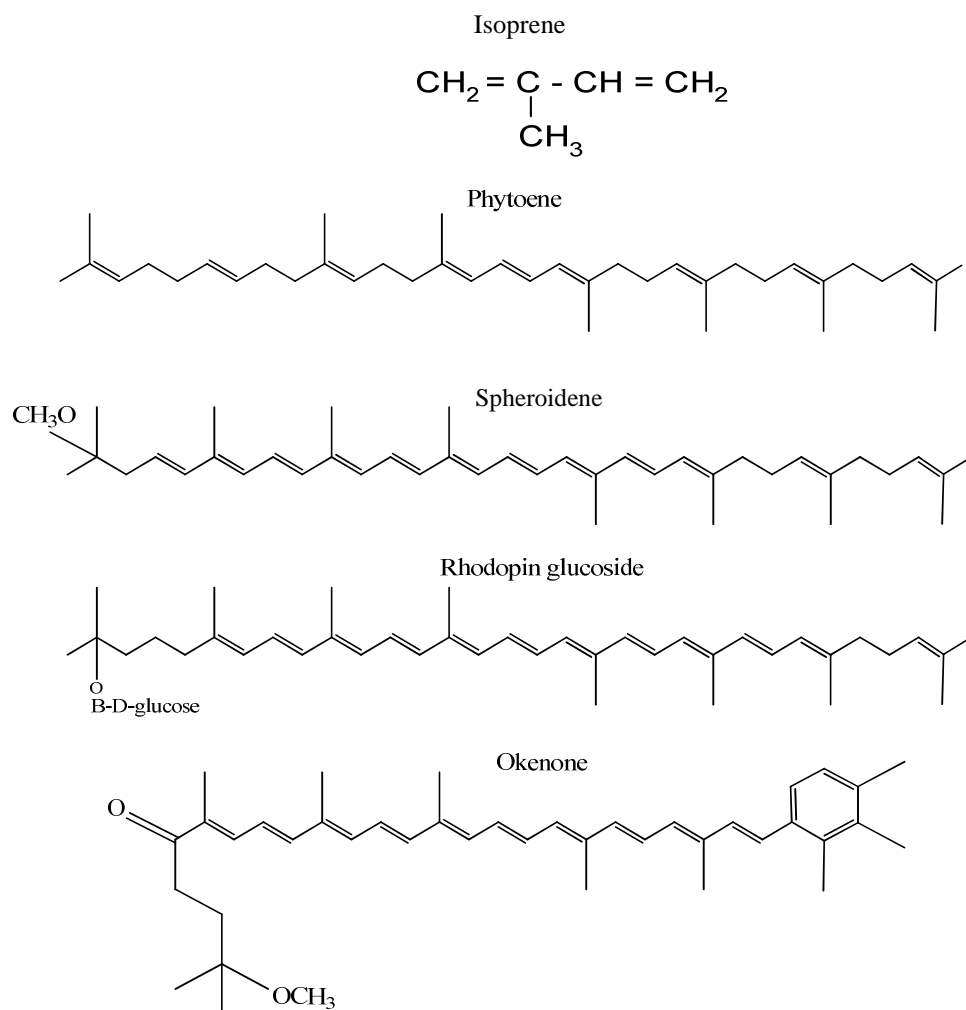


Figure 7. Chemical structure of several carotenoids found in purple bacteria (Cogdell, et al., 2006). Isoprene, is the precursor of all carotenoids,

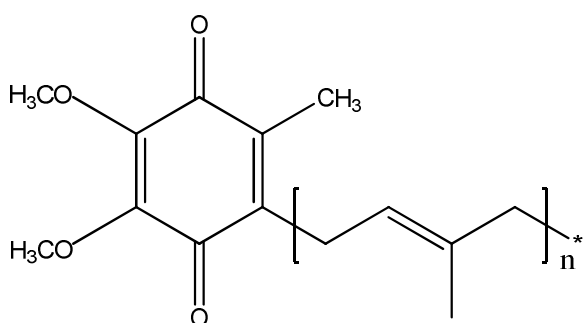


Figure 8. Chemical structure of quinone (Blankenship, 2002). n refers to the number of isoprenoid units that make up the chain ($n = 7 - 10$).

1.4. The energy transfer pathways in purple photosynthetic bacteria

In purple bacteria the photosynthetic apparatus consists of RC and several LH complexes. The RC can absorb photons directly but usually energy is transferred to them from the light harvesting complexes (Cogdell, et al., 1999). In all species of purple bacteria, the reaction centre is surrounded by a light harvesting complex called LH1. This complex is, therefore, called the LH1-RC “core” antenna complex (Brunisholz, et al., 1986; Cogdell, et al., 1996). In some types of purple bacteria a different type of complex, called LH2, is also present and is arranged peripherally around the ‘core’ complex. (Cogdell, et al., 1999).

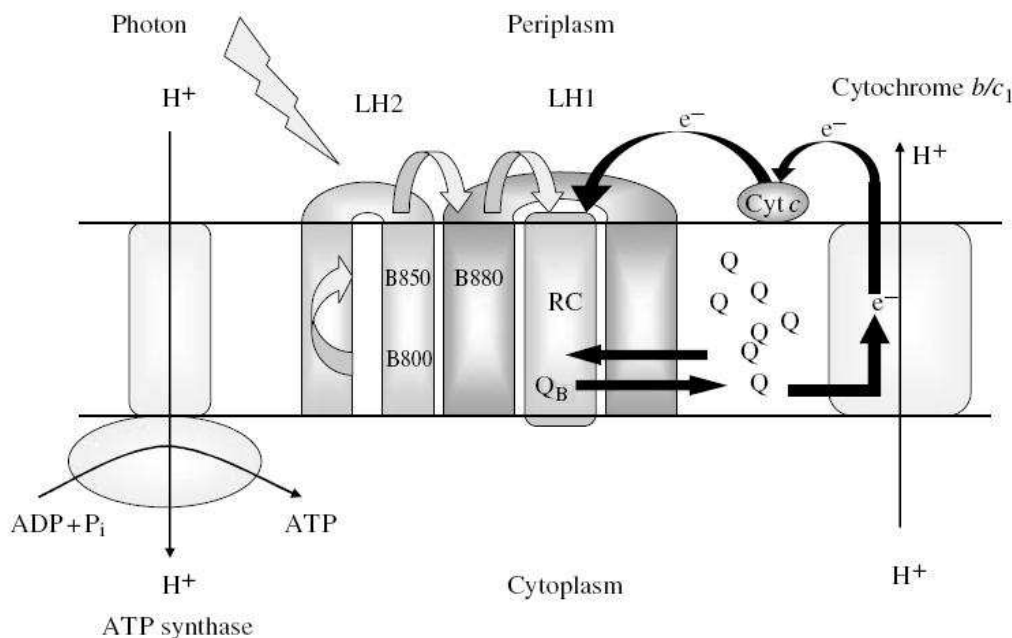


Figure 9. Schematic illustration of the photosynthetic membrane in purple bacteria (Cogdell, et al., 2006). Energy transfer is indicated by the grey arrows, whereas the redox reactions involved in cyclic electron transfer are shown by the black arrows. The cyclic movement of electrons results in a proton gradient across the membrane. This proton motive force is used to drive ATP synthesis.

The overall light reactions in the purple bacterial photosynthetic membrane are shown in Figure 9. The process begins when a photon is absorbed by a Bchl or Carotenoid molecule and this energy is then transferred to the RC. The RC then catalyses a series of electron transfer reactions resulting in charge separation across the membrane. This initial

charge separation reaction is then stabilised by subsequent slower electron transfer reactions where (ubi or mena) quinone is reduced to (ubi or mena) quinol. The reduced quinone molecule diffuses out from the RC-LH1 'core' into the lipid bilayer and migrates to its oxidation site on the cytochrome *b/c₁* complex. The reduced quinone is oxidised and the released proton pumped across the membrane. The electrons are used to reduce cytochrome *c*, which in turn reduces the RC to complete the electron transfer cycle. The resulting proton gradient drives the generation of ATP through the action of ATP synthase (Berg, et al., 2002).

The two well-defined pigment-protein complexes, LH2 and RC-LH1 'core' complex, of *Rhodopseudomonas (Rps.) acidophila* 10050 can be separated on a sucrose gradient and be clearly seen by eye (Cogdell, et al., 1983). This is illustrated in Figure 10 together with the absorption spectrum of each band. LH2 has two strong NIR absorption bands at ~800 and 850 nm, whereas the RC-LH1 complex has a single strong absorption band at ~875 nm. These complexes are also referred to as B800-850 and B875 respectively (Cogdell, et al., 1985). In both complexes, the Bchls and carotenoid in the LH parts are non-covalently bound to two hydrophobic proteins called α and β .

1.5. Bacterial light harvesting complexes

Light harvesting complexes, also known as antenna complexes, have evolved to increase the effective cross-sectional area for light absorption for each reaction centre. Antenna complexes supply each reaction centre with enough photons to allow photosynthesis to occur at reasonable rates over a wide range of incident light intensities. They also absorb a wider spectral range of light than the RC itself, so that more of the incident solar spectrum can be used efficiently (Cogdell, et al., 1999).

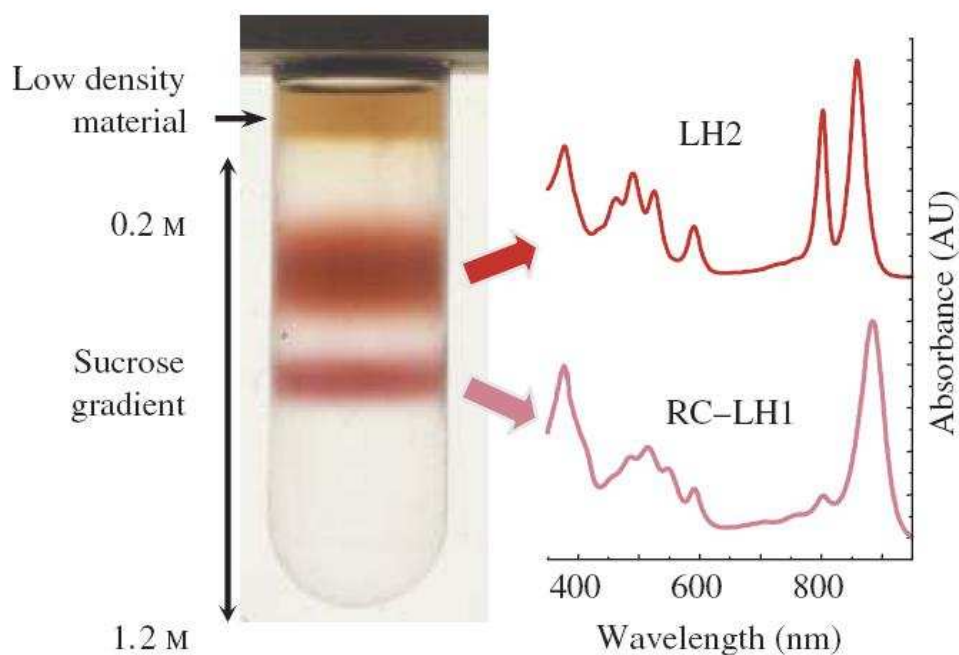


Figure 10. A sucrose gradient is the first step to isolate and purify of LH2 and RC-LH1 complexes from *Rps. acidophila* strain 10050 (Cogdell, et al., 1983). The solubilised photosynthetic membrane was layered onto the top of sucrose-density gradient and centrifuged overnight resulting in the separation of the two major light harvesting complexes. These two complexes have different colours as different carotenoids are bound. The LH2 complex has absorption bands at 800 – 850 nm, while the RC-LH1 complex has absorption band at 875 nm.

1.5.1. The structure of the light harvesting complexes

The 3D crystal structure to 2.5 Å resolution of the LH2 complex from *Rps. acidophila* strain 10050 has been elucidated (McDermott, et al., 1995). This high resolution structure has allowed the structure and function of LH2 to be discussed in detail. The LH2 complex is a circular nonamer, composed by nine pairs of α and β apoproteins (McDermott, et al., 1995). Overall, the complex consists of 18 polypeptides, 27 Bchl *a* and 9 carotenoids molecules. The structure of LH2 complex from *Rps. acidophila* 10050 can be seen in Figure 11. The complex can be split into nine heterodimers, each containing one α and one β apoprotein, one Bchl *a* molecule that absorbs at 800 nm (B800), two Bchl *a* molecules that absorb at 850 nm (B850) and one carotenoid.

The α -apoprotein helices make up the inner wall of the complex, while the β -apoprotein helices form the outer wall. The C-termini of both apoproteins are located at the periplasmic side of the photosynthetic membrane and the N-termini are at the cytoplasmic side (Brunisholz and Zuber 1992; Zuber and Cogdell, 1995). Both of these termini bend over, almost parallel to the membrane surfaces, to enclose the pigments.

The eighteen B850 Bchl *a* molecules are situated in between the helices, with the plane of their bacteriochlorin rings perpendicular to the proposed membrane surface. In each $\sigma\beta$ apoprotein pair, one Bchl is bound to the α -polypeptide and one to the β polypeptide. The nine B800 Bchl *a* molecules are situated between the β polypeptides. The positions of the B800 Bchl rings are almost parallel to the membrane surface (Figure 11).

The major carotenoid molecule present in the LH2 of *Rps. acidophila* strain 10050 is rhodopin glucoside, although there are small amount of others (Gardiner, 1992). These nine molecules are in van der Waals contact with the apoproteins and the Bchl *a* molecules. The carotenoid is found between the two B850 Bchl *a* molecules in the monomer and also comes into the close contact with the B800 Bchl *a* molecules. The carotenoids must be in close contact with the Bchl *a* molecules in order to quench their triplet states, before they can continue to make singlet oxygen.

The RC-LH1 ‘core’ complexes were first visualized in EM pictures of membranes from *Rps. viridis*, more recently referred to as *Blastochloris (Bl). viridis*, and confirmed to have a hexagonal shape surrounding the RC (Hiraishi, 1997, Jay et al. 1984; Miller, 1979, 1982, Stark et al. 1986). Since then, many other studies have suggested a variety of shapes for LH1; circular (Karrasch et al. 1995; Gerken et al. 2003), square (Stahlberg et al. 1998), ‘S’-shaped (Scheuring et al. 2004a, 2005; Siebert et al. 2004), elliptical (Scheuring et al. 2003; Fotiadis et al 2004) or even just arcs (Bahatyrova et al 2004).

Basically, there are at least two distinct classes of RC-LH1 ‘core’ complexes. In one class they are monomeric, namely one RC surrounded by one LH1 complex. Examples of this class are in the RC-LH1 complex of *Rsp. rubrum* and *Rps. palustris* (Gall, 1994; Karrasch et al. 1995). The second class are dimeric, consisting of two RC-LH1 units. This kind of RC-LH1 ‘core’ complex can be found in *Rb. sphaeroides* (Siebert et al. 2004).

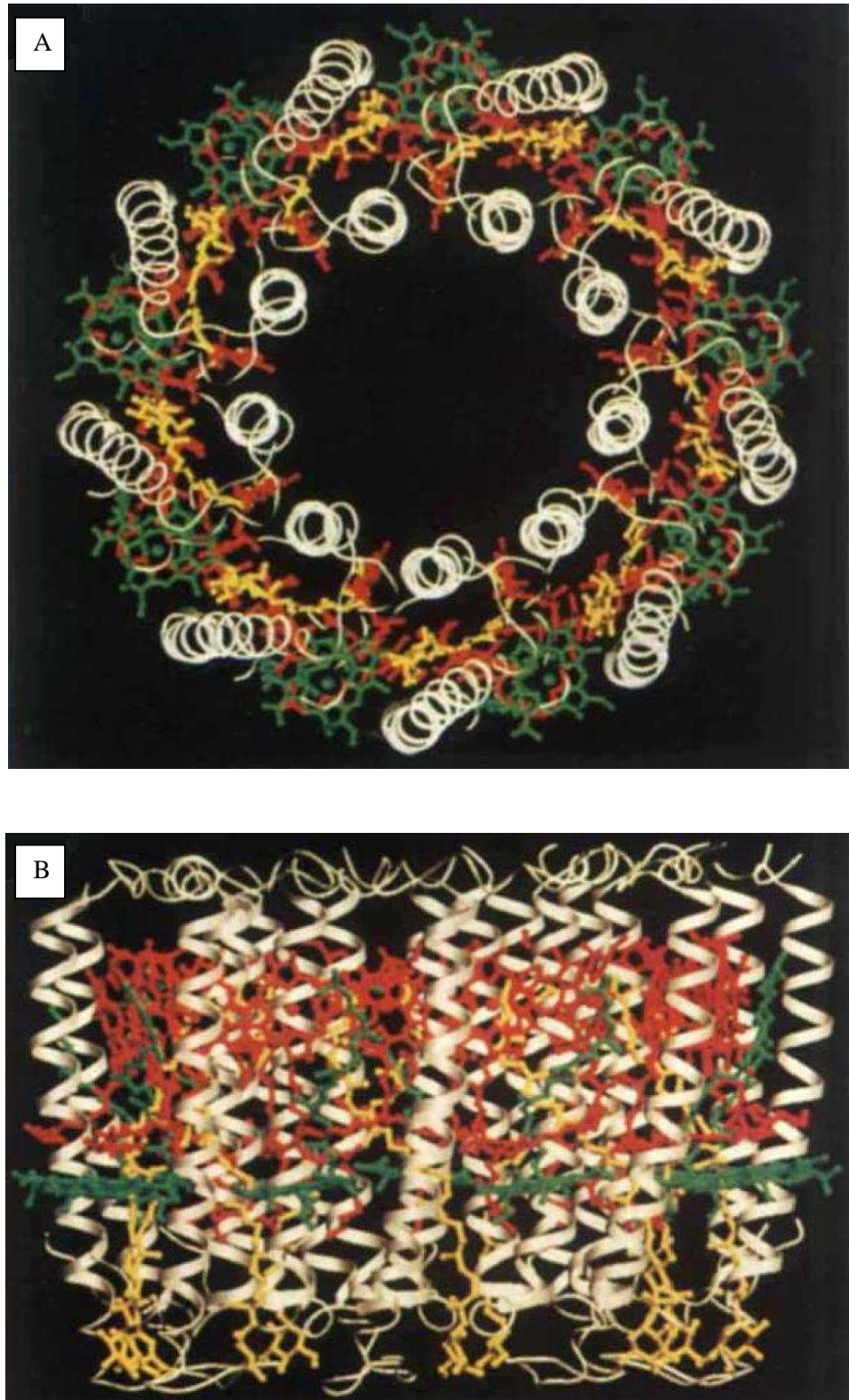


Figure 11. Illustration of the LH2 complex of *Rps. acidophila* (McDermott, et al., 1995). (a) The cytoplasmic side of the complex. Protein units are shown as white (inner as α while outer as β apoprotein), B800 Bchl *a* green, B850 Bchl *a* red, and carotenoid yellow. (b) A view of the complex perpendicular to the symmetry axis.

More recent studies have suggested that LH1 structure in some species, such as *Rb. sphaeroides*, *Rsp. Rubrum* or *Bl. viridis*, is rather flexible (Westerhuis et al. 2002; Scheuring et al. 2003, 2004; Bahatyrova et al. 2004b; Fotiadis et al. 2004). There is a further shape complication found in *Rb. sphaeroides*, which is related to the presence of a protein called PufX. When this protein is present in *Rb. sphaeroides* RC-LH1 complex, they are dimeric, whereas in a PufX⁻ phenotype the RC-LH1 complex is monomeric (Siebert et al. 2004). PufX is a protein that has a role in the photosynthetic cyclic electron transport. It was suggested that the presence of this protein is required to allow a sufficiently rapid redox connection between quinol reduced by the RC with the site of its oxidation on the cytochrome *b/c₁* complex (Barz et al. 1995).

It was further suggested that PufX provides a portal to allow the quinol to escape through the LH1 helices (Cogdell et al. 1996; Parkes-Loach et al. 2001). An illustration of this model can be seen in Figure 12. Further study is needed to reveal detail of the 'core' complex structure, since no high resolution structure of an RC-LH1 complex is available at the moment.

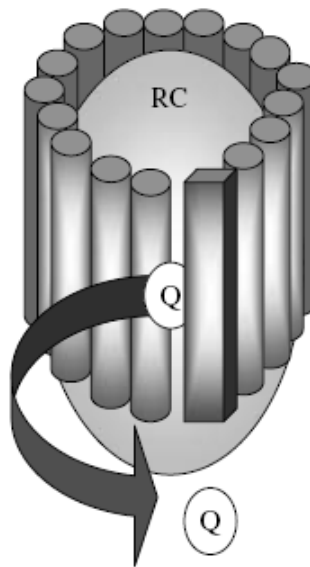


Figure 12. A schematic model of LH1 that is interrupted by PufX protein (Cogdell, et al., 2006). The PufX protein creates a gap in the α/β -dimers rings to allow the migration of quinol (Q) through the LH1 ring into the photosynthetic membrane.

1.5.2. How the light harvesting complexes function

The LH complexes function in a series of photophysical and photochemical processes. These processes are the result of the interaction of light with matter. There are three energy states that play important roles in the photophysics of organic molecules, namely, the ground state (S_0), the lowest singlet excited state (S_1) and the lowest triplet excited state (T_1). The absorption of light by a molecule results in the promotion of an electron from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO), Figure 13, to produce the first excited singlet state, S_1 (Kavarnos, et al., 1986). The singlet excited state (S_1) can decay to the ground state through several radiative and non-radiative processes. All of these processes are noted in a Jablonski diagram (see Figure 14).

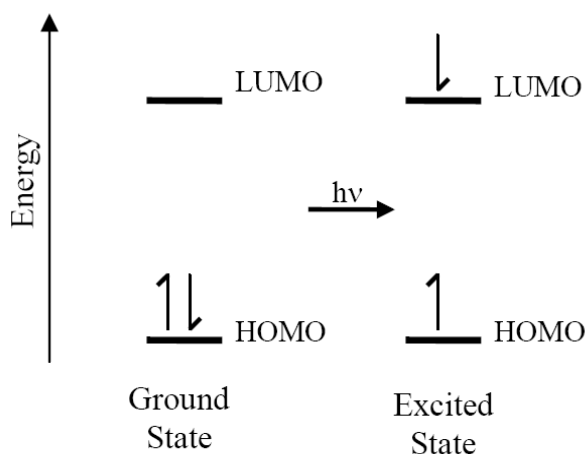


Figure 13. Electronic ground and excited state levels of a typical molecule (Garrison, 2005). An electron will be promoted from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO), creating an excited state, as the result of light absorption by the molecule.

Radiative decay consists of emissive processes, fluorescence (Fl) or phosphorescence (Ph), whereas non-radiative processes consist of internal conversion (IC) and intersystem crossing (ISC) (Kavarnos, et al., 1986). Fluorescence ($S_1 \rightarrow S_0 + h\nu$) is a spin allowed process in which the molecule relaxes to the ground state via emission of a photon of light. Phosphorescence ($T_1 \rightarrow S_0 + h\nu$) is a forbidden radiative process involving emission of light from the triplet state (T_1). IC is non-radiative decay process where energy is lost through vibrational motion ($S_2 \rightarrow S_1 + \text{Heat}$; $S_1 \rightarrow S_0 + \text{Heat}$), and involves a transition between

states of the same spin. ISC is also a non-radiative decay process but it performs a forbidden transition that incorporates an electron spin flip of the electron ($S_1 \rightarrow T_1$; $T_1 \rightarrow S_1$; $T_1 \rightarrow S_0$) (Kavarnos, et al., 1986). The Bchl molecule, for example, when excited may participate in energy transfer or electron transfer processes that compete with the normal modes of relaxation.

Energy transfer is the process whereby energy is transferred from an excited molecule to a second ground state molecule. Electron transfer is a process where an electron is transferred from a donor, excited molecule, to an acceptor molecule (Kavarnos, et al., 1986). When light drives the electron transfer process, it is usually called Photoinduced Electron Transfer (PET). A schematic figure of both processes is illustrated in Figure 15.

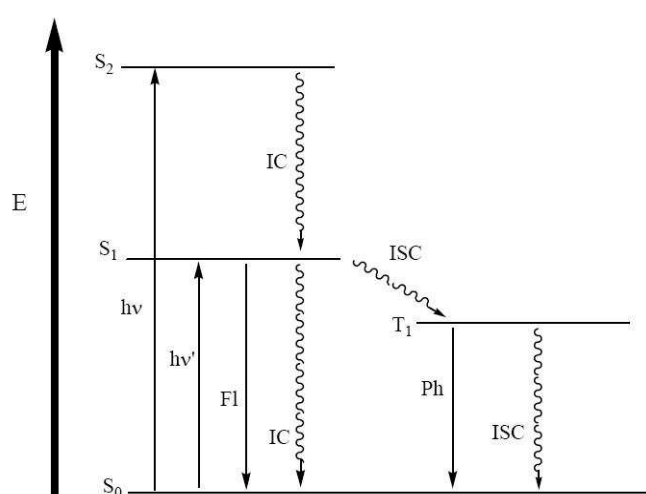


Figure 14. A Jablonski diagram showing the electronic states of a molecule (Garrison, 2005). The vertical scale represents energy and the arrows represent the type of possible transitions between different electronic states. Ground state molecule (S_0) is promoted to an excited state (S_1 or S_2) by absorbing light energy. Non-radiative decay mechanisms (Internal conversion (IC) and Intersystem crossing (ISC)) return the molecule to the ground state whereas radiative transitions (Fluorescence (Fl) or Phosphorescence (Ph)) also decay excited molecule by emitting light.

The light-absorbing pigments are non-covalently bound to two types of integral membrane proteins, forming either reaction centre or antenna complexes (Cogdell, et al., 1999). As mentioned previously in section 1.4, the function of the pigments depends on the protein and where they are bound. LH polypeptides hold the pigments in the correct orientation so that they are able to perform light-harvesting. In this case the protein

‘scaffold’ allows these pigments to absorb light energy and then efficiently and rapidly transfer that energy to the RC (Brunisholz, et al., 1985).

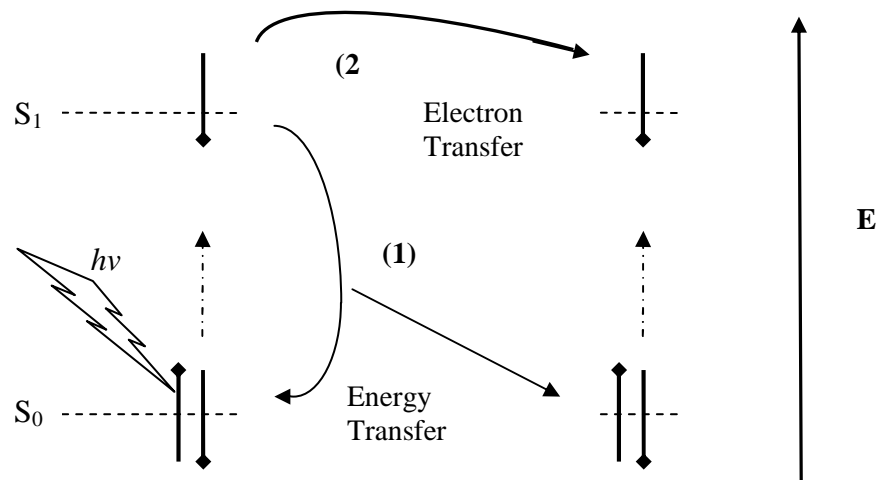


Figure 15. A schematic model of energy and electron transfer. If a photon is absorbed at the correct wavelength, an electron is excited from the ground state (S_0) to the singlet state (S_1). (1) Energy transfer occurs if the singlet state electron returns to the ground state by releasing the energy which is then captured by a second ground state molecule. (2) Electron transfer is process where the singlet state electron is transferred from donor into acceptor molecule (Bchl as an example).

The pathway of energy transfer in the purple bacterial antenna complexes is as follows: B800→B800, B800→B850, B850→B850, LH2→LH1 (B850→B875) and LH1→RC. Initially, when a LH2 B800 Bchl *a* molecule absorbs a photon the molecule is excited to the first excited singlet state (B800*). B800→B800 singlet-singlet energy transfer can occur. From B800*, energy can then be transferred to the B850 molecules that are excitonically coupled in a ring (Bergstrom et al, 1988; Sundstrom and van Grondelle, 1995).

Rapid energy migration around the B850 ring is functionally important in the bacterial photosynthetic membrane. The energy migrates around the B850 molecules very rapidly >100-200 fs (van Grondelle et al, 1994). The rate of energy transfer from *Rps. sphaeroides* and *Rb. acidophila* LH2 (B850) → LH1 (B875) has been time resolved (Sundstrom et al, 1986; van Grondelle et al, 1987, 1984; Zhang et al, 1992b; Hess et al, 1995; Nagarajan, et al., 1997). This energy transfer takes place in 2-5 ps. The energy transfer from B875 → RC takes between 20-50 ps (Sundstrom et al, 1986; Visscher et al. 1989; Zhang et al. 1992a;

Beekman et al. 1994; Freiberg et al. 1996). These energy transfer steps are illustrated in Figure 16.

Carotenoids function as accessory LH pigments and have a stabilising role in the complex (Lang and Hunter, 1994). However, the most important role of carotenoids in LH complexes is photoprotection. In the absence of carotenoids, over excitation of Bchl *a* can lead to the production of Bchl *a* triplets and on to the production of singlet oxygen (Foote, 1976), as seen in Figure 17. Singlet oxygen is very toxic to the cells. When carotenoids are present, energy transfer between Bchl *a* triplets and carotenoid leads to the production of carotenoid triplets rather than singlet oxygen (Monger, et al., 1976; Frank, et al., 1991; Bittl, et al., 2001).

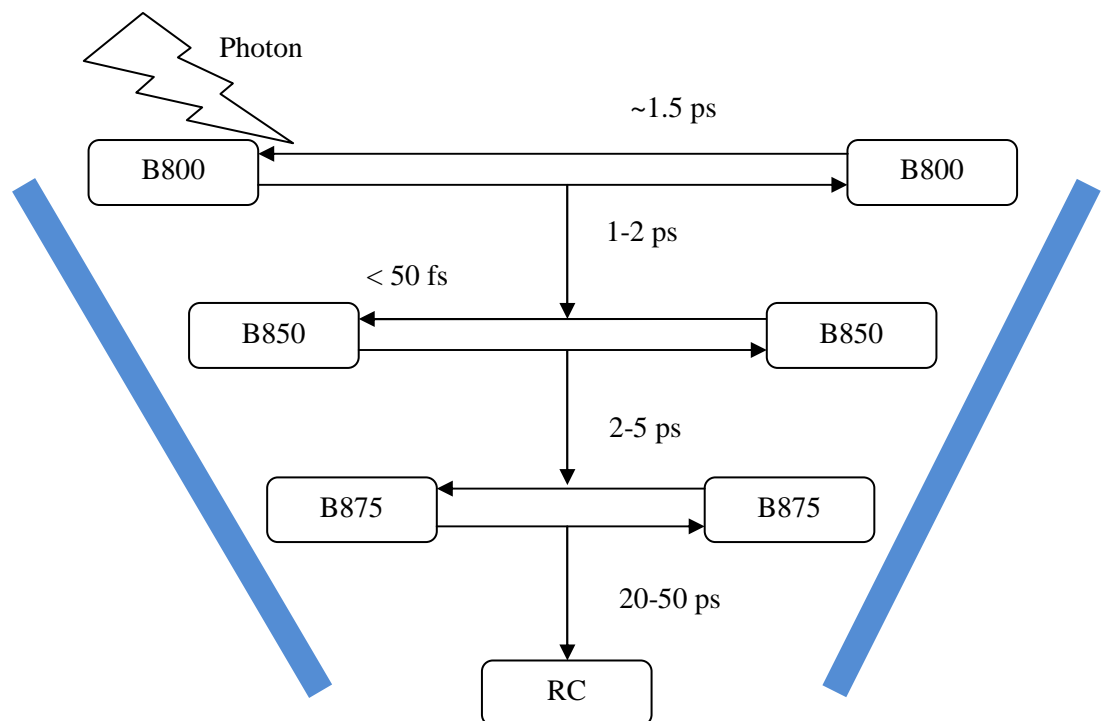


Figure 16. Illustration of energy transfer times in the light harvesting complex. Transfer between B800 takes about 500 fs. Transfer from B800 to B850 takes about 1-2 ps. Transfer between B850 takes about 100-200 fs. Transfer from B850 to B875 (LH1) takes about 2-5 ps. Then the energy migrates to RC in about 20-50 ps. The energy “funnels” its way to the RC, through the LH (blue line).

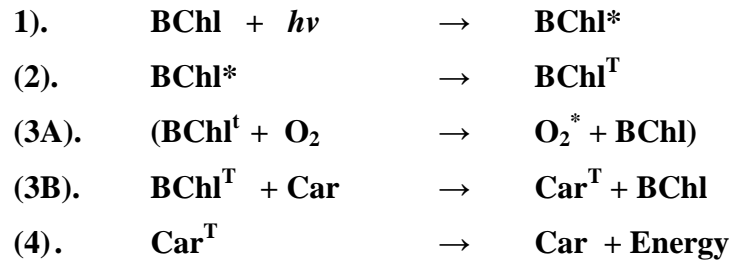


Figure 17. The photoprotective reactions of carotenoids in quenching singlet oxygen (Frank, et al., 1996). (1) After absorption of light, Bchl is promoted to an excited singlet stage. Over excitation of Bchl leads to the production of Bchls triplets (2), possibly resulting in the production of singlet O₂ (3A). However, this singlet O₂ production is prevented in the presence of Car, which can quench the Bchl triplets (3B). The triplet energy contained in Car^T is then released (4).

1.6. The reaction centre

The RC is a pigment-protein complex, in which energy from LH is used to drive the charge-separation reaction across the photosynthetic membrane (Cogdell, et al., 2008). The RC consists of both (bacterio) chlorophylls and other electron transfer cofactors, such as quinones or iron sulphur centres, along with hydrophobic peptides. The RCs contain a “special pair” of chlorophylls, which have unique properties because of their environment in the RC protein (Blankenship, 2002). Anoxygenic bacteria, such as purple bacteria, have only one type of RC. In contrast, higher plants, algae, cyanobacteria, (oxygenic photosynthetic organisms) have two RC called Photosystem 1 (PS1) & Photosystem 2 (PS2). Purple bacteria contain a pheophytin-quinone type RC similar to PS2.

1.6.1. The structure of the purple bacterial reaction centre

Most purple bacterial reaction centres contain three protein subunits; Heavy (H), Medium (M), Light (L) and several have an additional bound cytochrome *c* (C) (Lancaster, 2008). The RC from *Bl. viridis*, has this additional cytochrome subunit and the structure is shown in Figure 18. The L and M subunits have five transmembrane helices each, whereas the H subunit has only one. The L and M subunits in all bacterial RC have a similar overall protein fold (Lancaster, 2008) and are located both at the periplasmic and cytoplasmic

sides of the membrane. The H subunit is located on the cytoplasmic side of the membrane alone.

The C subunit (Figure 18), consists of five segments that bind covalently four iron-containing heme groups. These are; the N-terminal segment, the first heme binding segment, a connecting segment, a second heme-binding segment and the C-terminal segment. The two segments that contain binding sites for the heme groups make up the core of the cytochrome, while the other three non-heme binding segments contain little secondary structure. The first heme binding segment contains the binding sites for heme-1 and heme-2 whereas the second contains the binding sites for heme-3 and heme-4 (Lancaster, 2008).

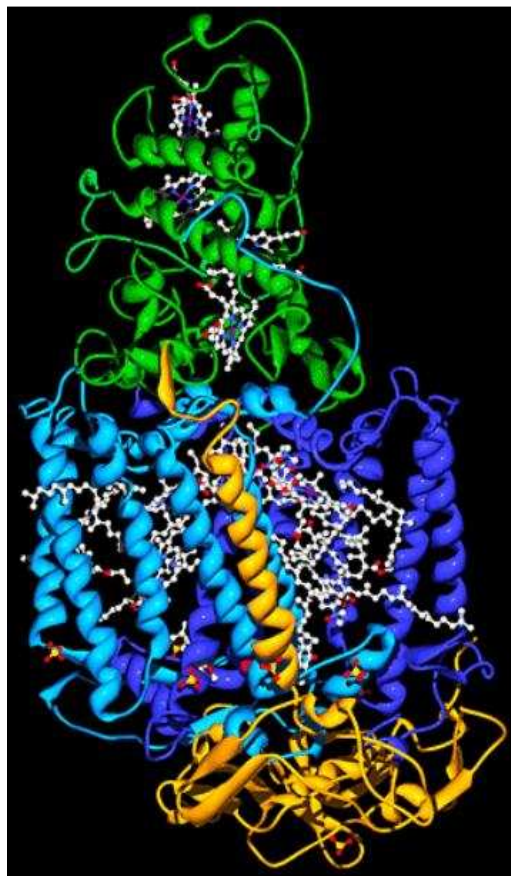


Figure 18. Structure of the *Bl. viridis* reaction centre (Deisenhofer, 2009) consisting of a cytochrome subunit in green and the M, L and H subunits in light blue, blue and yellow, respectively. Molecules that carry out the reaction centre function, the cofactors, are represented as ball-and-stick models; carbon atoms in white, nitrogen in blue, oxygen in red and sulphurs in yellow. The cofactors arrangement alone is shown in Figure 19.

The RC cofactors are surrounded and held in place by the protein subunits and are arranged into two symmetry related branches (Figure 19). The branch mainly located near the L subunit is active in electron transfer and is usually called the ‘A’ branch, whereas the non active, M-subunit side is called the ‘B’ branch (Lancaster, 2008). Both branches start from a ‘special pair’ (P) of Bchl molecules located near the periplasmic side of the membrane. Then each branch contains a single monomeric Bchl, followed by a Bpheo molecule and ends with a quinone, either mena- or ubiquinone. A non-heme iron is located between the two quinones and a carotenoid molecule is in van der Waals contact with the monomeric Bchl on B branch.

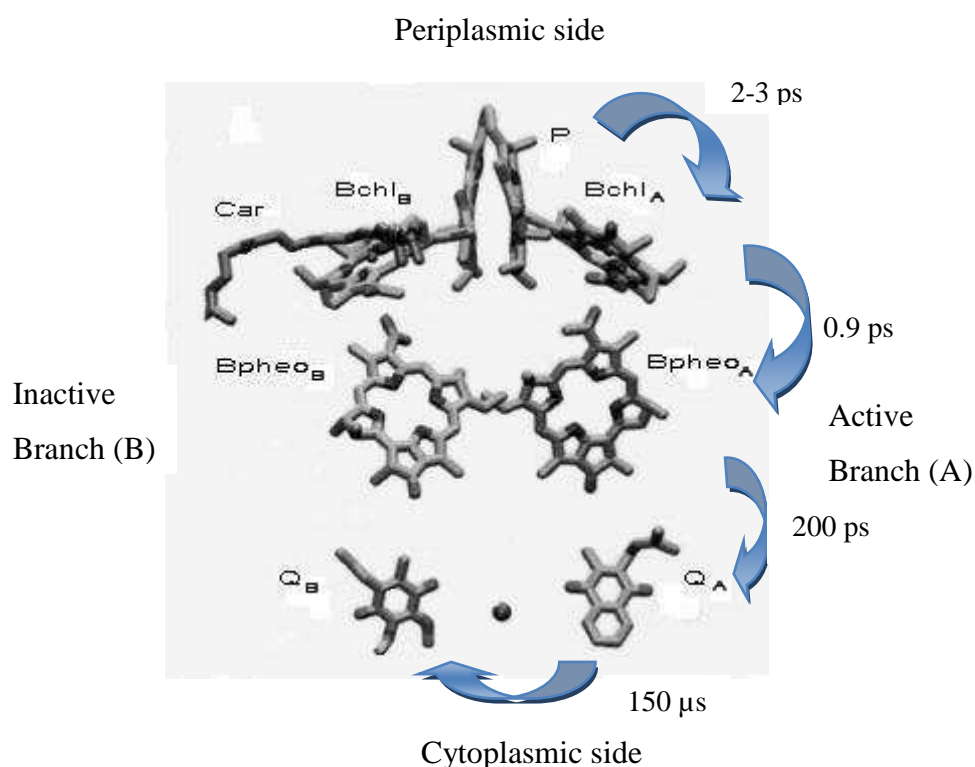


Figure 19. RC cofactor arrangement (Deisenhofer, et al., 1989), Special pair (P), Bacteriochlorophyll (B_A, B_B), Bacteriopheophytin (Bpheo_A, Bpheo_B), Quinone (Q_A, Q_B). The tails of the cofactor molecules have been removed for clarity. The electron transfer reactions are indicated by the blue arrows.

1.6.2. How the reaction centres function

Energy is transferred from the LH system into the RC and is used to excite P to P*. Electron transfer down the A-branch between P* and B_A occurs in about 2-3 ps. Next, the electron hops quickly in about 0.9 ps to Bp_{heo}_A and is then transferred on in the next step to Q_A at the cytoplasmic side in about 200 ps. Each of these electron transfers successfully slows down the back reaction by 3 orders of magnitude leaving P⁺. The positive charge left on P is then neutralized by electron flow from C. Q_B then requires a second turnover to take up two H⁺ from the cytoplasmic side of the membrane to form Q_BH₂, which diffuses out into the lipid bilayer and away from the RC.

1.7. Genetics

The photosynthesis gene organization and its regulation in the purple photosynthetic bacteria are well understood. Most of the genetic information needed to construct the photosynthetic apparatus in purple bacteria is contained in a photosynthetic gene cluster (PGC), a ~41-46 kb stretch of contiguous DNA located in the circular chromosome of the cell. This collection of genes codes for 38 open reading frames (ORFs) that comprise the additional information needed, beyond normal metabolism, to permit the organism to grow using photosynthesis (Blankenship, 2002).

The complete genome sequences of the purple bacterial species *Rhodobacter (Rb.) capsulatus*, *Rb. sphaeroides* and *Rps. palustris* have been determined. In essence, they have an identical PGC (Choudhary, et al., 2000). The PGC arrangements of *Rb. sphaeroides* is shown in Figure 20. There are several sub-clusters of genes, coding for bacteriochlorophyll biosynthesis, carotenoid biosynthesis and the structural proteins of the photosystem. The genes encoding the reaction centre L and M proteins and the α and β apoproteins of the LH1 complex are contained in an operon near the 3' end of the PGC known as *puf* (photosynthetic unit-fixed). The gene for the RC H subunit is located near the other end, known as the *puhA* gene. The LH2 antenna complexes are coded by the *puc* operon, which is not part of the PGC but located approximately 20 kb upstream of the *puhA* gene (Choudhary, et al., 2000).

The photosynthetic unit (PSU) can be defined as the ratio of total Bchl_s to RC BChl in the ICM (Law, et al., 2008). There are two factors that affect the synthesis of ICM, namely

oxygen and light (Lascelles, 1959; Holt, et al., 1965; Marrs, et al., 1973). ICM synthesis will be induced when cells are switched to anaerobic conditions. In this condition, the cytoplasmic membrane will begin to invaginate (Chory, et al., 1984). The regulation of how the PGC is expressed has been extensively studied (Zeilstra-Ryalls, et al., 1997A).

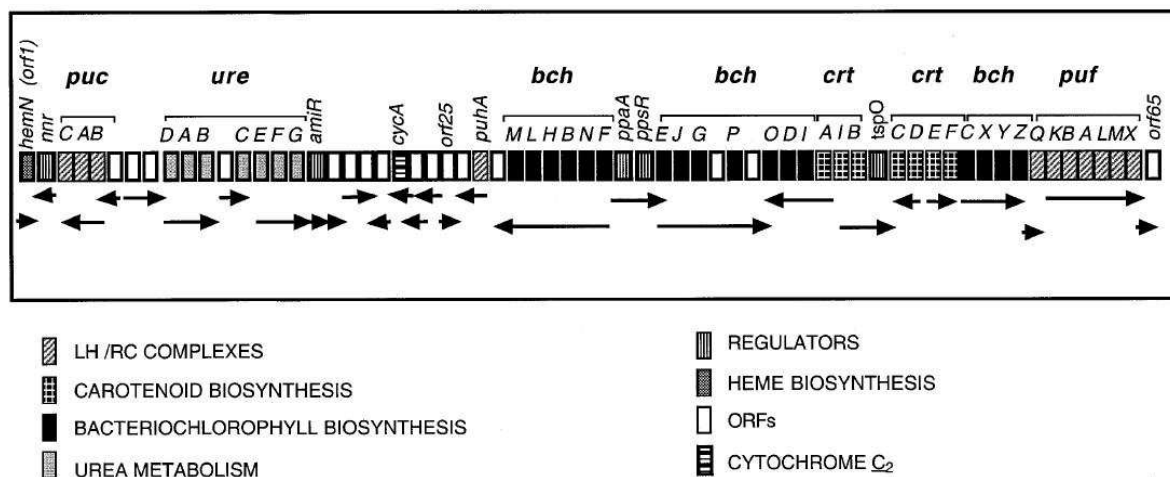


Figure 20. Physical and genetic map of the photosynthetic gene cluster of *Rb. sphaeroides* 2.4.1. (Choudhary, et al., 2000) The number of *orfs* is shown from left to right. The arrows show the likely direction of the transcription of genes/*orfs*.

Oxygen tension is the dominant factor for the PSU synthesis. Basically there is an all or nothing switch that is controlled by oxygen tension (Bauer, 1995; Young & Beatty, 2003). Once the oxygen tension falls below a critical level, invagination of the cytoplasmic membrane and the synthesis of LH apparatus will take place (Takemoto & Lascelles, 1973). From this point, the light intensity determines the cellular levels of ICM and amount of the different LH complexes in the PSU. More antenna complexes will be produced if the light intensity is lowered. Recently, the colour of the incident light has also been found to regulate the synthesis of the PSU. This regulation involves bacterial phytochromes in photosynthetic purple bacteria (Giraud et al. 2002, 2004, 2005; Armitage & Hellingwerf, 2003; Jaubert et al. 2004; Kyndt et al. 2004; Evans et al. 2005). In anaerobic conditions, cells will normally synthesise photosynthetic pigments but this will be halted immediately if the growing conditions switch to aerobic (Cohen-Bazire, et al., 1957). A similar situation occurs to cells when the light intensity is changed.

1.7.1. Regulation of photosynthetic unit expression mainly in the level of gene-expression

The transcriptional regulation of purple bacterial photosynthetic genes involves several classical two-component bacterial regulatory systems. These systems consist of a DNA-binding protein component and a sensor kinase (Zeilstra-Ryalls, et al., 1998; Oh, et al., 2000). In *Rb. sphaeroides* the transcriptional activation of phototropic growth is performed by several proteins. These are PrrBA (photosynthetic response regulator), FnrL (fumarate nitrate regulator), PpsR (photopigment suppression), AppA (activation of photopigment and *puc* expression), SpB (*sphaeroides puf*-binding) and TspO (tryptophan-rich sensory protein). The proteins that are regulated by oxygen tension are PrrBA, FnrL, PpsR and AppA whereas SpB and TspO are regulated by light. These proteins are summarise in Table 1

The two-component PrrBA (photosynthetic response regulator) system (Eraso, et al., 1996) consists of PrrB and PrrA. PrrB is a membrane-spanning sensor histidine kinase/phosphatase (Oh, et al., 2001) that responds to reduced oxygen tension by undergoing autophosphorylation. The phosphate moiety is then transferred onto the second component called PrrA. PrrA is a cytosolic response regulator protein. PrrA is then activated, resulting in gene expression either by binding to DNA directly or by forming transcriptional complexes involving other protein factors (Eraso, et al., 1995).

FnrL is a DNA-binding factor that binds to a recognition sequence, which resembles the FNR consensus sequence of *E. coli* (Zeilstra-Ryalls, et al., 1997A; Zeilstra-Ryalls, et al., 1997B). This sequence is similar to that found before the genes, which synthesize Bchl, carotenoids and the *puc* operon (that encodes LH2 protein). Under low oxygen tension, FnrL binds to this sequence and induces gene expression.

In contrast to FnrL, the PpsR represses the expression of the Bchl, carotenoids and the *puc* operon genes, under aerobic conditions (Gomelsky, et al., 2000; Gomelsky, et al., 1995; Penfold, et al., 1994). The PpsR repressor activity is not only influenced by the oxygen tension itself, it is also responsive to light intensity mediated by AppA (Gomelsky, et al., 1997; Gomelsky, et al., 1998). In this way, the cellular abundance of LH2 complex can be regulated.

As mentioned earlier, so far only two light-responsive factors have been identified. The first of these is SpB, which controls the RC and LH1 protein expression by binding upstream of the *puf* operon (Zeilstra-Ryalls, et al., 1998; Buggy, et al., 1994). This factor

does not regulate LH2 or Bchl (Shimada, et al., 1996). The second factor is TspO. Unfortunately, the details of how they work are still unknown (Zeng, et al., 2001).

Table 1. List of *Rb. sphaeroides* genes known or proposed to play a role in regulation of PSU gene expression (Law, et al., 2008).

Rb. sphaeroides genes	Known of proposed gene product function
prrB	Sensor kinase
prrA	Transcription regulator
ppsR	Transcription regulator
ppbC	Light-sensitive transcription regulator
himA	Subunits of IHF (DNA-binding protein)
appA	Unknown
mgpS	Unknown
prrC	Transcription regulator

Furthermore, it has been recently recognized that bacteriophytochromes (BphPs) are also involved in light regulation in photosynthetic and non-photosynthetic bacteria (Hughes et al, 1997; Davis et al, 1999; Giraud et al, 2002; Armitage et al, 2003; Karniol et al, 2003; Jaubert et al, 2004; Kyndt et al, 2004; Evans et al, 2005). BphPs sense light in the red/far-red region of the spectrum. BphPs have various roles in the regulation of Car synthesis (Davis et al, 1999), chalcone synthase (Jiang et al. 1999), LH complex biosynthesis (Giraud et al, 2002), and phototaxis (Armitage & Hellingwerf, 2003; Kyndt et al, 2004). BphPs genes have been identified in the following photosynthetic bacteria: *Rb. sphaeroides*, *Rps. palustris*, *Rhodospirillum (R.) rubrum*, *R. centenum*, *Thermochromatium (Tc.). tepidum* and *Bradyrhizobium* ORS278 (Jiang et al. 1999; Giraud et al, 2002; Kyndt et al, 2004). More work is required before the exact function of the BphPs will be clearly understood.

1.8. Polymerase chain reaction

The polymerase chain reaction (PCR) is an enzymatic method of amplifying nucleotide sequences. It can increase exponentially a single copy of a gene within a matter of hours. PCR is a powerful technique that has rapidly become one of the most widely used in molecular biology because it is quick, inexpensive and simple. PCR amplifies specific nucleotide fragments from minute quantities of source nucleotide material, even when it has relatively poor quality (Becker-Andre, et al., 1989). This method was originally only used with DNA molecules as templates. The method has now been improved so that RNA can also be used as a template. The basic principle of DNA amplification using PCR and the development of the PCR technology will be described in this section, as it is one of the major experimental methods used in this thesis.

1.8.1 Basic principle of PCR

There are four main components involved in the PCR process, namely DNA templates, oligonucleotide primers, deoxyribonucleotides triphosphates (dNTP), and DNA polymerase (Erich, 1989). The other component, which is also important, is the reaction buffer. DNA templates are the fragments of DNA that are to be amplified. Primers are short oligonucleotide sequences (around 15 – 25 bp), which are used to begin the DNA synthesis. dNTP are the nucleotides consisting of dATP, dCTP, dGTP and dTTP. DNA polymerase is the enzyme that catalyzes the DNA chain reaction (Gelfand, et al., 1990).

A PCR cycle consists of three steps. The process starts with the denaturation of the DNA template, where DNA is heated in order to separate the strands. Denaturation leads to the splitting of double strands into single strands under the high temperature condition, for example, 95-98 °C for 1-2 minutes. The temperature is then decreased to around 55 °C, in order to anneal the primers to the DNA templates (Gu, 1995). Two oligonucleotide primers, which flank the DNA segment to be amplified, will form hydrogen bonds with the template in the complementary region. These primers hybridize to opposite strands of the target sequence and are thus oriented. Therefore, the synthesis of DNA by the polymerase will proceed across the region between the primers, effectively doubling the amount of that DNA segment (Lawyer, et al., 1989). The extension is conducted by increasing the temperature, for example, to 72 °C for 1.5 minutes.

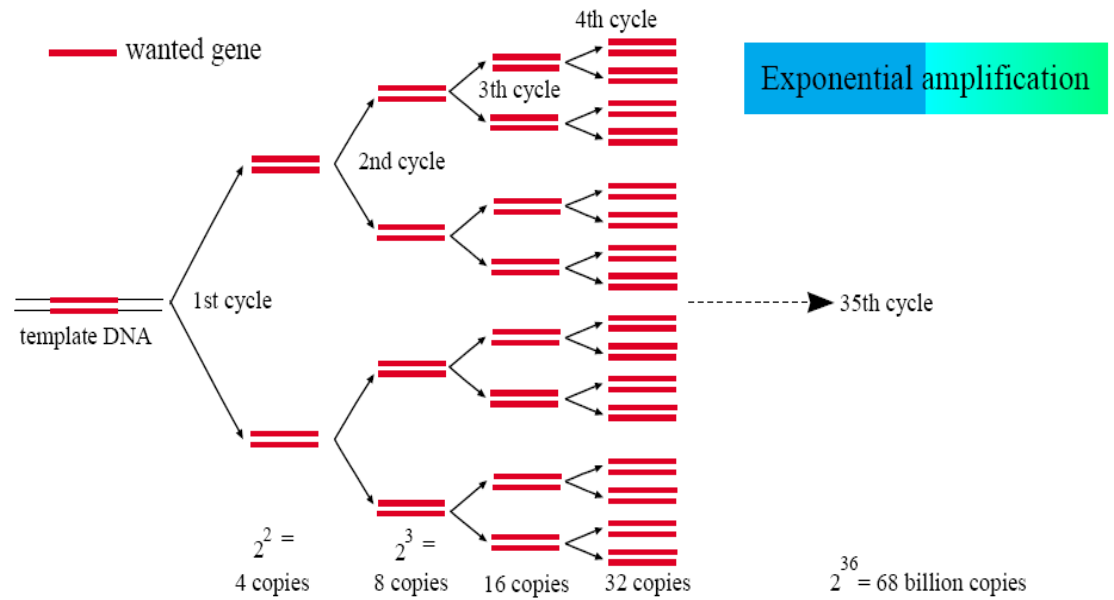


Figure 21. DNA amplification with PCR (Yuwono, 2006). The red lines represent the wanted gene, which is wanted to be amplified. The reaction product increase exponentially within matter of hours.

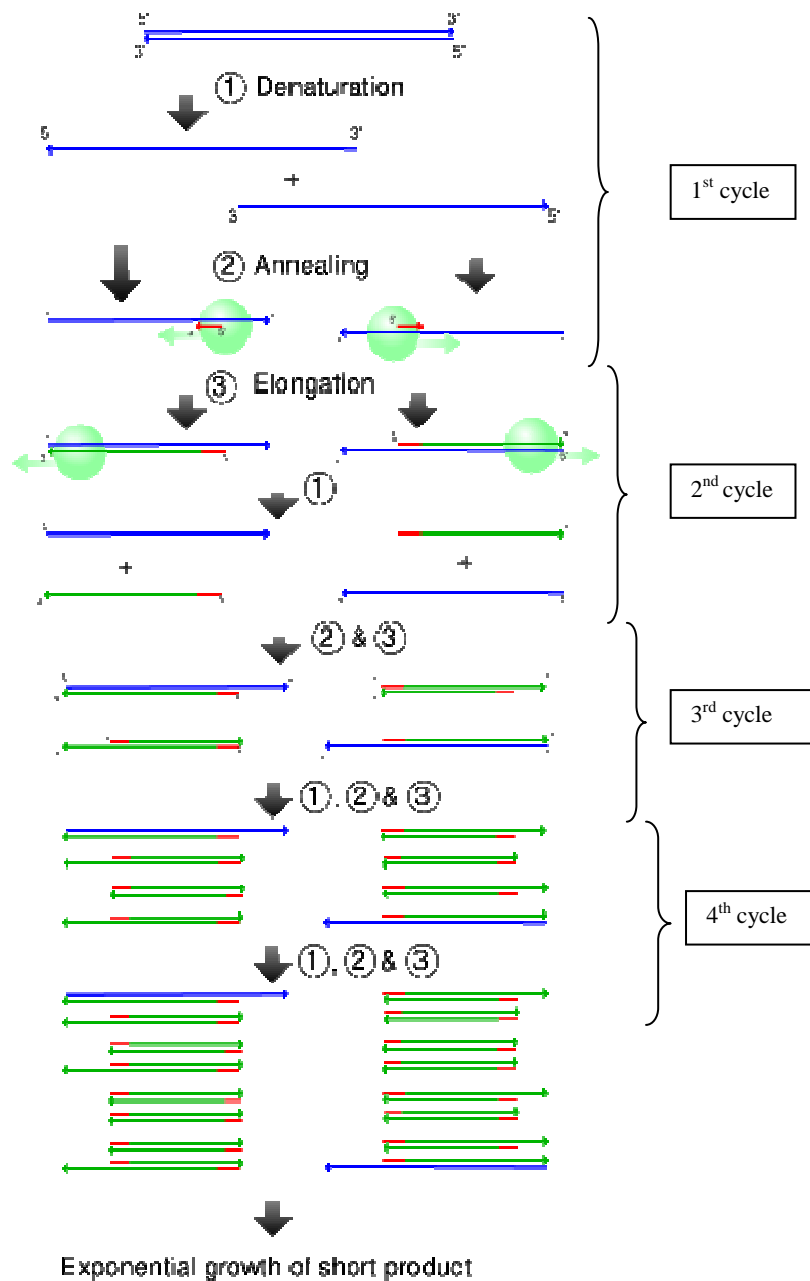


Figure 22. Schematic drawing of the PCR cycle (Yuwono, 2006). (1) Denaturation, (2) Annealing and (3) Elongation. Four cycles are shown here. The blue lines represent the DNA template to which primers (red) anneal. Primers are extended by the DNA polymerase (light green circles), to give shorter DNA products (green lines), which themselves are used as templates as PCR progresses.

Moreover, since the extension products are also complementary to and capable of binding primers, each successive cycle essentially doubles the amount of DNA synthesized in the previous cycle. The result is in exponential accumulation of the specific target fragment, approximately 2^n , where n is the number of cycle (Lawyer, et al., 1989). These reactions will usually be repeated up to 25 to 30 cycles, forming a lot more new DNA molecules than the amount of DNA template used. The number of amplification cycles required depends on the initial concentration of the DNA template in the reaction solution (Kubista, et al., 2005).

The basic principle of PCR scheme can be seen in Figure 21 and Figure 22. A water bath can be used to create each PCR condition. However, this manual process may lead to the incubation time inaccuracies (Lawyer, et al., 1989). Therefore, programmable equipment with higher incubation time accuracy has been developed.

1.8.2 DNA polymerase

The Klenow fragment was the DNA polymerase used in the early development stage of PCR. The Klenow fragment is a DNA polymerase that has lost its 5'→3' exonuclease activity (Saiki, et al., 1988). There are several disadvantages of the Klenow fragment, such as thermolability, only a moderate polymerase rate and low processivity (Mullis, et al., 1989). Processivity is the capability of a polymerase enzyme to catalyze many consecutive reactions without releasing its substrate (Berg, et al., 2002).

As making single strands of DNA from double strands requires a high temperature of incubation (Gu, 1995), an enzyme which has a greater resistance to heat would be more suitable. The modern alternative to the Klenow fragment is *Taq* DNA polymerase, which is extracted from the bacterium *Thermus aquaticus* (Saiki, et al., 1988). This strain of this bacterium usually does not have a restriction endonuclease (Brow, 1990), and it produces a *Taq* DNA polymerase with a molecular weight of around 95 kD that is heat stable (Brow, 1990). This enzyme has a high DNA polymerase activity (150 nucleotides per second for each enzyme molecule) but has no activity as a 3'→5' exonuclease (Gelfand, et al., 1990). It is very active at pH 9.0 (20 °C) and has an optimum temperature of around 75 °C – 80 °C. The half-life of *Taq* DNA polymerase at 95 °C is 40 minutes (Brow, 1990).

The *Taq* DNA polymerase's heat-resistant property has allowed the optimization of each step in the PCR process that originally required the Klenow fragment (Gelfand, et al., 1990). These modifications have simplified the process, making it amendable to

automation, and have substantially improved the overall performance of the reaction by increasing the fidelity, yield, sensitivity, and length of target that can be amplified (Saiki, et al., 1988). It is also possible to use *Taq* DNA polymerase for sequencing purposes (Brow, 1990).

A unique property of *Taq* DNA polymerase is that this enzyme is capable of adding one nucleotide, especially dATP, on the 3' end of a DNA fragment even if there is no template. Therefore, the PCR fragment products usually have blunt end. This has important implications because the PCR fragment product can then be ligated into a plasmid vector without using a DNA ligase enzyme. In general though, one of the weaknesses of *Taq* DNA polymerase is a tendency to mismatch when combining nucleotides; thereby causing a mutation in the amplified gene (Gelfand, et al., 1990).

Modern DNA polymerases are also used nowadays other than *Taq* DNA polymerase. An example of such DNA polymerase, which was also used in this experiment, is Phusion® DNA polymerase. It has processivity of approximately twice greater than *Taq* DNA polymerase and also has the ability to produce higher yields.

The other factor that should be considered in order to conduct PCR successfully is the Mg^{2+} concentration. Mg^{2+} concentrations influence the activity of *Taq* DNA polymerase. Its activity is usually maximal at an $MgCl_2$ concentration of 2 mM with dNTP concentrations of 0.7-0.8 mM. Higher concentrations of Mg^{2+} above 2.0 mM, will inhibit *Taq* DNA polymerase activity (Higuchi, et al., 1992). *Taq* activity will also decrease if the total concentration of dNTP used reaches 4-6 mM (Gelfand, et al., 1990).

1.8.3. Real time PCR

As an analytical technique, the original PCR method had some serious limitations. By first amplifying the DNA sequence and then analyzing the product, quantification was exceedingly difficult since the PCR gave rise to essentially the same amount of product independent of the initial amount of DNA templates molecules that were present (Higuchi, et al., 1992). The product was quantified at the end point of reaction with electrophoresis. Unfortunately the results were not very precise.

Real-time PCR has revolutionized the quantitative detection of DNA and RNA. It was pioneered by Higuchi in 1992 who constructed a system that detects PCR products as they accumulate (Kubista, et al., 2005). Theoretically, there is a quantitative relationship between amount of starting target sequence and amount of PCR product at any given cycle.

The overall process of real-time PCR is described in Figure 23. The exponential phase in PCR process is the optimal point for analyzing data (Kubista, et al., 2005). The benefit of this real time capability is it allows the researcher to determine the amount of starting DNA in the sample before the amplification by PCR (Valasek, et al., 2005).

In addition to conventional PCR, Real-time PCR needs a fluorescent reporter that binds to the product formed and reports its presence by fluorescence (Higuchi, et al., 1992). The reporter generates a fluorescence signal that reflects the amount of product formed. The signal is captured by a detector that measures the amount of PCR product formed during the process. In his initial work, Higuchi used the common nucleic acid stain ethidium bromide as the fluorescent reporter. Ethidium bromide becomes fluorescent upon intercalating into DNA (Higuchi, et al., 1992). Ethidium bromide is carcinogenic and that has become a serious problem, but now other types of dyes are used. There are several fluorescence reporters known such as molecular beacon and exonuclease probes (Kubista, et al., 2005). An example such a fluorescence reporter can be seen in Figure 24.

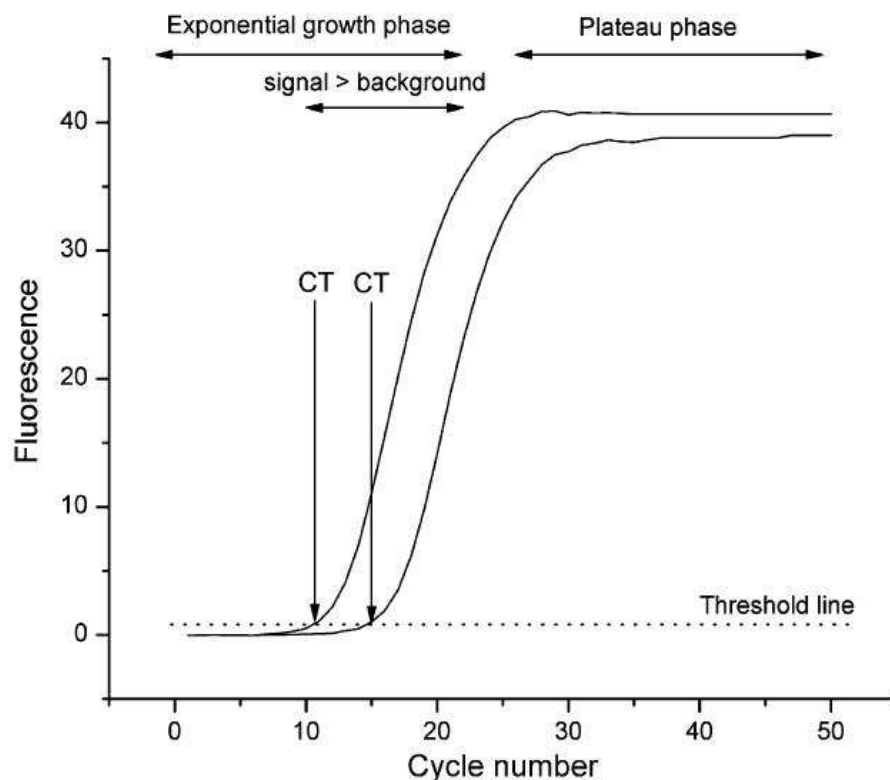


Figure 23. The figure shows real-time PCR response figure between fluorescence and number of amplification cycle. Fluorescence is correlated as the amount of amplicons. Exponential phase is where amplification occurs optimally, whereas the plateau phase is where amplification stops. CT is the number of cycles required to reach the threshold (Kubista, et al., 2005).

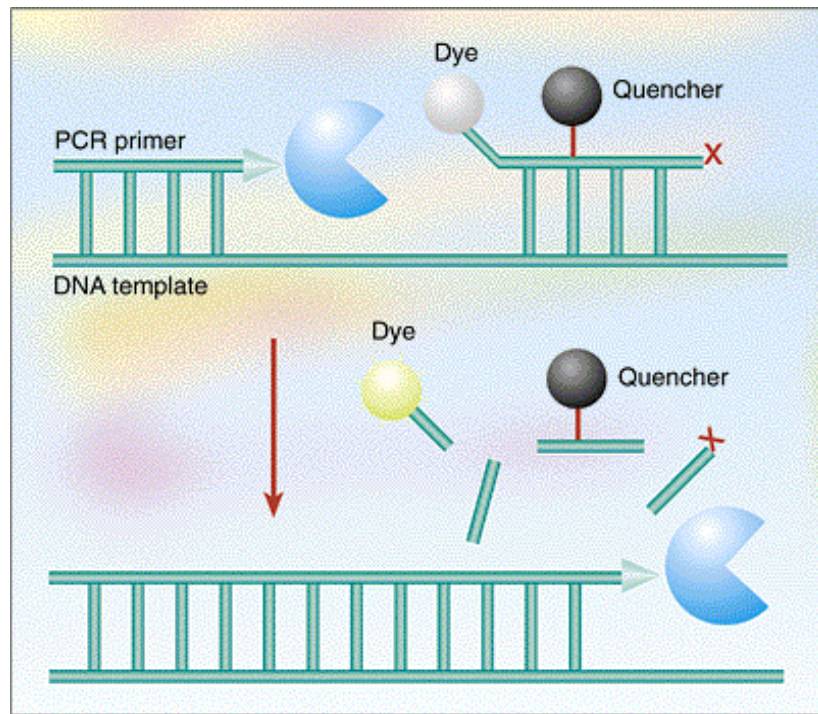


Figure 24. Exonuclease probe: A single-stranded DNA probe is labelled with a fluorescent dye and a fluorescent quencher. The proximity of the donor and quencher prevents fluorescence. A fluorescence signal is generated if the DNA polymerase runs into the hybridized probe during the extension process, because the polymerase treats the probe as an obstacle and cleaves it into pieces. That separates the quencher from the donor, allowing fluorescence (Yuwono, 2006).

1.8.4. Reverse transcription

The reverse Transcriptase PCR (RT-PCR) technique was developed to analyze mRNA. The PCR method is not able to use RNA as a template, therefore, a reverse transcription reaction must be carried out at the beginning of the PCR process. An enzyme called reverse transcriptase is added (Valasek, et al., 2005) that produces double stranded cDNA from mRNA (Berg, et al., 2002), the cDNA can then be used as the template in the PCR process.

1.9. DNA Sequencing

The definitive description of any DNA molecule, from a short fragment in a DNA clone to a whole human chromosome, is its sequence. The analysis of DNA structure and its role in gene expression have been markedly facilitated by the development of DNA sequencing. DNA sequencing methods have been around for several years but have been significantly improved since the late 70's (Brown, 1990). There are two types of DNA sequencing methods. The first method is DNA chain termination, which is known as Sanger's dideoxy method, taken from the inventor's name F. Sanger. The second method uses chemical degradation, which is known as Maxam-Gilbert method, also named after its inventor's A. Maxam and W. Gilbert (Hartwell, et al., 2008). Gilbert and Sanger developed their techniques, which both have similar accuracy approaching 99%. They won the Nobel Prize for the contribution to DNA sequencing technology. But now the Sanger method is the most common technique because it is can be readily automated.

1.9.1. The Sanger dideoxy method

The basis of the dideoxy method is to reveal the order of base pair in an isolated DNA molecule by controlled termination of DNA replication (Berg, et al., 2002). There are two steps to the Sanger method of sequencing. The first step is the generation of a complete series of single-stranded sub fragments, which are complementary to a portion of the DNA template under analysis. Basically, the DNA fragments are generated by interrupting the enzymatic replication, in other words, further chain elongation is blocked. The second step is the analysis of the mixture of DNA sub-fragments obtained by using polyacrylamide gel electrophoresis. The analysis allows the separation of DNA molecules differing in length by only a single nucleotide (Hartwell, et al., 2008). Therefore, the sequence of nucleotides can be read in order.

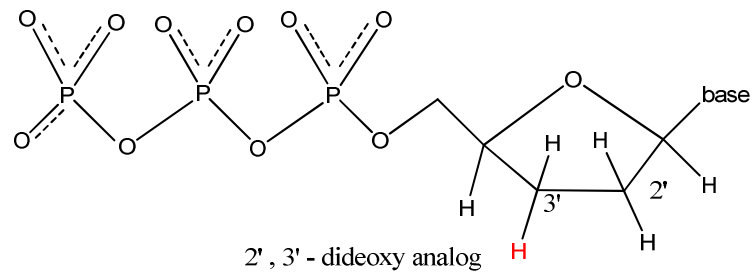


Figure 25. The structure of dideoxy analog; hydrogen (shown in red) replaced hydroxyl so that further elongations will be prevented.

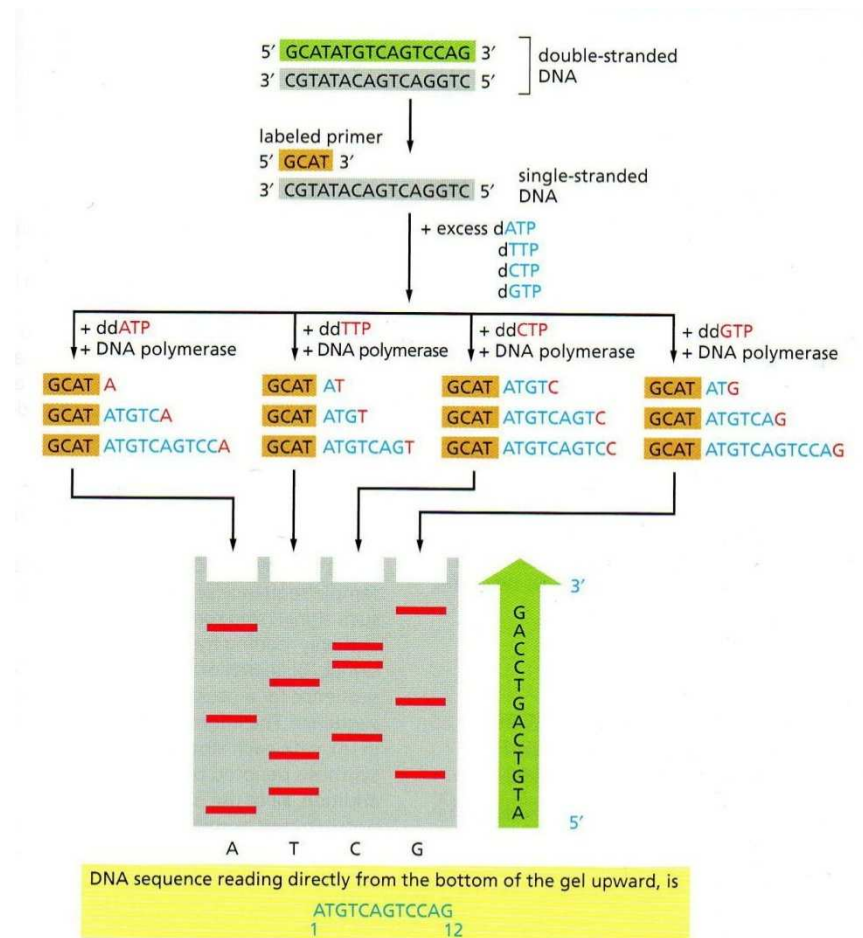


Figure 26. Strategy for chain-termination method for DNA sequencing (Albert, et al., 2008). Four different chain-terminating dideoxyribonucleotides triphosphate (dATP, dTTP, dGTP and dCTP , shown in red letters) are used in four separate DNA synthesis reactions with copies of the same single-stranded DNA template (gray). The addition of the dideoxy analog along with the deoxyribonucleotides triphosphate will interrupt DNA fragment replication. Thus, each reaction produces a set of DNA copies that terminate at different points of the sequence. These products are then separated by electrophoresis in four parallel lanes of polyacrylamide gel (bottom gray square). The DNA sequence reads from the bottom of the gel upward.

The first step begins by preparing the purified, denatured DNA with oligonucleotide primers that are complementary to a particular site on the DNA. The incorporation of each chain-terminating dideoxynucleotide triphosphate analog (dATP, dTTP, dCTP and dGTP) will produce separate fragments each with a dideoxynucleotide triphosphate at the end. Polymerization from the primer strand continues until, by chance, the dideoxynucleotide is incorporated (Hartwell, et al., 2008). An example of a dideoxy analog chemical structure can be seen in Figure 25.

The second step is as important as the first one. The fragments in each of four aliquots are separated according to their size by polyacrylamide gel electrophoresis. The resolution of the gel is such that you can distinguish DNA molecules that differ in length by only a single base. As shown in Figure 26, the appearance of a DNA fragment of a particular length demonstrates the presence of a particular nucleotide at that position in the strand (Berg, et al., 2002). Once the sequence of the newly synthesized DNA is known, it is a simple matter to convert this sequence into a complementary sequence of the template strand under analysis.

Instead of having single radioactive label on the primer oligonucleotide, the automated DNA sequencing process now uses chain-terminating dideoxynucleotides that each has a different colour from a fluorescent dye (Hartwell, et al., 2008). As a result, the four separate reactions are no longer needed. All four can be combined in a single reaction mixture that can then be analyzed in a single lane on a gel with the use of the automated sequencer (Figure 27). Automated sequencer detector will capture the fluorescence from each dideoxynucleotide. The detector records the colour order of the passing band that then can be translated into sequence data by a computer (Berg, et al., 2002; Hartwell, et al., 2008).

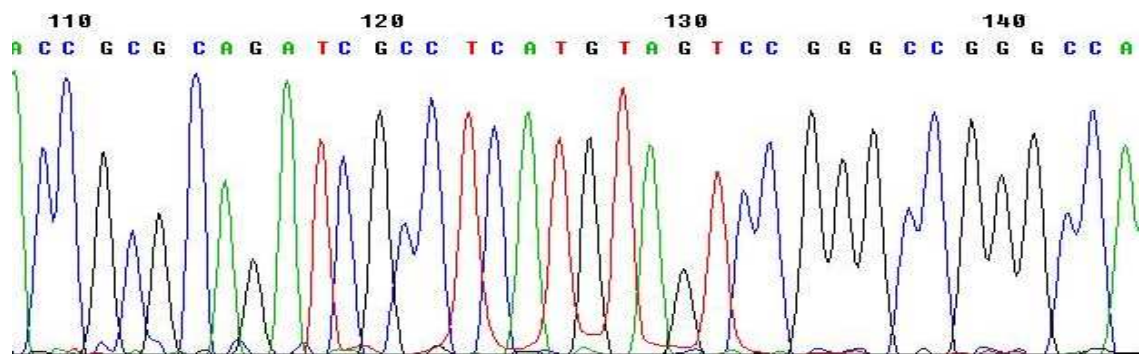
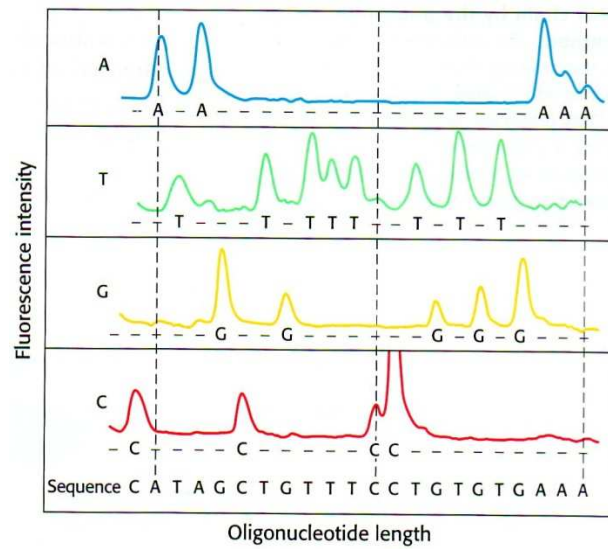


Figure 27. Automated fluorescence detection of oligonucleotide fragments produced by the dideoxy method. Upper: Four chain-terminating dideoxynucleotides with four different colours. Bottom: The combination of four chain-terminating mixtures at four wavelengths.

1.10. Aim of this project

The 3D structure of the *Bl. viridis* photosynthetic RC was the first membrane protein structure determined by X-ray crystallography (Deisenhofer, et al., 1984; Deisenhofer, et al., 1985). At the same time, Michel *et al* determined the amino acid sequence of the reaction centre polypeptides (Michel, 1982; Michel, et al., 1985; Michel, et al., 1986; Wiessner, et al., 1990). However, recent X-ray crystallographic studies on high resolution data from this RC (Roszak, et al., unpublished observations) have found that several amino acids in the sequence do not fit well in the electron density. This suggests that there may be errors in the original sequence data. The aim of this project, therefore, is to re-sequence the genes of the β and α subunits of the LH1 and the L, M, H and Cytochrome *c* subunits of the photosynthetic RC from *Bl. viridis* to establish if new sequence data produces a better fit to the recently obtained new electron density maps.

Chapter 2 Materials and methods

Unless stated otherwise all the chemicals used in this research were of Analar grade. Molecular biological enzymes and kits were obtained from Sigma, Promega and Qiagen, and used as per the manufacturer instructions. Full details of the chemicals and kits are summarised in Appendix.

2.1. General culturing method and cell storage

Bacterial cultures were maintained in agar stabs in the laboratory until required. The cultures were checked on a regular basis to ensure purity.

2.2. Growth of photoheterotrophic cultures

Bl. viridis bacteria were grown anaerobically in an illuminated, temperature controlled room (30 °C), in succinate based media (Bose S, 1963) media in screw-cap bottles.

In order to ensure the starting culture was pure, a liquid culture was grown up from an agar stab, serially diluted and plated out on agar plates before being placed in anaerobic jar. The cells are grown for about 1 week before nicely separated single colonies appear. Single, well separated colonies were then picked off from the plates and used to inoculate agar stabs, thereby providing fresh stocks of pure cultures.

To start a good quality liquid culture, fresh media was poured aseptically onto the top of the agar in the stab, sealed and incubated at 30 °C in the growth room for 3-4 days. The bacteria then grew into the liquid media. In turn, this small liquid culture was used to inoculate a larger culture bottle (500 ml). The 500 ml cultures were grown for approximately 2-3 days. Figure 28 shows various stages of the growth of the cultures.

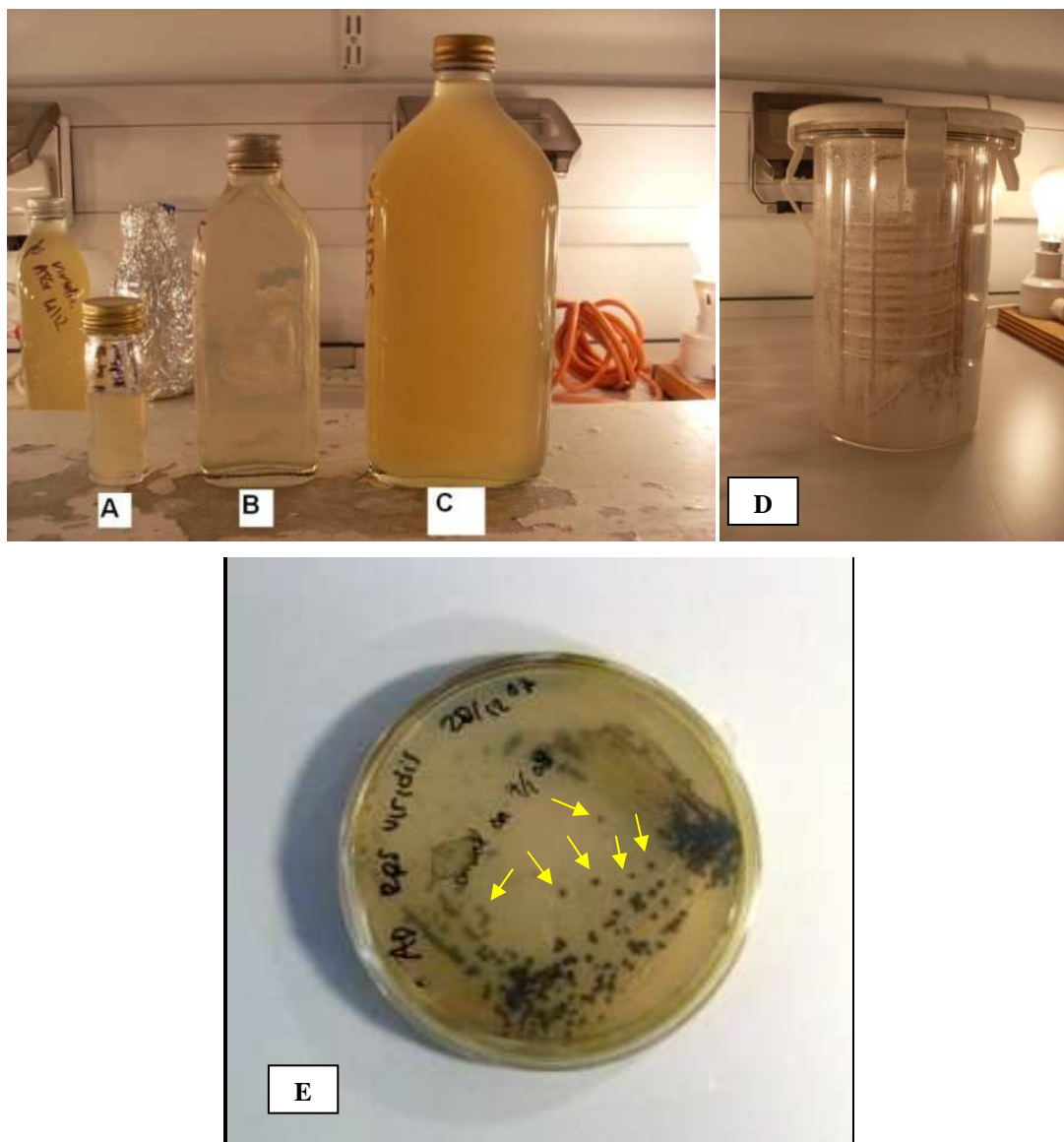


Figure 28. Bacterial cultures in bottles, an anaerobic jar and on a plate. The pure cultures are maintained in agar stabs (A), before being transferred into the 100 ml bottle (B) and then onto the 500 ml bottle (C). An anaerobic jar (D) where single colonies on plates (E) were grown. Single colonies are indicated by yellow arrows.

2.3. Cell harvesting

After 500 ml of cells had grown to sufficient density they were harvested by centrifugation (10 min, $8,000 \times g$, 4°C). The clear supernatant was discarded leaving a pellet of bacterial cells. This pellet was re-suspended in minimum volume of TE buffer (see appendix) using a hand homogeniser. TE buffer contains EDTA, which removes magnesium ions that are essential for preserving the overall structure of the cell envelope

and also inhibits cellular enzymes that could degrade DNA (Brown, 1990). Therefore, the initial TE buffer usage at this stage will protect the DNA. The re-suspended cells were then either used directly for DNA extraction or stored at -20°C until required.

2.4. Genomic DNA isolation

The bacterial cells are enclosed in a cytoplasmic membrane and surrounded by a rigid cell wall. This barrier has to be disrupted in order to release the cell components. Techniques for breaking open bacterial cells (lysis) can be divided into physical methods and chemical methods. Physical lysis causes cell disruption by a mechanical force but also shears the genomic DNA, whereas chemical lysis is brought about by chemical agents and causes less damage to the DNA (Brow, 1990). Chemical lysis generally involves one agent attacking the cell wall and another disrupting the cell membrane.

Only 200 μl of harvested cell suspension is required for the isolation of genomic DNA. In addition to DNA, a bacterial cell extract also contains significant quantities of protein. The standard way to remove protein from a cell extract is to add phenol or a 1:1 (v/v) mixture of phenol and chloroform. These organic solvents precipitate protein but leave the nucleic acid (DNA) in aqueous solution (Brow, 1990).

The re-suspended cell pellet was transferred into a 1.5 ml Eppendorf tube. An equivalent volume of the phenol-chloroform 1:1 (v/v) mixture was then added into the tube, vortexed briefly and incubated for 5 minutes. The resulting viscous solution was microfuged (4 min, $13,000 \times g$) and the clear supernatant then carefully loaded into a new 1.5 ml Eppendorf tube and the volume measured.

In the presence of salt (monovalent cations such as Na^+) and at a temperature of -20°C or less, absolute ethanol will efficiently precipitate polymeric nucleic acids. The DNA obtained is then able to be collected by simple centrifugation.

The DNA contained in supernatant from the previous step was precipitated with 2.5 volumes of 96% ethanol, together with 0.1 volumes of 3 M sodium acetate, pH 5.3. The sample was then cooled to -20°C for 2 hours (or 4°C overnight) and the pellet (DNA) obtained by centrifugation ($13,000 \times g$) in the cold room. The supernatant was discarded and the DNA pellet was washed once with 70% ethanol before being re-suspended in sterile water. The dissolved DNA was then ready to be used for analysis. How this procedure was optimised will be described in Chapter 3.

2.5. Quantification of the DNA by gel electrophoresis

When necessary, gel electrophoresis was used to check the size and the concentration of the DNA obtained. Fragments of DNA, containing the gene of interest, were size fractionated on 1.2% agarose gel in TAE buffer (400mM Tris-Acetate, 10mM EDTA, pH 8.3 at 23⁰C). In order to stain the DNA, Syber© safe DNA gel stain was used. Gels were run under a constant voltage of 100 V and 400 mA for 30 minutes. Two ladders, either with 100 bp or 1 kb steps were used to estimate the size of the DNA and, by comparison of the fluorescence intensity, to approximately determine the DNA concentration. The ladders were obtained from New England Biolabs (UK) Ltd and are shown in Figure 29. If the size and purity of the DNA fragment was satisfactory, it could then be used for the next step.

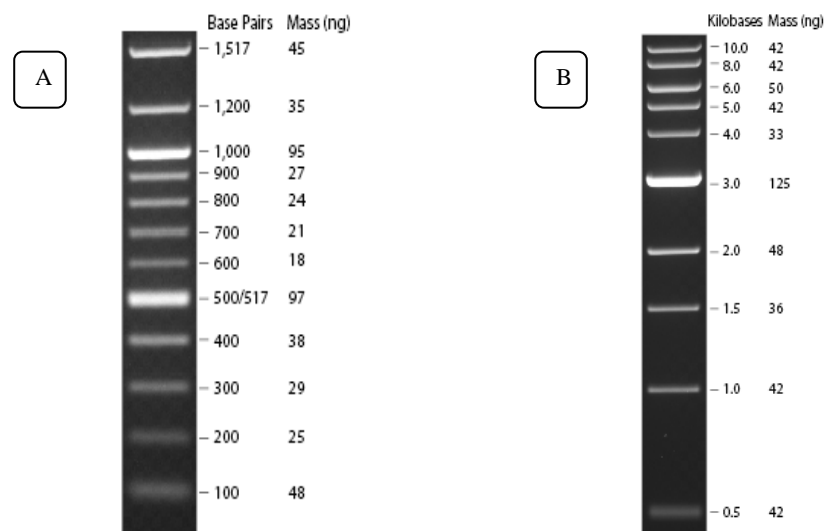


Figure 29. Gel electrophoresis results of ladder used. Ladders were named based on size difference between each fragment, (A) 100 bp ladder and (B) 1kb ladder. These pictures were obtained from New England Biolabs (UK) Ltd.

2.6. Polymerase chain reaction

The DNA amplification was performed by PCR using a Thermo Hybrid thermal cycler PX2. A PCR reaction mix (15 µl) was prepared for each reaction. Each reaction mix contains buffer (supplied with the polymerase), dNTPs, water, primers (forward and reverse), DNA polymerase and the DNA template. The PCR conditions were optimised by varying factors such as the annealing temperature, the elongation time and the primer

concentrations. Primers were designed based on the previously published DNA sequence of *Bl. viridis* (Michel, et al., 1985; Michel, et al., 1986; Wiessner, et al., 1990).

Details about the optimisation of the PCR conditions and the primer design are given in Chapter 3. Primers were obtained from Sigma. Finnzymes Phusion® DNA polymerase was obtained from New England Biolabs (UK) Ltd. The purification of DNA fragments, where required, was performed using a Qiaquick® Spin kit obtained from Qiagen.

2.7. DNA cloning

PCR products that could not be sequenced directly were cloned into JM109© High Efficiency *Escherichia coli* (*E. coli*) cells with the pGEM®-T Easy Vector System. The vectors are supplied so that they contain 3' terminal thymidine at both cutting ends. These single 3'-T overhangs at the insertion site enhance the ligation efficiency of the PCR product into plasmids by preventing re-circularisation of the vector. The pGEM®-T Easy Vector also contains T7 and SP6 RNA polymerase sites (see Figure 30 and Figure 31 for details) flanking the cloning region within the α -peptide coding region of the enzyme β -galactosidase. Insertional inactivation of the α -peptide allows recombinant clones to be directly identified by colour screening (blue - white) on indicator plates with the use of IPTG and X-gal. White colonies contain inserts, whereas blue colonies do not. The validity of the insert in the *E. coli* cells can be confirmed by PCR using T7 and SP6 primers.

Another method to identify whether the insertion process has been successful is by Ruscony test. This is a quick method to provide information about the presence of inserts in the *E. coli* cells. This method isolates the plasmid and measures its size. The *E. coli* cells are re-suspended in 15 μ l of Ruscony mix (see the appendix for details). Then the mixture is incubated at room temperature ($\pm 26^{\circ}\text{C}$) for 10 to 15 minutes. Either phenol (2 μ l) or phenol-chloroform (4 μ l) is added, and then mixed with hand homogeniser. Next, 10 μ l of chloroform is added, and the mixture is again homogenised and centrifuged (2 minutes, $6,000 \times g$) to separate the liquid into 2 phases. The plasmid will be contained in the upper phase. The upper phase is loaded onto a gel (section 2.5) and the plasmid size determined. PGEM®-T Easy Vector has a size of about 3 kb, therefore, the presence of the correct insert in the plasmid can be easily identified. Finally, the DNA plasmid can be directly isolated from the *E. coli* by using the QIAGEN® Plasmid Purification kit.

pGEM®-T Easy Vector

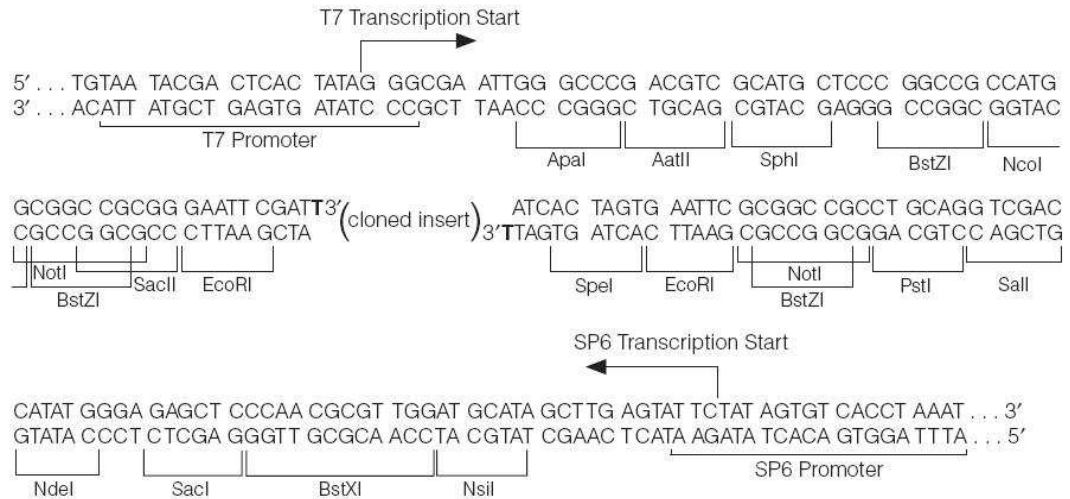


Figure 30. The promoter and multiple cloning sequences of the pGEM®-T Easy vectors (picture obtained from pGEM®-T Easy vectors handbook). The top strand of the sequence shown corresponds to the RNA synthesized by T7 RNA polymerase. Bottom strand corresponds to the RNA synthesized by SP6 RNA polymerase

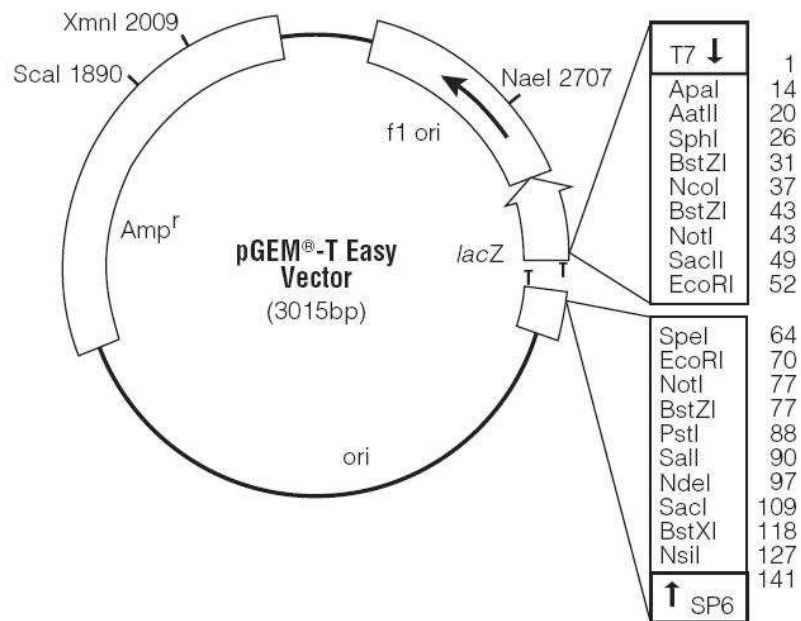


Figure 31. pGEM®-T Easy vector circle map and sequence reference points (picture obtained from pGEM®-T Easy vectors handbook)

2.8. DNA sequencing

DNA was sequenced by the Molecular Biology Support Unit (MBSU) DNA sequencing service at the University of Glasgow. The fragments were sequenced by Sanger's dideoxy method by using a MegaBASE 1000 sequencer from GE healthcare. The raw sequence data alignment was performed by ClustalW (EBI, 2008) and the amino acid translation by Molbiol 6 frame translator (Molbiol, 2008).

2.9. Experiment method overview

A flowchart of the experiment methods used is given in Figure 32. This flowchart therefore shows an outline of the complete project.

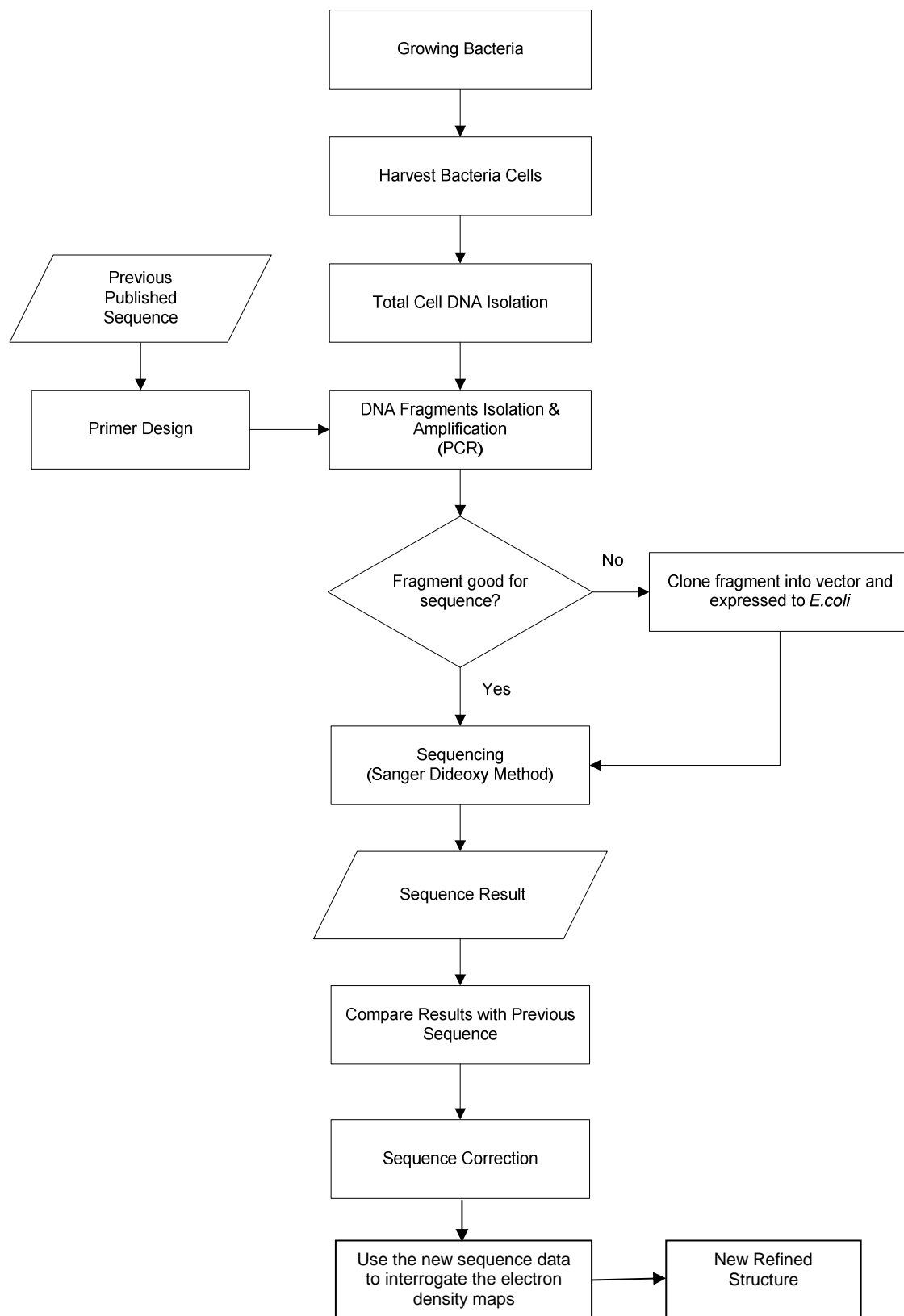


Figure 32. Overview of the experimental methodology. The sequence results were then compared with the previous published version in order to obtain the corrected sequence version.

Chapter 3 **Re-sequencing the genes of light harvesting complex and reaction centre from *Bl. viridis***

3.1. Introduction

The 3D structure of *Bl. viridis* RC was the first membrane protein structure determined by X-ray crystallography (Deisenhofer, et al., 1984; Deisenhofer, et al., 1985). In order to fit the polypeptide sequences into the electron density Michel had previously sequenced the genes encoding the RC and LH1 polypeptides (Michel, 1982). The *Bl. viridis* strain used in this research is DSM 133, the same as used by Michel.

The genes encoding the β and α subunits of the LH complex and the L, M, and C subunits of the photosynthetic RC from *Bl. viridis* are organized into the *pufBALMC* operon (Michel, et al., 1986), illustrated in (Michel, 1982). The H subunit is encoded separately in the *puh* operon (Michel, 1982). The DNA sequence of genes encoding the β (*pufB*), α (*pufA*), L (*pufL*), M (*pufM*), C (*pufC*) and H (*puhA*) were determined by manual DNA sequencing (Michel, 1982; Michel, et al., 1985; Michel, et al., 1986; Wiessner, et al., 1990). On the upstream part of the *puf* operon, the *pufB* and *pufA* genes are separated by 17 bp nucleotides. The *pufA* gene is separated from the *pufL* gene by 122 bp (Michel, et al., 1985), while *pufM* is separated from *pufL* by 11 bp (Michel, et al., 1986). The gene encoding the *pufC* is located at the end of the gene encoding *pufM* and overlaps it by 1 bp (Wiessner, et al., 1990). As often happen with new scientific developments and technical inventions, an accidental observation leads to the beginning of an experiment. A recent 2.0Å resolution structure, obtained whilst studying the binding of the carotenoid in the *Bl. viridis* RC has found that several amino acids positions in the protein sequence do not fit well into the electron density map (Roszak, et al. unpublished observations). It was apparent from the new electron density maps that the original *Bl. viridis* RC amino acid sequence may contain some errors and that the clashes that result could only be resolved by re-sequencing the *puf* and *puh* operons. Examples of these clashes are shown in Figure 34 and 35. In this case the electron density maps produced by the new high, resolution structure have been fitted with the Michel polypeptide sequence and the poor fits are rather obvious.

The examples presented Figure 34 and Figure 35 illustrate there are problems at positions M70, M164, C77, C277 and C43. The newly obtained electron density maps indicate that the amino acid residue assigned to each of these positions is probably not correct. The published amino acid sequence assigns M70, M164, C77, C277 and C43 to Leu, Thr, Ile, Leu and Ala, respectively. This work described in this thesis re-sequences the *puf* and *puh* operons from *Bl. viridis* and establishes new polypeptide sequences for these genes. The re-sequencing of these genes and a comparison between the old and new sequences is presented in Chapter 3 and how these differences enable us to better interpret the electron density maps is discussed in Chapter 4.

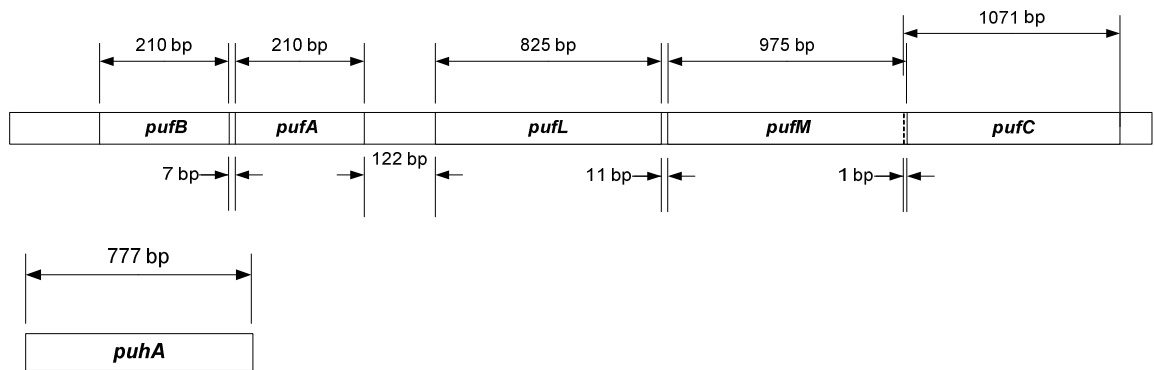


Figure 33. The genetic map of the *puf* operon DNA region between *pufB*, *pufA*, *pufL*, *pufM* and *pufC* (Wiessner, et al., 1990). *PuhA* position is not in the same operon with the others.

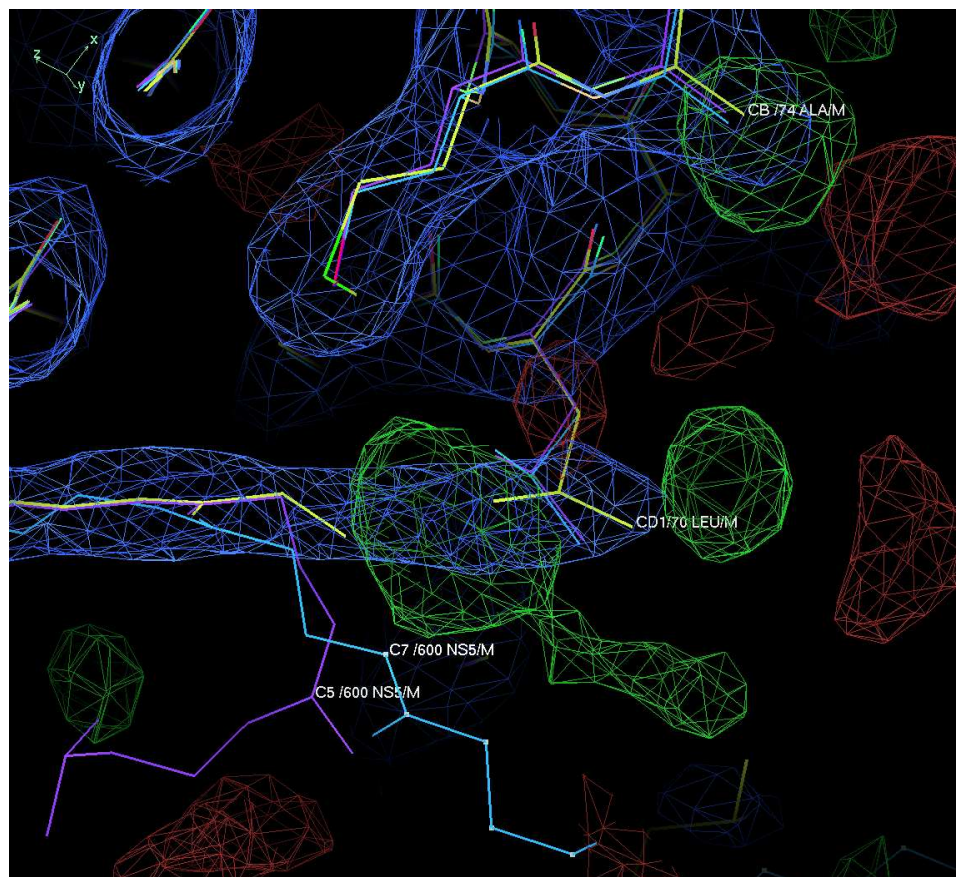


Figure 34. The $2Fo-Fc$ electron density at 1σ level (in blue) and the $Fo-Fc$ density at 3σ level (in green-red) at the carotenoid binding site. The preliminary RC model is shown in yellow (with shortened carotenoid) while two published models, PDB codes 1DXR and 1I5N are given in magenta and cyan respectively. Residue Leu M70 clashes with the apparent carotenoid density residue Ala M74 appears to not satisfy the existing density (Roszak, et al., unpublished observations)

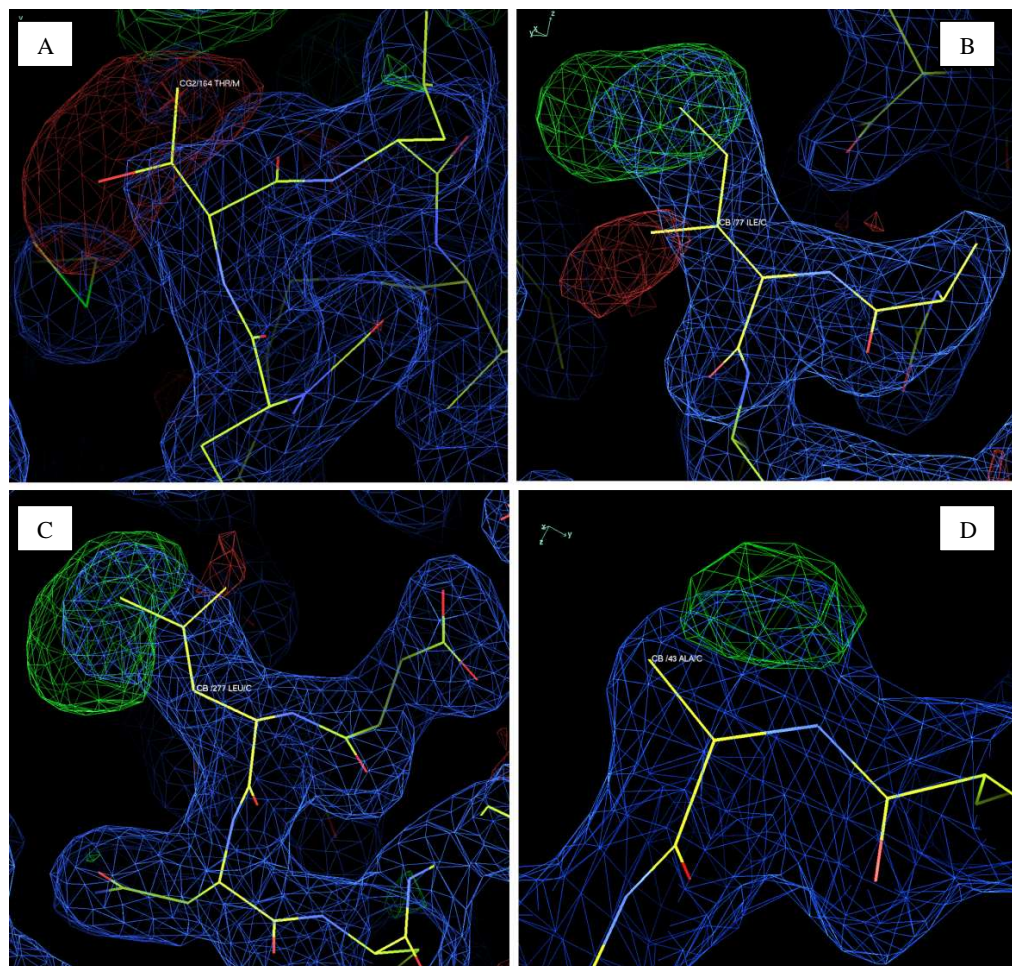


Figure 35. Four examples of putative mistakes in the published amino acid sequence identified based on the electron density maps. Colour identification as in Figure 34 (Roszak, et al., 2008). A) The side chain of Thr residue at M164 lies outside the electron density (red). B) Ile residue at C77 shows inappropriate positions where one methyl group overlaps with the density (red) while another density reveal at the other side (green). C) One of the methyl group on the side chain of Leu residue at C277 also lies outside the density (red) and other place expect more side chain (green), d) A larger amino acid residue structure is expected (green) rather than Ala for C43.

3.2. Isolation of genomic DNA from *Bl. viridis*

Initial attempts to isolate genomic DNA from *Bl. viridis* using the original protocol (described below) were not successful and gave only poor quality fragmented DNA. The experiments performed to improve and adapt this original method for *Bl. viridis* are described in this section. The final, improved protocol for isolating bacterial genomic DNA is outlined in Chapter 2.

Bl. viridis cells were harvested by as described in section 2.2.1. The supernatant was discarded and the cell pellet re-suspended in 1.5 ml Eppendorf tubes in a minimal volume of liquid, either sterile water or TE was used for comparison.

For the next step, 200 μ l of the harvested suspension was mixed 1:1 (v/v) with phenol-chloroform. The sample was vortexed briefly, incubated for 5 minutes at RT and microfuged to separate the phases. The supernatant was then removed into a fresh 1.5 ml Eppendorf tube and the volume measured. In the original protocol, the clear supernatant was washed, twice with the same volume of chloroform in order to remove the residual phenol. For comparison, a single wash step was also tried as it was hoped that fewer manipulations would decrease DNA fragmentation. The resulting viscous solution was microfuged (4 min, $13,000 \times g$), the supernatant carefully removed into a fresh 1.5 ml Eppendorf tube and the volume measured.

The DNA sample was precipitated as described in Section 2.4 and the pellet re-suspended in sterile water. To check the quantity and quality of the obtained DNA, it was run on an agarose gel as described in Section 2.5.

The initial protocol did not yield good quality, high-molecular weight DNA, even after repeating the experiment several times. Cultures were also re-grown and all chemical stocks were re-made fresh. A gel illustrating the DNA obtained is given in Figure 36 and the composition of each lane is given in Table 2. In this Figure, it is apparent that re-suspending the cells in TE buffer gives a greater yield of fragmented DNA after the final ethanol precipitation step compared with the experiment that just used water. Why this should be is not clear.

All lanes in Figure 36 had sheared DNA present judging from the large fluorescence intensity at the bottom of the gel. It is only at the beginning of the isolation where any difference is apparent. It can be seen that the lanes containing the DNA sample from cells re-suspended with TE (E, F and G, H) contain rather more DNA than the lanes containing

the DNA from cells re-suspended in H₂O (A, B and C, D), therefore, from this point onward cells were always re-suspended after pelleting in TE buffer.

Literature studies on different DNA isolation methods suggested other steps that may be worthwhile. The first isolation trial (described previously in Figure 36) was conducted without the use of any liquid N₂, SDS, Proteinase K and lysozyme. Liquid N₂ and SDS (2%) are different methods used in order to physically disrupt the cell wall and to digest proteins released from the cell lysate. These steps can be used before the phenol-chloroform extraction. I therefore modified my protocol to include these steps.

Harvested cells (200 µl) inside an Eppendorf tube were frozen in liquid N₂ for 1 to 2 minutes, then, promptly heated to 80 °C in water bath for 2 minutes. This freeze/thaw cycle was repeated three times. Next, 25 µl of SDS (2%), 50 µl of Proteinase K (20 mg/ml in TE buffer) and 80 µl of lysozyme (10 mg/ml) were added into the cell suspension. The mixture was gently inverted a few times by hand and then incubated for 1 hour at 37 °C. Lysozyme digests the polymeric compounds that make up the cell wall whereas Proteinase K digests proteins. After this incubation the protocol was continued with the phenol-chloroform treatment as described above for first isolation trial. The results between the second isolation trial (lane A (1µl DNA solution added) and B (5µl)) and the first trial (lane C (1µl DNA solution added) and D (5µl)) are compared in Figure 37.

In Figure 37, a comparison between the isolation that used liquid N₂, SDS, Proteinase K and lysozyme (Lane A and B) and the isolation that did not use these treatments (Lane C and D) is presented. The result from both trials was that no high-molecular weight DNA was present. There is high intensity fluorescence present at the bottom of the gel in Lanes C and D suggesting that the DNA had been degraded, however, there was no sign of any DNA fragments at all in Lanes A and B. This result was worrying and so further experiments were performed to see where the problem was occurring.

It was necessary to determine at which step in the method the DNA was being degraded or lost. Samples were taken after each step in the protocol and carefully analysed by gel electrophoresis (Lanes E to H in Figure 37). Lane E is the sample obtained after N₂, SDS, Proteinase K and lysozyme incubation, lane F is the sample after phenol-only extraction, lane G was sample after the first chloroform wash and lane H was the sample after the second chloroform wash.

Lane E in Figure 37 shows that liquid N₂, SDS, Proteinase K and lysozyme treatment results in the presence of high-molecular weight DNA but at low concentration. This is evident from the low intensity band above 10 kb found in the lane (marked with a yellow

arrow). Although it was difficult to load the sample following the phenol-only extraction into lane F, high-molecular weight DNA can still be visualised in lane F (Figure 37) suggesting that this step did not result in any degradation. However, there is large fluorescence intensity present from sheared DNA in lane G after the chloroform extraction. In lane H no DNA fluorescence bands can be seen at all, suggesting that the second chloroform treatment results in the loss (degradation?) of all DNA in the sample. This is probably because it is degraded to such an extent that it does not precipitate during the ethanol precipitation step.

From Figure 37, it can be seen that the fragmentation of the DNA occurs during the chloroform extraction (lane G). For this reason, therefore, DNA isolation was attempted with minimal chloroform washing. The sample was extracted using only phenol: chloroform mixture (1:1) bought from Sigma. The liquid N₂, SDS, Proteinase K and lysozyme treatments were removed from the protocol as they did not provide any significant improvement to the final result (see lane A, B and C, D in Figure 37). Doing this also minimises unnecessary steps that may decrease DNA yield. As a result of these modifications, good quality high-molecular weight DNA fragments could be obtained.

In (Figure 38) high-molecular weight DNA can be seen in all slots with no significant differences between the two duplicate samples (lanes A, B and lanes C, D). For quantification purposes, the volume added to lanes A, C (loading 0.5 µl) and lanes B, D (loading 2 µl) was varied. Then fluorescence intensity of the high-molecular weight DNA was compared with that from the DNA markers. The concentration of the total *Bl. viridis* genomic DNA obtained by this method, based on comparison with DNA marker fluorescence, was ~80 ng/µl. Now I could proceed to the PCR steps.

Table 2. The various volumes used for gel electrophoresis shown in Figure 36. The effect of using water compared with TE buffer was tested.

	H ₂ O				TE			
	A	B	C	D	E	F	G	H
Volume of DNA solution (ul)	1	2	1	2	1	2	1	2
Re-suspended cells used (ul)	80		100		80		100	
TE Buffer	-		-		200		200	
Water	200		200		-		-	

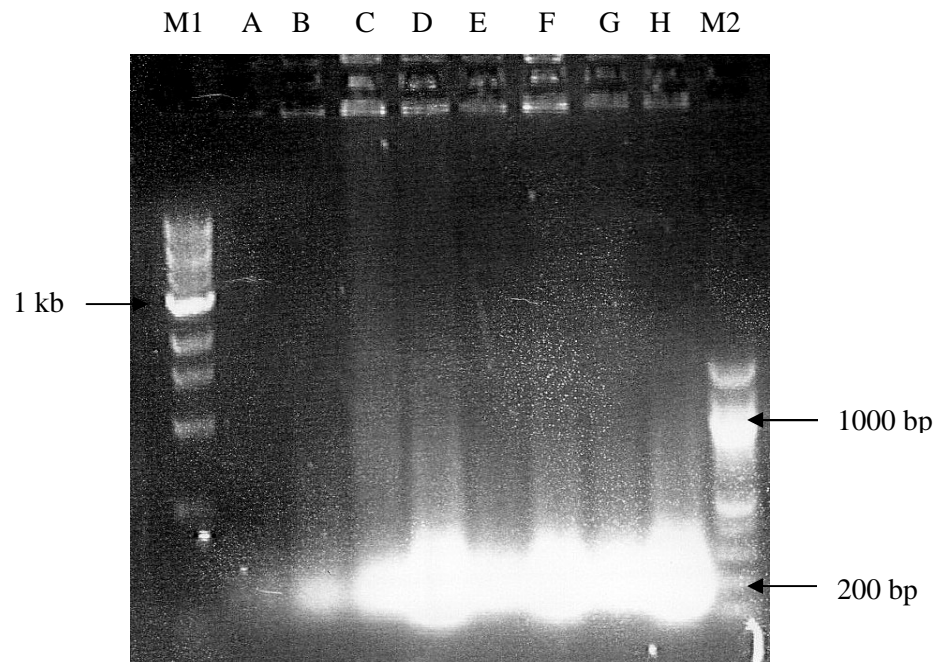


Figure 36. An example of gel electrophoresis result of DNA isolation that compares the use of TE buffer and water. The protocol used for each slot can be seen on Table 2. It is evident that no high-molecular weight DNAs were obtained from this experiment. The sizes of markers used are 1 μ l of a 10 kb ladder (M1) and a 1.5kb ladder (M2) (obtained from New England Biolabs).

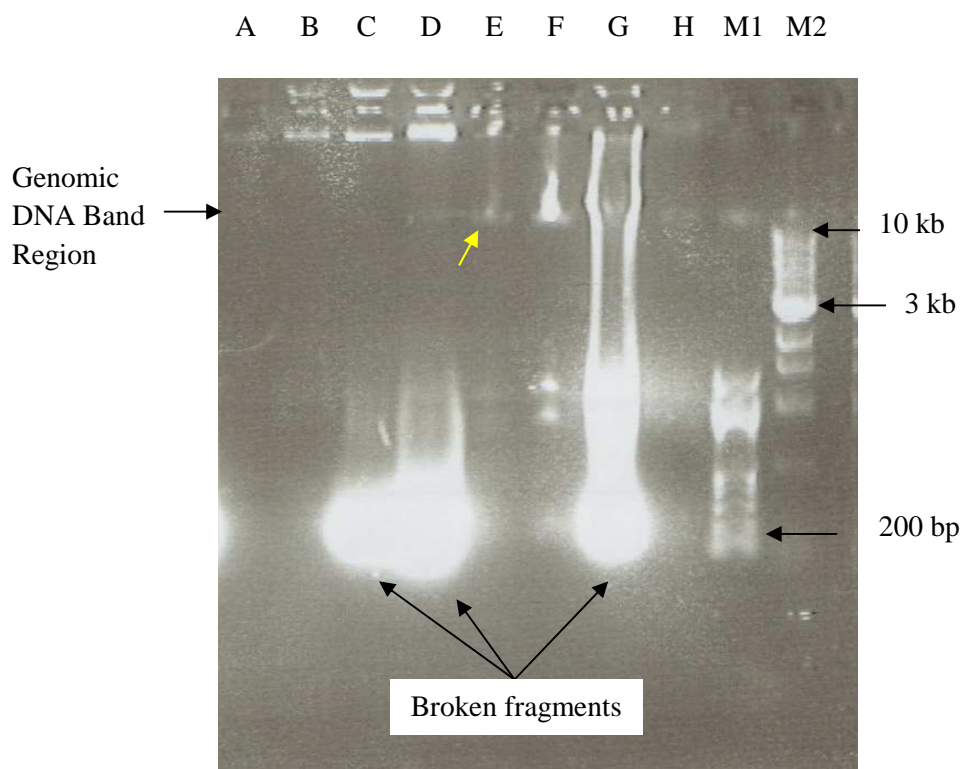


Figure 37. Gel electrophoresis analysis of each step analysis of the DNA isolation. DNA samples isolated using the liquid N_2 , SDS, Proteinase K and lysozyme treatment is given in Lane A (1 μ l) and B (5 μ l). Lanes C (1 μ l) and D (5 μ l) are without these treatments. The DNA extraction and precipitation was conducted as described in section 2.4. The results of sampling after each step are shown in lanes E to H. Sample after liquid N_2 , SDS, Proteinase K and lysozyme incubation (E), sample after phenol extraction (F), sample after first chloroform wash (G) and sample after second chloroform wash (H) are shown. Only a small amount of DNA can be seen in lane E. The size of markers used (M1) were 1 μ l of a 10kb ladder and (M2) a 1.5 kb ladder (obtained from New England Biolabs).

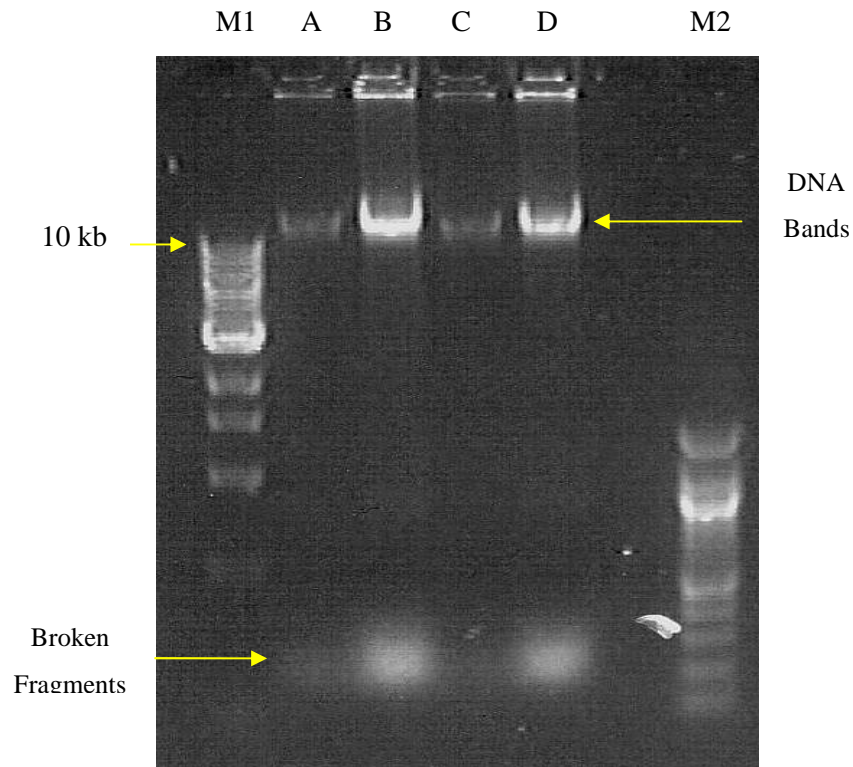


Figure 38. Gel electrophoresis of total genomic DNA obtained when the chloroform wash step was omitted. For this experiment, liquid N₂, SDS, Proteinase K and lysozyme treatments were also not used. The use of chloroform was also limited to one phenol-chloroform extraction only. Genomic DNA bands (more than 10 kb) can be clearly visualized in this picture. Lanes A, B and C, D are duplicates from the same initial sample. Lanes A and C have 0.5 μ l while lanes B and D have 2 μ l sample. The markers used were 1 μ l of a 10kb ladder (M1) and 1 μ l of a 1.5kb ladder (M2) (obtained from New England Biolabs).

3.3. Primers and fragment amplification

The first important step in the process of amplifying DNA by PCR is to design good quality primers. Primers are short oligonucleotide sequences (around 15 – 25 bp), which are used to begin the DNA synthesis. The GC content of the primers will affect its behaviour such as the annealing temperature or the possibility of producing secondary structure, which is unwanted. The sequence of as many as 500 bases can optimally be determined by the Sanger dideoxy method (Berg, et al., 2002). Therefore, the position of the primers must be planned according to the sequencer's capability. Each primer has to cover each of the approximately 500 bp sections in the gene to be sequenced. Primers used

for sequencing have to be positioned before, in the middle and after the expected fragment. The idea was to have all genes fully covered.

There are several important factors to be taken into consideration when designing good primers; they are GC content, fragment length and melting temperature. Neglecting these factors may result in bad amplification and bad sequencing results. The worst scenario will happen if there are no good positions, in the sequence, to design primers due to such factors as too much GC content. The properties of the primers were checked by using the online oligonucleotide properties calculator, which was obtained from the Northwestern University, web tools website (Northwestern University, 2008).

The primers were named based on the position of the primers in the gene and their role, either as forward or reverse primers. For instance, *pufBvirF* is primer that was designed for *Bl. viridis pufB* gene, positioned before the start codon of the genes that encodes β -subunit and will replicate in forward direction (5' to 3') whereas *pufCvirR* is primer that starts at the last base of gene that encodes the C subunit and replicates in a reverse complement direction. The reverse primer was designed to be complementary to the reverse complement strand (5' to 3') of the nucleotide sequence. Another example of a reverse primer is *pufL104virR*. This primer starts at base number 104 in the gene that encodes the L subunit and will replicate in a reverse complement direction (5' to 3').

Fourteen primers were designed for the PCR and the sequencing purposes (Table 3) based upon the previous sequence information obtained from Michel *et al* (1985 and 1986). The positions of these designed primers relative to the organisation of the *puf* and *puh* operons can be seen in Figure 39. The primers that were used in the fragment amplification will be described in section 3.3.1 whereas those used for sequencing in Section 3.3.2.

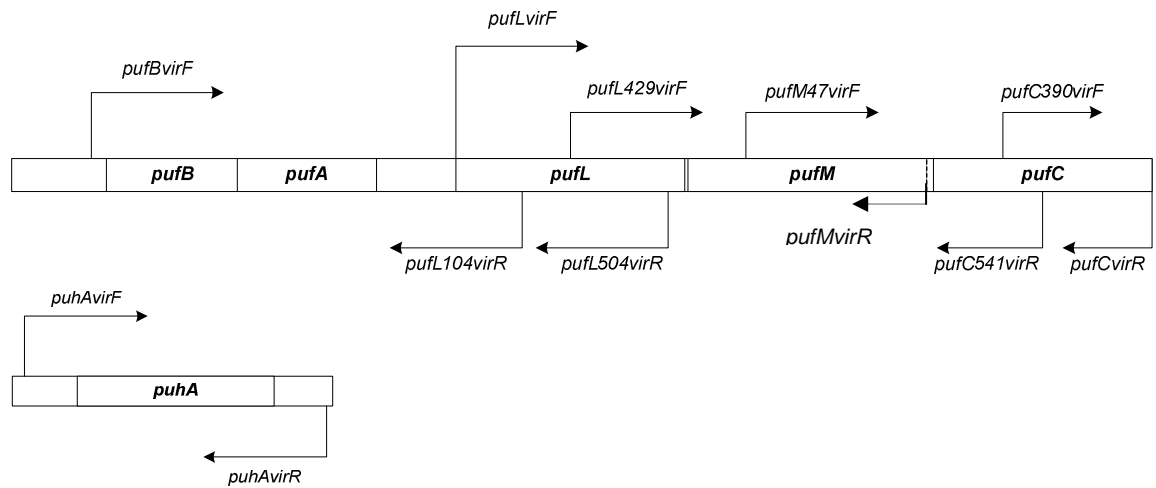


Figure 39. Illustration of the primers position in the genes map. Marked position on DNA sequence can be seen on appendix. Primers were checked first by amplification for its quality before used for sequencing. Amplified fragments that have poor quality primers for sequencing has to be cloned into a vector. Then, the cloned product was sequenced by T7 and SP6 primers.

Table 3. List of all primers involved in PCR amplification and sequencing. In total, there are 14 primers that were designed for this research.

No	Oligo Name	Length (bp)	Melting Temperature (oC)	GC (%)	Dimer	2ndry
1	<i>pufBvirF</i>	21	74.9	57	No	Moderate
2	<i>pufLvirF</i>	24	76.5	60	No	None
3	<i>pufL104virR</i>	22	67.8	50	No	None
4	<i>pufL429virF</i>	21	70.6	62	No	None
5	<i>pufL504virR</i>	21	63	52	No	None
6	<i>pufM47virF</i>	20	67.9	60	No	None
7	<i>pufM478virF</i>	21	71.4	62	None	Weak
8	<i>pufMvirR</i>	20	64.5	50	No	None
9	<i>pufC138virF</i>	19	51	53	No	None
10	<i>pufC390virF</i>	18	72.6	67	yes	None
11	<i>pufC541virR</i>	18	67.2	61	yes	None
12	<i>pufCvirR</i>	18	61.3	50	No	None
13	<i>puhAvirF</i>	18	60.2	53	No	None
14	<i>puhAvirR</i>	17	59.5	50	No	None

3.3.1. Amplification of the gene fragments

There were 10 primers used to amplify the required gene fragments. The combinations of primers used, the predicted and the final length of the amplified products are shown in Table 4.

Table 4. List of primers (in pairs) used in the amplifications. There were 10 primers involved in producing 5 fragments that were used as sequencing templates.

No	Primers Name	Gene/s contained	Temperature (oC)		GC (%)	Predicted Length	Obtained Length
			Melting	Annealing			
1	<i>pufBvirF</i>	<i>pufB</i> , <i>pufA</i> and most of <i>pufL</i>	74.9	55	57	1226	1226
	<i>pufL504virR</i>		63		52		
2	<i>pufLvirF</i>	<i>pufL</i> and <i>pufM</i>	76.5	58	60	1831	1831
	<i>pufMvirR</i>		64.5		50		
3	<i>pufM478virF</i>	End part of <i>pufM</i> & <i>pufC</i>	71.4	55	62	1564	1564
	<i>pufCvirR</i>		61.3		50		
4	<i>pufC138virF</i>	end part of <i>pufC</i>	51	55	53	933	933
	<i>pufCvirR</i>		61.3		50		
5	<i>puhAvirF</i>	<i>puhA</i>	60.2	45	53	823	823
	<i>puhAvirR</i>		59.5		50		

The PCR procedure used is described in the following example for *pufC*. *PufM478virF* and *pufCvirR* primers were employed. The expected length from the PCR process was 1,564 bp (Figure 40, Figure 41, and Figure 42). However, before a high concentration of this fragment could be obtained through PCR amplification, optimisation of the protocol had to be conducted. In this case the effect of primer concentration was tested.

Three concentrations of reverse primers were used (900 nM, 600 nM and 300 nM). The concentration of forward primer was kept the same (900 nM) for each different reverse primer concentration. The basic PCR conditions were kept the same. The annealing temperature used was 50 °C and the extension time was 50 seconds. Details of the protocol and the gel electrophoresis analysis of the results can be seen in Table 5 and Figure 40. Several DNA bands were found in each lane in the gel. There is band with the expected size that can be seen in lane corresponding to the highest reverse primer concentration, 900

nM (lane 1, Figure 40). The expected length of *pufC* fragment is 1,564 kb. Poor yields of the expected product were found in other lanes. The highest reverse primer concentration was, therefore, used for the next trial (900 nM).

The next parameter that was optimised was the annealing temperature. In some primers that have a GC content of more than 60%, low temperature can lead to the production of secondary products. In the case of the first trial with *pufC* (in Figure 40), secondary products were detected in all lanes. As can be seen in Table 3 and Table 4, the GC content of the primers is at acceptable level (~60 %). However, secondary products were clearly a problem. For this reason, the annealing temperature was increased by 5 °C to 55 °C. This modification resulted in a clear fluorescence band for a DNA product with the exact size of expected *pufC* fragment (Figure 41) with, in this case, no significant secondary products.

After the optimised protocol had been determined, large scale amplification of the *pufC* fragment was performed in order get enough for sequencing. The result of this *pufC* amplification can be seen in the gel in Figure 42. The concentration of the product obtained from this amplification was, approximately, 50 ng/μl. The amplification product was then purified using the Qiaquick® spin kit before being sent for sequencing.

All required gene fragments for use as sequencing templates were successfully amplified by PCR. These amplification products are summarised in Table 4. The next step was to test the primers that were going to be used for sequencing.

Table 5. PCR protocol for the isolation of the *pufC* fragment. *pufM478virF* was the forward primer and *pufCvirR* the reverse primer. There were 3 variations of reverse primer concentrations (900, 600 and 300 nM) tested. The volumes of reverse primer (in μl) are highlighted in yellow. The concentration of DNA template used was 8 ng/ μl . Best results were only found with Master mix 1, where the expected product was produced. However, further optimisation of the annealing temperature was required.

Cycler program:

Initial denaturation : 98 °C, 1 minute
 Denaturation : 98 °C, 30 seconds
 30x Annealing : 50 °C, 45 seconds
 Extension : 72 °C, 50 seconds
 Final extension : 72 °C, 5 minute
 Hold : 40 °C, ~

Master mix composition:

PCR mix volume : 15 μl
 Forward primer : *pufM478virF*; 900 nM
 Reverse primer : *pufCvirR*; 900 nM, 600 nM, 300 nM

Primer concentration	Reverse		900 nM	600 nM	300 nM
	Base Compositions per sample	μl vol. per sample	MASTER Mix 1 (μl)	MASTER Mix 2 (μl)	MASTER Mix 3 (μl)
Sample amount		1	2	2	2
Buffer HF 5x conc.		3.00	6.00	6.00	6.00
dNTPs 10mM	0,2mM	0.30	0.60	0.60	0.60
Water		10.15	18.52	19.40	20.60
Primer f 20 μM	200	0.15	1.34	1.34	1.34
Primer r 20 μM	200	0.15	1.34	0.90	0.45
DNA pol. Phusion (2U/ μl)		0.10	0.20	0.20	0.20
Template DNA (8ng/ μl)		1.00	2.00	2.00	2.00
	<i>Suma</i>	<i>15.00</i>	<i>30.00</i>	<i>30.00</i>	<i>30.00</i>

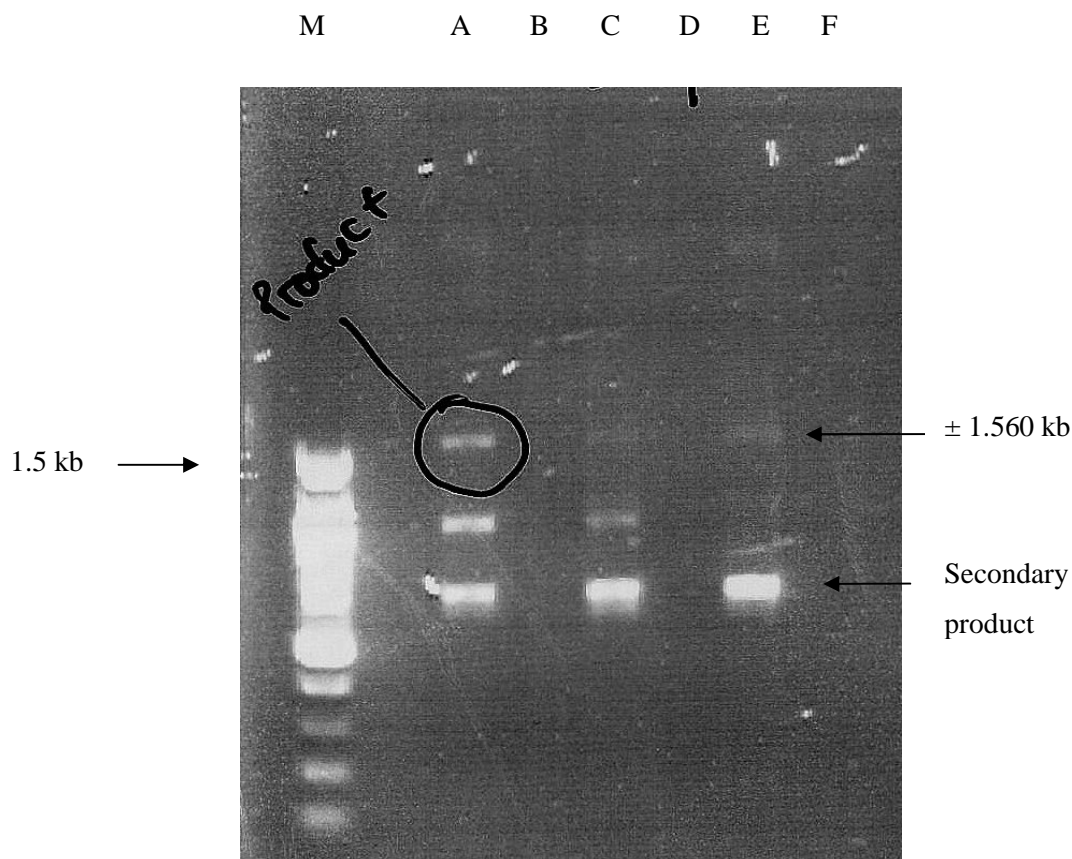


Figure 40. Gel electrophoresis analysis of the amplification of the *pufC* fragment (1,561 bp). *PufM478virF* is the forward primer and *pufCvirR* is the reverse primer. Three variations in reverse primer concentration were used in this amplification (A=900 nM, C= 600 nM and E=300 nM). High intensity of secondary products was seen with each variation. Slot numbers B, D and F are negative controls where the reactions were carried out in the absence of the genomic DNA. The highest intensity of the expected product was found in lane A, where highest reverse primer concentration was used. Clearly further optimisation is required. The marker (M) used was 1 μ l of a 1.5kb ladder (obtained from New England Biolabs).

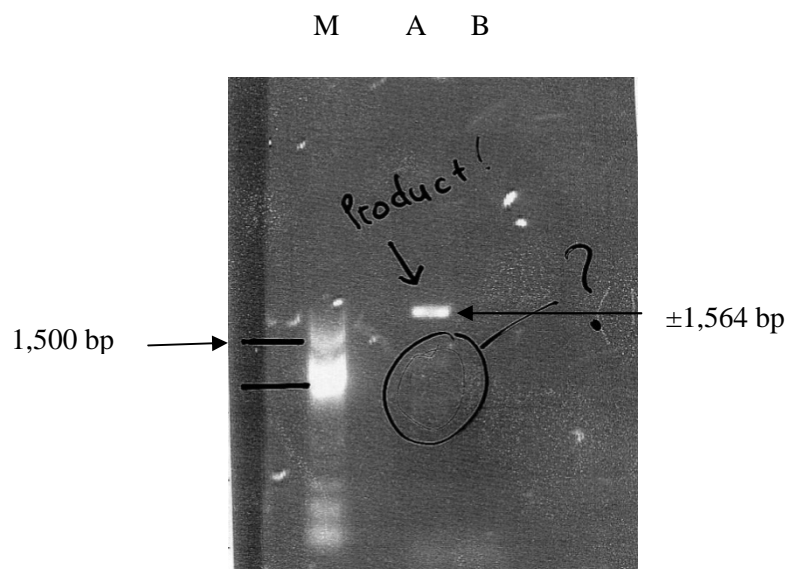


Figure 41. Gel electrophoresis analysis of the amplification of the *pufC* fragment. The fragment has size of around 1,564 bp. lane A used the optimised protocol (annealing temperature is 55 °C and highest reverse primer concentration is 900 nM). Lane B is the negative control with no added genomic DNA. The marker used (M) was 1 µl of the 10kb ladder (obtained from New England Biolabs).

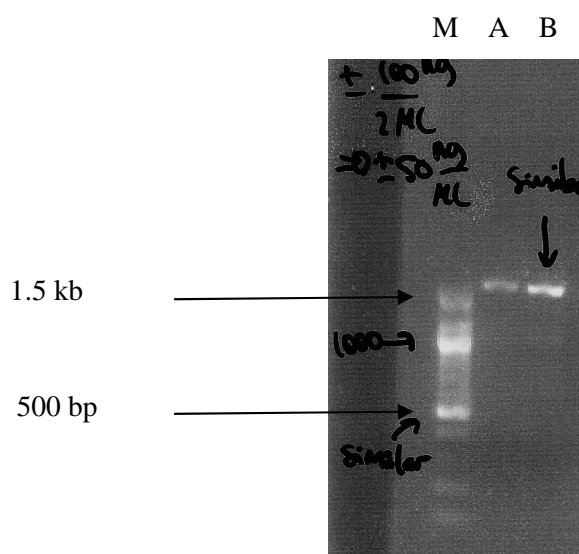


Figure 42. Gel electrophoresis analysis of the large scale amplification of the *pufC* fragment (1,564 bp). Two different volumes were used to measure the concentration (A) 1 µl and (B) 2 µl. The fluorescence intensity of the product in lane B was slightly more intense than the bands in the ladder (500 bp, 97 ng/ µl). The concentration in lane B was therefore estimated to be approximately 100 ng/ 2 µl (50 ng/ µl). The marker used (M) was 1 µl of the 1.5kb ladder (obtained from New England Biolabs).

3.3.2. Testing the primers to be used for sequencing

There were 10 primers designed to be used in the sequencing. The quality of these primers was tested using PCR prior to sending samples to the sequencing service. The same procedure as for the fragment amplification was used. If the primers pairs did not produce the exact size of fragment predicted then the sequencing will not be a success. The primer combinations, the target of the genes to be sequenced, the predicted and obtained length of the amplification products are shown in Table 6.

Almost all combinations successfully amplified the required genes for sequencing. There was only one combination, *pufC390virF* and *pufCvirR*, which was unsuccessful. The predicted size for this pair was 680 bp. *PufC390virF* was designed to amplify the sequence at the end of *pufC*. Optimisation of the PCR protocols was tried, but the results always showed the presence of secondary products (Figure 43). When the *pufC390virF* primer/*pufC* DNA combination was sent for sequencing the result was very poor. This is shown in Figure 44. This Figure compares an example of a successful sequencing run with this bad one.

Unfortunately, a single primer could not replace *pufC390virF* and provide the coverage needed. Therefore, another method involving cloning was used to overcome this problem.

Table 6. List of primers involved in the sequencing. There were 10 primers involved in the sequencing. Most of primers could be successfully used in the sequencing reaction. The one pair that failed is highlighted in red.

No	Primers Name	Target Genes to be sequenced	Temperature (oC)		GC (%)	Predicted Length	Obtained Length
			Melting	Annealing			
1	<i>pufBvirF</i>	<i>pufB, pufA pufL</i>	74.9	55	57	827	827
	<i>pufL104virR</i>		67.8		50		
2	<i>pufBvirF</i>	<i>pufL</i>	74.9	55	57	1226	1226
	<i>pufL504virR</i>		63		52		
3	<i>pufL429virF</i>	<i>pufL, pufM, pufC</i>	70.6	55	62	1940	1940
	<i>pufC541virR</i>		67.2		61		
4	<i>pufM47virF</i>	<i>pufM, pufC</i>	67.9	55	60	933	933
	<i>pufMvirR</i>		64.5		50		
5	<i>pufC390virF</i>	<i>pufC</i>	72.6	55	67	680	1,800; 1,000; 800; 680; 250
	<i>pufCvirR</i>		61.3		50		

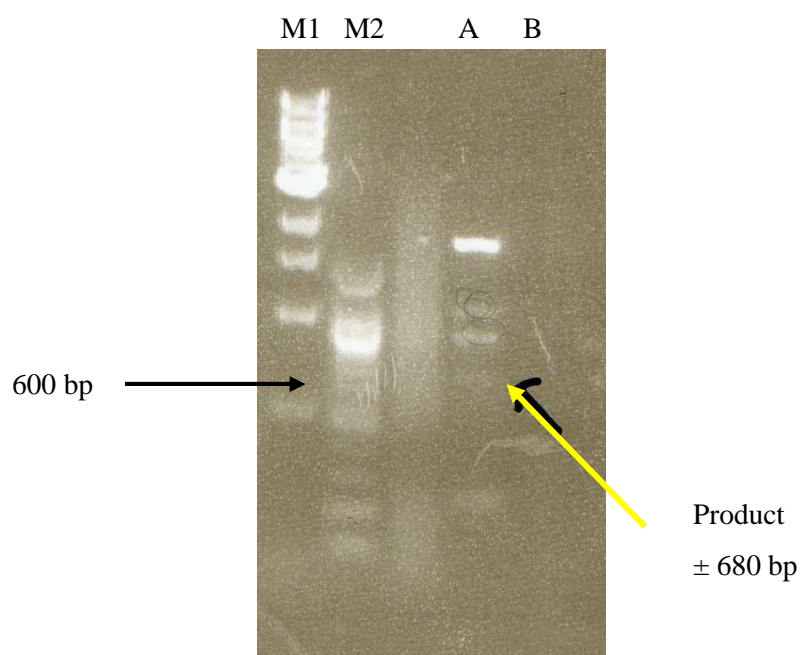


Figure 43. Gel electrophoresis analysis of the result of testing *pufC390virF* and *pufCvirR* primers. The presence of secondary products is clear. The size of the expected product (Lane A) was 680 bp (pointed by yellow arrow). Lane B is the negative control. The markers used were 1 μ l of a 10kb ladder (M1) and 1.5kb ladder (M2) (obtained from New England Biolabs).

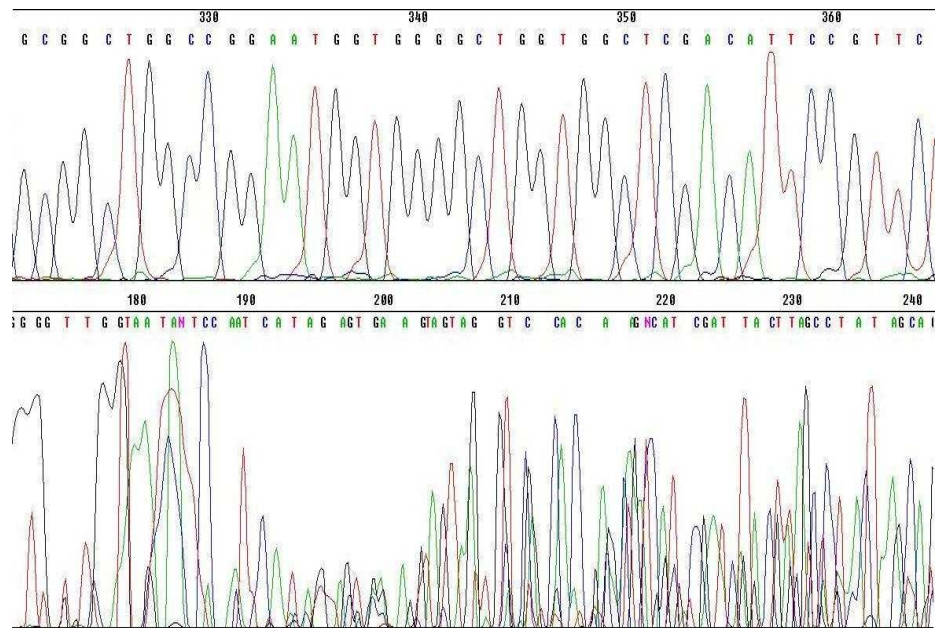


Figure 44. Examples of sequencing results that obtained by Sanger's dideoxy method. Above: a good sequence result. Bottom: a poor sequence result obtained with. Poor sequences have untidy peaks resulting in difficulties in reading the order of the bases.

3.3.3. PCR products that required cloning before they could be sequenced

As described above there were two parts of the genes to be sequenced that needed to be cloned before they could be sequenced. These were the end part of *pufC* where the primers designed for the sequencing reaction failed and the whole of *pufA* where due to the lack of previously available sequence data it was not possible to design primers that would give full coverage of the gene sequence (See Figure 45).

The combination of *pufC138virF* and *pufCvirR* did produce an amplification product that contained the end part of *pufC*. This length of this fragment was 933 bp, see Figure 46. This fragment was inserted into the vector and then cloned into *E. coli*. Therefore, the T7 and SP6 primers could be used to sequence the cloned fragment. More detail about cloning will be discussed in section 3.4.

GAATTCCC **CGGCTAGACAGTTTCTG**CCACGGAGGTTCTT**ATG**TATCACGGCGCTCTCGCTCAACATTTAGA
 CATCGCACAACCTCGTATGGTACGCGCAGTGGCTGGTCATCTGGACGGTTGTCTGCTGTACCTCCGCCGTGAG
 GACCGTCGCGAAGGCTACCCGCTGGTTCGAGCCGCTTGGTCTCGTCAAGCTGGCGCCGGAAGACGGCCAGGTCT
 ACGAGCTGCCCTATCCCAAGACGTTCTGTGCTCCCGCACGGCGGCACCGTCACGGTTCCGCGTCGTCTCGGA

AACCCGCGAACTGAAGCTGGCGCAGACCGACGGCTTTCGAAGGCGCCCCGCTGCAGCCGACCGGCAATCCGCTG
 GTCGACGCCGTCGGCCCCGGCTTCGTATGCCGAGCGTGCGGAAGTGGTCGACGCCACGGTTGACGGCAAGGCCA
 AGATCGTCCCGCTGCGCGTTTGCACCGACTTCTCGATCGCGGAAGGCGACGTCGATCCGCGTGGCCTGCCGGT
 GGTGCGCTGACGGCGTCGAGGCTGGTACGGTTACCGACCTCTGGGTCGACCGCTCGGAGCACTATTTCCGC
 TACCTCGAGCTCTCGGTGGCCGGCAGCGCCCGCACCGCGCTGATCCCGCTCGGCTTCTGCGACGTCAAGAAGG
 ACAAGATCGTCGTGACGTGATCCTGTCCGAGCAGTTCGCCAACGTGCCGCGTCTGCAGAGCCGCGACCAGAT
 CACGCTGCGCGAAGAAGACAAGGTGTGGCCTACTACGCTGGCGGTCTGCTCTACGCGACGCCGGAGCGTGCG
 GAATCGTTGCTGTGAACGACGATTTTCGA

puhAvirF: CGGCTAGACAGTTTTCTGC

puhAvirR: GTTGCTGTGAACGACGA

Figure 45. Previous published *puhA* DNA sequence obtained from GeneBank under accession number; *Bl. viridis*, X02659. Start codon (ATG) and stop codon (TGA) is highlighted in yellow. Forward primer is highlighted in green while reverse primer in gray. This *puhA* gene has fragment size of 800 bp. Around first part of 40 to 100 bp of every primer sequence will have bad result (see Figure 44). Therefore, cloning this fragment into a vector will be necessary.

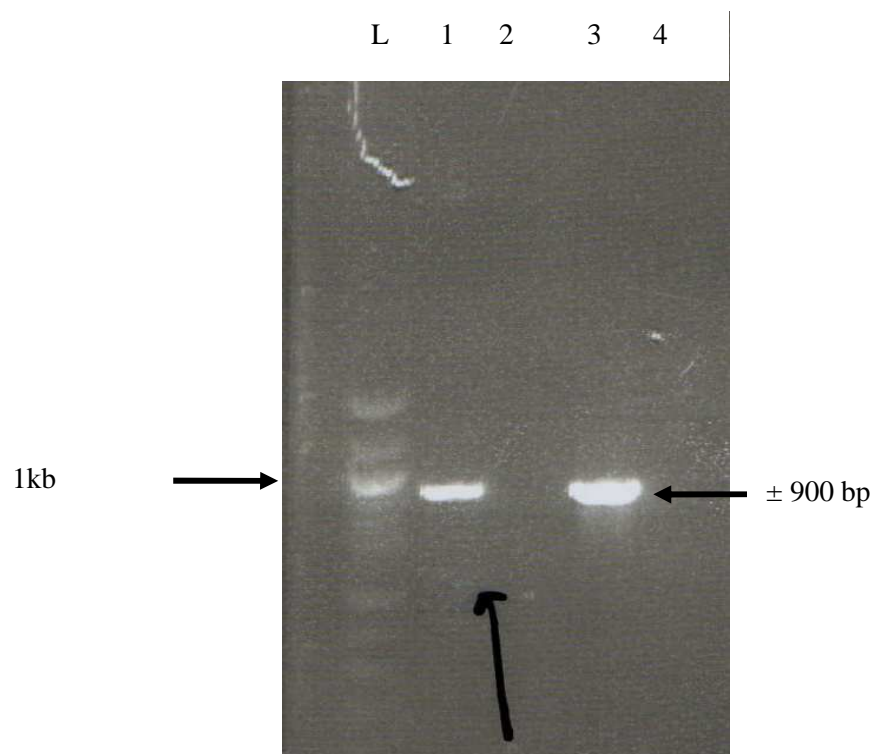


Figure 46. DNA Fragment resulting from the amplification using the *pufC138virF* and *pufCvirR* primers. PCR reaction was performed using *pufC* fragment as the template. The product length (Lane A and C) was expected at around 933 bp. Lanes B and D are negative controls. The marker used (M) was 1 µl of 10 kb ladder (obtained from New England Biolabs).

The other case where a fragment could not be sequenced directly was the *puhA* gene. The position of primers, which were designed to amplify the *puhA* gene for cloning, can be seen in Figure 45.

The optimised protocol for the amplification of the *puhA* gene is given in Table 7. The result of this optimised PCR reaction is shown in Figure 47. The result of the large scale amplification of the *puhA* gene is shown in Figure 48. This amplified product was then purified by using the Qiaquick® spin. The concentration of this fragment was approximately 50 ng/μl.

Table 7. PCR protocol for *puhA* fragment isolation. *puhAvirF* was used as forward primer whereas *puhAvirR* as reverse primer. There were 3 variations of primers concentrations (400, 600 and 900 nM). Concentration of DNA template used was 8 ng/μl. All variations have great results (Figure 47). Therefore, the lowest primer concentration could be used for the amplification purpose.

Cycler program:

Initial denaturation : 98 °C, 1 minute
 Denaturation : 98 °C, 30 seconds
 30x Annealing : 45 °C, 45 seconds
 Extension : 72 °C, 30 seconds
 Final extension : 72 °C, 5 minute
 Hold : 40 °C, ~

Master mix composition:

PCR mix volume : 15 μl
 Forward primer : *puhAvirF*; 400 nM, 600 nM, 900 nM
 Reverse primer : *puhAvirR*; 400 nM, 600 nM, 900 nM

as ion	Primer			400 nM	600 nM	900 nM
	Base Compositions per sample	μl vol. per sample	MASTER Mix 1	MASTER Mix 2	MASTER Mix 3	
Sample amount		1	2	2	2	
Buffer HF 5x conc.		3.00	6.00	6.00	6.00	
dNTPs 10mM	0,2mM	0.30	0.60	0.60	0.60	
Water		10.15	20.00	19.40	18.50	
Primer f 20μM	200	0.15	0.60	0.90	1.35	
Primer r 20μM	200	0.15	0.60	0.90	1.35	
DNA pol. Phusion (2U/ μl)		0.10	0.20	0.20	0.20	
Template DNA (8ng/ μl)		1.00	2.00	2.00	2.00	
	Suma	15.00	30.00	30.00	30.00	

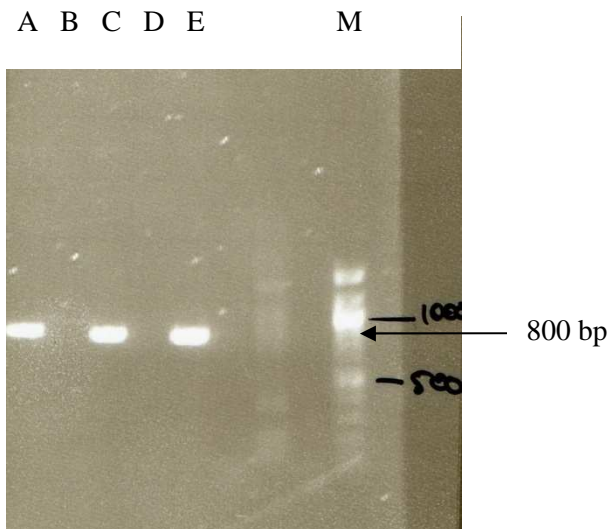


Figure 47. Gel electrophoresis analysis of the result of protocol optimisation of the amplification of *puhA* (823 bp). Variations of primers concentration were used in the amplification (400 nM; lane A, 600 nM; lane C and 900 nM; lane E). All products have clear and correct size. Unfortunately, the primers, which were used for amplifications could not be used for sequencing purposes. Therefore, cloning the fragment into plasmid will be necessary. The marker used (M) was 1 μ l of 1.5 kb ladder (obtained from New England Biolabs).

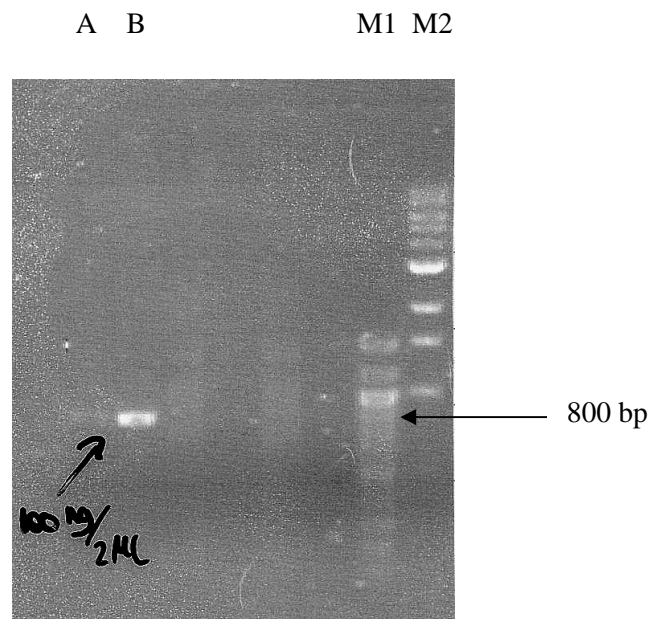


Figure 48. Gel electrophoresis analysis of the result of the large scale *puhA* amplification. Fair intensity is found in lane A (1 μ l) while sharp intensity in lane B (2 μ l). Product intensity in lane B looks similar with the 3 kb band in the ladder. The product has concentration of approximately 100 ng / 2 μ l (50 ng/ μ l). Protocol used can be seen on Table 7. The markers used were 1 μ l of a 1.5 kb (M1) and 10 kb (M2) ladder (obtained from New England Biolabs).

3.4. Fragment cloning

As also mentioned in previous section (3.3), the fragments that required cloning in this study were end part of *pufC* and *puhA*. This section will describe how this cloning was performed using *puhA* as an example. The cloning method can be divided into ligation and transformation. Ligation is the part where PCR product, *puhA* for example, is inserted into the cloning vehicle, the pGEM-T Easy® vector system. Transformation is the insertion of the cloning vehicle into the *E. coli* cells. The key components in ligation step are the vector, the DNA ligation enzyme (ligase) and the PCR product. The ligation reaction was performed by using a kit obtained from Promega. The ligation protocol can be seen in Table 8.

The position where the PCR product will be inserted can be seen in Figure 30 and Figure 31 in Chapter 2.7. This pGEM®-T Easy vector has a size of 3 kb, including the T7 and SP6 replication site (178 bp). Thus in the case of *puhA*, the ligation product should give the new plasmid a total length of 3,8kb. Amplification using T7 and SP6 as primers should then produce a fragment with the size of about 1kb (823bp for *puhA* + 178bp for T7 and SP6).

In the transformation step, the plasmid containing *puhA* was inserted into *E. coli* cells by heat shock, for 45 to 50 seconds, using a water bath set at 42 °C. The ‘transformation’ cultures were then incubated at 37 °C for 1.5 hours. These putative transformed cultures were aseptically inoculated onto LB media plates and incubated overnight (16-24 hours) at 37 °C for blue-white colour screening. Figure 49 shows the result of this test. 9 white colonies were selected from these two plates and then aseptically inoculated onto a new plate in order to obtain bigger amounts of the potential ‘transformant’ cultures.

The presence of the correct insert in the plasmid has to be confirmed, either by the Ruscony method or by PCR. Since the Ruscony method was quicker it was used first. This was then confirmed by PCR, using T7 and SP6 as primers.

All 9 white colonies were subjected to the Ruscony test as described in Chapter 2 (Section 2.7). The result of these tests is shown in the gel presented in Figure 50. Only one of these colonies actually contained the correct plasmid (lane I). This can be seen from the size of the band, which is 800 bp bigger than the others. The other lanes show plasmids that have sizes around 3 kb (lanes A to H) indicating the absence of the required insert.

After the Ruscony test, the nine selected colonies were also used as the templates in the PCR reactions. The PCR protocol used is given in Table 9. The results of these PRC

reactions are shown in Figure 51. Lanes A-H showed the presence of inserts but their sizes are too small to have included *puhA*. Lane I, which should have contained the correct insert based on the Ruscony test, strangely did not show an insert. However when the PCR reaction with this sample was given a longer extension time (lane K) the correct insert size was found. This confirms that *puhA* had indeed been cloned in this white colony. The plasmid was then isolated and purified using the QIAGEN® Plasmid purification kit. The plasmid has an approximate concentration of 100 ng/ µl (Figure 52) and was ready to be sequenced.

This procedure was also used successfully to clone the end of *pucC*.

Table 8. Ligation protocol used for *puhA* cloning. The *puhA* fragment was obtained from a previous PCR amplification (Figure 48). Prior to using the *puhA* fragment the ligation protocol was checked using the manufacturer's test inserts. This test gave the expected results (data not shown).

Ligation Protocol - pGEM-T Easy Vector system			
Vir- <i>puhA</i> PCR fragments			
Date: 28/04/2008	Vir <i>puhA</i> 40ng/ul	Positive control	negative control (background)
Buffer 2x Rapid	10.0	10.0	10.0
pGEM-T Easy Vector (50ng/ul)	2.0	1.0	1.0
PCR product	6.0	0.0	0.0
Control DNA	0.0	4.0	0.0
T4 DNA ligase	2.0	2.0	2.0
water	0.0	3.0	4.0
total volume	20.0	20.0	20.0
Incubate at 37°C for 2 h, or at 4 °C overnight			

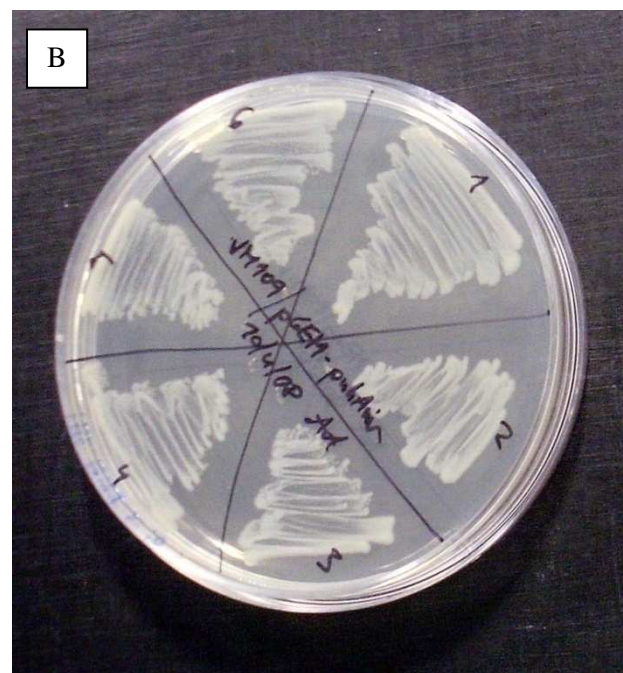
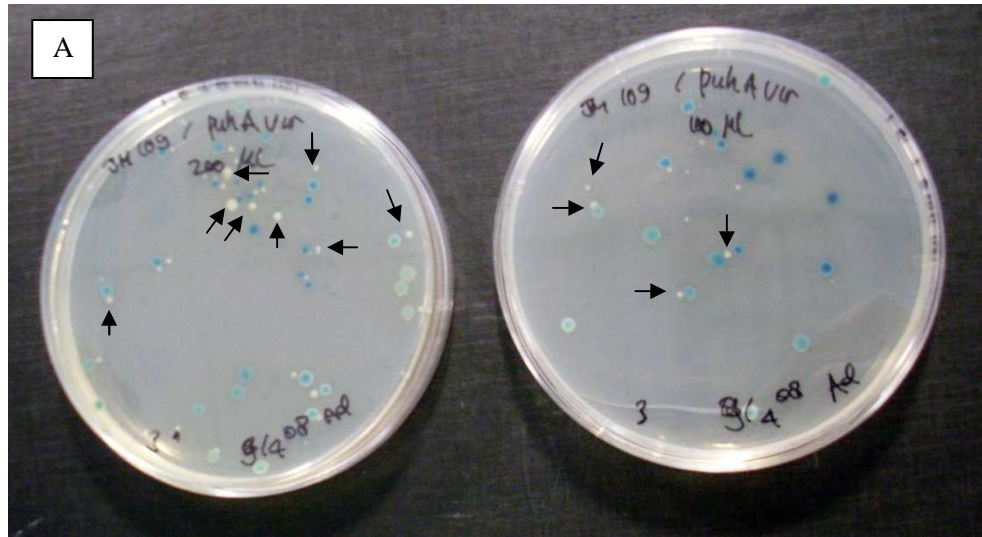


Figure 49. Pictures of the plates of the blue-white colour screening to test for successful transformation of *puhA* in the pGEM© T-Easy plasmid vector. (A) White colonies, pointed out by black arrows, show positive clones. The blue colonies are negative clones. (B) The plate containing the amplified selected white colonies.

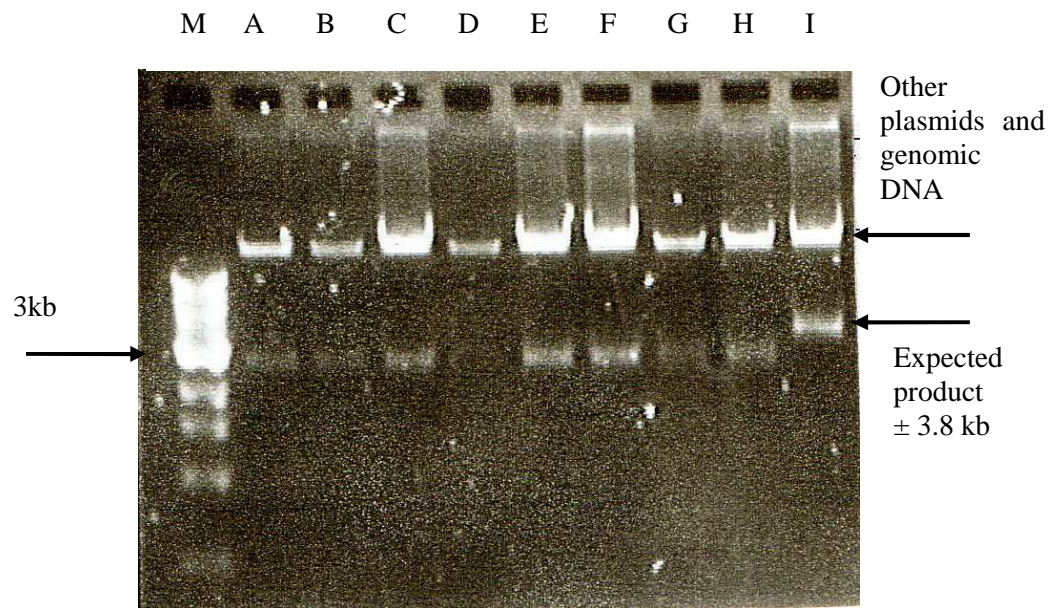


Figure 50. A gel showing the results of the Ruscony test on the white colonies, to test for the presence of the *puhA* insert. Negative clones are found in Lanes A to H. There is only one positive clone (Lane I). The size of this band is around 3.8 kb. The marker (M) used 1 μ l of a 10 kb ladder (obtained from New England Biolabs).

Table 9. The PCR protocol for amplification of the *puhA* insert. T7 was used as forward primer and SP6 as reverse primer. The concentration of both primers was 200 nM.

Cycler program:

Initial denaturation : 95 °C, 2 minute s
 Denaturation : 94 °C, 20 seconds
 30x Annealing : 45 °C, 30 seconds
 Extension : 72 °C, 1.5 minutes
 Final extension : 72 °C, 5 minute
 Hold : 4 °C, ~

Master mix composition:

PCR mix volume : 15 µl
 Forward primer : T7, 200 nM
 Reverse primer : SP6, 200 nM

	Base Compositions per sample	µl vol. per sample	MASTER Mix
Sample amount		1	9
Buffer HF 5x conc.		1.5	13.5
dNTPs 10mM	0,2mM	0.45	4.05
Water		10.15	91.35
Primer f 20µM	200	0.15	1.35
Primer r 20µM	200	0.15	1.35
DNA pol. Phusion (2U/ µl)		0.10	.9
Cells		0.00	0
	<i>Suma</i>	<i>15.00</i>	<i>135.00</i>

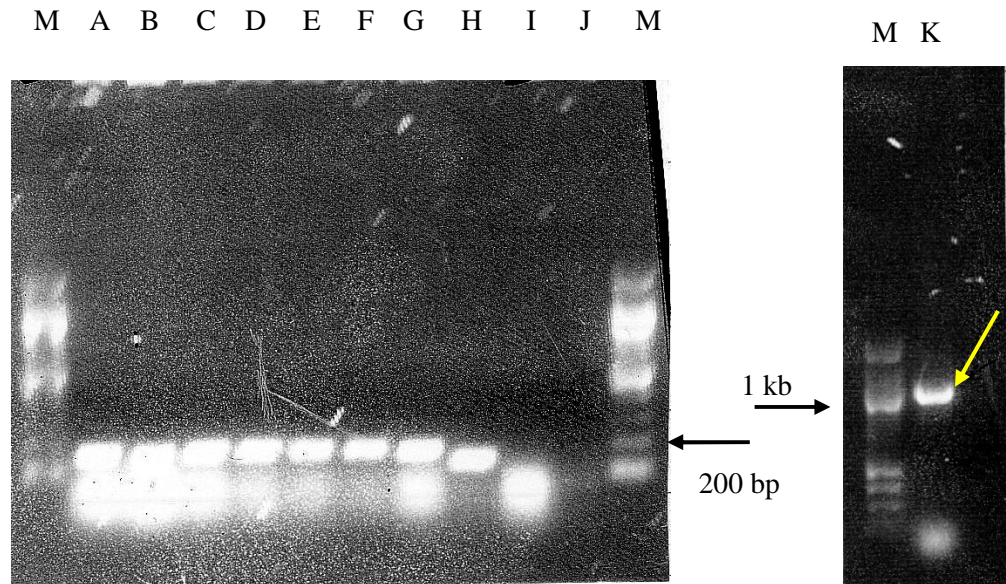


Figure 51. A Gel showing the analysis of the results of the PCR amplification of the inserts of the plasmids that gave the white colonies shown in Figure 49. Nine samples (A to I), which gave the white colonies were amplified using the T7 and SP6 primers. Lanes A to H showed small inserts of about 200bp that probably come from a direct joining of the two primers. The negative control is in lane J. The w amplification from the white colony that showed the positive Ruscony test gave no product. This clone was then tested with a longer elongation time (2 min) and the result of this is shown in lane K. The expected fragment (yellow arrow) is present at around 1 kb (800 bp for puhA + 200 bp for T7 and SP6 area). The marker (M) used was 1 μ l of a 1.5 kb ladder (obtained from New England Biolabs).

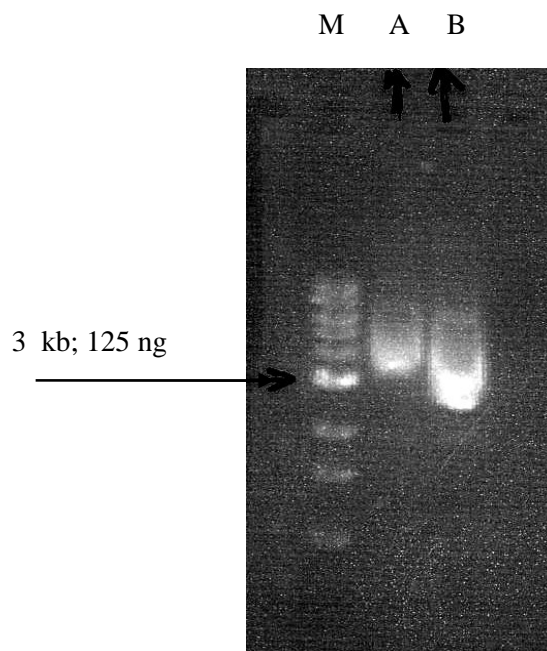


Figure 52. A gel showing the result of a larger scale isolation of the plasmid that contained the *pufA* insert. The volume of the samples was (A) 1 μ l and (B) 2 μ l. The marker (M) used was 1 μ l of 10 kb ladder (obtained from New England Biolabs). The concentration of the plasmid, estimated from lane A, is 100ng/ μ l

3.5. Nucleotide sequences of *pufB*, *pufA*, *pufL*, *pufM*, *pufC* and *pufA*

The nucleotide sequences obtained were translated into the corresponding amino acid sequences, which were then compared with the previous sequences determined by Michel (1982, 1985 and 1986). The nucleotide sequences were translated using the Molbiol 6 frame translation program (Molbiol, 2008) and aligned by the ClustalW2 program (EBI, 2008). The previously published sequences were obtained from their GenBank accession numbers as follows; *Bl. viridis* M55261 for beta and alpha subunits, X03915 for L and M subunits, X02659 for the H subunit and X05768 for the cytochrome *c* subunit.

As mentioned previously in section 3.3, for each sequencing reaction both forward and reverse primers were used. This then produces data from both strands of the DNA allowing an internal check for each sequence i.e., they are complimentary and have to match.

The region of the genomic DNA that contains the *pufB* and *pufA* genes was sequenced using the *pufBvirF* and *pufL104virR* primers. The complete sequence of both of these genes was obtained using the *pufBvirF* primer. The sequence obtained using the reverse

primer (*pufL104virR*) yielded the identical (complimentary) sequence. These sequences are shown in Figure 53. A comparison between this new sequence data and the previous published sequence is shown in Figure 54. In this case the new sequence confirms the previous sequence.

The region of genomic DNA that contains the *pufL* gene (825 bp) was sequenced next. The primers involved in sequencing *pufL* gene were *pufL504virR*, *pufLvirF* and *pufL429virF*. The 5'-end of *pufL* gene (start codon area) was sequenced by the same primers that sequenced the *pufB* and *pufA* genes (described above). Then the middle part of the *pufL* gene was sequenced using the *pufL504virR* and *pufLvirF* primers (Figure 55). The sequences of the 5'-end and the middle of *pufL* were good quality. Finally, the 3'-end of *pufL* was sequenced using the *pufL429virF* primer (Figure 56). It can be seen in Figure 56 that the start of the *pufL* sequence was of poor quality. However the later part of the gene did sequence well and had sufficient overlap with the data shown in Figure 55 to allow the two sequences to be successfully joined. The new sequence of *pufL* was identical to the previous sequenced determined by Michel et al (1982).

The primers involved in sequencing the region of the genomic DNA that contains *pufM* gene (975 bp) were *pufL429virF*, *pufM47virF* and *pufC541virR*. The 5'-end of *pufM* gene (start codon area) was sequenced using the *pufL429virF* primer, while the middle part of the gene was sequenced using *pufM47virF* (see Figure 56). These two sequences were of high quality and could be easily overlapped. The 3'-end part of *pufM* gene was sequenced using the reverse primer *pufC541virR* (Figure 57). The 3'-end of *pufM* also sequenced well and provided an excellent overlap with the middle part of the gene. The new sequence showed 5 differences when compared with the previous sequence.

The sequence obtained using the *PufC541virR* primer described in the last paragraph also contains the first part of *pufC* (start codon area). The *PufC* gene consists of 1,071 bp (see below) but using this primer only allowed the first 496 bp to be sequenced. The rest of the *pufC* gene was then sequenced using the T7 and SP6 primers (see Figure 57) from the remaining fragment that had been cloned into the pGEM© T-easy vector (described in Section 3.4). The new sequence of *pufC* had 8 differences when compared to the previous sequence. All of *pufA* was sequenced using the T7 and SP6 primers from *pufA* that had been cloned into the pGEM T-easy vector (Figure 58). There were two differences in the new sequence when compared with the previous one. All new assembled sequences have been translated into their corresponding amino acids and aligned with the previously published sequences for comparison (see Figures 59 - 62).

An overall summary of the differences between the new sequence data and the previous ones is provided in Table 10. The consequences for the structure of the *Bl. viridis* RC when these changes are fitted into the electron density map are presented in Chapter 4.

The DNA sequence of *pufB* and *pufA* obtained using *pufBvirF* (706 bp) as the primer

CGGCGTGGTTTTCGGCCACGTCGCATCCCTGCCGCGTGGGATATGCGCGGT	}	<i>pufB</i>
GACGGCCGCAAGGCCCGCATTGATGAGGGTATCAAAATGCTGACTTGAA		
ACCGAGCCTGACAGGGTTGACGGAAGAAGAGGCGAAGGAGTTCCACGGGA		
TCTTCGTGACCTCGACGGTTCTGTACCTCGCGACCGCCGTGATCGTTCAC		
TACCTGGTGTGGACGGCTCGTCCGTGGATCGCTCCCATCCCGAAGGGCTG		
GGTGAATCTGGAAGGCGTCCAGTCGGCGCTTTTCGTATCTGGTCTGAGCGG	}	<i>pufA</i>
GGAGGACTGACACATGGCTACCGAATATCGCACTGCTTCCTGGAAGCTCT		
GGCTGATCCTGGATCCGCGCCGCGTTCTGACCGCTCTGTTCTGTCTACCTG		
ACGGTCATCGCTCTGCTCATCCACTTCGGTCTGCTCAGCACCGATCGTCT		
GAAGTGGTGGGAATTCCAGCGCGGCTTCCGAAGGCGGCGTCGCTCGTGG		
TCGTTCCGCGGCTGTGCGCTGAGCCTGAAGCCCATTTTCGGGTGAATGGT	}	<i>pufL</i>
CGGGTGGTGGGCGAGGCCTAATGTGGCTTTGCCCGCCACCCAAATTCAGG		
GGCGAGGTGGCGCTGGCAACGCCCCCTCAAGCGGAGGACAGAGCAATGGCA		
CTGCTCAGCTTTGAGAGAAAGTATCGCGTCCGCGGGGGGACGCTGATCGG		
TGGGGA		

The DNA sequence of *pufB* and *pufA* obtained using *pufL104virR* (648 bp) as the primer

AAGGCCCGCATTGATGAGGGTATCAAAATGCTGACTTGAAACCGAGCCT	}	<i>pufB</i>
GACAGGGTTGACGGAAGAAGAGGCGAAGGAGTTCCACGGGATCTTCGTGA		
CCTCGACGGTTCTGTACCTCGCGACCGCCGTGATCGTTCCTACCTGGTG		
TGGACGGCTCGTCCGTGGATCGCTCCCATCCCGAAGGGCTGGGTGAATCT		
GGAAGGCGTCCAGTCGGCGCTTTTCGTATCTGGTCTGAGCGGGAGGACTG		
ACACATGGCTACCGAATATCGCACTGCTTCCTGGAAGCTCTGGCTGATCC	}	<i>pufA</i>
TGGATCCGCGCCGCGTTCTGACCGCTCTGTTCTGTCTACCTGACGGTCATC		
GCTCTGCTCATCCACTTCGGTCTGCTCAGCACCGATCGTCTGAACTGGTG		
GGAATTCCAGCGCGGCTTCCGAAGGCGGCGTCGCTCGTGGTCTGTTCCGC		
CGGCTGTGCGCTGAGCCTGAAGCCCATTTTCGGGTGAATGGTCTGGGTGGTG		
GGCGAGGCCTAATGTGGCTTTGCCCGCCACCCAAATTCAGGGGCGAGGTG	}	<i>pufL</i>
GCGCTGGCAACGCCCCCTCAAGCGGAGGACAGAGCAATGGCACTGCTCAGC		
TTTGAGAGAAAGTATCGCGTCCGCGGGGGGACGCTGATCGGTGGAGAC		

Figure 53. Sequencing results of *pufB* and *pufA* genes using both the *pufBvirF* and *pufL104virR* primers. The sequence obtained using *pufBvirF* (upper) covers 706 bp and that using *pufL104virR* (bottom) covers 648 bp. The start and stop codons were highlighted in yellow and the complete genes in red. Both sequences when translated into the corresponding amino acid sequence (see Figure 54) are identical. These sequencing reactions also revealed the first part of *pufL*.

PufB

```
>New    MADLKPSLTGLTEEEAKEFHGIFVTSTVLYLATAVIVHYLVWTARPWIAPIPKGWNLEG 60
>Old    MADLKPSLTGLTEEEAKEFHGIFVTSTVLYLATAVIVHYLVWTARPWIAPIPKGWNLEG 60
*****

>New    VQSALSYLV- 69
>Old    VQSALSYLV- 69
*****
```

PufA

```
>New    MATEYRTASWKLWLILDPRRVLTAFLVYLTVIALLIHFGLLSTDRLNWWFQRGLPKAAS 60
>Old    MATEYRTASWKLWLILDPRRVLTAFLVYLTVIALLIHFGLLSTDRLNWWFQRGLPKAAS 60
*****

>New    LVVVPVAVG- 69
>Old    LVVVPVAVG- 69
*****
```

Figure 54. Comparisons between the newly obtained and the previously published amino acid sequences of the PufB and PufA. Matching residues are indicated by a star (*) and it can be seen that the new and old sequences are identical.

The DNA sequence of *pufL* obtained using *pufL504virR* (439 bp) as the primer

```
ATG GCACTGCTCAGCTTTGAGAGAAAGTATCGCGTCCGCGGGG GACGCT
GATCGGTGGAGACTTGTTTCGATTTCTGGGTGGGGCCGTACTTTGTCGGCT
TCTTCGGAGTTTCGGCAATCTTCTTCATTTTCCTCGGCGTTAGTTTGATC
GGCTACGCGGCGTCACAAGGGCCACCTGGGATCCCTTCGCCATCAGCAT
CAATCCCGCCCCGACCTGAAGTACGGGCTCGGCGCGGCGCGCTGCTCGAG
GGCGGCTTCTGGCAGGCGATCACCGTCTGCGCTCTTGGTGCGTTTCATTTT
GTGGATGCTCCGTGAGGTGCGAAATTTCCCGCAAGCTCGGAATTGGTTGGC
ACGTCCCGCTGGCCTTCTGCGTTCCGATCTTCATGTTCTGCGTTCTGCGAG
GTTTTTCGCGCGCTGCTTCTCGGTTCTGTTGGGTCATGCGTTTCCCTACGG
CATCCTGAGCCATCTCG
```

The DNA sequence of *pufL* obtained using *pufLvirF* (492 bp) as the primer

```
G GACGCTGATCGGTGGAGACTTGTTTCGATTTCTGGGTGGGGCCGTACTTT
GTCGGCTTCTTCGGAGTTTCGGCAATCTTCTTCATTTTCCTCGGCGTTAG
TTTGATCGGCTACGCGGCGTCACAAGGGCCACCTGGGATCCCTTCGCCA
TCAGCATCAATCCGCCCCGACCTGAAGTACGGGCTCGGCGCGGCGCGCTG
CTCGAGGGCGGCTTCTGGCAGGCGATCACCGTCTGCGCTCTTGGTGCGTT
CATTTTCGTGGATGCTCCGTGAGGTGCGAAATTTCCCGCAAGCTCGGAATTG
GTTGGCACGTCCCGCTGGCCTTCTGCGTTCCGATCTTCATGTTCTGCGTT
CTGCAGGTTTTTCGCGCGCTGCTTCTCGGTTCTGTTGGGTCATGCGTTTCC
CTACGGCATCCTGAGCCATCTCG TTGGGTGAACAACCTTCGGGTATCAGT
ACCTTAAGTGGCACTACAACCCGGGACACATGTCGTCCGTTTCGTTTCCTG
TTTGTGAACGCGATGGCGCTGGGTCTGCACGGCGGTCTCATCCTGTCCGT
TGCCAACCCCGCGATGGCGACAAGGTGAAGACGGCGGAACACGAGAACC
AGTACTTCCGTGACGTGCTCGGCTACTCGATCGGCGCGCTCAGCATTCAC
CGTCTGGGCTGTTCTTGGCGTGAACATCTTCTGACGGGCGCCTTTGG
CACCATCGCTAGCGGTCCGTTCTGGACCCGCGGCTGGCCGGAATGTTGGG
GCTGGTGGCTC
```

Figure 55. Sequencing results of the *pufL* gene using *pufL504virR* (upper) and *pufLvirF* (bottom). The sequence obtained using *pufL504virR* covers 468 bp and that using *pufLvirF* covers 747 bp. The start codon of *pufL* is highlighted in yellow and the completed *pufL* gene in red. The overlap

regions between these two primers are highlighted in green and purple. The highlighted grey base indicates the sequence continued by primer *pufL429virF* (Figure 56).

The DNA sequence obtained using *pufL429virF* (808 bp) as the primer

TGAGCCATCTCGATTGGGTCACAGCTNNANGTATCAGTACCTTAACTGGC	}	Difficult to read
ACTACAANNNGGGACACATGTCGTCCGTTTCGTTTCCTGTTTGTGAACGCG		
ATGGCGCTGGGTCTGCACGGCGGTCTCATCCTGTCGGTTGCCAACCCCGG		
CGATGGCGACAAGGTGAAGACGGCGGAACACGAGAACCAGTACTTCCGTG		
ACGTGTCGTGGCTACTCGATCGGCGCGCTCAGCATTCACCGTCTGGGCCTG		
TTCCTGGCGTGAACATCTTCCTGACGGGCGCCTTTGGCACCATCGCTAG	}	<i>pufL</i>
CGGTCCGTTCTGGACCCGCGGCTGGCCGGAATGGTGGGGCTGGTGGCTCG		
ACATTCCGTTCTGGAGCTTAAAGGGGCGCACAATGGCTGATTATCAAATA	}	<i>pufM</i>
TCTACAAGCAGATTACAGCCCCGCGCCCGCATATCACTGTCTCCGGCGAG		
TGGGGCGACAACGATCGCATCGGTAAGCCCTTCTATTCTTACTGGCTGGG		
CAAGATCGGTGACGCGCAGATCGGGCCGATCTATCTGGGTGCTTCGGGAA		
TCGCGGCGTTTCGCCTTCGGCGCGACCGCGATCCTGATCATCGGGTTCAAC		
ATGCTGGCCGAAGTCCACTTCGACCCGCTGCAGTTCTTCCGCCAGTTCTT		
CTGGCTCGGCCTTACCCGCCGAAGGCGCAGTACGGCATGGGCATCCCCG		
CGCTTCATGATGGCGGCTGGTGGCTTATGGCCGGCCTGTTTCATGACGCTG		
TCGCTTGGCTCCTGGTGGATCCGGGGTGTACTCGCGGGCCCGTGCCTCTC		
GGGCCTT		

The DNA sequence obtained using *pufM47virF* (737 bp) as the primer

GCAGATCGGGCCGATCTATCTGGGTGCTTCGGGAATCGCGGCGTTTCGCC	}	<i>pufM</i>
TTCCGGCGCGACCGCGATCCTGATCATCGGGTTCAACATGCTGGCCGAAGT		
CCACTTCGACCCGCTGCAGTTCTTCCGCCAGTTCTTCTGGCTCGGCCTCT		
ACCCGCCGAAGGCGCAGTACGGCATGGGCATCCCGCCGCTTCATGATGGC		
GGCTGGTGGCTTATGGCCGGCCTGTTTCATGACGCTGTCGCTTGGCTCCTG		
GTGGATCCGGGTGTACTCGCGGGCCCGTCTCTCGGCCTTGGTACCCACA		
TCGCGTGGAATTCGCTGCGGCGATCTTCTTCGTGCTGTGCATCGGTTGC		
ATCCATCCGGCGCTTGTGCGGCAGCTGGTCGGAAGGCGTTCCGTTCCGGCAT		
CTGGCCGCACATCGACTGGCTGACCGCGTTCTCGATCCGCTACGGCAACT		
TCTACTATTGCCCGTGGCACGGGTTCTCGATCGGCTTCGCCTACGGCTGC		
GGCCTTGTTCGCGGCTCACGGTGCACCATCCTGGCCGTCGCTCGGTT		
TGGCGGCGATCGCGAAATCGAGCAGATCACCGACCGCGGCACCGCGGTGG		
AGCGTGCGGCTCTGTTCTGGCGCTGGACGATCGGCTTCAACGCCACGATC		
GAGTCTGTCCATCGCTGGGCTGGTTCTTCTCGCTGATGGTGATGGTGTC		
TGCGAGCGTCGGTATTCTTCTGACCGGCACCTTCGT		

Figure 56. Sequencing results of *pufM* gene using *pufL429virF* (upper) and *pufM47virF* (bottom). The sequence for 3'-end of *pufL* as well as 5'-end of *pufM* were obtained using *pufL429virF*. The stop codon of *pufL* and start codon for *pufM* are highlighted in yellow and the completed *pufM* gene in red. The overlap region with *pufLvirF* (Figure 55) is highlighted in gray whereas base that overlaps with *pufM47virF* (bottom) is highlighted in dark green and dark blue. The highlighted cyan and bright green bases indicate the sequence continued by *pufL429virF* primer (Figure 57).

The DNA sequence obtained using *pufC541virR* (668 bp) as the primer

```

GGGGCTGGTTCTTCTCGCTGATGGTGTCTGCGAGCGTCGGTATT
CTTCTGACCGGCACCTTCGTGACAACCTGGTACCTCTGGTGTGTCAAGCA
TGGCGCTGCGCCGACTATCCGGCTTATCTCCCGCCACGCCTGATCCTG
CCTCGCTGCGGGAGCACCGAAATGAACAGCTGATTGTTAATTCGGTTG
CGACTGTCGCGCTGGCGTCTCTCGTGGCCGGTTGTTTCGAGCCGCCGCCG
GCGACCACGACCCAGACTGGTTTCCGCGGGCTGTCGATGGGTGAGGTTCT
TCACCCGGCGACCGTGAAGGCCAAGAAGGAGCGTGACGCGCAGTATCCGC
CGGCGCTGCCGGCGGTGAAGGCCGAAGGCCGCCGGTGTGCGAGGTCTAC
AAGAACGTCAAGGTGCTCGGAAATCTGACCGAGGCCGAGTTCCTGCGGAC
CATGACGGCGATGACGGAATGGGTGTGCGCGGAGGAAGGCTGCACGTACT
GCCACGACGAGAACAACCTCGCTTCAGAGGCCAAGTACCCGTACGTGGTG
GCGCGTCGCATGCTCGAGATGACGCGTGCATCAACACCAACTGGACGCA
GCACGTGCGCCAGACCGGTGTGACCTGCTACACCTGCCACCGTGGCACGC
CGCTCCCGCCGTACGTC

```

} *pufM*

} *pufC*

The DNA sequence obtained using *SP6* (608 bp) as the primer

```

AAAGCCAAGAAGGAGCGTGACGCGCAGTATCCGCCGGCGCTGCCGGCGGT
GAAGGCCGAAGGCCGCCGGTGTGCGAGGTCTACAAGAACGTCAAGGTGC
TCGGAATCTGACCGAGGCCGAGTTCCTGCGGACCATGACGGCGATGACG
GAATGGGTGTGCGCGGAGGAAGGCTGCACGTACTGCCACGACGAGAACAA
CCTCGCTTCAGAGGCCAAGTACCCGTACGTGGTGGCGCGTGCATGCTCG
AGATGACGCGTGCATCAACACCAACTGGACGACGACGTGCGCCAGACC
GGTGTGACCTGCTACACCTGCCACCGTGGCAGCCCGCTCCCGCCGTACGT
CCGGTACCTGGAGCCGACGCTGCCCTGAACAATCGTGAGACGCCGACCC
ACGTGAGCGGGTTGAGACCCGTTCCGGGCTACGTGCTGCGCTCGCGAAG
TACACGGCCTACTCGGCTCTGAACTACGATCCGTTACGATGTTCTCGC
GAACGACAAGCGTCAGGTCCGTGTGGTGGCGCAGGACGGCGCTCCCGCTT
GTCGGCGTCAGCCGCGGCAAGGAACGTGCGCCGCTGTGCGACGCCTATGC
GACCTTC

```

} *pufC*

The DNA sequence obtained using *T7* (634 bp) as the primer

```

GCCACCGTGACGCGCGCTCCCGCCGTACGTCCGGTACCTGGAGCCGACGC
TGCCCCCTGAACAATCGTGAGACGCCGACCCACGTGCGAGCGGGTTGAGACC
CGTTCCGGGTACGTGCTGCGCTCGCGAAGTACACGGCCTACTCGGCTCT
GAACTACGATCCGTTACGATGTTCTCGCGAACGACAAGCGTCAGGTCC
GTGTGGTGCCGACGCGCGCTCCCGCTTGTGCGCGTCAGCCGCGGCAAG
GAACGTGCGCCGCTGTGCGACGCCTATGCGACCTTCGCTGATGATGAG
CATCTCTGACTCGCTCGGAACCAACTGCACGTTCTGCCATAACGCGCAGT
CGTTTCGAGACCTGGGTAAGAAGAGCACGCCGACGCGCCATCGCTTGG
TGGGGCATTGCGATGGTTCTGTGACATGAACATGAACATATCTCGCTCCGCT
GAACACCGTGCTGCCGCGCAGCCGCTTTGGCCGCCAGGGTGAGGCTCCGC
AGGCCGACTGCCGACCTGCCACAGGGTGTGACGAAGCCGCTGTTCCGT
GCGTCCCGTCTCCAGGATTATCCGAGCTGGGCCCCGATCAAGGCTGCTGC
GAAGTAATAATCGATTCCCGCGCCGCCATGCGGCC

```

} *pufC*

Figure 57. Sequencing results of the *pufC* gene using the following primers *pufC541virR* (upper), *SP6* (middle) and *T7* (bottom). The sequence for 3'-end of *pufM* as well as 5'-end of *pufC* were obtained using *pufC541virR*. The stop codon of *pufC* and start codon for *pufM* are highlighted in yellow and the completed *pufC* gene is in red. The overlap regions between the three sequences obtained with these three primers are highlighted in olive green and magenta. In the sequence data from the run with *pufC541virR* the stop codon for *pufM* and the start codon for *pufC* overlap. This overlap region is highlighted in yellow.

The DNA sequence of *puhA* obtained using T7 (634 bp) as the primer

GGCGGCGCGGGAATCGATTTCGGCTAGACAGTTTTCTGCCACGGAGGTTTC
TTATGATACACGGCGCTCTCGCTCAACATTTAGACATCGCCAACTCGTA
TGGTACGCGCAGTGGCTGGTCATCTGGACGGTTGTCTGCTGTACCTCCG
CCGTGAGGACCGTCGCGAAGGCTACCCGCTGGTCGAGCCGCTTGGTCTCG
TCAAGCTGGCGCCGGAAGACGGCCAGGTCTACGAGCTGCCCTATCCAAG
ACGTTTCGTGCTCCCGCACGGCGGACCGTCACCGTTCCGCGTCGTCTGCC
CGAGACCCGCGAGCTGAAGCTCGCGCAGACCGACGGCTTCGAGGGCGCCC
CGCTGCAGCCGACCGGAATCCGCTGGTCGACGCCGTTGGCCCCGGCTTCG
TATGCCGAGCGCGCGGAAGTGGTCGACGCCACGGTTGACGGCAAGGCCAA
GATCGTCCCGCTTCGTGTTGCGACCGACTTCTCGATCGCGGAAGGCGACG
TCGATCCGCGTGGCCTGCCGGTGGTTGCCGCCGACGGCGTCGAAGCCGGT
ACGGTTACCGACCTCTGGGTGACCGCTCGGAGCACTATTTCCGCTACCT
CGAGCTCTCGGTGGCCGGCAGCGCCCGACCGCGCTGATCCCG

The DNA sequence from *puhA* obtained using SP6 (608 bp) as the primer

ACCGTCACCGTTCCGCGTCGTCTCCGAGACCCGCGAGCTGAAGCTCGCG
CAGACCGACGGCTTCGAGGGCGCCCCGCTGCAGCCGACCGGAATCCGCTGG
TCGACGCCGTTGGCCCCGGCTTCGTATGCCGAGCGCGCGGAAGTGGTCGACGC
CACGGTTGACGGCAAGGCCAAGATCGTCCCGCTTCGTGTTGCGACCGACTTC
TCGATCGCGGAAGGCGACGTCGATCCGCGTGGCCTGCCGGTGGTTGCCGCCG
ACGGCGTCGAAGCCGGTACGGTTACCGACCTCTGGGTGACCGCTCGGAGCA
CTATTTCCGCTACCTCGAGCTCTCGGTGGCCGGCAGCGCCCGCACCGCGCTG
ATCCCGTCGGCTTCTGCGATGTCAAGAAGGACAAGATCGTCGTGACGTCGA
TCCTGTCCGATCAGTTTCGCCAACGTGCCGCGTCTGCAGAGCCGCGACCAGAT
CACGCTGCGCGAAGAAGACAAGGTGTCGGCCTACTACGCGGGCGGTCTGCTC
TACGCGACGCCGAGCGTGCAGGAAGCGTTGCTGTGAACGACGAATCACTAGT
GAATTCGCGGCCCGCTGCAGGTTCGACCATAT

Figure 58. Sequencing results of the *puhA* gene using the following primers T7 (upper) and SP6 (bottom). The sequence for 5'-end of *pufC* were obtained using T7 while for 3'-end using SP6. The start and stop codon are highlighted in yellow and the completed *pufC* gene is in red. The overlap regions between the sequences obtained using these two primers are highlighted in bright green and pink.

PufL

```

New    MALLSFERKYRVRGGTLIGGDLFDFWVGPIYFVGFFGVSAIFFIFLGVSLIGYAASQGPTW 60
Old    MALLSFERKYRVRGGTLIGGDLFDFWVGPIYFVGFFGVSAIFFIFLGVSLIGYAASQGPTW 60
      *****

New    DPFAISINPPDLKYGLGAAPLLEGGFWQAITVCALGAFISWMLREVEISRKLIGWHVPL120
Old    DPFAISINPPDLKYGLGAAPLLEGGFWQAITVCALGAFISWMLREVEISRKLIGWHVPL120
      *****

New    AFCVPIMFCVLQVFRPLLLGSWGHAFPYGILSHLDWVNNFGYQYLNWHYNPGHMSSVSF180
Old    AFCVPIMFCVLQVFRPLLLGSWGHAFPYGILSHLDWVNNFGYQYLNWHYNPGHMSSVSF180
      *****

New    LFNAMALGLHGGLILSVANPGDGDKVKTAEHENQYFRDVGYSIGALSIIHRLGLFLASN240
Old    LFNAMALGLHGGLILSVANPGDGDKVKTAEHENQYFRDVGYSIGALSIIHRLGLFLASN240
      *****

New    IFLTGAFGTIASGPFWTRGWPEWGWGLDIPFWS- 274
Old    IFLTGAFGTIASGPFWTRGWPEWGWGLDIPFWS- 274
      *****

```

Figure 59. Comparison between the newly obtained and the previously published amino acid sequence of PufL. Matching residues are indicated by a star (*) and it can be seen that these sequences are identical.

PufM

```

New    MADYQTIYTQIQARGPHITVSGEWGDNDRIGKPFYSYWLKGIGDAQIGPIYLGASGIAAF 60
Old    MADYQTIYTQIQARGPHITVSGEWGDNDRYGKPFYSYWLKGIGDAQIGPIYLGASGIAAF 60
      *****

New    AFGATAILII GFNMLAEVHFDPLQFFRQFFWLGLYPPKAQYGMGIPPLHDGGWWLMAGLF120
Old    AFGSTAILIILFNMAAEVHFDPLQFFRQFFWLGLYPPKAQYGMGIPPLHDGGWWLMAGLF120
      *** *****

New    MTLSLGSWWIRVYSRARALGLGTHIAWNFAAAIFFVLCIGCIHPALVGSWSEGVPFGIWP180
Old    MTLSLGSWWIRVYSRARALGLGTHIAWNFAAAIFFVLCIGCIHPALVGSWSEGVPFGIWP180
      *****

New    HIDWLTAFSIRYGNFYPCWGHGFSIGFAYGCGLLFAAHGATILAVARFGGDREIEQITDR240
Old    HIDWLTAFSIRYGNFYPCWGHGFSIGFAYGCGLLFAAHGATILAVARFGGDREIEQITDR240
      *****

New    GTAVERAALFWRWTIGFNATIESVHRWGWFFSLMVMVSASVGILLTGTFVDNWYLCVKH300
Old    GTAVERAALFWRWTIGFNATIESVHRWGWFFSLMVMVSASVGILLTGTFVDNWYLCVKH300
      *****

New    GAAPDYPAYLPATPD PASLPGAPK- 324
Old    GAAPDYPAYLPATPD PASLPGAPK- 324
      *****

```

Figure 60. Comparison between the newly obtained and the previously published amino acid sequence of PufM. Matches between the old and new sequences are indicated by star (*) whereas differences are given by a space and highlighted in yellow. In total, 5 differences have been found for this polypeptide at positions M29, M63, M70, M74 and M164.

PufC

```
New    MKQLIVNSVATVALASLVAGCFEPPPATTTQTGFRGLSMGEVLHPATVKAKKERDAQYPP 60
Old    MKQLIVNSVATVALASLVAGCFEPPPATTTQTGFRGLSMGEVLHPATVKAKKERDAQYPP 60
      *****

New    ALPAVKAEGPPVSQVYKNVKVLGNLTEAEFLRTMTAMTEWVSPEEGCTYCHDENNLASEA120
Old    ALAAVKAEGPPVSQVYKNVKVLGNLTEAEFLRTMTAITEWVSPQEGCTYCHDENNLASEA120
      ** *****

New    KYPYVVARRMLEMTRAINTNWTQHVAQTGVTCYTCHRGTPLPYVRYLEPTLPLNNRETP180
Old    KYPYVVARRMLEMTRAINTNWTQHVAQTGVTCYTCHRGTPLPYVRYLEPTLPLNNRETP180
      *****

New    THVERVETRSGYVVRLAKYTAYSALNYDPFTMFLANDKRQVRVVPQTALPLVGVSRGKER240
Old    THVERVETRSGYVVRLAKYTAYSALNYDPFTMFLANDKRQVRVVPQTALPLVGVSRGKER240
      *****

New    RPLSDAYATFALMMSISDSLGTNCTFCHNAQSFESWGKKSTPQRAIAWWGIRMVRDMNMN300
Old    RPLSDAYATFALMMSISDSLGTNCTFCHNAQTFESWGKKSTPQRAIAWWGIRMVRDLNMN300
      *****

New    YLAPLNTVLPPASRLGRQGEAPQADCRTCHQGVTKPLFGASRLQDYPELGPIKAAAK 356
Old    YLAPLNTALSLPPASRLGRQGEAPQADCRTCHQGVTKPLFGASRLKDYPELGPIKAAAK 356
      *****
```

Figure 61. Comparison between the newly obtained and the previously published amino acid sequence of PufC. The matches between the old and new sequences are indicated by a star (*) whereas the differences are given by a space and highlighted in yellow. In total, there were 8 differences found in this polypeptide at positions C43, C77, C84, C252, C277, C287, C288 and C323.

PuhA

```
New    MYHGALAQHLDIAQLVWYAQWLVIWTVVLLYLRRDDRREGYPLVEPLGLVKLAPEDGQVY60
Old    MYHGALAQHLDIAQLVWYAQWLVIWTVVLLYLRRDDRREGYPLVEPLGLVKLAPEDGQVY60
      *****

New    ELPYPKTFVLPHGGTVTVPRRRPETRELKLAQTDGFEGAPLQPTGNPLVDAVGPASYAER 120
Old    ELPYPKTFVLPHGGTVTVPRRRPETRELKLAQTDGFEGAPLQPTGNPLVDAVGPASYAER 120
      *****

New    AEVVDATVDGKAKIVPLRVATDFSIAEGDVDPRGLPVVAADGVEAGTVTDLWVDRSEHYF 180
Old    AEVVDATVDGKAKIVPLRVATDFSIAEGDVDPRGLPVVAADGVEAGTVTDLWVDRSEHYF 180
      *****

New    RYLELSVAGSARTALIPLGFCDVKKDKIVVTSILSDQFANVPRLQSRDQITLREEDKVSA 240
Old    RYLELSVAGSARTALIPLGFCDVKKDKIVVTSILSEQFANVPRLQSRDQITLREEDKVSA 240
      *****

New    YYAGGLLYATPERAEALL 258
Old    YYAGGLLYATPERAESLL 258
      *****
```

Figure 62. Comparison between the newly obtained and the previously published amino acid sequence of PuhA. The matches between the old and new sequences are indicated by a star (*) whereas the differences are given by a space and highlighted in yellow. In total, there were 2 differences found in this gene at positions H216 and H256.

Table 10. Summary of all amino acid differences obtained from the new sequence data.

Gene	Position	Amino Acids	
		New	Old
PufB	No differences		
PufA	No differences		
PufM	M29	Ile	Val
	M63	Ala	Ser
	M70	Gly	Leu
	M74	Leu	Ala
	M164	Ala	Thr
PufC	C43	Pro	Ala
	C77	Met	Ile
	C84	Glu	Gln
	C252	Ser	Thr
	C277	Met	Leu
	C287	Thr	Ala
	C288	Val	Ser
	C323	Gln	Lys
PuhA	H216	Asp	Glu
	H256	Ala	Ser

Chapter 4 Fitting the new sequence data with the high resolution structure of *Bl. viridis* reaction centres

It is apparent that there are residue positions for which there are clashes within the electron density map when the previously published polypeptide sequences of the RC polypeptides are fitted into the 2.0 Å newly refined high resolution structure (Roszak, et al., unpublished observations). The new sequence data presented in Chapter 3 has enabled a careful re-examination of these residue positions. This chapter presents the results of this re-examination. All electron density maps shown in this chapter are contoured at the 3σ level.

4.1. M-subunit

The new sequence of the M-subunit has 5 amino acid changes compared with the previously published sequence data (see Table 10). How each of these changes fit into the new structural data is described below.

The carotenoid (1, 2-dihydroneurosporene) binding site in *Bl. viridis* RC is formed mainly residues in the M-subunit. Figure 63A illustrates how the assignment of Leu in the previous polypeptide sequence at position M70 produces a clash with the carotenoid if it is fitted as a flat, planar molecule. As a result the carotenoid conformation in the *Bl. viridis* RC is strange when compared with that found in other RC structures or expected from resonance Raman spectra (Robert, 1999). As can be seen in the electron density, Leu at the M70 position does not fit well into the new density map. The new sequence, however, shows that the residue at position M70 position is in fact Gly and not Leu. The presence of the much smaller Gly residue removes the steric hindrance that was caused by the bulky Leu residue and provides a much better fit into the density. This then allows a fully planar conformation of the carotenoid that also fits better into the electron density (Figure 63B). This new carotenoid conformation is also much more similar to the carotenoid structures found in the reaction centres of *Rb. sphaeroides* (Roszak et al, 2004). These carotenoid structures are compared in Figure 64. The previous structures of the RC carotenoid from *Bl. viridis* have a strong kink at one end compared with that seen in *Rb. sphaeroides*. Fitting the new sequence data at this point removes the kink and in the new structure the

carotenoid conformation is more reasonable. Interestingly when the sequences of RC's from several other purple bacterial species, such as *Rvi. gelatinosis*, *Chr. tepidum*, *Rsp. molischianum*, *Rsb. denitrificans* and *Rb. sphaeroides*, are compared the position corresponding to M70 is also occupied by a Gly residue (see Figure 65). This further substantiates the new sequence data.

In the newly refined structure at position M74 there is an area of positive density in the side chain part of the residue. This clearly shows that the assignment of Ala in the previous polypeptide sequence at position M74 does not fit well into the electron density (see Figure 66A). The new sequence data places a Leu residue at this position. The assignment of Leu, with its two additional methyl groups, instead of Ala, at position M74 results in a much better fit into the density (see Figure 66B). Now there is no additional unfilled density.

When the previously published residue at position M63 was fitted into the new electron density map it did not give a good result (Figure 67A). Ser seems to be too big. The new sequence places Ala at this position and this assignment produces a much better fit than Ser (Figure 67B).

The old sequence has a Thr residue at the position M164. Again clearly this Thr residue does not fit well into the density. The rather the bulky side chain results in a large area of negative density (Figure 68A). The new sequence data assigns Ala to this position. The Ala side chain is smaller, as it lacks the hydroxyl group and an additional carbon atom, and as a result of this Ala provides a much better fit within the electron density (see Figure 68B).

The previous sequence suggested that there is a Val at position M29. This residue can be fitted nicely to the new electron density map (Figure 69A). However, the new amino acid sequence suggests that this position is Ile. Unfortunately this assignment of Ile does not improve the fit to the new electron density (Figure 69B). Due to this mismatch the sequence of this part of the M subunit was repeated. The same sequence was obtained the second time. Furthermore, careful inspection of the sequences of the M subunits from the other species of purple bacteria shows enough variability at this residue so that no decision can be made about the correctness of the new sequence at this position based on possible homologies (Figure 65). There is however the possibility that there was an error in the PCR reaction and this should probably be re-checked.

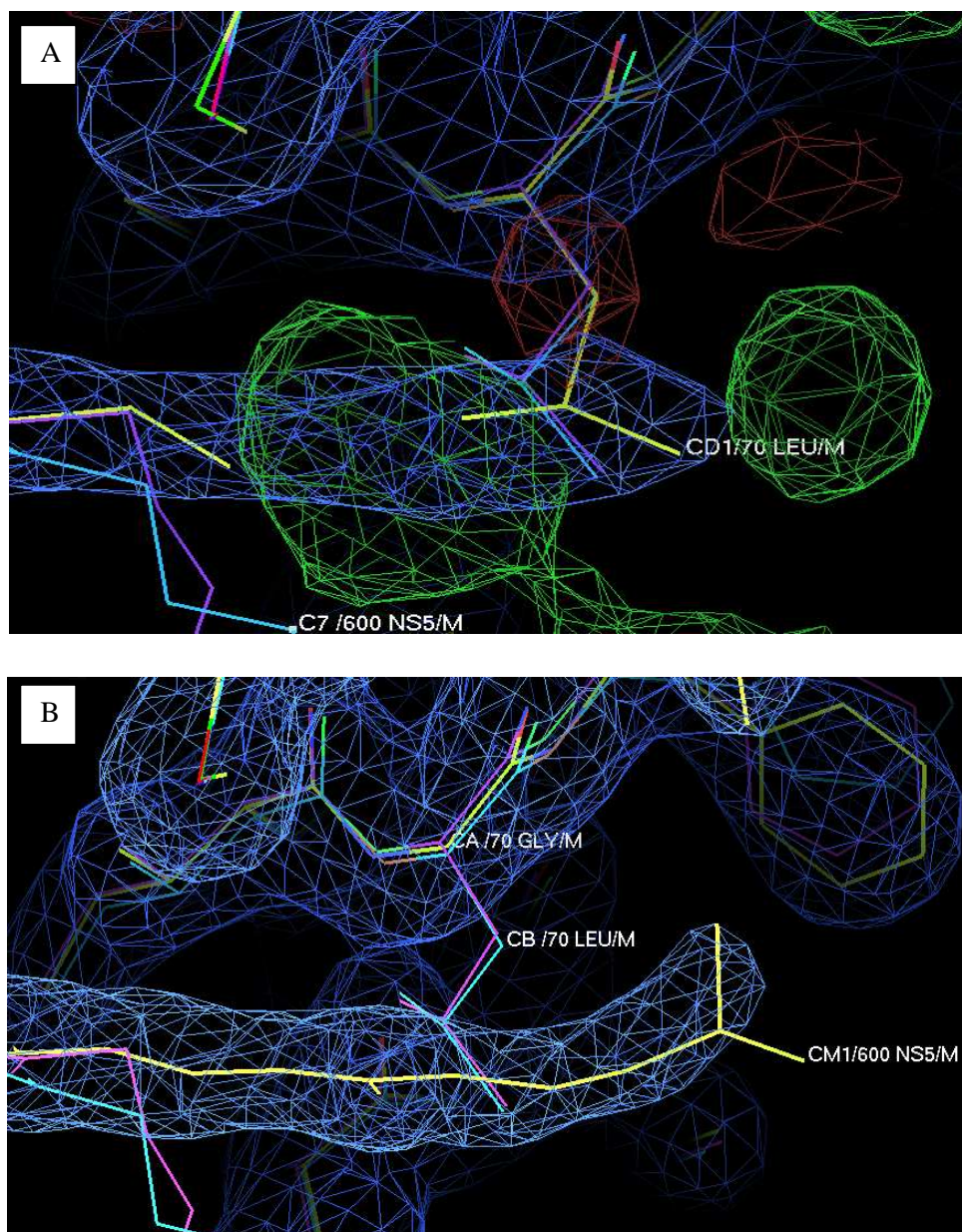


Figure 63. The 2*Fo*-*Fc* electron density at 1σ level (in blue) and the *Fo*-*Fc* density at 3 σ level (in green-red) at the carotenoid binding site (Roszak, et al., unpublished observations). The fit of new RC polypeptide sequence is shown in yellow, while two previously published models, PDB codes 1DXR (Lancaster, et al., 2000) and 1I5N (Mustafi, et al., 2006), are shown in magenta and cyan respectively. (A) Negative density was found at the position where the Leu side chain is supposed to be present (red). The Leu residue at M70 position does not fit within the new density map. Furthermore, some density was found suggesting that the carotenoid has a planar structure (green). However, the carotenoid was previously modelled with a kink to avoid clashing with the bulky Leu residue. (B) The new sequence has Gly instead of Leu at position M70. This residue removes the steric hindrance and enables a much better fit to be obtained within the density. This in turn enables the carotenoid to be modelled as a flat, planar molecule that also fits nicely within the density.

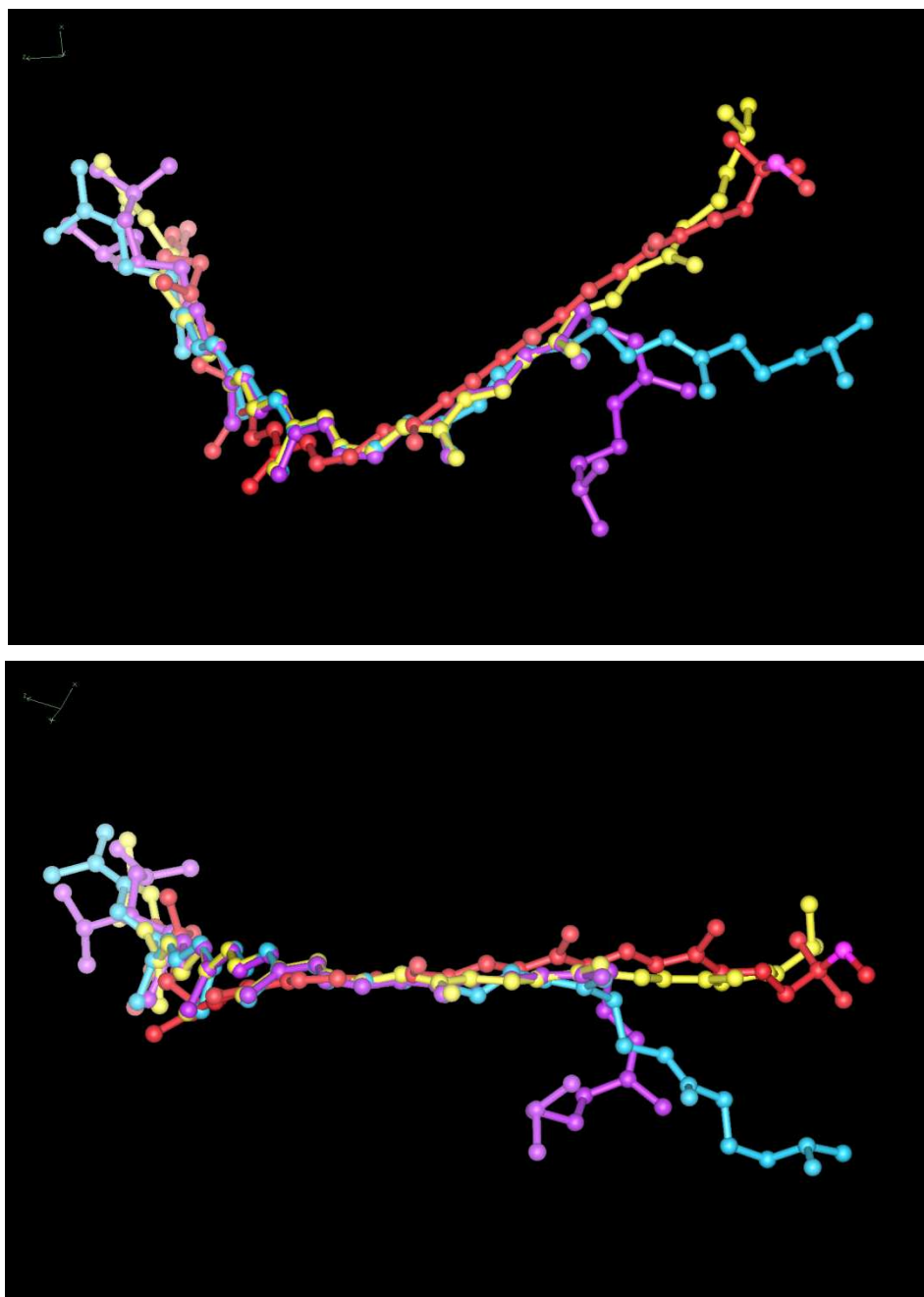


Figure 64. Comparison of carotenoid structures from different crystal structures, Ball-and-stick models of the 1,2-dihydroneurosporene (newly refined conformation as found in the reaction centre from *B. viridis* (Roszak et al., unpublished observations) shown in yellow, and two conformations from the published structures of reaction centre from *B. viridis*, PDB codes 1DXR (Lancaster, et al., 2000) in cyan, and 1I5N (Li, et al., 2006) in magenta) superimposed with the model of spheroidene (in red) found in the wild-type reaction centre from *Rb. sphaeroides* strain 2.4.1 (Fujii, et al., unpublished observations). This match of carotenoid molecules was obtained by superposition of reaction centre protein chains surrounding the carotenoid binding site.


```

                                M29
Bl. viridis (old)  --MADYQTIYTQIQARGP-HITVSGEWGD--NDRVG-KPFYSYWLKGIGDAQIGPIYLGA 54
Bl. viridis (new) --MADYQTIYTQIQARGP-HITVSGEWGD--NDRIG-KPFYSYWLKGIGDAQIGPIYLGA 54
Rvi. gelatinosus  --MAEYQNIFTRVQVAGPAHMGVPLPEQD--SPRTGKKPWQIHLGLRGLMAQIGPIYLG 56
Chr. tepidum      --MPEYQNIFTAVQVRAPAPYGPVPLPKGN--LPRIG-RPIFSYWLKGIGDAQIGPIYLG 55
Rsp. molischianum --MTEYQNIFNRVQIRGETYPGIALPDGN--DERFG-KGFFFYVWGVKANAQIGPIYLG 55
Rsb. denitrificans MMYPEYQNIFTQVQVRGTPEMGMDAGNNMMEERVG-KPFFSTLAGLFGNGQIGPIYFGW 59
Rb. sphaeroides   --MAEYQNIFSQVQVRGPADLGMTEDVN--LANRSG-VGPFSTLLGWFGNAQLGPYILGS 55
                .:*.*: .: * : * * * .. :*: *:*

                                car.
                                M63 M70 M74
Bl. viridis (old)  SGIAAFAFGSTAILIILFNMAAEVHFDPLQFFRQFFWLGLYPPKAQYGMGIP-PLHDGGW 113
Bl. viridis (new) SGIAAFAFGATAILLIIGFNMLAEVHFDPLQFFRQFFWLGLYPPKAQYGMGIP-PLHDGGW 113
Rvi. gelatinosus  LGILSIVFGSLAIMIIGFNMLASVGNWPIEFFRQFFWLALPEPPSPKYGLKLP-PLNDGGW 115
Chr. tepidum      TGTLSIFFGLVAISIIGFNMLASVHWDVFQFLKHFFWLGLEPPPPQYGLRIP-PLSEGGW 114
Rsp. molischianum WGVASALCFGIAAAVVGNLFLAQVGDWVQFVRQFFWLGLEPPLPKYGLSFP-PLAEGGW 114
Rsb. denitrificans TSIVAFGTGIAWVIVGFNMLAQVGSIPQFIRQLFWLALPEPPSPEYGLSMP-PLNDGGW 118
Rb. sphaeroides   LGVLSLFSGLMWFFTIGIWFYQAGWNPVFLRDLFFFSLEPPAPEYGLSFAAPLKEGGL 115
                . : : : : . . : *.:*: * * .:*. :. * * :*:

Bl. viridis (old)  WLMAGLFMTLSLGSWWIRVYSRARALGLGTHIAWNFAAAIFFVLCIGCIHP TLVGSWSEG 173
Bl. viridis (new) WLMAGLFMTLSLGSWWIRVYSRARALGLGTHIAWNFAAAIFFVLCIGCIHPALVGSWSEG 173
Rvi. gelatinosus  WLMAGLFLTISILLWVVMYTRARALGMGTHVAFWAFWAAIWLVLVLFIRPVLMGSWSEA 175
Chr. tepidum      WLIAGLFLTISILLWVRTYKRAEALGMSQHLWAFWAAIWFYLVLFIRPVMMGSWAKA 174
Rsp. molischianum WLITGFFLTASILCWVVRVYTRARALGLGTHVAFWAFWAAIWFYLSLGFIRPILMGSWGEA 174
Rsb. denitrificans YIIASFLLVSVMTWLLRAYLLAEQHKMGKHIWGFAAAVWLFLVLGLFRPILMGSWSEA 178
Rb. sphaeroides   WLIASFMMFVAVSWWGRTYLRAQALGMGKHTAWAFLSAIWLWMVLGFIIRPILMGSWSEA 175
                :*: .: * : * * . : * * * :*: : : * :*: :*: :*:

Bl. viridis (old)  VPFGIWPIDWLTAFSIRYGNFYCPWHGFSIGFAYGCGLLFAAHGATILAVARFGGDRE 233
Bl. viridis (new) VPFGIWPIDWLTAFSIRYGNFYCPWHGFSIGFAYGCGLLFAAHGATILAVARFGGDRE 233
Rvi. gelatinosus  VPFGIFPHLDWTAAFSLRYGNLFPNPHALSIAFLYGATLILAVSRFGGERE 235
Chr. tepidum      VPFGIFPHLDWTAAFSIRYGNLYNPNPHALSIAFLYGSALLFAMHGATILSVSRFGGDRE 234
Rsp. molischianum VPFGIFPHLDWTAASIRYGNFYNPNPHALSIAFLYGSALLFAMHGATILAVSRFGGDRE 234
Rsb. denitrificans VPYGFIFPHLDWTTAFSIRYGNLYNPNPHCLSIVFLYGSVLLFCMHGGTILAVTRYGGDRE 238
Rb. sphaeroides   VPYGFIFSHLDWTNNFSLVHGNLFPNPHGLSIAFLYGSALLFAMHGATILAVSRFGGERE 235
                ***: .: * : * : * * : * * * * .: * * * :*: * :*:

Bl. viridis (old)  IEQITDRGTAVERAALFWRWTIGFNATIESVHRWGWFSLMVMVSASVGILLTGTFTVDNW 293
Bl. viridis (new) IEQITDRGTAVERAALFWRWTIGFNATIESVHRWGWFSLMVMVSASVGILLTGTFTVDNW 293
Rvi. gelatinosus  LEQIADRGTASERAQLFWRWTMGFNATIESIHRWAWWFAVLCPLCGGIGILLSGTVVDNW 295
Chr. tepidum      IDQITHRGTAEGAALFWRWTMGFNATIESIHRWAWWCAVLTVITAGIGILLSGTVVDNW 294
Rsp. molischianum LEQIVDRGTGVERAMLFWRWQLGFNATIESFHRWGYWFAVLTVLTAAGIGILLTGTFTVDNW 294
Rsb. denitrificans LEQIYDRGTATERAALFWRWTMGFNATMEGIHRWAWWFAVLTPITGGIGILLTGTFTVDNW 298
Rb. sphaeroides   LEQIADRGTAAERAALFWRWTMGFNATMEGIHRWAIWMAVLVTLTGGIGILLSGTVVDNW 295
                :*: .: * * * * * : * * * .: : : : :*: * :*:

Bl. viridis (old)  YLWCVKHGAAPDYPAYLPATPD PASLPGAPK-- 324
Bl. viridis (new) YLWCVKHGAAPDYPAYLPATPD PASLPGAPK-- 324
Rvi. gelatinosus  YLWAVKHGVPSPYPAVFAPTIDPATLQGVK--- 325
Chr. tepidum      YLWAVKHGMAPAYPEVVTAVNPYETAAEVMQ-- 325
Rsp. molischianum YLWAVKHSHFVSPYPVFN-VADPALTSQVAG-- 324
Rsb. denitrificans FIWAQEHFAPMYDGSYGIEDYGSYEAFIGKEN 331
Rb. sphaeroides   YVWQNHGMAPLN----- 308
                :*: .: *

```

Figure 65. Multiple sequence alignment of M-subunit RCs of several species. DNA sequence data were obtained from the gene bank under accession numbers: *Bl. viridis* (old), X03915; *Rvi. gelatinosus*, AB034704; *Chr. tepidum*, D85518; *Rsp. molischianum*, D50654; *Rsb. denitrificans*, CP000362; *Rb. sphaeroides*, CP000143. The new *Bl. viridis* sequence data is presented in bold letters. Identical residues are shown with an asterisk (*). The positions where differences have been found in the new sequence are highlighted in yellow. Only *Bl. viridis* (old) has a Leu (L) residue at M70 (highlighted in green). Just as in the cases of all the sequences from the other species, the newly obtained sequence has Gly (G) at M70.

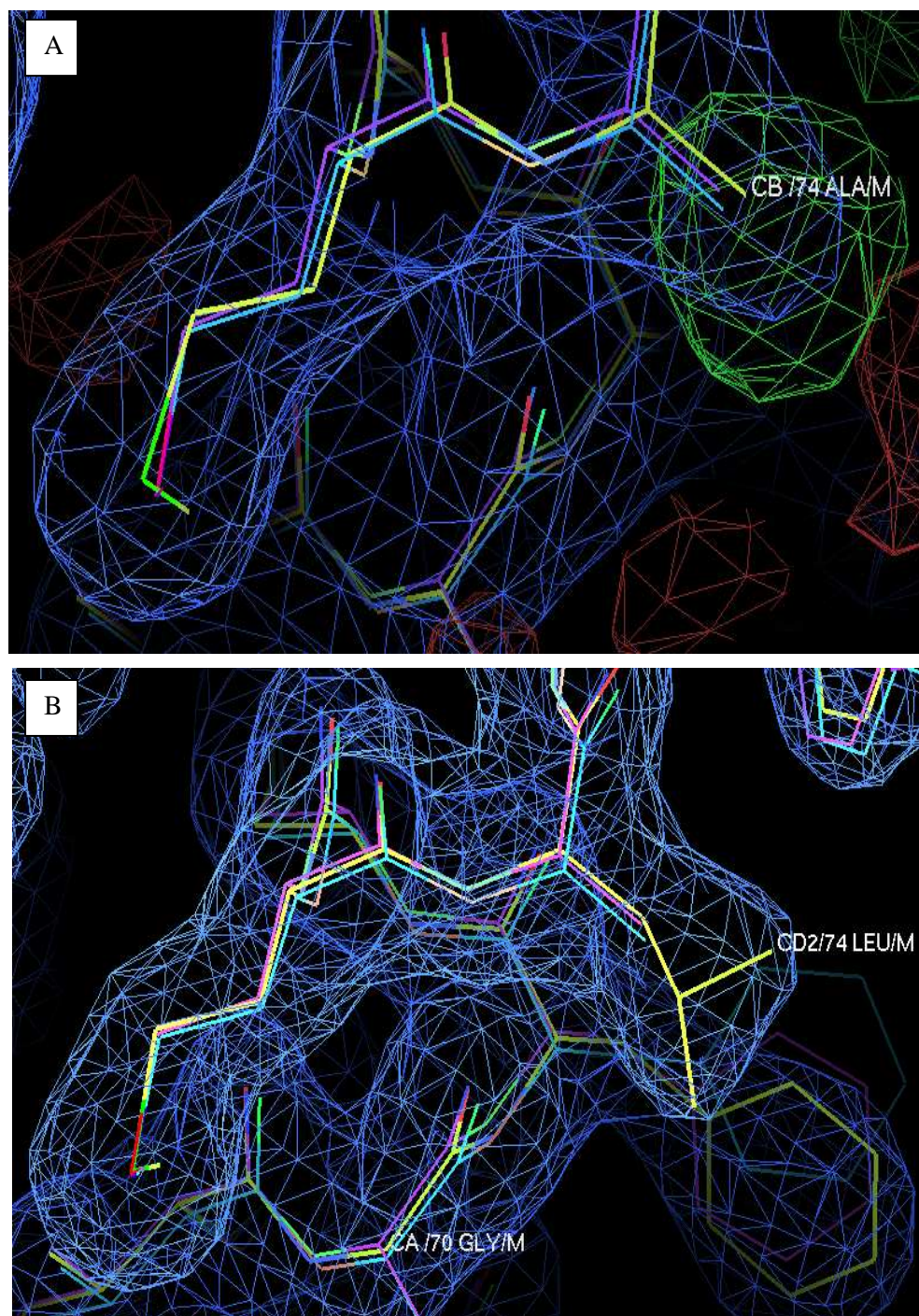


Figure 66. Electron density fitted with the old and the new sequence data at position M74. The colouring scheme is the same as previous figure. (A) Previously unfitted electron density is shown in green. The new density map suggests that Ala residue at M74 position does not provide a good fit due to the appearance of this positive density (green). (B) The new amino acid sequence for position M74 confirms places a Leu at position M74. Leu nicely fits the new, 2.0Å electron density map and removes the missing positive density.

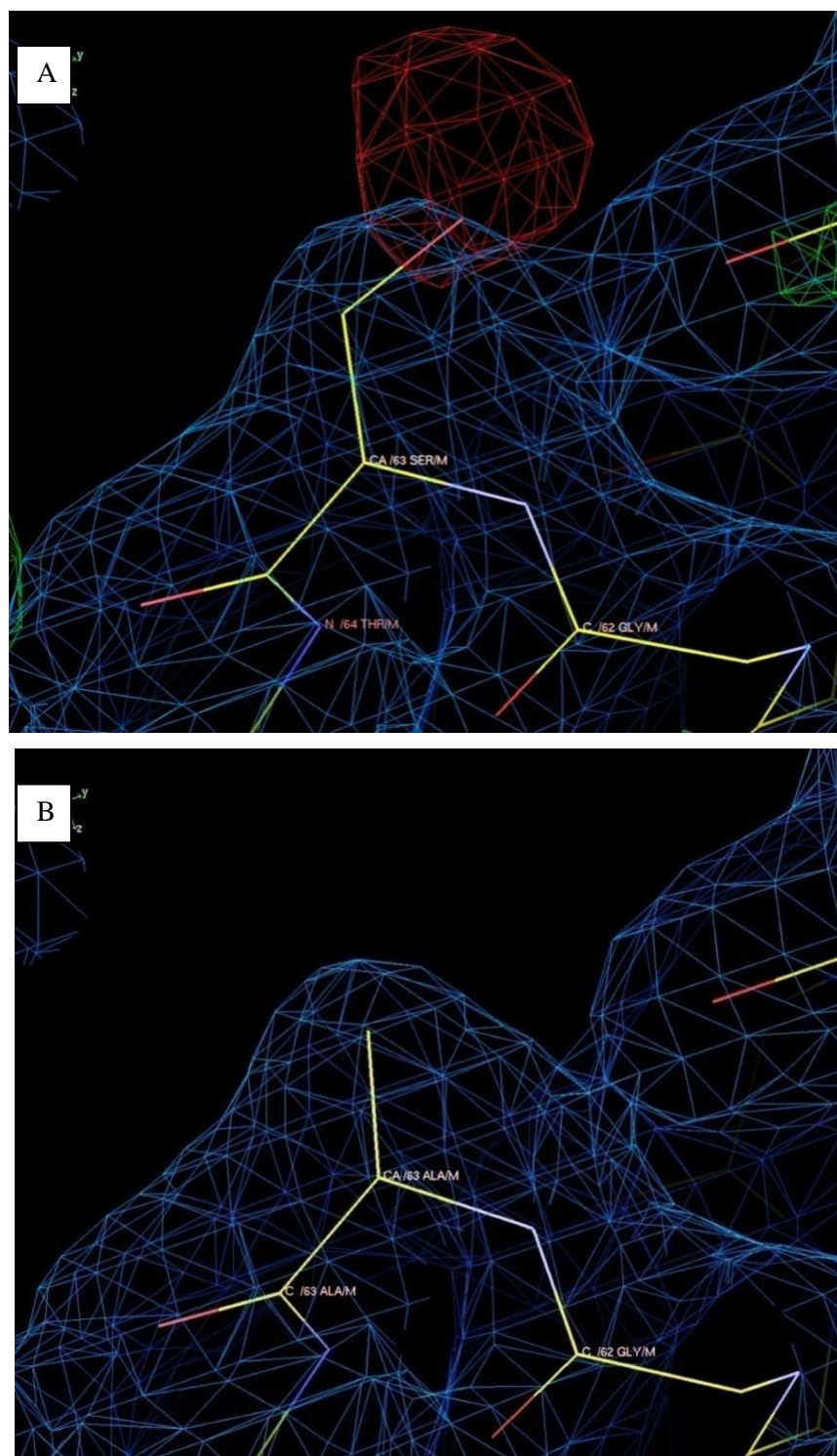


Figure 67. Electron density fitted with the old and the new sequence data at position M63. The colouring scheme is the same as the previous figure. (A) There is a large area of red, negative density in the map when Ser from the previously published amino acid sequence is fitted. (B) The new amino acid sequence data changes the sequence at position M63 from Ser to Ala. This new, smaller residue nicely fits the electron density map, removing the red negative density.

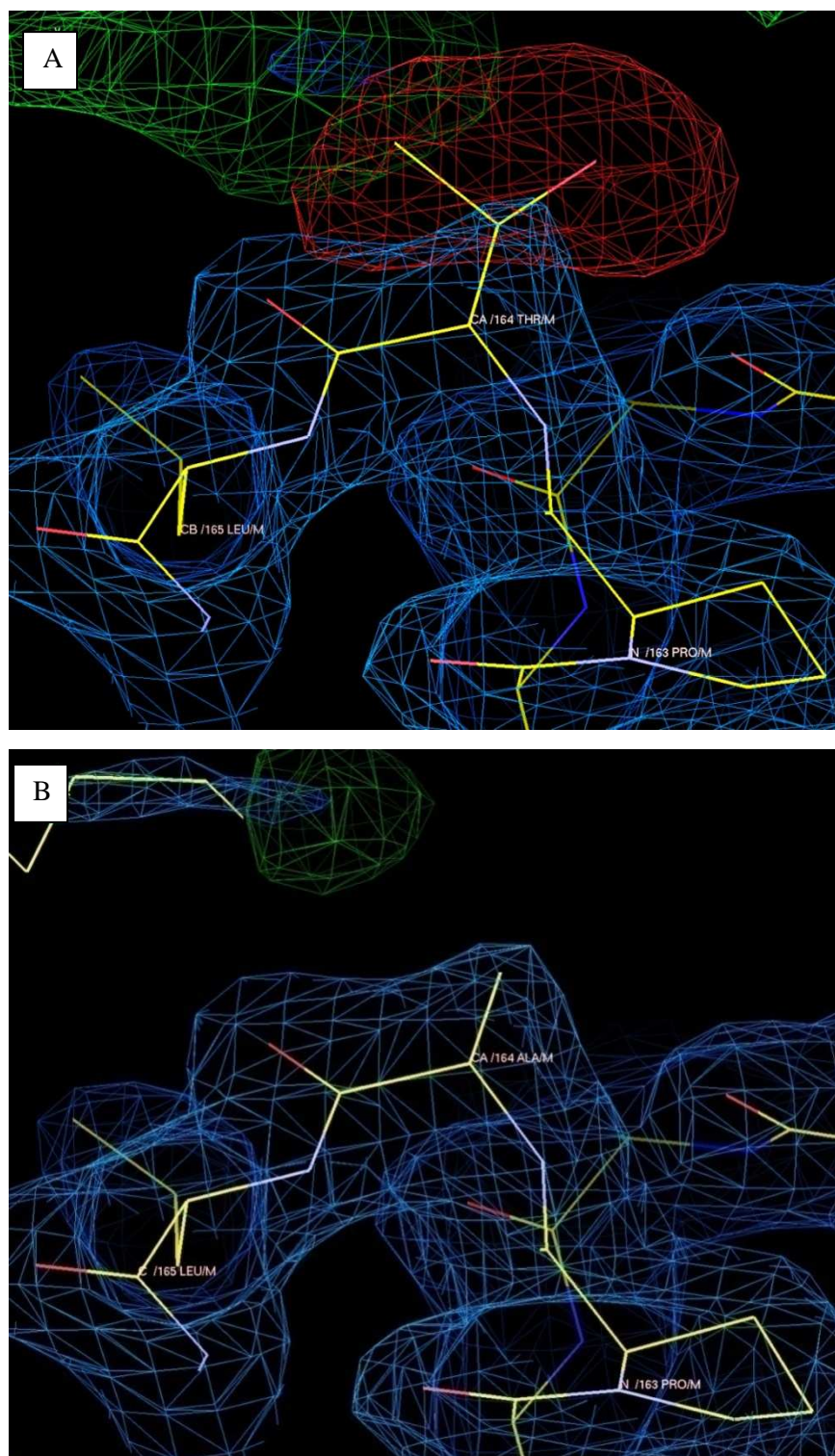


Figure 68. Electron density fitted with the old and the new sequence data at position M164. The colouring scheme is the same as the previous figure. (A) When the density is fitted with the old sequence data a large area of negative density is apparent (red). The amino acid at this position was Thr in the previous sequence (B) The new amino acid sequence assignment for position M164 is Ala, which fits much better into the new electron density map.

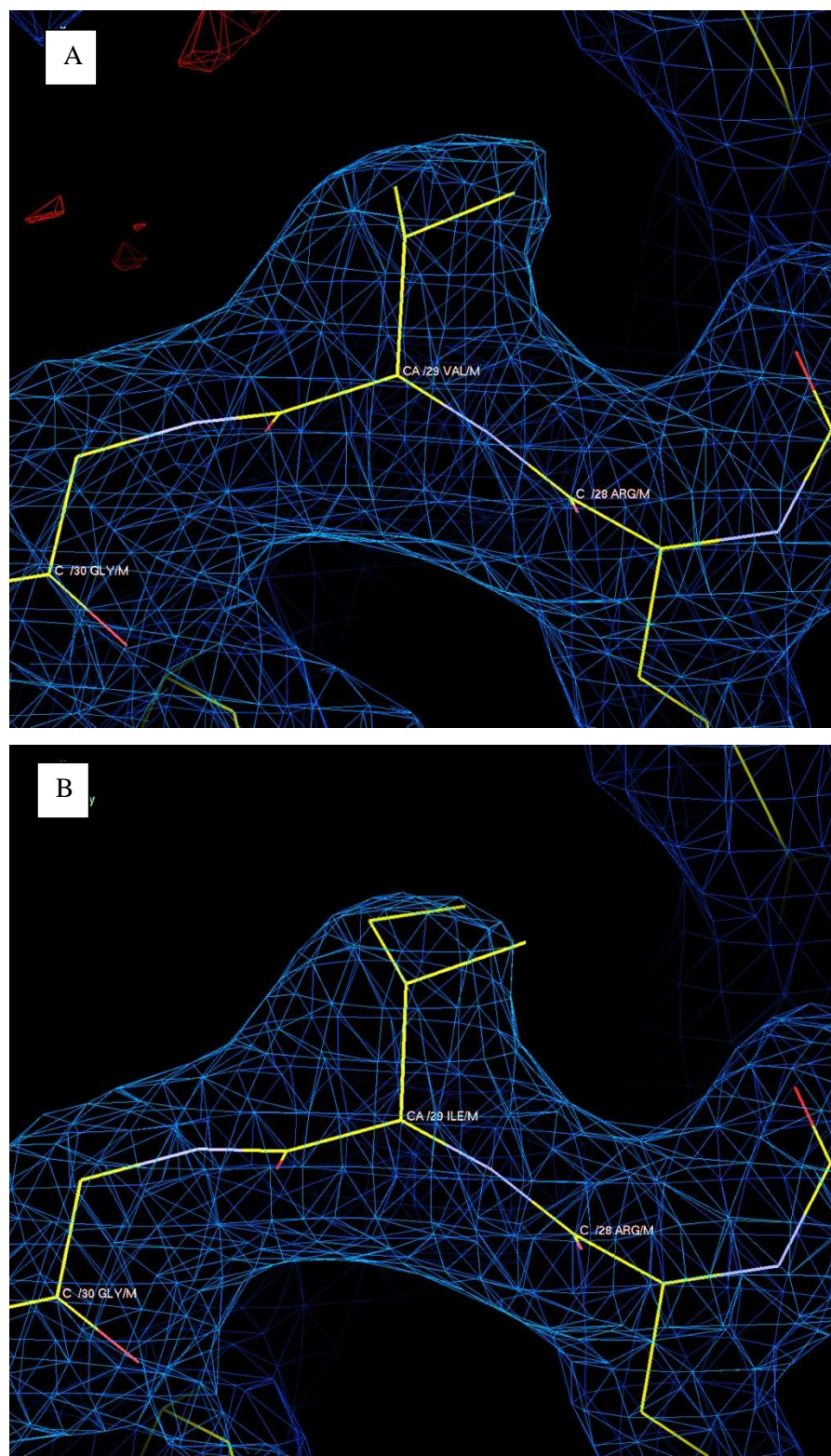


Figure 69. Electron density fitted with the old and the new sequence data at position M29. The colouring scheme is the same as the previous figure. (A) The previously published sequence has Val at this position and it is apparent that this residue fits very nicely into the new density. However, the new amino acid sequence suggests that Ile sits at this position. (B) Ile at position M29 could not fit well into the new 2.0 Å electron density map. The additional side chain is not contained within the new density.

4.2. C-subunit

There are 8 amino acid positions in the C-subunit that were changed as a result of the new sequencing data. Seven of the new assignments at C43, C77, C84, C252, C277, C287 and C323 (see Table 10) each resulted in much better fits to the 2.0 Å electron density maps. However, the change at position C288, though it fits well into the new electron density, only appears to result in a rather minor improvement.

The new electron density map at position C43 does not fit well if this residue is Ala (Figure 69A). There is extra non-fitted positive density suggesting that C43 should be a larger residue. The new sequence places a Pro at this position. The cyclic structure of Pro fits very nicely into the electron density map (Figure 69B), and ‘fills’ it appropriately.

In the new electron density map, the previous assignment of Ile position C77 does not fit well. This new sequence places a Met residue at this position. One of the methyl groups (-CH₃) of Ile lies outside of the electron density map (see Figure 70A) and there unfilled density elsewhere. The new sequence shows that this position is Met. Met contains a large aliphatic side chain that includes a thioether (-S-) group (Berg, et al., 2002). The structure of Met fits into the new electron density very well (see Figure 70).

At position C252, Thr present in the previous sequence does not fit well into the new electron density map (Figure 71A). Part of Thr lies outside the density. The new sequence places Ser at C252. Ser at C252 fits the new density much better than Thr (see Figure 71B). Ser lacks a methyl group on its side chain compared to Thr.

The previous sequence had suggested that C277 was a Leu residue. One of methyl groups on the side chain of Leu clearly lies outside of the new electron density while there is also extra, unfilled, density present (see Figure 72A). The new amino acid sequence places Met at this position and this residue fits very nicely to the new electron density (Figure 72B).

The electron density at C287 is clearly does not fit an Ala residue (see Figure 73A). The new sequence places a Thr residue at this position. Thr fits into the density very well as can be seen in Figure 73B. Another similar result was found for next residue C288. The electron density at this position C288 seems more appropriate for bigger amino acid rather than the Ser found in the previous sequence (see Figure 74A). The new sequence replaces Ser with Val. Difference between these two residues is the size of the side chain. Val has of two methyl groups whereas Ser has only one hydroxyl group. Val fits the electron density map much better than Ser (see Figure 74B).

The new electron density shows that that C323 is not Lys. The side chain of Lys lies outside the electron density (see Figure 75A). The new sequence found that C323 is Gln. In compared with Lys, Gln has shorter side chain. Gln fits very nicely in the new electron density (Figure 75B).

Finally, based on the new electron density map at position C84, the assignment of Gln from the previously published sequence seems to produce a reasonable fit (Figure 76A). There is only a small amount of negative density. However, the newly obtained sequence places Glu in this position. The difference between Gln and Glu is rather small. Glu can also be fitted very nicely into the new electron density (Figure 76B). In this case the negative density is a little smaller. The sequencing of this region of C was repeated with the same result.

4.3. H subunit

There were only 2 changes found in the new sequence of the H subunit compared with the previous sequence. These are at positions H216 and H256. The residues placed at these positions in the old sequence both appear to be too big to fit into the new electron density maps (see Figure 77A and Figure 78A, respectively). The previous assignments of Glu at H216 and at H256 do not, therefore, seem to be correct. The new sequencing result places an Asp at H216 (Figure 77B). Asp has is smaller than Glu. Asp fits very nicely to the new density. The new sequence replaces Ser at H256 with Ala. Again the smaller Ala much better fits to the new electron density (78B).

4.4. Conclusions

To summarise, in all but one case the new sequences of the RC polypeptides fit better to the high-resolution electron density map of the *Bl. viridis* RC. The new sequence data therefore has enabled a more accurate RC crystal structure to be produced. In particular the new structure reveals that the conformation of the RC carotenoid in *Bl. viridis* is very similar to that in other RC's and that this means that its structure is now also consistent with the spectroscopic data. All alterations in the structure apart from the one in the vicinity of the carotenoid do not impact the conformation of the *Bl. viridis* RC cofactors.

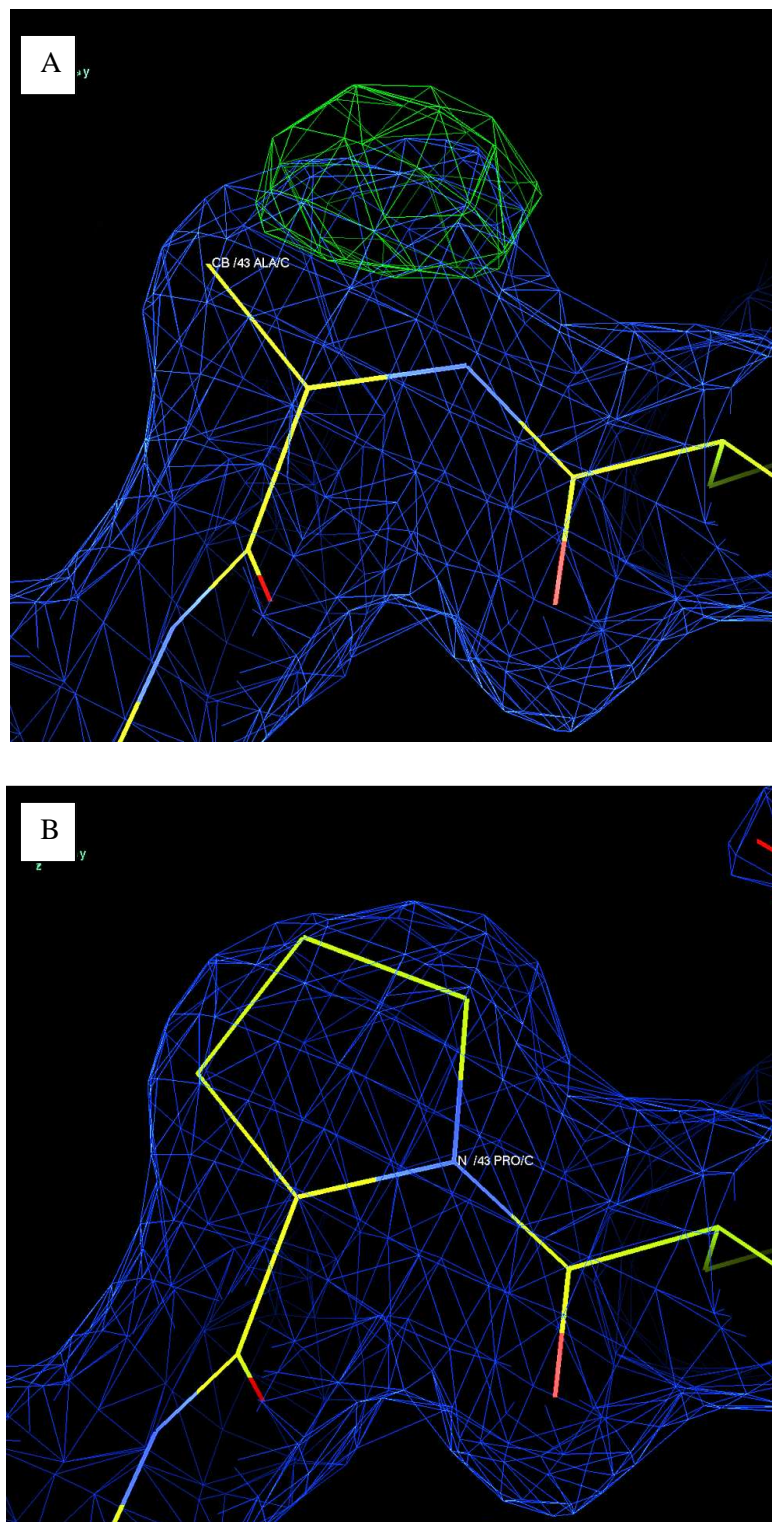


Figure 69. Electron density fitted with the old and the new sequence data at position C43. The colouring scheme is the same as previous figure. (A) There is a large putative mistake in the previous amino acid sequence revealed in green. New density has found that Ala residue at position C43 position was not satisfying. The new amino acid sequence at position C43 confirms that Ala is not present. (B) The new assignment of Pro at position C43 is nicely fits to the new, 2.0Å electron density map.

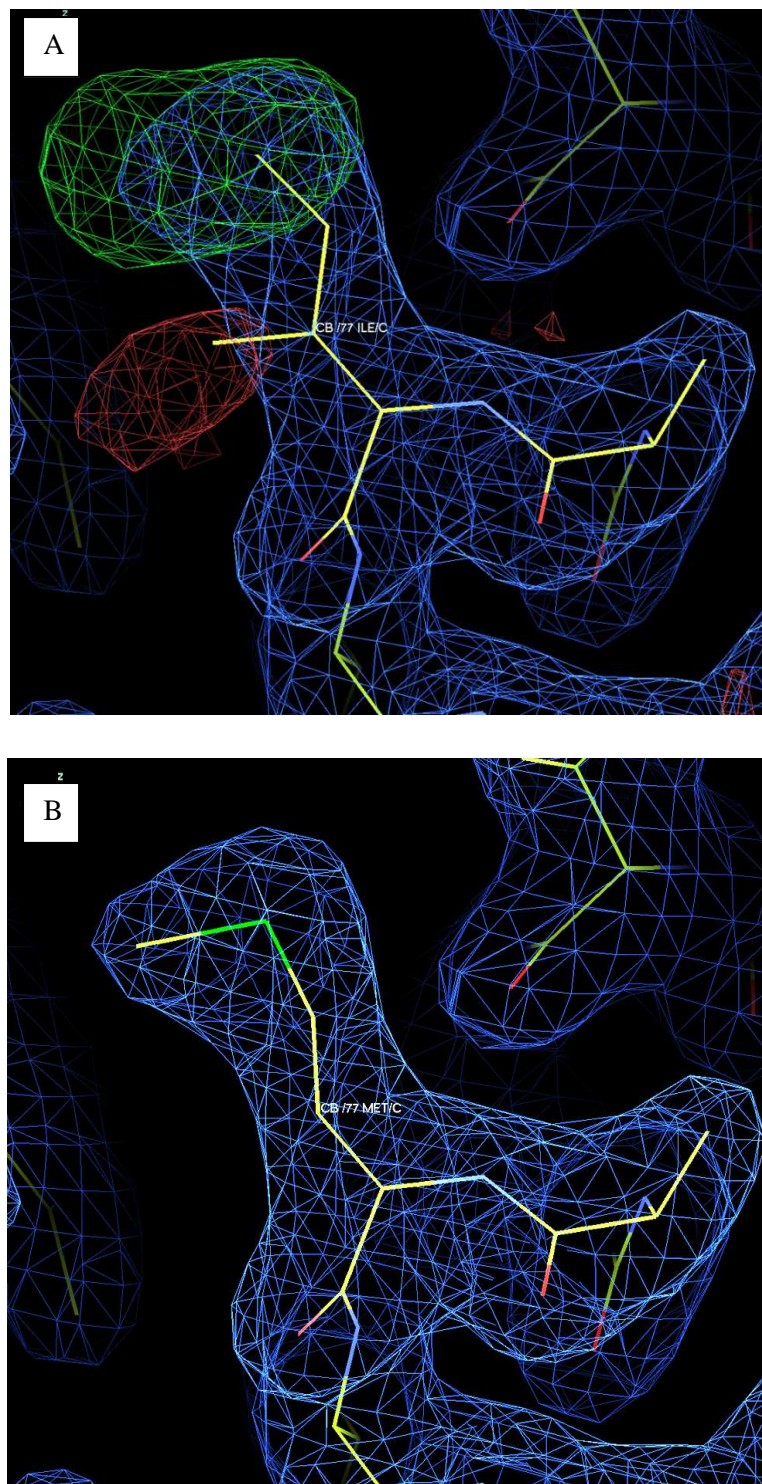


Figure 70. Electron density fitted with the old and the new sequence data at position C77. The colouring scheme is the same as previous figure. (A) When Ile residue at C77 position is fitted into the new density, there is a large unfilled region of positive density (green) and a large area of negative density is apparent (red). The newly obtained sequence suggests that this position is Met. The new assignment of Met at the C77 (B) has nicely fits to the new electron density map.

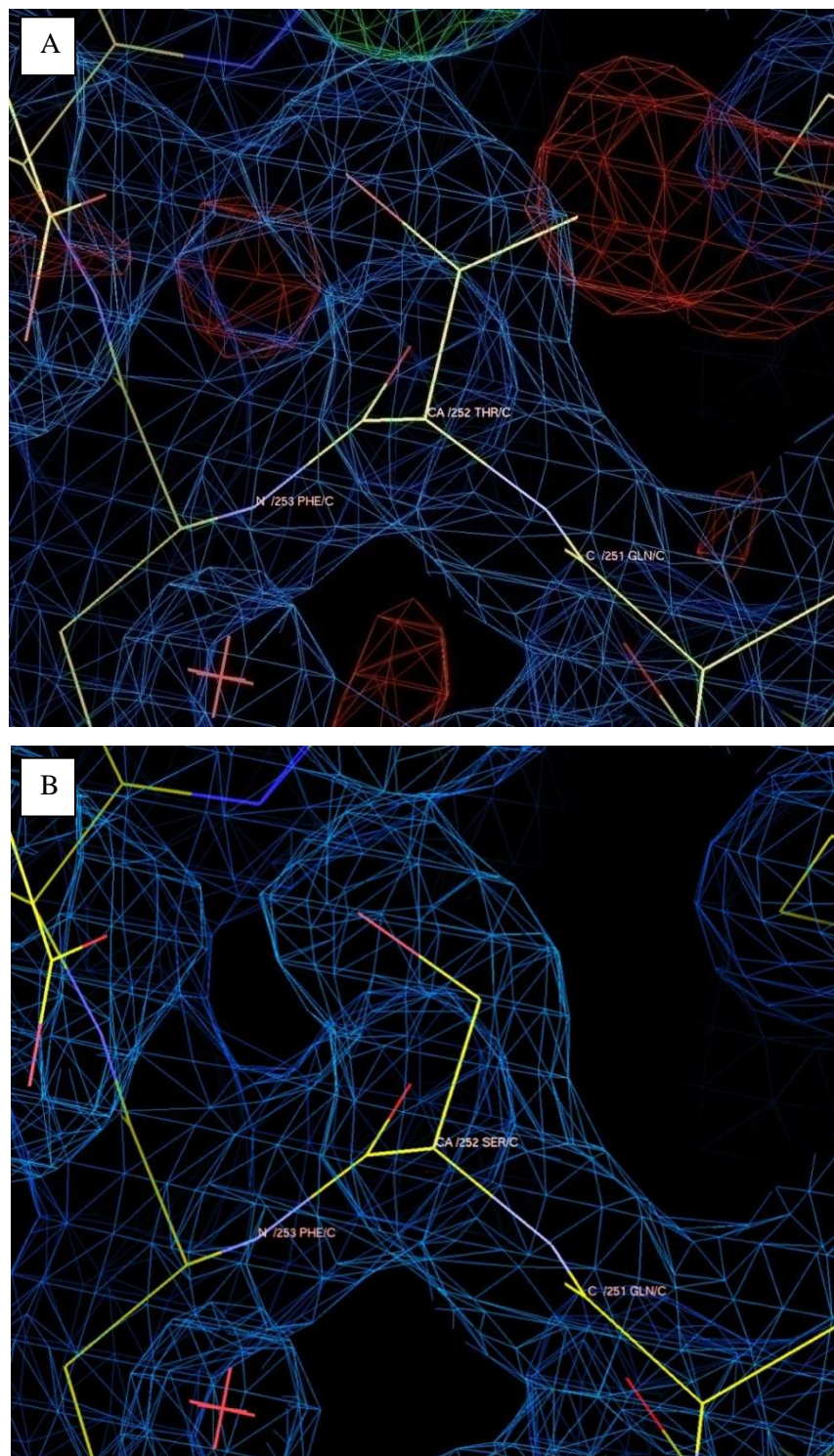


Figure 71. Electron density fitted with the old and the new sequence data at position C252. The colouring scheme is the same as previous figure. (A) When Thr residue from the previous sequence was fitted into the new density, a large area of negative density is apparent. (B) The new amino acid sequence assignment for position C252 is Ser, which fits very nicely to the new electron density map.

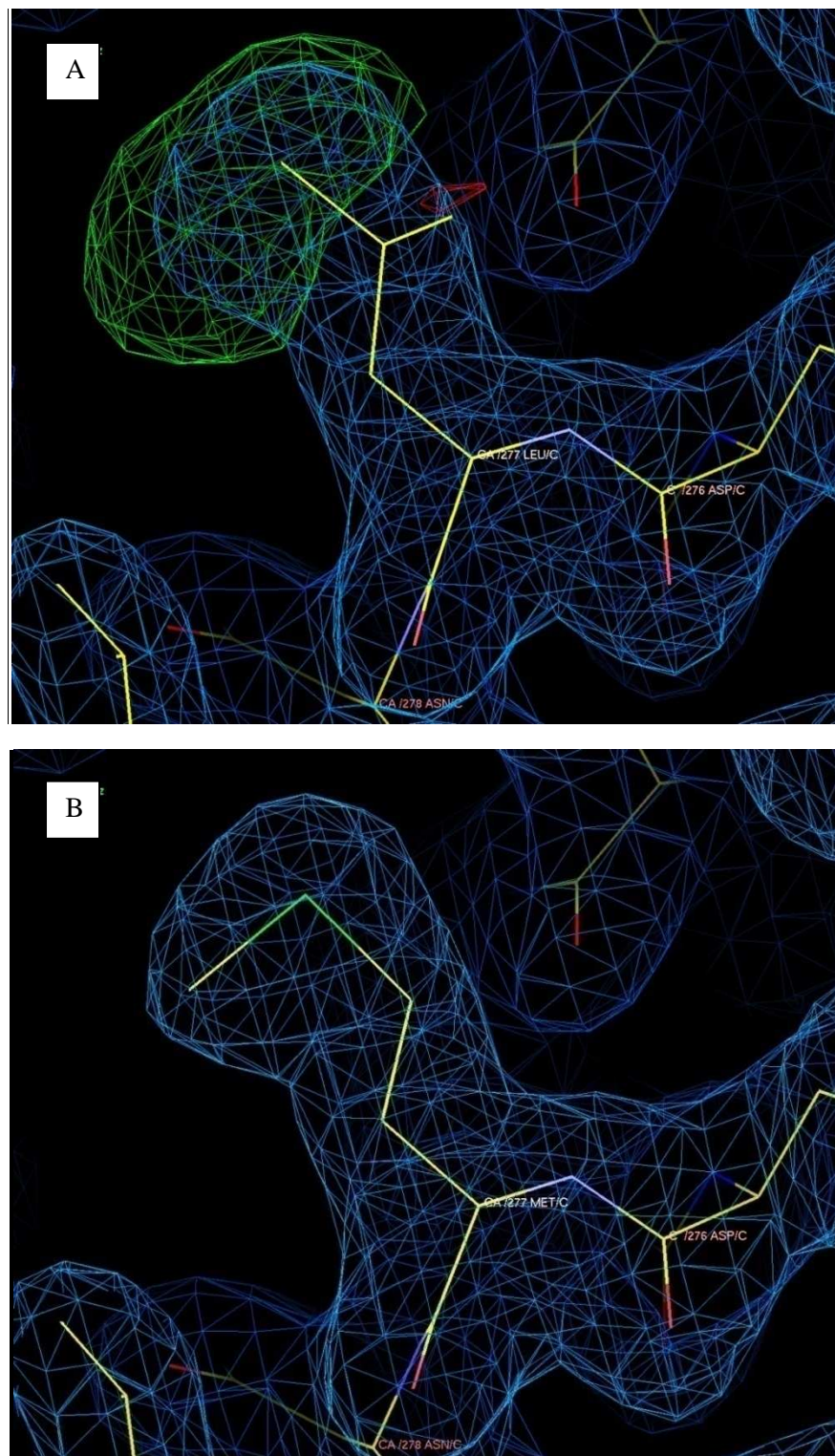


Figure 72. Electron density fitted with the old and the new sequence data at position C277. The colouring scheme is the same as previous figure. (A) There is a large unfilled region of positive density (green) and small negative density (red) appeared when Leu from the previously published sequence is fitted into the new density. The new amino acid sequence for this position is Met. The new assignment of Met fits very nicely to the new electron density.

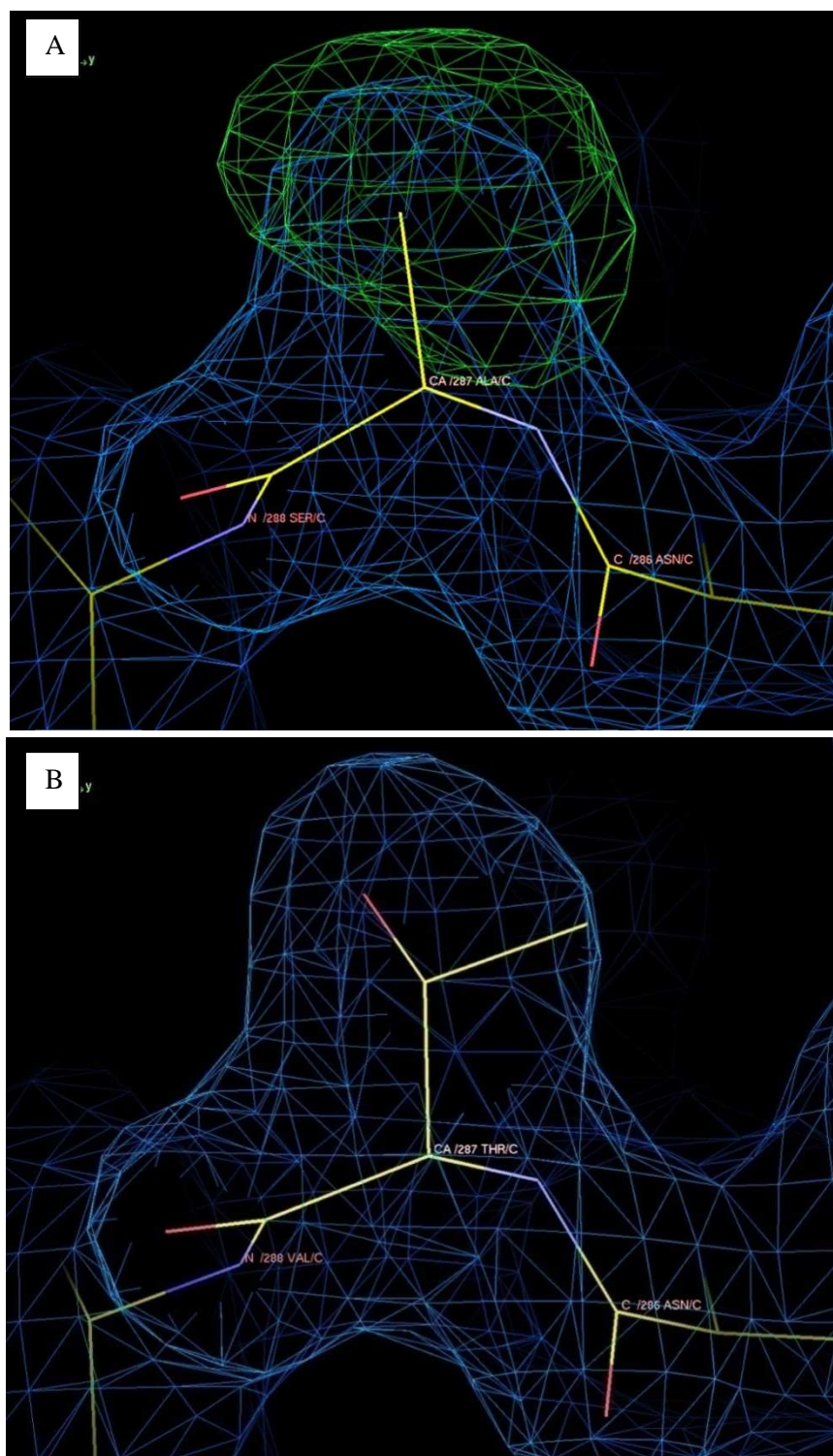


Figure 73. Electron density fitted with the old and the new sequence data at position C287. The colouring scheme is the same as previous figure. (A) The assignment of Ala from the previously published amino acid sequence to the new density resulting in the appearance of a large unfilled positive density (green). (B) The new amino acid sequence assignment for position C287 is Thr, which fits very nicely to the new electron density map.

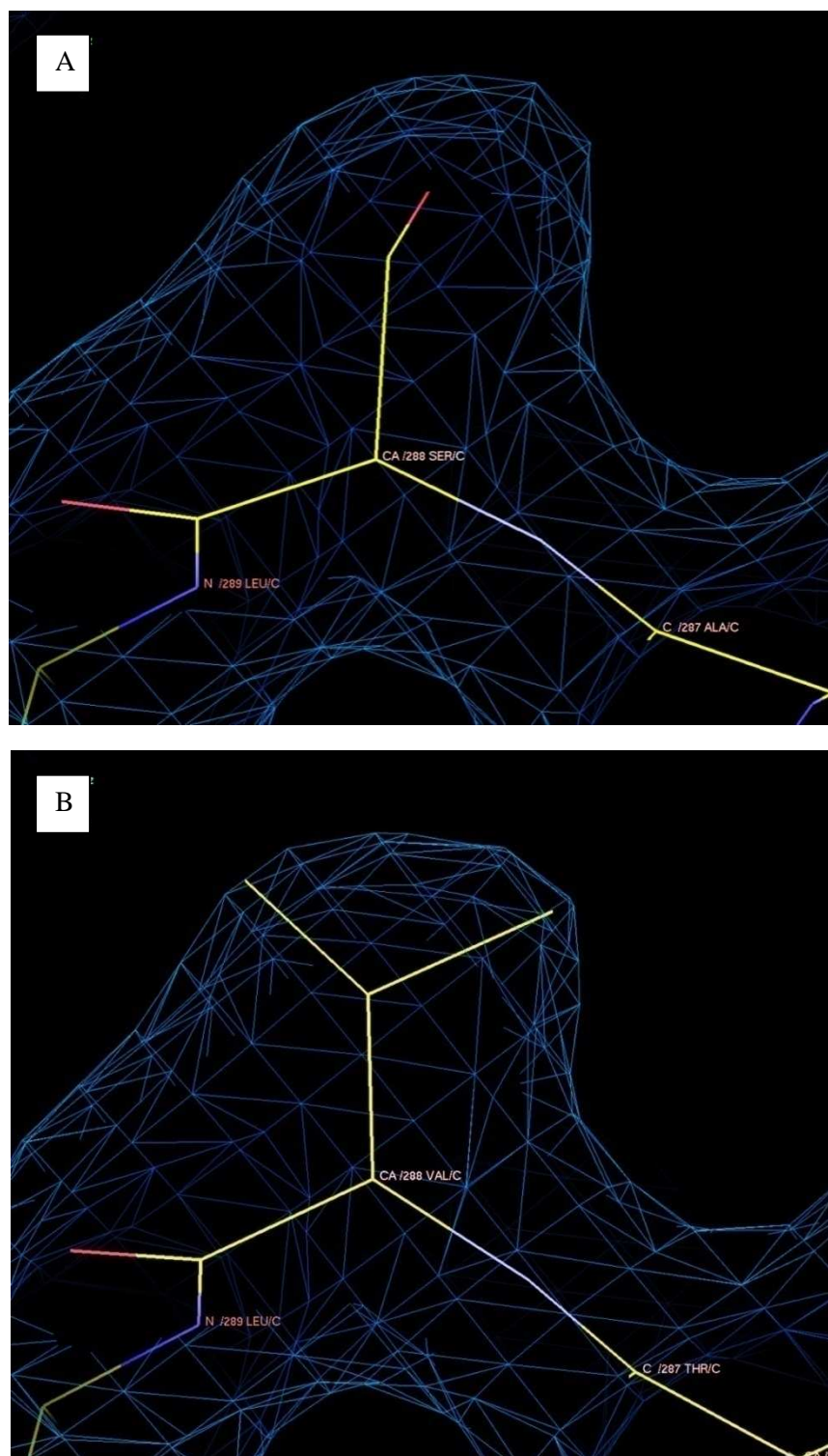


Figure 74. Electron density fitted with the old and the new sequence data at position C288. The colouring scheme is the same as the previous figure. (A) The previously published sequence has Ser at this position and it is apparent that this residue fits very nicely into the new density. However, the new amino acid sequence suggests that Val sits at this position. (B) Val at position C288 could not fit well into the new 2.0 Å electron density map. The additional side chain is not contained within the new density.

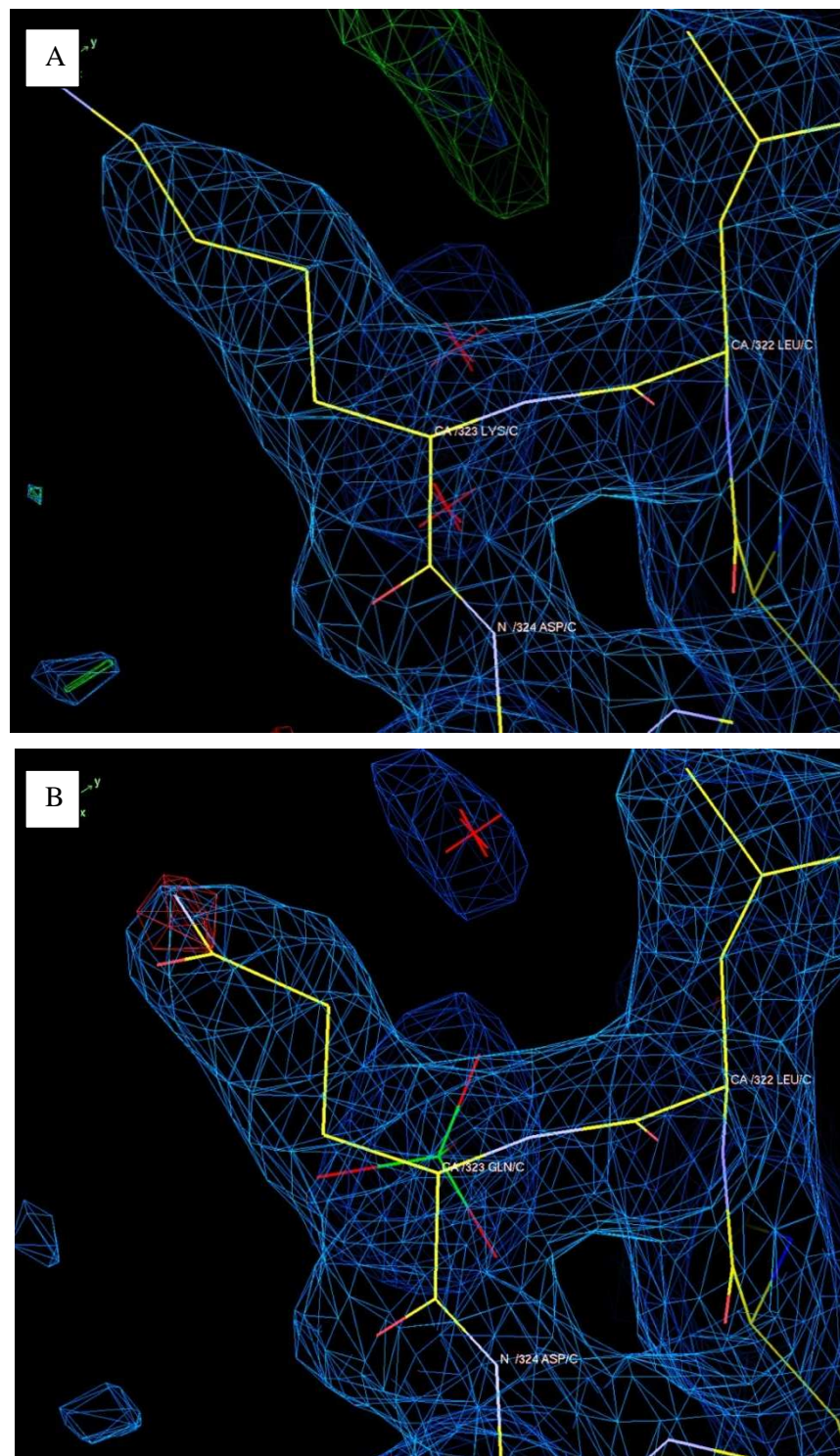


Figure 75. Electron density fitted with the old and the new sequence data at position C323. The colouring scheme is the same as previous figure. (A) The assignment of Lys residue from the previously published amino acid sequence does not fit well to the new electron density. There is too much carbon chain emerges from the electron density. (B) The new amino acid sequence assignment Gln at position C323 fits nicely to the new electron density map.

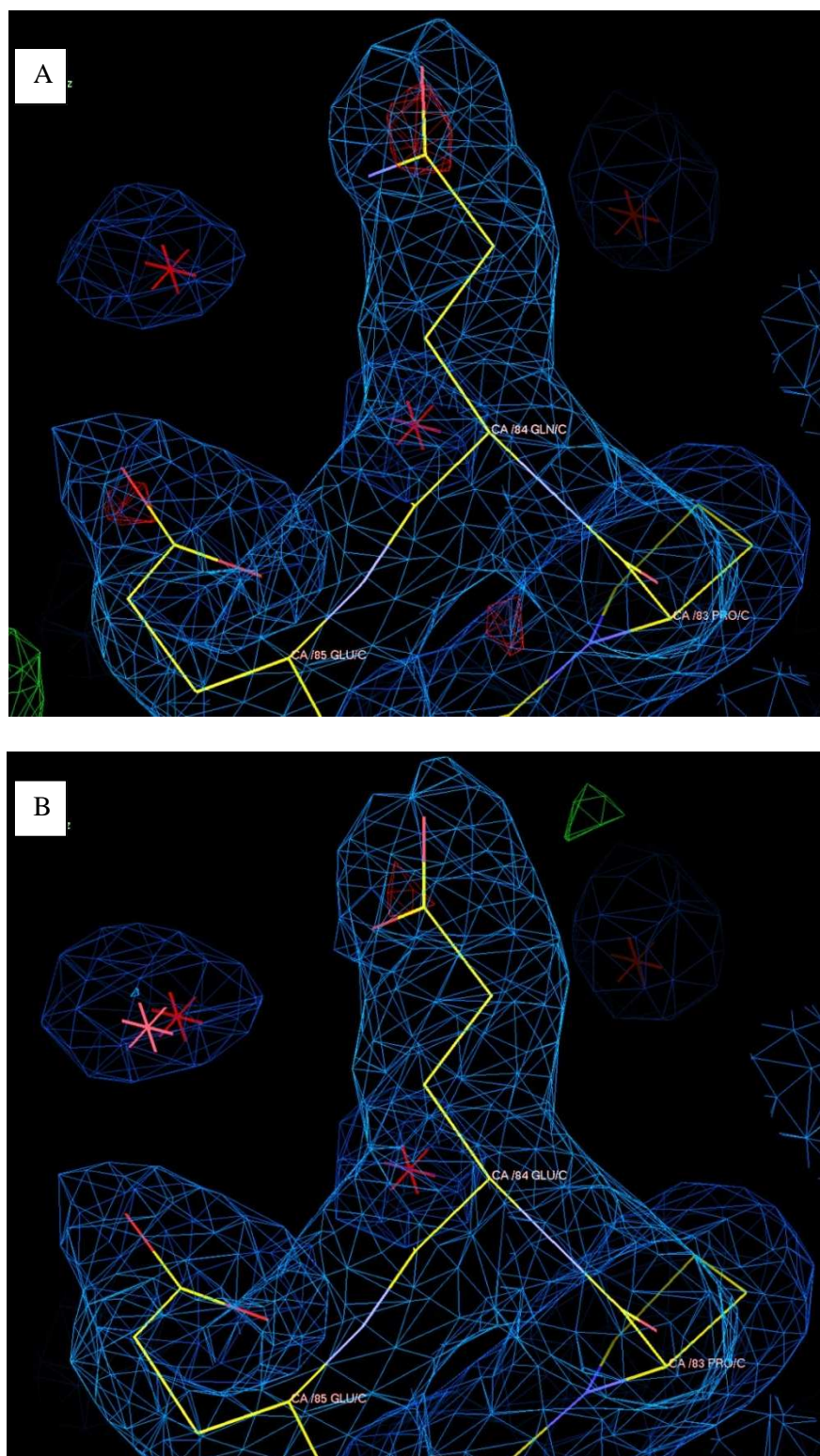


Figure 76. Electron density fitted with the old and the new sequence data at position C84. The colouring scheme is the same as previous figure. (A) The assignment of Gln from the previously published sequence seems to produce a reasonable fit. The new sequence data confirms that this position belongs to Glu. (B) The new assignment of Glu residue at the C84 position also fits nicely to the new electron density map.

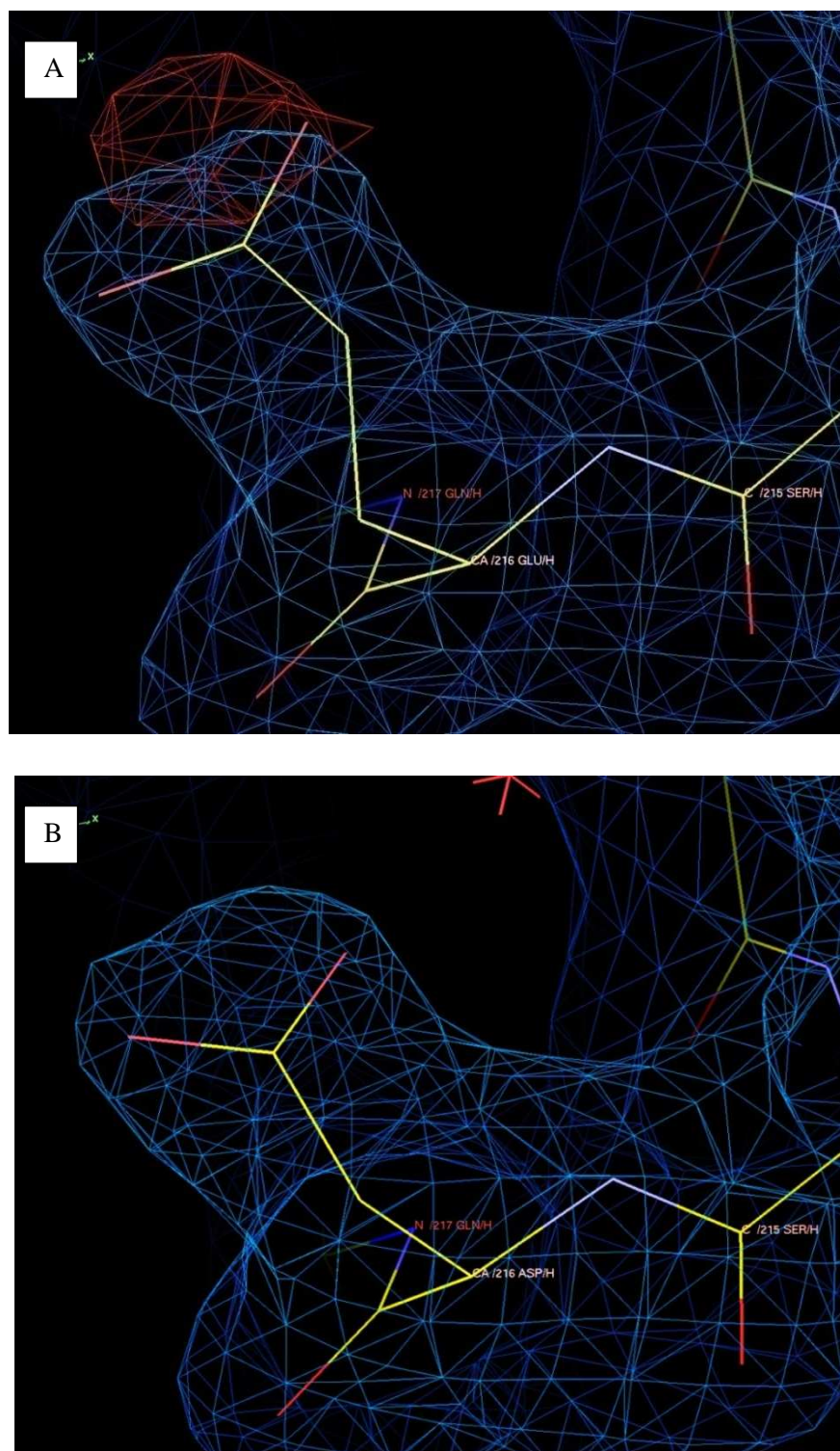


Figure 77. Electron density fitted with the old and the new sequence data at position H216. The colouring scheme is the same as previous figure. (A) There is a large area of red, negative density in the map when Glu from the previously published amino acid sequence is fitted. (B) The new amino acid sequence data changes the sequence at position H216 from Glu to Asp. This new, smaller residue nicely fits the electron density map, removing the red negative density.

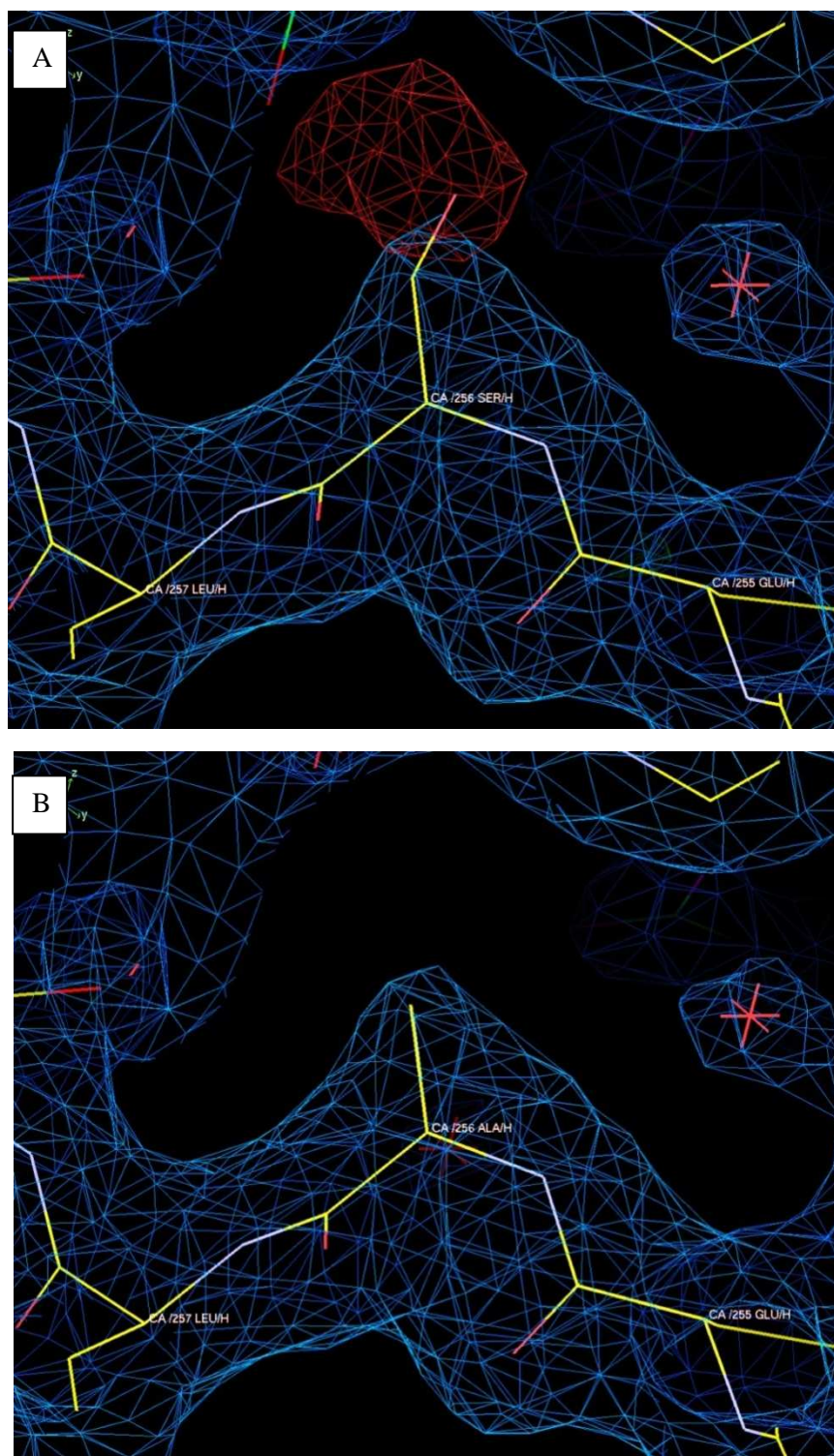


Figure 78. Electron density fitted with the old and the new sequence data at position H256. The colouring scheme is the same as previous figure. (A) There is a large area of red, negative density in the map when Ser from the previously published amino acid sequence is fitted. (B) The new amino acid sequence data changes the sequence at position H216 from Ser to Ala. This new, smaller residue nicely fits the electron density map, removing the red negative density.

References

- Albert, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P., 2008. *Molecular Biology of The Cell*. 5th Ed. New York: Garland Science.
- Rodriguez-Amaya, D. B., 2001. "A guide to carotenoid analysis in foods." Washington: ILSI Press
- Armitage, J. P., & Hellingwerf, K. J., 2003. "Light induced behavioral responses ('phototaxis') in prokaryotes." *Photosynth. Res.* 76: 145–155.
- Bahatyrova, S., Frese, R. N., Van der Werf, K. O., Otto, C., Hunter, C. N., & Olsen, J. D., 2004. "Flexibility and size heterogeneity of the LH1 light harvesting complex revealed by atomic force microscopy: functional significance for bacterial photosynthesis." *J. Biol. Chem.* 279: 21327–21333.
- Balows, A., Truper, H. G., Dworkin, M., Harder, W., & Schliefer, K., 1992. *The Prokaryotes*. Vol. 2. Berlin: Springer-Verlag.
- Barz, W. P., Vermeglio, A., Francia, F., Venturoli, G., Melandri, B.A. & Oesterhelt, D., 1995. "Role of the PufX protein in photosynthetic growth of *Rhodobacter sphaeroides*. 2. PufX is required for efficient ubiquinone/ubiquinol exchange between the reaction center QB site and the cytochrome bc1 complex." *Biochem.* 34: 15248–15258.
- Bauer, C., 1995. "Regulation of photosynthesis gene expression." In *Anoxygenic Photosynthetic Bacteria*, vol. 2. *Advances in Photosynthesis* (Blakenship, R. E., Madigan, M. T., & Bauer, C. E., eds.), pp. 1221–1234. Dordrecht: Kluwer Academic Publishers.
- Becker-Andre, & Hahlbrock, K., 1989. "Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by PCR aided transcription assay." *Nucl. Acids Res*: 17: 9437-9446.
- Beekman, L. M. P., Van Mourik, F., Jones, M. R., Visser, H. M., Hunter, C. N., & Van Grondelle, R., 1994. "Trapping kinetics in mutants of the photosynthetic purple bacterium *Rhodobacter sphaeroides*: Influence of the charge separation rate and consequences for the rate-limiting step in the light-harvesting process." *Biochem.* 33: 3143–3147.
- Berg, J. Tymoczko, M. J., & Stryer, L., 2002. *Biochemistry - 5th Ed*. New York: W. H. Freeman & Company.
- Bergstrom, H., Sundstrom, V., Van Grondelle, R., Gillbro, T., & Cogdell, R., 1988. "Energy-transfer dynamics of isolated B800–850 and B800–820 pigment-protein complexes of *Rhodobacter*

sphaeroides and *Rhodopseudomonas acidophila*.” *Biochim. et Biophys. Act. – Bioenerg.* 936: 90–98.

Bittl, R., Schlodder, E., Geisenheimer, I., Lubitz, W., & Cogdell, R. J., 2001. “Transient EPR and absorption studies of carotenoid triplet formation in purple bacterial antenna complexes.” *J. Phys. Chem.* 105: 5525–5535.

Blankenship, R. E, 2002. *Molecular Mechanisms of Photosynthesis*. Oxford: Blackwell Science.

Bose, S. K., 1963. In *Bacterial photosynthesis* (Gest, H., Ed). Yellow Springs: The Antioch Press. pp 501-511.

Britton, G., 1998. “Overview of carotenoid biosynthesis.” In *Carotenoids*. (Britton, G., Liaaen-Jensen, S., Pfander, H., eds.) Basel: Birkenhauser. pp. 133-180.

Brow, M. A. D, 1990. “Sequencing with Taq DNA polymerase.” In *PCR Protocols. A Guide to Methods and Applications*, (Innis, M. A., Gelfand, D. H., Snisnsky J. J., & White, T. J., eds.) San Diego: Academic Press. pp.189-196.

Brown, T., Hunter, C. N., Kneale, W. G., & Kennard, O., 1986. “Molecular structure of the G-A base pair in DNA and its implications for the mechanism of transversion mutations.” *Proc. Natl. Acad. Sci.* 83: 2402-2406.

Brown, T. A., 1990. *Gene Cloning - An Introduction*. Manchester: Chapman & Hall.

Brunisholz, R.A., & Zuber, H., 1992. ”Structure, function and organization of antenna polypeptides and antenna complexes from the three families of Rhodospirillaneae.” *J. Photochem & Photobiol.* 113–140.

Brunisholz, R. A., Zuber, H., Valentine, J., Lindsay, G. J., Wooley, K. J., & Cogdell, R. J., 1986. “The membrane location of the B890-complex from *Rs. rubrum* and the effect of carotenoid on the conformation of its two apoproteins at the cytoplasmic surface.” *Biochim. Biophys. Acta.* 849: 295-303.

Brunisholz, R. A., Jay, F., Suter, F., & Zuber, H., 1985. “The light-harvesting polypeptides of *Rhodopseudomonas viridis*.” *Biol. Chem. Hoppe-Seyer.* 336: 87-89.

Buggy, J. J., Sganga, M. W., & Bauer, C. E., 1994. “Characterization of a light-responding trans-activator responsible for differentially controlling reaction centre and light harvesting-I gene expression in *Rhodobacter capsulatus*.” *J. Bacteriol.* 176: 6936-6943.

- Chory, J., Donahue, T., Varga, A. R., Staehelin, L. A., & Kaplan, S., 1984. "Induction of the photosynthetic membranes of *Rhodospseudomonas sphaeroides*: Biochemical and morphological studies." *J. Bacteriol.* 159: 540-554.
- Choudhary, M., & Kaplan, S., 2000. "DNA sequence analysis of the photosynthetic region of *Rhodobacter sphaeroides* 2.4.1." *Nucl. Acid.* 28: 862-867.
- Cogdell, R. J., Gall, A., & Kohler, J., 2006. "The architecture and function of the light-harvesting apparatus of purple bacteria: from single molecule to in vivo membranes." *Q.Rev. of Biophys.* 39:227-324.
- Cogdell, R. J., Gardiner, A. T., Hashimoto, H., & Brotosudarmo, T. H. P., 2008. "A comparative look at the first few milliseconds of the light reactions of photosynthesis." *Photochem. & Photobiol. Sci.* 7: 1150-1158.
- Cogdell, R. J., & Lindsay, G., 1998. "Can photosynthesis provide a "biological blueprint" for the desing of novel solar cells." *Tibtech.* 16: 521-527.
- Cogdell, R. J., Durant, I., Valentine, J., Lindsay, G., & Schmidt, K., 1983. "The isolation and partial characterization of the light-harvesting pigment-protein complement of *Rhodospseudomonas acidophila*." *Biochim. Biophys. Acta.* 722:427-435.
- Cogdell, R. J., Isaacs, N., Howard, T. D., McLuskey, K., Fraser, N. J., & Price, M. S., 1999. "How Photosynthetic Bacteria Harvest Solar Energy." *J. Bacteriol.* 181:3869-3879.
- Cogdell, R. J., Zuber, H., Thornber, J. P., Drews, G., Ginras, G., Niederman, R. A., Parson, W. W., & Feher, G., 1985. "Recommendation for the naming of photochemical reaction centres and light-harvesting pigment-protein complexes from purple photosynthetic bacteria." *Biochim. Biophys. Acta.* 806: 185-186.
- Cogdell, R. J., Fyfe, P. K., Barret, S. J., Prince, S. M., Freer, A. A., Isaacs, N. W., McGlynn, P., & Hunter, C. N., 1996. "The purple bacterial photosynthetic unit." *Photosynthesis. Res.* 48: 55-63.
- Cohen-Bazire, G., Sistrom, W. R., & Stanier, R. Y., 1957. "Kinetics studies of pigment synthesis by purple non-sulphur bacteria." *J. Cell. Comp. Physiol.* 49: 25-68.
- Davis, S. J., Vener, A. V., & Vierstra, R. D., 1999. "Bacteriophytochromes: Phytochrome-like photoreceptors from nonphotosynthetic eubacteria." *Sci.* 286: 2517-2520.
- Deisenhofer, J., 2009. *Nobel Laurates: Johan Deisenhofer, 1988 Nobel Prize in Chemistry.* Howard Huges Medical Institute.

http://www.hhmi.org/research/investigators/figs/deisenhofer_nobel1.html (accessed January 6, 2009).

Deisenhofer, J., & Michel, H., 1989. "The Photosynthetic Reaction Centre from the Purple Bacterium *Rhodopseudomonas viridis*." *EMBO J.* 8: 2149-2170.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., & Michel, H., 1985. "Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3 Å resolution." *Nature.* 318: 618-624.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., & Michel, H., 1984. "X-ray structure analysis of a membrane protein complex: electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction centre from *Rhodopseudomonas viridis*." *J. Mol. Biol.* 180: 385-398.

Drews, G., & Giesbrecht, 1966. "Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3 Å resolution." *Nature.* 318: 618-624.

Drews, G., & Golecki, J. R., 1995. "Structure, molecular organization and biosynthesis of membranes of purple bacteria." In *Anoxygenic photosynthetic bacteria*, (Blankenship, R. E., & Madigan, M. T., eds.) Dordrecht: Kluwer Academic Publisher. pp.231-257.

Duysens, L. N. M., 1952. *Transfer of excitation energy in photosynthesis*. Doctoral Thesis, State University of Utrecht.

EBI. *ClustalW2*, 2008. European Bioinformatics Institute.
<http://www.ebi.ac.uk/Tools/clustalw2/index.html>.

Eraso, J. M., & Kaplan, S., 1996. "Complex regulatory activities associated with the histidine kinase PrrB in expression of photosynthesis genes in *Rhodobacter sphaeroides* 2.4.1." *J. Bacteriol.* 178: 7037-7046.

Eraso, J. M., & Kaplan, S., 1995. "Oxygen-insensitive synthesis of the photosynthetic membranes of *Rhodobacter sphaeroides*: A mutant histidine kinase." *J. Bacteriol.* 177: 2695-2706.

Erich, H. A., 1989. "PCR technology: principle and applications for DNA amplifications." New York: Stockton Press.

Evans, K., Fordham-Skelton, A. P., Mistry, H., Reynolds, C. D., Lawless, A. M., & Papiz, M. Z., 2005. "A bacteriophytochrome regulates the synthesis of LH4 complexes in *Rhodopseudomonas palustris*." *Photosynth. Res.* 85: 169–180.

- Fotiadis, D., Qian, P., Philippsen, A., Bullough, P. A., Engel, A., & Hunter, C. N., 2004. "Structural analysis of the reaction center light-harvesting complex I photosynthetic core complex of *Rhodospirillum rubrum* using atomic force microscopy." *J. Biol. Chem.* 279: 2063–2068.
- Frank, H. A., & Cogdell, R. J., 1996. "Carotenoid in Photosynthesis." *Photochem. & Photobiol.* 3: 257-264.
- Frank, H. A., Young, J., Britton, G., & Cogdell, R. J., 2000. "The photochemistry of Carotenoids: Application in Biology." Kluwer Academic Publisher.
- Frank, H. A., Violette, C. A., Trautman, J. K., Shreve, A. P., Owens, T. G., & Albrecht, A. C., 1991. "Carotenoids in photosynthesis: structure and photochemistry." *Pure & Appl. Chem.* 63: 109-114.
- Freiberg, A., Allen, J. P., Williams, J. & Woodbury, N. W., 1996. "Energy trapping and detrapping by wild type and mutant reaction centers of purple non-sulfur bacteria." *Photosynth. Res.* 48: 309–319.
- Fujii, R., Adachi, Shin-ichi, A., Roszak, A. W., Gardiner, A. T., Cogdell, R. J., Isaacs, N. W., Koshihara, S., & Hashimoto, H., 2006. "Precise structure of a carotenoid bound to the reaction centre from photosynthetically cultured *Rhodobacter sphaeroides* 2.4.1 revealed by the time-resolved X-ray crystallography." Abstract
- Gardiner, A. T., Cogdell, R. J., & Takaichi, S., 1992. "The effect of growth conditions on the light-harvesting apparatus in *Rhodopseudomonas acidophila*." *Photosynth. Res.* 38: 159-167.
- Garrison, S. A., 2005. *Synthesis and photophysical characterization of porphyrin containing supramolecular systems: Structural issues for porphyrin photophysics and electron transfer*. Doctoral Thesis. The Graduate Faculty of The University of Akron.
- Gelfand, D. H., & White, T. J., 1990. "Thermostable DNA polymerase." In *PCR protocols. A Guide to Methods and Applications*. (Innis, M. A., Sninsky, J. J., & White, T. J., eds.) San Diego: Academic Press. pp. 129-141.
- Gerken, U., Lupo, D., Tietz, C., Wrachtrup, J., & Ghosh, R., 2003. "Circular symmetry of the light-harvesting 1 complex from *Rhodospirillum rubrum* is not perturbed by interaction with the reaction center." *Biochem.* 42: 10354–10360.
- Giraud, E., Fardoux, J., Fourier, N., Hannibal, L., Genty, B., Bouyer, P., Dreyfus, B., & Vermeglio, A., 2002. "Bacteriophytochrome controls photosystem synthesis in anoxygenic bacteria." *Nature.* 417: 202–205.

- Giraud, E., Zappa, S., Jaubert, M., Hannibal, L., Fardoux, J., Adriano, J. M., Bouyer, P., Genty, B., Pignol, D., & Vermeglio, A., 2004. "Bacteriophytochrome and regulation of the synthesis of the photosynthetic apparatus in *Rhodospseudomonas palustris*: pitfalls of using laboratory strains." *Photochem. Photobiol. Sci.* 3: 587–591.
- Giraud, E., Zappa, S., Vuillet, L., Adriano, J. M., Hannibal, L., Fardoux, J., Berthomieu, C., Bouyer, P., Pignol, D., & Vermeglio, A., 2005. "A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in *Rhodospseudomonas palustris*." *J. Biol. Chem.* 280: 32389–32397.
- Gomelsky, M., & Kaplan, S., 1998. "AppA, a redox regulator of photosystem formation in *Rhodobacter sphaeroides* 2.4.1, is flavoprotein identification of a novel FAD binding domain." *J. Biol. Chem.* 273: 35319–35325.
- Gomelsky, M., & Kaplan, S., 1995. "Genetic evidence that PspR from *Rhodobacter sphaeroides* 2.4.1 functions as a repressor of puc and bchF expression." *J. Bacteriol.* 177: 1634–1637.
- Gomelsky, M., Horne, I. M., Lee, H. J., Pemberton, J. M., McEwan, A. G., & Kaplan, S., 2000. "Domain structure, oligomeric state, and mutational analysis of PspR, the *Rhodobacter sphaeroides* repressor of photosystem gene expression." *J. Bacteriol.* 182: 2253–2261.
- Gomelsky, M. K., & Kaplan, S., 1997. "Molecular genetic analysis suggesting interactions between AppA and PspR in regulation of photosynthesis gene expression in *Rhodobacter sphaeroides* 2.4.1." *J. Bacteriol.* 179: 128–134.
- Gouterman, M., 1961. "Spectra of porphyrins." *J. Mol. Spec.* 6: 138–163.
- Gross, J., 1991. "Pigments in vegetables: Chlorophyll and Carotenoids." New York: van Nostrand Reinhold.
- Gu, J., 1995. "In situ PCR - An Overview. In situ Polymerase Chain Reaction and Related Technology (Gu, J., ed.). Boston: Birkhauser. pp. 1–21.
- Hartwell, H. L., Hood, L., Goldberg, L. M., Reynolds, E. A., Silver, M. L., & Veres, C. R., 2008. "Genetics - From Genes to Genomes." New York: McGraw-Hill.
- Hess, S., Chachisvilis, M., Timpmann, K., Jones, M. R., Fowler, G. J. S., Hunter, C. N., & Sundstrom, V., 1995. "Temporally and spectrally resolved subpicosecond energy transfer within the peripheral antenna complex (LH2) and from LH2 to the core antenna complex in photosynthetic purple bacteria." *Proc. Natl. Acad. Sci. USA.* 92: 12333–12337.

- Higuchi, R., Dollinger, G., Walsh, P. S., & Griffith, R., 1992. "Simultaneous amplification and detection of specific DNA-sequences." *Bio-Tech.* 10: 413-417.
- Hiraishi, A., 1997. "Transfer of the Bacteriochlorophyll *b*-Containing Phototropic Bacteria *Rhodopseudomonas viridis* and *Rhodopseudomonas sulfoviridis* to the Genus *Blastochloris* Gen. Nov." *Int. J. Syst. Bacteriol.* 47: 217-219.
- Holt, S. C. M., & Marr, A. G., 1965. "Effect of light intensity on the formation of intracytoplasmic membrane in *Rhodospirillum rubrum*." *J. Bacteriol.* 89: 1421-1429.
- Hughes, J., Lamparter, T., Mittmann, F., Hartmann, E., Gartner, W., Wilde, A., & Borner, T., 1997. "A prokaryotic phytochrome." *Nature.* 386, 663.
- Imhoff, J. F., 1995. "Taxonomy and Physiology of Phototropic Purple Bacteria and Green Sulfur Bacteria." In *Anoxygenic Photosynthetic Bacteria*, (Blankenship, R., Madigan, M., & Bauer, C., eds.). Netherland: Kluwer Academic Publishers. pp. 1-15.
- Imhoff, J. F., & Truper, H. G., 1989. The purple non sulphur bacteria. In *Bergey's Manual of Systematic Bacteriology*. (Stanley, J., Bryant, M., Pfenning, N., & Holt, J., eds.) Baltimore: Wiliam and Wilkins. pp. 1658-1661.
- Imhoff, J. F., 1984A. "Quinones of phototropic purple bacteria." *FEMS Microbiol. Letts.* 25: 85-89.
- Imhoff, J. F., 1984B. "Reassignment of the genus *Ectothiorhodospiraceae* Pelsh 1936 to a new family *Ectothiorhodospiraceae* fam. nov., and amended description of the *Chromatiaceae* bavendamm 1924." *Int. J. Syst. Bacteriol.* 34: 338-339.
- Jaffe, B., 1976. "Crucibles: The story of chemistry." 4th Ed. Dover, New York.
- Jaubert, M., Zappa, S., Fardoux, J., Adriano, J. M., Hannibal, L., Elsen, S., Lavergne, J., Vermeglio, A., Giraud, E., & Pignol, D., 2004. "Light and redox control of photosynthesis gene expression in *Bradyrhizobium* – dual roles of two PpsR." *J. Biol. Chem.* 279: 44407–44416.
- Jay, F., Lambillotte, M., Stark, W. & Muehlethaler, K., 1984. "The preparation and characterization of native photoreceptor units from the thylakoids of *Rhodopseudomonas viridis*." *EMBO J.* 3:773–776.
- Jiang, Z. Y., Swem, L. R., Rushing, B. G., Devanathan, S., Tollin, G., & Bauer, C. E., 1999. "Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes." *Science.* 285: 406–409.

- Karniol, B., & Viestra, R. D., 2003. "The pair of bacteriophytochromes from *Agrobacterium tumefaciens* are histidine kinases with opposing photobiological properties." *Proc. Natl. Acad. Sci. USA*. 100: 2807–2812.
- Karrasch, S., Bullough, P. A., & Ghosh, R., 1995. "The 8.5 Å projection map of the light-harvesting complex I from *Rhodospirillum rubrum* reveals a ring composed of 16 subunits." *EMBO J.* 14: 631–638.
- Kavarnos, G. J., & Turro, N. J., 1986. "Photosensitization by Reversible Electron Transfer: Theories, Experiment, Evidence, and Example." *Chem. Rev.* 86: 401-449.
- Keppen, O. L., & Gorlenko, V. M., 1975. "A new species of purple budding bacteria containing bacteriochlorophyll *b*." *Mikrobiol.* 44: 258-264.
- Kobayashi, M., Akiyama, M., Kano, H., & Kise, H., 2007. "Spectroscopy and structure determination. In *Chlorophylls and Bacteriochlorophylls Biochemistry, Biophysics, Functions and Applications*," (Grimm, B., Porra, R. J., Rudiger, W., & Scheer, H., eds.) Netherlands: Springer. pp. 79-94.
- Kobayashi, M., Ohashi, S., Iwamoto, K., Shiraiwa, Y., Kato, Y., and Watanabe, T., 2007. "Redox potential of chlorophyll *d* *in vitro*." *Biochim. Biophys. Acta*. 1767: 596-602.
- Kubista, M., Andrade, J. M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjoback, R., Sjogreen, B., Strombom, L., Stahlberg, A., & Zoric, N., 2006. "The real-time polymerase chain reaction. Molecular Aspect of Medicine." *Mol. Aspects. Med.* 3: 95-125.
- Kyndt, J. A., Meyer, T. E., & Cusanovich, M. A., 2004. "Photoactive yellow protein, bacteriophytochrome, and sensory rhodopsin in purple phototrophic bacteria." *Photochem. Photobiol. Sci.* 3: 519–530.
- Lancaster, C. R. D., Bibikova, M. V., Sabatino, P., Oesterhelt, D., & Michel, H., 2000. "Structural basis of the drastically increased initial electron transfer rate in the reaction centre from a *Rhodopseudomonas viridis* mutant described at 2.00-Å resolution." *J. Biol. Chem.* 275: 39364-39368
- Lancaster, D. R. C., 2008. "Structure of Reaction Centres in Anoxygenic Bacteria." In *Primary Processes of Photosynthesis-Part 2*. (Renger, G., ed.) Cambridge: RSC Publishing. pp. 5-56.
- Lang, F. S., & Oesterhelt, D., 1989. "Microaerophilic growth and induction of the photosynthetic reaction centre in *Rhodopseudomonas viridis*." *J. Bacteriol.* 171: 2827-2834.

- Lascelles, J., 1959. "Adaptation to form bacteriochlorophyll in *Rhodobacter sphaeroides*: Changes in activity of enzymes concerned in pyrrole synthesis." *Biochem. J.* 72: 508-515.
- Law, C. J., & Cogdell, R. J., 2008. "Structure and Function of Antenna System." In *Primary Process of Photosynthesis - Part 1*. (Renger, G., ed.) Cambridge: RSC Publishing. pp. 205-260.
- Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., & Gelfand, D. H., 1989. "Isolation, characterization and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*." *J. Biol. Chem.* 264: 6427-6437.
- Madigan, M.T., Martinko, J. M., & Parker, J., 1997. "Brock Biology of Microorganisms." Upper Saddle River, NJ: Prentice-Hall.
- Marrs, B. G., & Gest, H., 1973. "Regulation of bacteriochlorophyll synthesis by oxygen in respiratory mutants of *Rhodopseudomonas capsulata*." *J. Bacteriol.* 114: 1052-1057.
- McDermott, G., Prince, S. M., Freer, A. A., Hawthorntwaite-Lawless, M., Papiz, Z., Cogdell, R. J., & Isaacs, N. W., 1995. "Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria." *Nature.* 374: 517-521.
- Michel, H., 1982. "Three dimensional crystal of membrane protein complex. The photosynthetic reaction centre from *Rhodopseudomonas viridis*." *J. Mol. Biol.* 158: 567-572.
- Michel, H., Weyer, K. A., Gruenberg, H., & Lottspeich, F., 1985. "The 'heavy' subunit of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: isolation of the gene, nucleotide sequence and amino acid sequence." *EMBO J.* 4: 1667-1672.
- Michel, H., Weyer, K. A., Gruenberg, H., Dunger, D. O., & Lottspeich, F., 1986. "The 'light and 'medium' subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: isolation of the genes, nucleotide and amino acid sequence." *EMBO J.* 6: 1149-1158.
- Miller, K. R., 1982. "Three-dimensional structure of a photosynthetic membrane." *Nature.* 300: 53-55.
- Miller, K. R., 1979. "Structure of a bacterial photosynthetic membrane." *Proc. Natl. Acad. Sci. USA.* 76: 6415-6419.
- Mitchell, I. A., 2003. *Structural Studies On Purple Non-Sulphur Bacterial Pigment-Protein Complexes*. M. Sc. Thesis Departement of Chemistry. The University of Glasgow.
- Molbiol. *Molbiol 6 Frame Translation*. 2008. http://www.molbiol.edu.ru/eng/scripts/01_13.html.

- Monger, T. G., Cogdell, R. J., & Parson, W. W., 1976. "Triplet states of bacteriochlorophyll and carotenoids in chromatophores of photosynthetic bacteria." *Biochim. Biophys. Act. – Bioenerg.* 449: 136–153.
- Mullis, K. B., & Fallona, F. A., 1987. "Specific synthesis of DNA in vitro via polymerase-catalyzed chain reaction." *Methods Enzymol.* 155: 335-350.
- Li, L., Mustafi, D., Fu, Q., Tereshko, V., Chen, D. L., Tice, J. D., & Ismagilov, R. F., 2006. "Nanoliter microfluidic hybrid method for simultaneous screening and optimization validated with crystallization of membrane proteins." *Proc. Natl. Acad. Sci. USA.* 103: 19243-19248
- Nagarajan, V. & Parson, W. W., 1997. "Excitation energy transfer between the B850 and B875 antenna complexes of *Rhodobacter sphaeroides*." *Biochem.* 36: 2300–2306.
- Northwestern University. *Oligonucleotide Properties Calculator*. Vers. 3.23. Northwestern University. 15 May 2008. <http://www.basic.northwestern.edu/biotools/OligoCalc.html>.
- Oelze, J., & Drews, G., 1972. "Membrane of Photosynthetic Bacteria." *Biochim. Biophys. Acta.* 26: 209-239.
- Oh, J. I., Ko, I. J., & Kaplan, S., 2001. "The default state of the membrane-localized histidine kinase PrrB of *Rhodobacter sphaeroides* 2.4.1 is in the kinase-positive mode." *J. Bacteriol.* 183: 4044-4050.
- Oh, J. I., Eraso, J. M., & Kaplan, S., 2000. "Interacting regulatory circuits involved in orderly control of photosynthesis gene expression in *Rhodobacter sphaeroides* 2.4.1." *J. Bacteriol.* 182: 3081-3087.
- Old, W. R., & Primrose, B. S., 1994. *Principles of Genes Manipulation - An Introduction to Genetic Engineering*. 5th Ed. Oxford: Blackwell Science Ltd.
- Parkes-Loach, P. S., Law, C. J., Recchia, P. A., Kehoe, J., Nehrlich, S., Chen, J., & Loach, P. A., 2001. "Role of the core region of the PufX protein in inhibition of reconstitution of the core light-harvesting complexes of *Rhodobacter sphaeroides* and *Rhodobacter capsulatus*." *Biochem.* 40: 5593–5601.
- Parson, W. W., 2008. "Functional Patterns of Reaction Centres in Anoxygenic Photosynthetic Bacteria." In *Primary Processes of Photosynthesis-Part 2*, (Renger, G., ed.) Cambridge: RCS Publishing. pp. 57-85.

- Penfold, R. J. P., & Pemberton, J. M., 1994. "Sequencing, chromosomal inactivation and functional expression of ppsR, a gene which repress carotenoid and bacteriochlorophyll synthesis in *Rhodobacter sphaeroides*." *J. Bacteriol.* 176: 2869-2876.
- Pfennig, N., 1978. "General physiology and ecology of photosynthetic bacteria." In *The Photosynthetic Bacteria*, (Clayton, R. K., & Sistrom, W. R., eds.) New York: Plenum Publishing Corporation. pp. 3-18.
- Pfennig, N., 1969. "*Rhodopseudomonas acidophila*, sp. n., a New Species of the Budding Purple Nonsulfur Bacteria." *J. Bacteriol.* 99: 597-602
- Robert, B., 1999. "The Electronic Structure, Stereochemistry and Resonance Raman Spectroscopy of Carotenoids." In *The Photochemistry of Carotenoids*, (Frank, H. A., Young, A. J., Britton, G., & Cogdell, R. J., eds.) Netherlands: Kluwer Academic Publisher. pp. 189-201.
- Roszak, A.W., Gardiner, A. T., Moulisova, V., Putranto., A. D., Fujii, R., Hashimoto, H., Isaacs, N. W., Cogdell, R. J., 2008. "New insight into the structure of the carotenoid bound to the reaction centre from *Blastochloris viridis*." *Abstract*.
- Roszak, A. W., McKendrick, K., Gardiner, A. T., Mitchell, I. A., Isaacs, N. W., Cogdell, R. J., Hashimoto, H., & Frank, H., 2004. "Protein Regulation of Carotenoid Binding: Gatekeeper and Locking Amino Acids Residues in Reaction Centers of *Rhodobacter sphaeroides*." *Structure*. 12: 765-773.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., & Mullis, K. B., 1988. "Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA polymerase." *Science*. 230: 487-491.
- Shimada, H., Wada, T., Handa, H., Ohta, H., Mizoguchi, H., Nishimura, K., Masuda, T., Shioi, Y., & Takamiya, K., 1996. "A transcription factor with a leucine-zipper motif involved in light-dependent inhibition of expression of the puf operon in the photosynthetic bacterium *Rhodobacter sphaeroides*." *Plant Cell Physiol.* 37: 515-522.
- Scheuring, S., Francia, F., Busselez, J., Melandri, B. A., Rigaud, J. L., & Levy, D., 2004a. "Structural role of PufX in the dimerization of the photosynthetic core complex of *Rhodobacter sphaeroides*." *J. Biol. Chem.* 279: 3620–3626.
- Scheuring, S., Rigaud, J. L. & Sturgis, J. N., 2004b. "Variable LH2 stoichiometry and core clustering in native membranes of *Rhodospirillum rubrum*." *EMBO J.* 23: 4127–4133.

- Scheuring, S., Seguin, J., Marco, S., Levy, D., Robert, B. & Rigaud, J. L., 2003. "Nanodissection and highresolution imaging of the *Rhodopseudomonas viridis* photosynthetic core complex in native membranes by AFM." *Proc. Natl. Acad. Sci. USA*. 100: 1690–1693.
- Siebert, C. A., Qian, P., Fotiadis, D., Engel, A., Hunter, C. N., & Bullough, P. A., 2004. "Molecular architecture of photosynthetic membranes in *Rhodobacter sphaeroides*: the role of PufX." *EMBO J*. 23: 690–700.
- Stahlberg, H., Dubochet, J., Vogel, H. & Ghosh, R., 1998. "Are the light-harvesting I complexes from *Rhodospirillum rubrum* arranged around the reaction center in a square geometry?" *J. Mol. Biol.* 282: 819–831.
- Stark, W., Jay, F. & Muehlethaler, K., 1986. "Localization of reaction center and light harvesting complexes in the photosynthetic unit of *Rhodopseudomonas viridis*." *Arc. Microbiol.* 146: 130–133.
- Sundstrom, V., & Van Grondelle, R., 1995. "Kinetics of excitation transfer and trapping in purple bacteria." In *Anoxygenic Photosynthetic Bacteria*, vol. 2. *Advances in Photosynthesis* (Blankenship, R. E., Madigan M. T., & Bauer, C. E., eds.). pp. 350–372. Dordrecht: Kluwer Academic Publishers.
- Sundstrom, V., Van Grondelle, R., Bergstrom, H., Aakesson, E., & Gillbro, T., 1986. "Excitation-energy transport in the bacteriochlorophyll antenna systems of *Rhodospirillum rubrum* and *Rhodobacter sphaeroides*, studied by low-intensity picosecond absorption spectroscopy." *Biochim. Biophys. Act. – Bioenerg.* 851: 431–446.
- Takaichi, S., 1999. "Carotenoids and carotenogenesis in anoxygenic photosynthetic bacteria. In *The photochemistry of carotenoids: Advances in photosynthesis respiration.*" (Frank, H. A., Young, A. J., Britton, G., & Cogdell, R. J., eds.) Springer Netherlands. Pp. 39-69.
- Takemoto, J., & Lascelles, J., 1973. "Coupling between bacteriochlorophyll and membrane protein synthesis in *Rhodopseudomonas spheroides*." *Proc. Natl. Acad. Sci. USA*. 70: 799–803.
- Valasek, A. M., & Repa, J. J., 2005. "The power of real-time PCR." *Physiol Edu.* 29: 151-159.
- Van Grondelle, R., Bergstrom, H., Sunstrom, V., & Gillbro, T., 1987. "Energy transfer within the bacteriochlorophyll antenna of purple bacteria at 77 K, studied by picosecond absorption recovery." *Biochim. et Biophys. Act. – Bioenerg.* 894: 313–326.
- Van Grondelle, R., Dekker, J. P., Gillbro, T., & Sundstrom, V., 1994. "Energy transfer and trapping in photosynthesis." *Biochim. Biophys. Act. – Bioenerg.* 1187: 1–65.

- Van Neil, C. B., 1941. "The bacterial photosynthesis and their importance for the general problem of photosynthesis. In *Advance in Enzymology and Related Areas of Molecular Biology.* " (Nord, F. F., & Werkman, C. H., eds.) Interscience Publisher. pp. 263-328.
- Visscher, K. J., Bergstrom, H., Sundstrom, V., Hunter, C. N., & Van Grondelle, R. 1989. "Temperature dependence of energy transfer from the long wavelength antenna BChl-869 to the reaction center in *Rhodospirillum rubrum*, *Rhodobacter sphaeroides* (w. t. and M21 mutant) from 77 to 177 K, studied by picoseconds absorption spectroscopy." *Photosynth. Res.* 22: 211–217.
- Westerhuis, W. H. J., Sturgis, J. N., Ratcliffe, E. C., Hunter, C. N., & Niederman, R. A., 2002. "Isolation, size estimates, and spectral heterogeneity of an oligomeric series of light-harvesting 1 complexes from *Rhodobacter sphaeroides*." *Biochem.* 41: 8698–8707.
- Wiessner, C., Dunger, I., and Michel, H., 1990. "Structure and Transcription Genes Encoding the B1015 Light-Harvesting Complex Beta and Alpha Subunits and the Photosynthetic Reaction Centre L, M, and Cytochrome c Subunits from *Rhodopseudomonas viridis*." *J. Bacteriol.* 172: 2877-2887.
- Wilcox, J. N., 1993. "Fundamental Principle of *In Situ* Hybridization." *J. Histochem. & Cytochem.* 12: 1725-1733.
- Willems, A., Gillis, M., & de Ley, J., 1991. "Transfer of *Rhodocylus gelatinosis* to *Rubrivivax gelatinosis* gen. nov., comb nov., and phylogenetic relationships with *leptothrix*, *Sphaerotilus natans*, *Pseudomonas saccharophila*, and *Alcaligenes latus*." *Int. J. Syst. Bacteriol.* 41: 65-73.
- Young, C. S., & Beatty, J. T., 2003. "Multi-level regulation of purple bacterial light-harvesting complexes." In *Light-harvesting Antennas in Photosynthesis*, vol. 13. *Advances in Photosynthesis* (Green, B. B., & Parson, W. W., eds.) pp. 449–470. Dordrecht: Kluwer Academic Publishers.
- Yuwono, T., 2006. *Teori dan aplikasi Polymerase Chain Reaction*. Yogyakarta: Andi.
- Zhang, F. G., Gillbro, T., Van Grondelle, R., & Sundstrom, V., 1992a. "Dynamics of energy transfer and trapping in the light-harvesting antenna of *Rhodopseudomonas viridis*." *Biophys. J.* 61: 694–703.
- Zhang, F. G., Van Grondelle, R., & Sundstrom, V., 1992b. "Pathways of energy flow through the light harvesting antenna of the photosynthetic purple bacterium *Rhodobacter sphaeroides*." *Biophys. J.* 61: 911–920.
- Zeilstra-Ryalls, J. H., Gabbert, K., Kranz, R., & Kaplan, S., 1997A. "Analysis of the *fnrL* genes and its function in *Rhodobacter capsulatus*." *J. Bacteriol.* 179: 6422-6431.

- Zeilstra-Ryalls, J. H., & Kaplan, S., 1997B. "Aerobic and anaerobic regulation in *Rhodobacter sphaeroides* 2.4.1: the role of the *fnrL* gene." *J. Bacteriol.* 179: 6422-6431.
- Zeilstra-Ryalls, J. H., Gomelsky, M., Eraso, J. M., Yeliseev, A., O'Gara, J., & Kaplan, S., 1998. "Control of photosystem formation in *Rhodobacter Sphaeroides*." *J. Bacteriol.* 180: 2801-2809.
- Zeng, X. H., & Kaplan, S., 2001. "TspO as a modulator of the repressor/antirepressor (PpsR/AppA) regulatory system in *Rhodobacter sphaeroides* 2.4.1." *J. Bacteriol.* 183: 6355-6364.
- Zuber, H. and Cogdell, R. J. (1995). "Structure and organization of purple bacterial antenna complexes." In *Anoxygenic Photosynthetic Bacteria. Advances in Photosynthesis* (Blakenship, R. E., Madigan, M. T., & Bauer, C. E., eds.) Dordrecht: Kluwer Academic Publishers. pp. 315–348.
- Zubritzky, E., 1999. "Widespread interest in gene quantification and high-throughput assays are putting quantitative PCR back in the spotlight." *Analytical chemistry news and features*: 194A.

Appendix 1 – Growth Media

Succinate media for growth of *Bl. viridis*

Per 1 liter of media

Concentrated base (see below)	20 ml
1M Dipotassium hydrogen orthophosphate	10 ml
1M potassium dihydrogen orthophosphate	10 ml
10% (w/v) ammonium sulphate solution	5 ml
1M sodium succinate	10 ml
Growth factors (see below)	1 ml
Casamino acids	1 gr

The pH of the media was adjusted to 6.8 using HCL and then sterilized in the autoclave.

Concentrated Base

Per 1 liter of solution

Nitrilotriacetic acid	10 g
Magnesium sulphate	14.45 g
Calcium chloride.2H ₂ O	3.4 g
Ammonium molybdate	9.25 mg
Ferrous sulphate.7H ₂ O	99 mg
Nicotinic acid	50 mg
Aneurine hydrochloride	25 mg
Biotin	0.5 mg
Metos 44 (see below)	50 ml

Growth Factors per 1 litre solution

Biotin	0.02 g
Sodium hydrogen carbonate	0.5 g
Make up to 1 litre with deionised water to dissolved, then add	
Nicotinic acid	1 g
Aneurine hydrochloride	0.5 g
4-aminobenzoic acid	1 g
Boil to dissolve	

Metos 44

Per 100 ml of solution

Disodium EDTA	0.25 g
Zinc sulphate	1.095 g
Manganous sulphate.4 H ₂ O	0.154 g
Copper sulphate.5H ₂ O	39.2 mg
Cobaltous nitrate.6 H ₂ O	24.8 mg
Ferrous sulphate.7 H ₂ O	0.55 g
Di-sodium tetraborate.10 H ₂ O	17.7 mg

Make up to 100 ml with deionised water and add 2 drops concentrated sulphuric acid

SOC medium (Per 100 ml of solution)

Bacto®-tryptone	2.0 g
Bacto®-yeast extract	0.5 g
1M NaCl	1 ml
1M KCl	0.25 ml
2M Mg ²⁺ stock, filtersterilized	1 ml
2M glucose, filter-sterilized	1 ml

Add Bacto®-tryptone, Bacto®-yeast extract, NaCl and KCl to 97 ml distilled water. Stir to dissolve. Autoclave and cool to room temperature. Add 2M Mg²⁺ stock and 2M glucose, each to a final concentration of 20mM. Bring to 100 ml with sterile, distilled water. The final pH should be 7.0.

LB medium (per liter)

Bacto®-tryptone	10g
Bacto®-yeast extract	5g
NaCl	5g

Adjust pH to 7.0 with NaOH.

LB plates with ampicillin

Add 15g agar to 1 liter of LB medium. Autoclave. Allow the medium to cool to 50°C before adding ampicillin to a final concentration of 100µg/ml. Pour 30–35ml of medium into 85mm petri dishes. Let the agar harden. Store at 4°C for up to 1 month or at room temperature for up to 1 week.

LB plates with ampicillin/IPTG/X-Gal

Make the LB plates with ampicillin as above; then supplement with 0.5mM IPTG and 80µg/ml X-Gal and pour the plates. Alternatively, 100µl of 100mM IPTG and 20µl of 50mg/ml X-Gal may be spread over the surface of an LB ampicillin plate and allowed to absorb for 30 minutes at 37°C prior to use.

Appendix 2- Standard Solutions

TE buffer

10 mM Tris-Cl, pH 7.5

1 mM EDTA

From 1M stock of Tris-Cl (pH 7.5) and 500 mM stock of EDTA (pH 8.0)

10 ml 1M Tris-Cl pH 7.5 per L

2 ml 500 mM EDTA pH 8.0

1 M Tris (crystallized free base)

Tris (hydroxymethyl) aminomethane

FW 121.4 g/mol

60.57 g in 0.5L mq water

pH to 7.5 using HCl

0.5M EDTA

Diaminoethane tetraacetic acid (diNa)

FW 372.2 g/mol

18.6 g in 100ml mq water

pH to 8 using NaOH

>> EDTA will not soluble until pH reaches 8.0

>> Use vigorous stirring, moderate heat (microwave for 2mins) then adjust pH to 8.0;

Calculation example for 250ml of TE buffer:

- $1000 \text{ mM Tris} / 10 \text{ mM Tris (composition)} = 100 \text{ X}$
 $250 \text{ ml} / 100 = \underline{2.5 \text{ ml of Tris buffer use}}$
- $500 \text{ mM EDTA} / 1 \text{ mM EDTA (composition)} = 500 \text{ X}$
 $250 \text{ ml} / 500 = \underline{0.5 \text{ ml of EDTA use}}$

Ruscony Mix:

500 µl lysozyme solution (10 mg/ml)

500 µl 0.5 M EDTA

250 µl 1 M TRIS pH 8.0

100 µl RNase (10 mg/ml)

1.15 ml glycerol (87 %)

0.02 % Bromphenol Blue (end conc. 2 mg/10 ml)

7.5 ml of water

Total volume 10 ml; aliquots 1 ml and stored at -20 °C

IPTG stock solution (0.1M)

1.2g IPTG

Add water to 50ml final volume. Filter-sterilise and store at 4°C.

X-Gal (per 2ml)

100mg 5-bromo-4-chloro-3-indolyl-β-D-galactoside

Dissolve in 2ml N,N'-dimethylformamide. Cover with aluminum foil and store at -20°C.

2M Mg²⁺ stock

20.33g MgCl₂ • 6H₂O

24.65g MgSO₄ • 7H₂O

Add distilled water to 100ml. Filter-sterilise.

Appendix 3 – PCR product (insert): vector Molar Ratios calculation (obtained from pGEM®-T Easy vectors handbook)

To calculate the appropriate amount of PCR product (insert) to include in the ligation reaction, the following equation was used:

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert : vector molar ratio} = \text{ng of insert}$$

Example:

How much 0.5kb PCR product should be added to a ligation in which 50ng of 3.0kb vector will be used if 3:1 insert:vector molar ratio is desired?

$$\frac{50 \text{ ng of vector} \times 0.5\text{kb size of insert}}{3.0\text{kb size of vector}} \times \frac{3}{1} = 25\text{ng of insert}$$

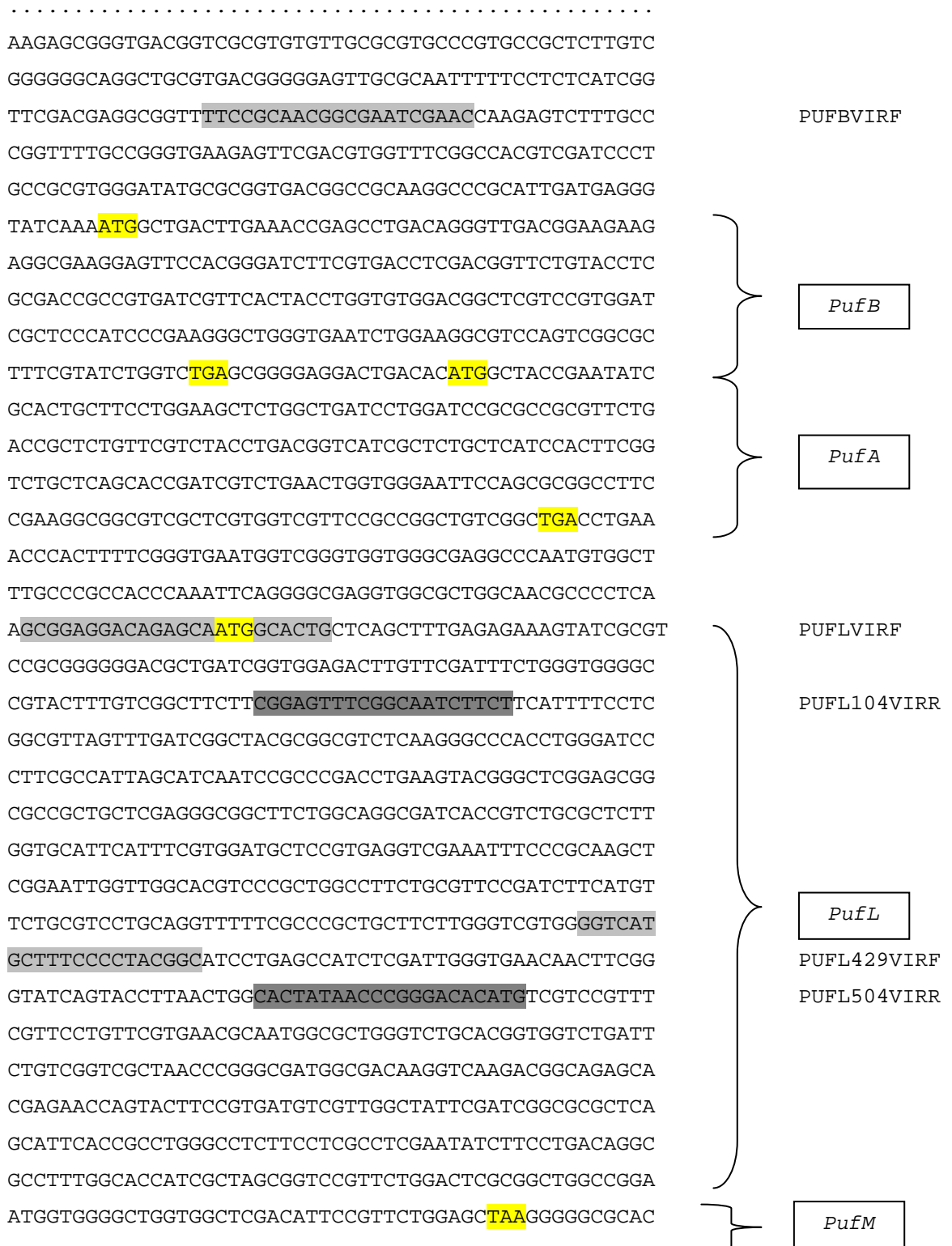
Answer: 25 ng of insert should be added to the ligation mixture.

Appendix 4 - Abbreviations and Masses of the amino acids

Name	3 letter abbreviation	1 letter abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

**Appendix 5 – Primer Design Position in Previous DNA sequence;
Start and stop codons are highlighted in yellow, Primers position are in
gray. (Michel, 1982; Michel, et al., 1985; Michel, et al., 1986; Wiessner,
et al., 1990)**

Puf operon (pufB, pufA, pufL, pufM and pufC)



AATGGCTGACTATCAAACATCTACACGCAGATTCAGGCCCCGCGCCCGC	PUFM47VIRF
ATATCACTGTCTCCGGCGAGTGGGGCGACAACGATCGCGTCGGTAAGCCC	<div> <div></div> <div> <div> <div>PuFM</div> </div> </div> </div>
TTCTATTCTTACTGGCTGGGCAAGATCGGTGACGCGCAGATCGGGCCGAT	
CTATCTGGGTGCTTCGGGAATTGCGGCGTTCGCTTTCGGTTCGACCGCGA	
TCCTGATCATCCTGTTCAACATGGCGGCCGAAGTTCACTTCGATCCTCTG	
CAGTTCTTCCGGCAGTTCTTCTGGCTCGGCCTTTATCCGCCGAAGGCGCA	
GTACGGCATGGGCATCCCGCCGCTTCATGATGGCGGCTGGTGGCTTATGG	
CCGGCCTGTTTCATGACGCTGTGCTCGGTTCTGGTGGATCCGGGTGTAC	
TCGCGGGCTCGTGCTCTCGGCCTCGGCACCCACATCGCGTGGAAC TTCG	
TGCGGCGATCTTCTTCGTGCTGTGTATTGGCTGCATCCATCCGACTCTGG	
TCGGCAGCTGGTCGGAAGGCGTTCGGTTCGGCATCTGGCCGCACATCGAC	
TGGCTGACTGCGTTCTCGATCCGCTACGGCAACTTCTACTATTGCCCGTG	<div> <div></div> <div> <div> <div>PuFM</div> </div> </div> </div>
GCACGGGTTCTCGATCGGCTTCGCCTATGGCTGCGGCCTGCTGTTTCGCGG	
CTCACGGCGCCACCATTCCTGGCCGTTGCCCGGTTTGGCGGCGATCGCGAA	
ATCGAGCAGATCACCGACCGCGGCACTGCGGTGGAGCGTGCGGCTCTGTT	
CTGGCGCTGGACGATCGGCTTCAACGCCACGATCGAGTCTGTCCATCGCT	
GGGGCTGGTTCTTCTCGCTGATGGTGATGGTGTCTGCGAGCGTCGGTATC	
CTTCTGACCGGCACGTTTCGTGCACAACTGGTACCTCTGGTGTGTCAAGCA	
TGGCGCTGCGCCGGACTATCCGGCTTATCTCCCGCCACGCCTGATCCTG	
CCTCGCTGCCGGGAGCACCGAAATGA AACAGCTGATTGTTAATTCGGTCG	
CGACTGTGCGCTGGCGTCTCTCGTGGCCGGTTGTTTCGAACCGCCGCCG	
GCTACCACGACCCAGACTGGTTTCCGCGGACTTTCGATGGGTGAGGTTCT	<div> <div></div> <div> <div> <div>PuFC</div> </div> </div> </div>
TCACCCGGCGACCGTGAAAGCCAAGAAGGAACGTGACGCTCAGTATCCGC	
CGGCTCTGGCCGCGGTGAAGGCGGAAGGCCCGCCCGTCTCGCAGGTCTAC	
AAGAACGTCAAGGTGCTCGGGAATCTGACCGAGGCCGAGTTCCTGCGGAC	
CATGACGGCGATCACGGAATGGGTGTGCGCTCAGGAAGGCTGCACGTACT	
GCCACGACGAGAACAACCTCGCTTCCGAGGCCAAGTATCCGTACGTGGTG	
GCGCGCCGCATGCTGGAAATGACGCGCGCTATCAACACCAACTGGACGCA	
GCACGTGCGCCAGACCGGTGTGACCTGCTACACCTGCCACCGTGGCACGC	
CGCTCCCGCCGTACGTCCGGTACCTGGAGCCGACGCTGCCGCTGAACAAT	
CGTGAGACGCCGACCCACGTGAGCGGGTTGAGACCCGTTCCGGGCTATGT	
CGTCCGCCTTGCGAAGTACACGGCCTACTCGGCTCTGAACTACGATCCGT	<div> <div></div> <div> <div> <div>PuFC</div> </div> </div> </div>
TCACGATGTTCTCGCGAACGACAAGCGTCAGGTCCGTGTTGTGCCGCAG	
ACGGCGCTTCCGCTGGTTCGGCGTCAGCCGCGGCAAGGAACGTGTCGCT	
GTCGGACGCCTACGCGACCTTCGCGCTGATGATGAGCATCTCTGACTCGC	
TCGGCACCAACTGCACGTTCTGCCATAACGCGCAGACCTTCGAGTCGTGG	
GGTAAGAAGAGCACGCCGACGCGCCATTGCGTGGTGGGGCATTCGGAT	
GGTTCGTGATCTGAACATGAACTATCTCGCTCCGCTGAACGCCCTCGCTGC	
CGGCCAGCCGTCTTGGCCGCCAGGGTGAGGCTCCGAGGCCGACTGCCGC	
ACCTGCCACCAGGGTGTGACGAAGCCGCTGTTCCGTGCGTCCCGGCTCAA	
GGATTATCCGGAGCTGGGCCGATCAAGGCTGCTGCGAAGTAA	
	PUFM478VIRF
	PUFMVIRR
	PUFC138VIRF
	PUFC390VIRF
	PUFC541VIRR
	PUFCVIRR

PuhA

GAATTCCC CGGCTAGACAGTTTTCTGCCACGGAGGTTCTTATGTATCAC
GGCGCTCTCGCTCAACATTTAGACATCGCACAACTCGTATGGTACGCGCA
GTGGCTGGTCATCTGGACGGTTGTCCTGCTGTACCTCCGCCGTGAGGACC
GTCGCGAAGGCTACCCGCTGGTCGAGCCGCTTGGTCTCGTCAAGCTGGCG
CCGGAAGACGGCCAGGTCTACGAGCTGCCCTATCCCAAGACGTTTCGTGCT
CCCGCACGGCGGCACCGTCACGGTTCCGCGTCGTTCGGAAAACCCGCG
AACTGAAGCTGGCGCAGACCGACGGCTTCGAAGGCGCCCCGCTGCAGCCG
ACCGGCAATCCGCTGGTCGACCCGTCGGCCCGGCTTCGTATGCCGAGCG
TGCGGAAGTGGTCGACGCCACGGTTGACGGCAAGGCCAAGATCGTCCCGC
TGCGCGTTGCGACCGACTTCTCGATCGCGGAAGGCGACGTCGATCCGCGT
GGCCTGCCGGTGGTTGCCGCTGACGGCGTCGAGGCTGGTACGGTTACCGA
CCTCTGGGTCGACCGCTCGGAGCACTATTTCCGCTACCTCGAGCTCTCGG
TGGCCGGCAGCGCCCGCACCGCGCTGATCCCGCTCGGCTTCTGCGACGTC
AAGAAGGACAAGATCGTCGTGACGTCGATCCTGTCCGAGCAGTTCGCCAA
CGTGCCGCGTCTGCAGAGCCGCGACCAGATCACGCTGCGCGAAGAAGACA
AGGTGTCGGCCTACTACGCTGGCGGTCTGCTCTACGCGACGCCGGAGCGT
GCGGAATCGTTGCTGTGAACGACGATTTTCGA

PUFAVIRF

PUHAVIRR

Appendix 6 – Raw Sequencing Data of Primers

pufBvirF

CAGAGCTTTGCCGGCTTTGCTGCGTGAANTANNACGGCGTGGTTTCGGCCACGTCGCATC
CCTGNCCGCGTGGGATATGCGCGGTGACGGCCGCAAGGCCCGCATTGATGAGGGTATCAA
AATGGCTGACTTGAAACCGAGCCTGACAGGGTTGACGGAAGAAGAGGCGAAGGAGTTCCA
CGGGATCTTCGTGACCTCGACGGTTCTGTACCTCGCGACCGCCGTGATCGTTCACTACCT
GGTGTGGACGGCTCGTCCGTGGATCGCTCCCATCCCGAAGGGCTGGGTGAATCTGGAAGG
CGTCCAGTCGGCGCTTTTCGTATCTGGTCTGAGCGGGGAGGACTGACACATGGCTACCGAA
TATCGCACTGCTTCCTGGAAGCTCTGGCTGATCCTGGATCCGCGCCGCGTTCTGACCGCT
CTGTTCGTCTACCTGACGGTCATCGCTCTGCTCATCCACTTCGGTCTGCTCAGCACCGAT
CGTCTGAACTGGTGGGAATTCCAGCGCGGCCCTTCCGAAGGCGGCGTCGCTCGTGGTCGTT
CCGCCGGCTGTGCGGTGACCTGAAGCCCATTTTCGGGTGAATGGTCGGGTGGTGGGCGAG
GCCTAATGTGGCTTTGCCCGCCACCCAAATTTCAGGGGCGAGGTGGCGCTGGCAACGCCCC
TCAAGCGGAGGACAGAGCAATGGCACTGCTCAGCTTTGAGAGAAAAGTATCGCGTCCGCGG
GGGACGCTGATCGGTGGGGAGACTTTGTTTTCGATTTCTTGGGTGGGGGCCGTACTTTGT
CGGCTTCTTCGGAGTTTTTCGGCAATCTTCTTCATTTTCTCGGCGTTAGTTTGATNCGGGG
CTACGCCGGGCGTCACACAAGGGCCCCAACCTGGAATCCCTTTTCGCCATCAGCATTCAAT
CCCGGCCCGGAACCTTGGAAGGTTAACGGGGGCTCCGGCGCCGGGCGGCCGCTTANCT
CGAAAGGGCCGGGNTTCTTGCCAGGCANATCANCCGGCTTAGGGCTCCTTGGGTGCGTT
CACATTCCACGGGGATTCCCCGGAAGATCAAAATTCCCGCGAGGCCTCGATTTGGTTGGC
ACGTCCCCGTGCCTTTTGGCTCACACTTTTGTTTTNGNCTTCACGGTTTTCCNCGTCTT
CTGCAGGGAGACT

PufLvifF

GCAACCCCCNTAGCTTTGAGAGAAAAGTATCGCCGTCCGGCGTTANGGGACGCTGATCGG
TGGAGACTTGTTTCGATTTTCTGGGTGGGGCCGTACTTTGTTCGGCTTCTTCGGAGTTTCGG
CAATCTTCTTCATTTTCTCGGCGTTAGTTTGATCGGCTACGCGGCGTCACAAGGGCCCA
CCTGGGATCCCTTCGCCATCAGCATCAATCCGCCCCGACCTGAAGTACGGGCTCGGCGCGG
CGCCGCTGCTCGAGGGCGGCTTCTGGCAGGCGATCACCGTCTGCGCTCTTGGTGCCTTCA
TTTTCGTGGATGCTCCGTGAGGTGAAATTTCCCGCAAGCTCGGAATTGGTTGGCACGTCC
CGCTGGCCTTCTGCGTTCCGATCTTCATGTTCTGCGTTCTGCAGGTTTTTCGCCCCGTGC
TTCTCGGTTTCGTGGGGTCATGCGTTTCCCTACGGCATCTGAGCCATCTCGATTGGGTGA
ACAACTTCGGGTATCAGTACCTTAACTGGCACTACAACCCGGGACACATGTCGTCCGTTT
CGTTCTCTGTTTGTGAACGCGATGGCGCTGGGTCTGCACGGCGGTCTCATCTGTGCGTTG
CCAACCCCGGCGATGGCGACAAGGTGAAGACGGCGGAACACGAGAACCAGTACTTCCGTG
ACGTGCTCGGCTACTCGATCGGCGCGCTCAGCATTCACCGTCTGGGCCTGTTCTTGGCGT
CGAACATCTTCTGACGGGCGCCTTTGGCACCATCGCTAGCGGTCCGTTCTGGACCCGCG

GCTGGGCCGGAATGTTGGGGCTGGTTGGCTCGACATTTCCCGTTCTGGAAGGCTAGGGGG
 GCGCACAATGGCTGATTATTCAAACCTTATCTAACACGCAGATTTCAGGCCGCGCCCCGCAT
 ATCCACTTGTCTTCCGCGAATTGGGGGCAACAACAATTGNANTCTGGGTTAAAGCCCTTC
 TATTTTAAACTGGGCTGGGACAAAACCTCGGGTGAAACGCCCAAATATTAGGGGCCGACTC
 TAANCNGGGNNACTACCGGGAAACTCCCCGGGTTTGCCNTTGGGGGGCACACGNGACCCCT
 GAACAATCCGGTTCNAAANTTGGCCAAGATCCTTAGCACCCGTNGATTTTCCNCANTTTT
 GGTGCTCTACCCCAGGAGGCGGAGAGTGGTGGACACCCCCCTATAAAGNGTGGGTTTAN
 GANCGTTACACACACTTACGNCAGAGAGNGCGACGGGAGAAAGGACACCTCTTNCAAGT

pufL104virR

AAGCGACAAGTACGGCCNCCCAGGATTAGAACAAGTCTCCACCGATCAGCGTCCCCCT
 GCGGACGCGATACTTTCTCTCAAAGCTGAGCAGTGCCATTGCTCTGTCTCCGCTTGAGG
 GCGGTTGCCAGCGCCACCTCGCCCCCTGAATTTGGGTGGCGGGCAAAGCCACATTAGGCCT
 CGCCCACCACCCGACCATTACCCGAAAATGGGCTTCAGGTCAGCCGACAGCCGGCGGAA
 CGACCACGAGCGACGCCGCTTCGGAAGGCCGCGCTGGAATTTCCACCAGTTTCAGACGAT
 CGGTGCTGAGCAGACCGAAGTGGATGAGCAGAGCGATGACCGTCAGGTAGACGAACAGAG
 CGGTGAGAACGCGGCGCGGATCCAGGATCAGCCAGAGCTTCAGGAAGCAGTGCGATATT
 CGGTAGCCATGTGTGTCAGTCCTCCCCGCTCAGACCAGATACGAAAGCGCCGACTGGACGCC
 TTCCAGATTACCCAGCCCTTCGGGATGGGAGCGATCCACGGACGAGCCGTCCACACCAG
 GTAGTGAACGATCACGGCGGTTCGCGAGGTACAGAACCGTCGAGGTCACGAAGATCCCCGTG
 GAACTCCTTCGCCTCTTCTTCGTCAACCCTGTGTCAGGCTCGGTTTCAAGTCAGCCATTTT
 GATACCCTCATCAATGCGGGCCTTGGCGGCCGTACCCGCGCATATCCACGCGGCAGGGA
 TCGACGTGGCCGAAACCACGCCGAACCTCTTACCCAGCAAAGCGG

pufL504virR

CAGTTAGGTACTGATACCGACAGTTGGTCGNNTCTAATCGAGAATGGCTCAGGATGCCGT
 AGGGTAAACGCATGACCCACGAACCGAGAAGCAGCGGGCGAAAAACCTGCAGAACGCGAG
 AACATGAAGATCGGAACGCAGAAGGCCAGCGGGACGTGCCAACCAATTCCGAGCTTGCGG
 GAAATTTGACCTCACGGAGCATCCACGAAATGAACGCACCAAGAGCGCAGACGGTGATC
 GCCTGCCAGAAGCCGCCCTCGAGCAGCGGCGCCGCGCCGAGCCCGTACTTCAGGTCGGGC
 GGGATTGATGCTGATGGCGAAGGGATCCCAGGTGGGCCCTTGTGACGCCGCGTAGCCGAT
 CAAACTAACGCCGAGGAAAATGAAGAAGATTGCCGAAACTCCGAAGAAGCCGACAAAGTA
 CGGCCCCACCCAGAAATCGAACAAGTCTCCACCGATCAGCGTCCCCCGCGGACGCGATA
 CTTTCTCTCAAAGCTGAGCAGTGCCATTGAGAAATGGCTCAGGATGCCGTAGGGAAACGC
 ATGACCCACGAACCGAGAAGCAGCGGGCGAAAAACCTGCAGAACGCAGAACATGAAGAT
 CGGAACGCAGAAGGCCAGCGGGACGTGCCAACCAATTCCGAGCTTGCGGGAAATTTGAC
 CTCACGGAGCATCCACGAAATGAACGCACCAAGAGCGCAGACGGTGATCGCCTGCCAGAA
 GCCGCCCTCGAGCAGCGGCGCCGCGCCGAGCCCGTACTTCAGGTCGGGCGGGATTGATGC

TGATGGCGAAGGGATCCCAGGTGGGCCCTTGTGACGCCGCTAGCCGATCAAACCTAACGC
CGAGGAAAATGAAGAAGATTGCCGAACTCCGAAGAAGCCGACAAAAGTACGGCCCCACCC
AGAAATCGAACAAGTCTCCACCGATCAGCGTCCCCCGCGGACGCGATACTTTCTCTCAA
AGCTGAGCAGTGCCAT

pufL429virF

ATCTGAGCATCTCGATTGGGTACAGTTTAAAGTATCAGTACCTTAACTGGCACTACAACC
TCGGGACACATGTGCTCCGTTTTCGTTCTGTTTGTGAACGCGATGGCGCTGGGTCTGCAC
GGCGGTCTCATCTGTGCGTTGCCAACCCCGGCGATGGCGACAAGGTGAAGACGGCGGAA
CACGAGAACCAGTACTTCCGTGACGTGCTGCGCTACTCGATCGGCGCGCTCAGCATTCAC
CGTCTGGGCCTGTTCTGCGCTCGAACATCTTCTGACGGGCGCCTTTGGCACCGATCGC
TAGCGGTCCGTTCTGGACCCCGCGGTGGCCGGAATGGTGGGGCTGGTGGCTCGACATTC
GTTCTGGAGCTAAGGGGGCGCACAATGGCTGATTATCAAACCTATCTACACGCAGATTCAG
GCCCCGCGCCCGCATATCACTGTCTCCGGCGAGTGGGGCGACAACGATCGCATCGGTAAG
CCCTTCTATTCTTACTGGCTGGGCAAGATCGGTGACGCGCAGATCGGGCCGATCTATCTG
GGTGCTTCGGGAATCGCGGCGTTTCGCTTCGGCGCGCGNCCGCGATCCTGATCATCGGGTT
CAACATGCTGGCCGAAGTCCACTTCGACCCGCTGCAGTTCTTCCGCCAGTTCTTCTGGCT
CGGCCTCTACCCGCCGAAGGCGCAGTACGGCATGGGCATCCCGCCGCTTCATGATGGCGG
CTGGTGGCTTATGGCCGGCCTGTTTCATGACGCTGTGCTTGGCTCCTGGTGGATCCGGGG
TGTA~~CT~~CTCGGGGGCCCGTGCCTCTCGGGCCTTTGGGTACCCAACATCGCGTGGAAACT
TCGCTTGGCGGGATCTTCTTTCTGTGCTGTGCATCGGGTTGCCTCCATCCGGCGCTTGTCTG
GGANGCTGGGTCTAGGAANGGNTNCCGTC

pufM47virF

AGTGGGCGACACGATCGCATCGCCTAGCGCCTTCNNATTCTTACTGGCTGGGCAACATCG
GTGATCGCGCAGATCGGGCCGATCTATCTGGGTGCTTCGGGAATCGCGGCGTTTCGCCCTC
GGCGCGACCGCGATCCTGATCATCGGGTTCAACATGCTGGCCGAAGTCCACTTCGACCCG
CTGCAGTTCTTCCGCCAGTTCTTCTGGCTCGGCCTCTACCCGCCGAAGGCGCAGTACGGC
ATGGGCATCCCGCCGCTTCATGATGGCGGCTGGTGGCTTATGGCCGGCCTGTTTCATGACG
CTGTGCTTGGCTCCTGGTGGATCCGGGTGTA~~CT~~CTCGCGGGCCCGTGCCTCTCGGCCTTGGT
ACCCACATCGCGTGGAACTTCGCTGCGGCGATCTTCTTCTGTGCTGTGCATCGGTTGCATC
CATCCGGCGCTTGTGCGCAGCTGGTCGGAAGGCGTTCCGTTTCGGCATCTGGCCGCACATC
GACTGGCTGACCGGTTCTCGATCCGCTACGGCAACTTCTACTATTGCCCGTGGCACGGG
TTCTCGATCGGCTTCGCCTACGGCTGCGGCCTTTTGTTCGCGGCTCACGGTGCGACCATC
CTGGCCGTGCTCGGTTTGGCGGCGATCGCGAAATCGAGCAGATCACCGACCGCGGCACC
GCGGTGGAGCGTGCGGCTCTGTTCTGCGCTGGACGATCGGCTTCAACGCCACGATCGAG
TCTGTCCATCGCTGGGGCTGGTTCTTCTGCTGATGGTGATGGTGTCTGCGAGCGTCGGT
ATTCTTCTGACCGGCACCTTCGTGCGACAACCTGGTAACCTCTGGTGTGTCAAGCATGGGC
GCTGCGCCGGGACTATCCGGGCTTAATCTCCCCGCCACGGCCTGATCCTGGCTCGCTGC

CGGAT

T7 primer for *pufC* fragment

CGATGGGCCGACGTGCGATGCTCCCGGCCGCGATGGCGGCGCGGGAATCGATTTTACTTTCG
CAGCAGCCTTGATCGGGCCCAGCTCCGGATAATCCTGGAGACGGGACGCACCGAACAGCG
GCTTCGTACACCCTGGTGGCAGGTGCGGCAGTCGGCCTGCGGAGCCTCACCTGGCGGC
CAAGACGGCTGGCCGGCAGCACGGTGTTTCAGCGGAGCGAGATAGTTCATGTTTCATGTCAC
GAACCATCCGAATGCCCCACCAAGCGATGGCGCGCTGCGGCGTGCTCTTCTTACCCAGG
TCTCGAACGACTGCGCGTTATGGCAGAACGTGCAGTTGGTTCCGAGCGAGTCAGAGATGC
TCATCATCAGCGCGAAGGTGCGCATAGGCGTCCGACAGCGGGCGACGTTCCCTTGCCGCGGC
TGACGCCGACAAGCGGGAGCGCCGTCTGCGGCACCAACACGGACCTGACGCTTGTCGTTTCG
CGAGGAACATCGTGAACGGATCGTAGTTCAGAGCCGAGTAGGCCGTGTACTTTCGCGAGGC
GGACGACGTAGCCCGAACGGGTCTCAACCCGCTCGACGTGGGTGCGCGTCTCACGATTGT
TCAGGGGCAGCGTCGGCTCCAGGTACCGGACGTACGGCGGGAGCGGCGTGACGGTGGCA
GGGTTGTAAGCCAGGGTCCACACCCGGGTCTGGGCCGAACGTGCTGCGTTCCCAGATTTG
GGTGTTCGATCGGCAACGCCGTTTCATCCTCGNAGGCATTGCGGAACGCGGCCAACACCAC
CGTTACGGGGTACTTGGGCCTCTTGAAGGCGAAGGTTTGTTCCTACGGTCAGTGGGCAAG
TACGTTGCAAGCCTTTCTCCTCCGGGNAACCCCAATTTCCGGTATTCGCCNGGTCCATTGG
TCCCGAAGGGAACANTCGGCNTCGGGTCAGANTTTCCGAGAGCCACCT

SP6 primer for *pufC* fragment (end part)

GATTCATGCTATCNTTCCAAGTGCCTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGG
CCGCGAATTCAGTAGTGATTACCGTGAAAGCCAAGAAGGAGCGTGACGCGCAGTATCCGC
CGGCGCTGCCGGCGGTGAAGGCGGAAGGCCCGCGGTGTGCGAGGTCTACAAGAACGTCA
AGGTGCTCGGAAATCTGACCGAGGCCGAGTTCCTGCGGACCATGACGGCGATGACGGAAT
GGGTGTGCGCCGAGGAAGGCTGCACGTACTGCCACGACGAGAACCAACCTCGCTTCAGAGG
CCAAGTACCCGTACGTGGTGGCGCTGCGATGCTCGAGATGACGCGTGCGATCAACACCA
ACTGGACGCAGCACGTGCCCCAGACCGGTGTGACCTGCTACACCTGCCACCGTGGCACGC
CGCTCCCGCCGTACGTCCGGTACCTGGAGCCGACGCTGCCCCCTGAACAATCGTGAGACGC
CGACCCACGTGAGCGGGTTGAGACCCGTTCCGGGCTACGTGCTCCGCCTCGCGAAGTACA
CGGCCTACTCGGCTCTGAACTACGATCCGTTTCAGATGTTCCCTCGCGAACGACAAGCGTC
AGGTCCGTGTGGTGCCGAGGACGGCGCTCCCGCTTGTCGGCGTCAGCCGCGGCAAGGAA
CGTCGCCCCGCTGTGCGACGCCTATGCGACCTTCGCCGCTGATGATGAGCATCTCTGACTC
GCTCGGGAACCAACTGCAACGTTCTGCCATAACGCGCAGGTCGTTGAGAACCCCTGGGGT
TAAAGAAAGAGCACACGCCGAAATCGCGCCATCGCTTGGGTGGGGCCATTTCGGGATGGTT
CGTGGACATTGAACATTAAGCTTATCTCTCCGNTTCAGTGAACACCCGTGCCTGCCGCCA
NACGTCTTTGGGCCGAGGTGAAGGCTCCGGAAGCGATTGACGGCACNCTTGAC

T7 primer for *puhA* fragment

CAATGGGCCGACGTCGCATGCTCCCGGCTGTCATGGCGGCGCGGGAATCGATTTCGGCTAG
ACAGTTTTCTGCCCACGGAGGTTCTTATGTATCACGGCGCTCTCGCTCAACATTTAGACA
TCGCCCAACTCGTATGGTACGCGCAGTGGCTGGTCATCTGGACGGTTGTCTGCTGTACC
TCCGCCGTGAGGACCGTCGCGAAGGCTACCCGCTGGTCGAGCCGCTTGGTCTCGTCAAGC
TGGCGCCGGAAGACGGCCAGGTCTACGAGCTGCCCTATCCCAAGACGTTCTGTGCTCCCGC
ACGGCGGCACCGTCAACGTTCCGCGTCGTCTGTCGCCGAGACCCGCGAGCTGAAGCTCGCGC
AGACCGACGGCTTCGAGGGCGCCCCGCTGCAGCCGACCGGCAATCCGCTGGTCGACGCCG
TTGGCCCCGGCTTCGTATGCCGAGCGCGCGGAAGTGGTCGACGCCACGGTTGACGGCAAGG
CCAAGATCGTCCCGCTTCGTGTTGCGACCGACTTCTCGATCGCGGAAGGCGACGTCGATC
CGCGTGGCCTGCCGGTGGTTGCCGCCGACGGCGTCGAAGCCGGTACGGTTACCGACCTCT
GGGTGACCGCTCGGAGCACTATTTCCGCTACCTCGAGCTCTCGGTGGCCGGCAGCGCCC
GCACCGCGCTGATCCCGCTTCGGCTTCTGCGAATGTCCANANGAAGGACCAAGATCGTCC
GTGACGTCGATCTGTCCGGATCAAGTTCGCCAACGNTGCCGCGTCTGCAGAGCCGCGAAC
CAGNATCACGCTGGCCGAAGAANACAAGGTGTGCGCCTATTAAGCGGGGGCTGTCATAAG
AGAAGCCGAAACCTGNGGAAGCCGTGTT

SP6 primer for puhA fragment

AATCTAAGCTATGCATCCAACGCGCTTGGGAGCNNTTACCCATATGGTCGACCTGCAGGC
GGCCGCGAATTCAGTAGTGATTCTGTCGTTACAGCAACGCTTCCGCACGCTCCGGCGTCG
CGTAGAGCAGACCGCCCGCTAGTAGGCCGACACCTTGTCTTCTTCGCGCAGCGTGATCT
GGTCGCGGCTCTGCAGACGCGGCACGTTGGCGAACTGATCGGACAGGATCGACGTCACGA
CGATCTTGTCCTTCTTGACATCGCAGAAGCCGAGCGGGATCAGCGCGGTGCGGGCGCTGC
CGGCCACCGAGAGCTCGAGGTAGCGGAAATAGTGCTCCGAGCGGTGACCCAGAGGTCGG
TAACCGTACCGGCTTCGACGCCGTCGGCGGCAACCACCGGCAGGCCACGCGGATCGACGT
CGCCTTCCGCGATCGAGAAGTCGGTCGCAACACGAAGCGGGACGATCTTGGCCTTGCCGT
CAACCGTGGCGTCGACCACTTCCGCGCGCTCGGCATACGAAGCCGGGGCAACGGCGTCGA
CCAGCGGATTGCCGGTCGGCTGCAGCGGGGCGCCCTCGAAGCCGTCGGTCTGCGCGAGCT
TCAGCTCGCGGGTCTCGGGACGACGACGCGGAACGGTGACGGTGCCGCCGTGCGGGAGCA
CGAACGTCTTGGGATAGGGCAGCTCGTAGACCTGGCCCGTCTTCCCGCGCCAAGACTTG
ACGAGACCAAGCCGGCTCGACCAGCGGGTAGCCCTTCGGACNAGGGGTCTCAACGGGGGA
AGGTACAGCAGGACAACCCGTCCAGATGACCAGGCACATCTGGCGTACACTAACAGATTG
GCAGATGTCTAAATNNTTAGATAAAGCGCNGGATACATAAAAACCTNCGTGGGCAGAAAAT
TGTTT

Appendix 7 – Amino acids translation for the new sequence

PufB: 210 nucleotides

Translation in forward direction:

DNA: atggctgacttgaaaccgagcctgacaggggtgacggaagaagaggcgaaggagttccacgggatc
AA : .M..A..D..L..K..P..S..L..T..G..L..T..E..E..E..A..K..E..F..H..G..I..

DNA: ttcgtagacctgacgggttctgtacctcgcgaccgccgtgatcgttcactacctggtgtggacggct
AA : .F..V..T..S..T..V..L..Y..L..A..T..A..V..I..V..H..Y..L..V..W..T..A..

DNA: cgtccgtggatcgctcccatcccgaagggctgggtgaatctggaaggcgtccagtcggcgctttcg
AA : .R..P..W..I..A..P..I..P..K..G..W..V..N..L..E..G..V..Q..S..A..L..S..

DNA: tatctgggtctga
AA : .Y..L..V..*..

PufA: 210 nucleotides

Translation in forward direction:

DNA: atggctaccgaatatcgactgcttctctggaagctctggctgatcctggatccgcgccgcgttctg
AA : .M..A..T..E..Y..R..T..A..S..W..K..L..W..L..I..L..D..P..R..R..V..L..

DNA: accgctctgttcgtctacctgacggatcatcgctctgctcatccacttcggtctgctcagcacccgat
AA : .T..A..L..F..V..Y..L..T..V..I..A..L..L..I..H..F..G..L..L..S..T..D..

DNA: cgtctgaactgggtgggaattccagcgcggccttccgaaggcggcgctcgctcggtcggttcggtccgccg
AA : .R..L..N..W..W..E..F..Q..R..G..L..P..K..A..A..S..L..V..V..V..P..P..

DNA: gctgtcggtga
AA : .A..V..G..*..

PufL: 825 nucleotides

Translation in forward direction:

DNA: atggcactgctcagctttgagagaaagtatcgcgctccgcggggggacgctgatcggtggagacttg
AA : .M..A..L..L..S..F..E..R..K..Y..R..V..R..G..G..T..L..I..G..G..D..L..

DNA: ttcgatttctgggtggggccgtactttgtcggttcttccggagtttcggcaatcttcttcattttc
AA : .F..D..F..W..V..G..P..Y..F..V..G..F..F..G..V..S..A..I..F..F..I..F..

DNA: ctcggcgttagtttgatcggtacgcggcgctcacaagggcccacctgggatcccttcgccatcagc
AA : .L..G..V..S..L..I..G..Y..A..A..S..Q..G..P..T..W..D..P..F..A..I..S..

DNA: atcaatccgccccgacctgaagtacgggctcggcgcgggcgccgctgctcgagggcggcttctggcag
AA : .I..N..P..P..D..L..K..Y..G..L..G..A..A..P..L..L..E..G..G..F..W..Q..

DNA: gcgatcacgctctgcgctcttgggtgcgttcattttcggtggatgctccgtgaggtcgaaatttcccg
AA : .A..I..T..V..C..A..L..G..A..F..I..S..W..M..L..R..E..V..E..I..S..R..

DNA: aagctcgggaattgggttggcacgtcccgcgtggccttctgcgttccgatcttcatgttctgcgttctg
AA : .K..L..G..I..G..W..H..V..P..L..A..F..C..V..P..I..F..M..F..C..V..L.

DNA: cagggtttttcgccccgctgcttctcggttcgtggggcatgcgtttccctacggcatcctgagccat
AA : .Q..V..F..R..P..L..L..L..G..S..W..G..H..A..F..P..Y..G..I..L..S..H.

DNA: ctcgattgggtgaacaacttcgggtatcagtaccttaactggcactacaacccgggacacatgtcg
AA : .L..D..W..V..N..N..F..G..Y..Q..Y..L..N..W..H..Y..N..P..G..H..M..S.

DNA: tccgtttctgttctgtttgtgaacgcgatggcgctgggtctgcacggcggtctcatcctgtcggtt
AA : .S..V..S..F..L..F..V..N..A..M..A..L..G..L..H..G..G..L..I..L..S..V.

DNA: gccaacccccggcgatggcgacaagggtgaagacggcggaacacgagaaccagtacttccgtgacgtc
AA : .A..N..P..G..D..G..D..K..V..K..T..A..E..H..E..N..Q..Y..F..R..D..V.

DNA: gtcggctactcgatcggcgcgctcagcattcaccgtctgggcctgttccctggcgctgaacatcttc
AA : .V..G..Y..S..I..G..A..L..S..I..H..R..L..G..L..F..L..A..S..N..I..F.

DNA: ctgacggggcgccctttggcaccatcgctagcgggtccgttctggacccgcggctggccggaatgttgg
AA : .L..T..G..A..F..G..T..I..A..S..G..P..F..W..T..R..G..W..P..E..W..W.

DNA: ggctgggtggctcgacattccgttctggagctaa
AA : .G..W..W..L..D..I..P..F..W..S..*.

PufM: 975 nucleotides

Translation in forward direction:

DNA: atggctgattatcaaactatctacacgcagattcaggccccgcggcccgcatatcactgtctccggc
AA : .M..A..D..Y..Q..T..I..Y..T..Q..I..Q..A..R..G..P..H..I..T..V..S..G.

DNA: gagtggggcgacaacgatcgcatcggttaagcccttctattcttactggctgggcaagatcggtgac
AA : .E..W..G..D..N..D..R..I..G..K..P..F..Y..S..Y..W..L..G..K..I..G..D.

DNA: gcgcagatcgggccgatctatctgggtgcttcgggaatcgcggttcgccttcggcgcgaccgcg
AA : .A..Q..I..G..P..I..Y..L..G..A..S..G..I..A..A..F..A..F..G..A..T..A.

DNA: atcctgatcatcgggttcaacatgctggccgaagtccacttcgacccgctgcagttcttccgccag
AA : .I..L..I..I..G..F..N..M..L..A..E..V..H..F..D..P..L..Q..F..F..R..Q.

DNA: ttcttctggctcggcctctacccgccgaaggcgagtacggcatgggcatcccgccgcttcatgat
AA : .F..F..W..L..G..L..Y..P..P..K..A..Q..Y..G..M..G..I..P..P..L..H..D.

DNA: ggcggctgggtggcttatggccggcctgttcatgacgctgtcgcttggtcctgggtggatccgggtg
AA : .G..G..W..W..L..M..A..G..L..F..M..T..L..S..L..G..S..W..W..I..R..V.

DNA: tactcgcgggcccgctgctctcggccttggtagccacatcgcggtggaacttcgctgcgggcatcttc
AA : .Y..S..R..A..R..A..L..G..L..G..T..H..I..A..W..N..F..A..A..A..I..F.

DNA: ttctgtgtgtgcatcggttgcatccatccggcgcttgtcggcagctggctcggaaggcgttccgttc
AA : .F..V..L..C..I..G..C..I..H..P..A..L..V..G..S..W..S..E..G..V..P..F.

DNA: ggcatctggccgcacatcgactggctgaccgcgttctcgatccgctacggcaacttctactattgc
AA : .G..I..W..P..H..I..D..W..L..T..A..F..S..I..R..Y..G..N..F..Y..Y..C.

DNA: ccgtggcacgggttctcgatcggttcgcctacggctgcggccttttgttcgcgggtcacggtgcg
AA : .P..W..H..G..F..S..I..G..F..A..Y..G..C..G..L..L..F..A..A..H..G..A.

DNA: accatcctggccgctcgctcggtttggcgggcgatcgcgaaatcgagcagatcacccgaccgcggcacc
AA : .T..I..L..A..V..A..R..F..G..G..D..R..E..I..E..Q..I..T..D..R..G..T.
DNA: gcgggtggagcgctgcggctctgttctggcgctggacgatcggttcaacgccacgatcgagtctgtc
AA : .A..V..E..R..A..A..L..F..W..R..W..T..I..G..F..N..A..T..I..E..S..V.

DNA: catcgctggggctggttcttctcgctgatgggtgatgggtgtctgcgagcgctcggtattcttctgacc
AA : .H..R..W..G..W..F..F..S..L..M..V..M..V..S..A..S..V..G..I..L..L..T.

DNA: ggcaccttcgctcgacaactggtagctctgggtgtgtcaagcatggcgctgcgccggactatccggct
AA : .G..T..F..V..D..N..W..Y..L..W..C..V..K..H..G..A..A..P..D..Y..P..A.

DNA: tatctccccgccacgcctgatcctgcctcgctgccgggagcaccgaaatga
AA : .Y..L..P..A..T..P..D..P..A..S..L..P..G..A..P..K..*.

PufC: 1071 nucleotides

Translation in forward direction:

DNA: atgaaacagctgattgttaattcggttgcgactgtcgcgctggcgctctctcgctggccgggttgtttc
AA : .M..K..Q..L..I..V..N..S..V..A..T..V..A..L..A..S..L..V..A..G..C..F.

DNA: gagccgccgcccggcgaccacgaccagactgggtttccgcgggctgtcgatgggtgaggttcttcac
AA : .E..P..P..P..A..T..T..T..Q..T..G..F..R..G..L..S..M..G..E..V..L..H.

DNA: ccggcgaccgtgaaagccaagaaggagcggtgacgcgcagtatccgccggcgctgccggcggtgaag
AA : .P..A..T..V..K..A..K..K..E..R..D..A..Q..Y..P..P..A..L..P..A..V..K.

DNA: gcggaaggcccgcgggtgtcgaggtctacaagaacgtcaaggtgctcgaaaatctgaccgaggcc
AA : .A..E..G..P..P..V..S..Q..V..Y..K..N..V..K..V..L..G..N..L..T..E..A.

DNA: gagttcctgcggaccatgacggcgatgacggaatgggtgtcgccggaggaaggctgcacgtactgc
AA : .E..F..L..R..T..M..T..A..M..T..E..W..V..S..P..E..E..G..C..T..Y..C.

DNA: caccgacgagaacaacctcgcttcagaggccaagtacccgtacgtgggtggcgctgcgatgctcgag
AA : .H..D..E..N..N..L..A..S..E..A..K..Y..P..Y..V..V..A..R..R..M..L..E.

DNA: atgacgcgtgcgatcaacaccaactggacgcagcacgtcgccagaccggtgtgacctgctacacc
AA : .M..T..R..A..I..N..T..N..W..T..Q..H..V..A..Q..T..G..V..T..C..Y..T.

DNA: tgccaccgtggcacgcccgtccccgccgtacgtccggtacctggagccgacgctgcccctgaacaat
AA : .C..H..R..G..T..P..L..P..P..Y..V..R..Y..L..E..P..T..L..P..L..N..N.

DNA: cgtgagacgcccagcccacgtcgagcggttgagacccgttcgggctacgtcgtccgcctcgcggaag
AA : .R..E..T..P..T..H..V..E..R..V..E..T..R..S..G..Y..V..V..R..L..A..K.

DNA: tacacggcctactcggtctgaactacgatccgttcacgatgttcctcgcgaaacgacaagcgctcag
AA : .Y..T..A..Y..S..A..L..N..Y..D..P..F..T..M..F..L..A..N..D..K..R..Q.

DNA: gtcctgtggtgccgcagacggcgctcccgttgcggcgctcagccgcggcaaggaacgtcgccccg
AA : .V..R..V..V..P..Q..T..A..L..P..L..V..G..V..S..R..G..K..E..R..R..P.

DNA: ctgtcggacgcctatgcgaccttcgcgctgatgatgagcatctctgactcgctcggaaaccaactgc
AA : .L..S..D..A..Y..A..T..F..A..L..M..M..S..I..S..D..S..L..G..T..N..C.

DNA: acgttctgccataacgcgcagtcgttcgagacctggggtaagaagagcacgccgcagcgcgccatc
AA : .T..F..C..H..N..A..Q..S..F..E..T..W..G..K..K..S..T..P..Q..R..A..I.

DNA: gcttggtggggcattcggatggttcgtgacatgaacatgaactatctcgctccgctgaacaccgtg
AA : .A..W..W..G..I..R..M..V..R..D..M..N..M..N..Y..L..A..P..L..N..T..V.
DNA: ctgccggccagccgtcttgcccgccagggtagggtccgcaggccgactgccgcacctgccaccag
AA : .L..P..A..S..R..L..G..R..Q..G..E..A..P..Q..A..D..C..R..T..C..H..Q.

DNA: ggtgtgacgaagccgctggttcggtgcgtcccgtctccaggattatccggagctgggcccgatcaag
AA : .G..V..T..K..P..L..F..G..A..S..R..L..Q..D..Y..P..E..L..G..P..I..K.

DNA: gctgctgcgaagtaa
AA : .A..A..A..K..*.

PuhA: 777 nucleotides

Translation in forward direction:

DNA: atgtatcacggcgctctcgctcaacatttagacatcgcccaactcgtatggtacgcgcagtggtg
AA : .M..Y..H..G..A..L..A..Q..H..L..D..I..A..Q..L..V..W..Y..A..Q..W..L.

DNA: gtcattctggacggttgtcctgctgtacctccgccgtgaggaccgtcgcaaggctaccgcgtggtc
AA : .V..I..W..T..V..V..L..L..Y..L..R..R..E..D..R..R..E..G..Y..P..L..V.

DNA: gagccgcttggtctcgctcaagctggcgccggaagacggccaggtctacgagctgccctatcccaag
AA : .E..P..L..G..L..V..K..L..A..P..E..D..G..Q..V..Y..E..L..P..Y..P..K.

DNA: acgttcgtgctcccgccacggcgccaccgtcaccggtccgcgctcgctcgctcccgagaccgcgagctg
AA : .T..F..V..L..P..H..G..G..T..V..T..V..P..R..R..R..P..E..T..R..E..L.

DNA: aagctcgcgcagaccgacggcttcgagggcgccccgctgcagccgaccggcaatccgctggctcgac
AA : .K..L..A..Q..T..D..G..F..E..G..A..P..L..Q..P..T..G..N..P..L..V..D.

DNA: gccgttggccccggcttcgtatgccgagcgcgcggaagtggctcgacgccacggttgacggcaaggcc
AA : .A..V..G..P..A..S..Y..A..E..R..A..E..V..V..D..A..T..V..D..G..K..A.

DNA: aagatcgtcccgttcgtgttgcgaccgacttctcgatcgcggaaggcgacgtcgatccgcgtggc
AA : .K..I..V..P..L..R..V..A..T..D..F..S..I..A..E..G..D..V..D..P..R..G.

DNA: ctgccgggtggttgccgcccagcggcgtcgaagccggtacgggtaccgacctctgggtcgaccgctcg
AA : .L..P..V..V..A..A..D..G..V..E..A..G..T..V..T..D..L..W..V..D..R..S.

DNA: gagcactatctccgctacctcgagctctcggtggccggcagcgcgcccgaccgcgctgatcccgctc
AA : .E..H..Y..F..R..Y..L..E..L..S..V..A..G..S..A..R..T..A..L..I..P..L.

DNA: ggcttctgcgatgtcaagaaggacaagatcgctcgtagctcgatcctgtccgatcagttcgccaac
AA : .G..F..C..D..V..K..K..D..K..I..V..V..T..S..I..L..S..D..Q..F..A..N.

DNA: gtgccgcgtctgcagagccgcgaccagatcacgctgcgcgaagaagacaaggtgtcggcctactac
AA : .V..P..R..L..Q..S..R..D..Q..I..T..L..R..E..E..D..K..V..S..A..Y..Y.

DNA: gcgggcgggtctgctctacgcgacgccggagcgtgcggaagcgttgctgtga
AA : .A..G..G..L..L..Y..A..T..P..E..R..A..E..A..L..L..*.